Hum Genet
DOI 10.1007/s00439-011-1135-1

ORIGINAL INVESTIGATION

OPRM1 and *EGFR* contribute to skin pigmentation differences between Indigenous Americans and Europeans

Ellen E. Quillen · Marc Bauchet · Abigail W. Bigham · Miguel E. Delgado-Burbano · Franz X. Faust · Yann C. Klimentidis · Xianyun Mao · Mark Stoneking · Mark D. Shriver

Received: 6 September 2011 / Accepted: 14 December 2011

Abstract Contemporary variation in skin pigmentation is the result of hundreds of thousands years of human evolution in new and changing environments. Previous studies have identified several genes involved in skin pigmentation differences among African, Asian, and European populations. However, none have examined skin pigmentation variation among Indigenous American populations, creating a critical gap in our understanding of skin pigmentation variation. This study investigates signatures of selection at 76 pigmentation candidate genes that may contribute to skin pigmentation differences between Indigenous Americans and Europeans. Analysis was performed on two samples of Indigenous Americans genotyped on genome-wide SNP arrays. Using four tests for natural selection—locus-specific branch length (LSBL), ratio of heterozygosities (ln*RH*), Tajima's *D* difference, and extended haplotype homozygosity (EHH)—we identified 14 selection-nominated candidate genes (SNCGs). SNPs in each of the SNCGs were tested for association with skin pigmentation in 515 admixed Indigenous American and European individuals from regions of the Americas with high ground-level ultraviolet radiation. In addition to *SLC24A5* and *SLC45A2*, genes previously associated with European/non-European differences in skin pigmentation, *OPRM1* and *EGFR* were associated with variation in skin pigmentation in New World populations for the first time.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-011-1135-1) contains supplementary material, which is available to authorized users.

E. E. Quillen · M. D. Shriver
Department of Anthropology, Pennsylvania State University, University Park, PA 16802, USA

Present Address:
E. E. Quillen (✉)
Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX 78227, USA
e-mail: equillen@txbiomedgenetics.org

M. Bauchet · M. Stoneking
Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

A. W. Bigham
Department of Anthropology, University of Michigan, Ann Arbor, MI 48109, USA

M. E. Delgado-Burbano
División Antropología, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, Paseo del Bosque s/n., La Plata B1900FWA, Argentina

M. E. Delgado-Burbano
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

F. X. Faust
Department of Anthropology, Universidad del Cauca, Popayán 949020, Colombia

Y. C. Klimentidis
Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL 35294, USA

X. Mao
Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA

Published online: 24 December 2011

Introduction

A complex and highly variable trait, human skin pigmentation shows a striking amount of variation within and among human populations. Skin color is predominantly determined by the pigment melanin and varies based on the amount and type of melanin synthesized in organelles called melanosomes as well as the number, size, and distribution of melanosomes in the epidermis (Thong et al. 2003). Human mapping studies and research in model organisms have led to the inclusion of more than 70 candidate genes for pigmentation in the Online Mendelian Inheritance in Man (OMIM) database. Because of the intense research on pigmentation in mouse models, the genes involved in the pigmentation pathway are well characterized; yet, their role in human skin color variation is not fully understood.

As *Homo sapiens* spread across the globe, pigmentation was subject to a number of evolutionary forces including new environmental pressures, mutation, drift, and sexual selection. Of these, natural selection has been the predominant research focus. A strong, positive correlation between the intensity of ultraviolet radiation (UVR) and skin pigmentation suggests that local adaptation is a primary contributor to the modern geographic distribution of the phenotype (Relethford 1997; Jablonski and Chaplin 2010). The intensity of this selection is evident in European and East Asian populations where genes related to skin pigmentation consistently show extreme deviations from neutrality (Voight et al. 2006; Lao et al. 2007; Myles et al. 2008). It is a reasonable hypothesis, then, that genes exhibiting strong evidence of selection contribute to differences in skin pigmentation between the Indigenous American and European parental populations of contemporary, admixed New World populations.

Previous studies have evaluated selection at skin pigmentation genes in Old World populations, but failed to consider the Americas. Indigenous populations in the American tropics, while less pigmented than tropical populations in Africa and South Asia, are more deeply pigmented than modern inhabitants of northern and central Asia (Jablonski and Chaplin 2000). This study represents a first step in detecting the genes involved in pigmentation differences between New World and Old World populations by identifying pigmentation genes showing consistent deviations from neutral expectations in multiple tests of selection. Additionally, it is the first to test the effects of selected genes on variation in skin color.

Materials and methods

Analysis was performed in two population samples which were genotyped on two different genome-wide single nucleotide polymorphism (SNP) arrays. The first dataset included 88 Indigenous Americans (Quechua from Peru, Aymara from Bolivia, Maya from the Yucatan Peninsula in Mexico, and Mixtec, Tlapanec, and Nahua from Guerrero, Mexico) who were compared to the CEPH Europeans ($n = 60$) and East Asians (Chinese + Japanese, $n = 90$) from the HapMap Project Phase I + II (The International HapMap Consortium 2003). Subjects were genotyped at 906,600 SNPs on the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Inc., Santa Clara, CA) as described in Bigham et al. (2010). The second dataset comprised publicly available genome-wide SNP data from the HGDP-CEPH samples genotyped on the Illumina HumanHap650 K Beadchips (Illumina, Inc., San Diego, CA) which included over 650,000 SNPs (Li et al. 2008). From this set, only Indigenous American ($n = 156$), East Asian ($n = 232$), and European ($n = 156$) individuals were included in analyses following the classification of populations by Li et al. (2008). In both samples, the Indigenous American populations are predominantly localized to the high-UVR regions of the Americas.

Many methods have been developed to identify regions of the genome diverging from neutrality. Frequently referred to as tests of selection, the methods vary in their ability to assess selection at different time-depths and their robustness to population demography (in particular, population size changes which feature prominently in the recent evolutionary history of the Americas). For the best resolution of Indigenous American specific changes—which would have occurred in the last 20,000 years, locus-specific branch length (LSBL) (Shriver et al. 2004), the ratio of heterozygosities (ln*RH*) (Schlötterer and Dieringer 2005), Tajima's *D* difference (Bigham et al. 2010; Tajima 1989) and extended haplotype homozygosity (EHH) (Sabeti et al. 2007) were considered. For each of the analyses, Indigenous American populations were grouped and compared to populations from East Asia, the most closely related available sample population. In the case of LSBL, where a three-way comparison is necessary, Indigenous Americans were compared to both East Asian and European populations. Each SNP with a call rate >95% was considered for LSBL and ln*RH,* and was included in the Tajima's *D* difference sliding window analysis that used 10 kb sliding windows with 5 kb offsets. Because EHH is calculated for genic regions, only SNPs within 200 kb of pigmentation candidate genes were considered with automated identification of the core haplotype regions.

Seventy-six candidate genes for skin pigmentation from the OMIM database were identified based on key-word searches for "color" and "pigmentation"—*ADAM17, ADAMTS20, AP3B1* (*ADTB3A*), *AP3D1, ASIP, ATP7B, ATRN, BCL2, BLOC1S1, BLOC1S3, BNC2, BRCA1, CITED2, CNO, CYP1A2, CYP1B1, CYP2C8, CYP2C9,*

DAG1, DCT, DTNBP1, ECE1, EDN3, EDNRB, EGFR, ERCC2, FGFR2, GNA11, GNAQ, GPNMB, HPS1, HPS3, HPS4, HPS5, HPS6, KIT, KITLG (*MGF*), *KRT1, KRT2A, LMX1A, LYST, SLC45A2* (*MATP*), *MC1R, MCOLN3, MGRN1, MITF, MLPH, MREG* (*DSU*), *MYO5A, MYO7A, OCA2, OPRM1, PAX3, PLDN, POMC* (*α-MSH*), *RAB27A, RAB38, RABGGTA, SFXN1, SILV, SLC24A5, SLC7A11, SNAI2, SNAPAP, SOX10, SOX18, STK11, TBX19* (*ACTH*), *TFAP2A, TYR, TYRP1, VPS13B* (*COH1, CHS1*), *VPS33A, WNT1, WNT3A,* and *ZIC2*. In an effort to capture near-by regulatory elements, because selective sweeps generally affect large expanses of DNA, and to lessen bias against detection of selection in small genes, the gene regions were classified as an interval encompassing 200 kb upstream and downstream of the gene itself. Gene locations were defined by the UCSC Genome Browser Build 35 (Kent et al. 2002). For ln*RH*, LSBL, and Tajima's *D* difference, SNPs or windows in the top (or bottom, depending on the nature of the test) 5% of the distribution of all SNPs were considered statistically significant. For all three tests, only a single tail was considered to focus specifically on directional selection in the ancestors of contemporary Indigenous Americans. When calculating EHH, cores were automatically defined by Sweep (Sabeti et al. 2002) and the three most common cores were used for each gene. In practice, there were rarely more than three core haplotypes at substantial frequencies in the Indigenous Americans. Supplementary file 1 contains all ln*RH*, LSBL, and Tajima's *D* difference scores for the pigmentation candidate genes.

Since multiple lines of evidence were evaluated, genes were considered to show substantial evidence of selection if there were significant results on two or more tests spanning across a region of three or more SNPs or two windows. Genes meeting this threshold are referred to as selection-nominated candidate genes (SNCGs). Based on this criteria, 14 SNCGs were identified—*ADAM17, AP3D1, ASIP, ATRN, EGFR, HPS1, KIT, MYO5A, OPRM1, PAX, PLDN, POMC, RAB27A*, and *SILV*. Figure 1 depicts the LSBL, ln*RH*, and Tajima's *D* difference results for *OPRM1* as compared to *SLC24A5*, a gene which

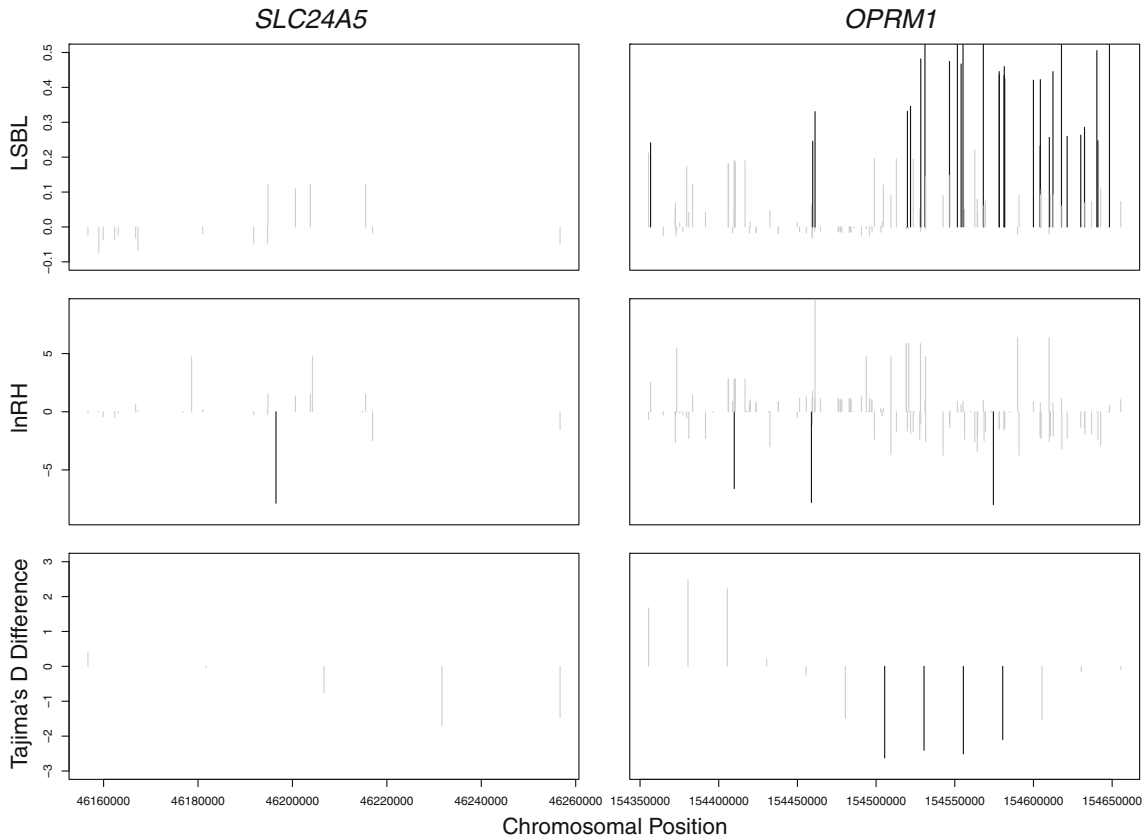

Fig. 1 Plots of deviation from neutrality at *OPRM1* and *SLC24A5* in Indigenous Americans. Scores for LSBL, ln*RH*, and Tajima's *D* difference are plotted for regions extending 200 kb on either side of the gene. *Black bars* represent SNPs (or windows in the case of Tajima's *D* difference) with scores that fall in the 5% tail of genome-wide results. While *SLC24A5* is among the most strongly selected genes in European populations, there is no evidence of selection in the Indigenous Americans. The contribution of *SLC24A5* to pigmentation differences in the admixed sample reflects the European/non-European distribution of rs1426654. In contrast, *OPRM1* shows strong evidence of selection in Indigenous Americans. Only the data genotyped on the larger Affymetrix 6.0 panel are shown here, but there is substantial concordance with the HGDP data for these genes

shows no evidence of selection in Indigenous American populations, but strong evidence of selection in European populations (Lamason et al. 2005; Norton et al. 2007).

Even if all SNCGs are assumed to have undergone selection in Indigenous Americans, this does not prove that the favored polymorphisms contribute to population-level differences in pigmentation. To test the effects of the SNCGs on pigmentation differences between Indigenous American and Old World populations, one to three SNPs per gene were genotyped in an admixed Hispanic sample using a custom Sequenom iPLEX Gold assay (Table 1). To afford the maximum ability to detect admixture linkage association, only SNPs showing high allele frequency differences (δ) between the European and Indigenous American populations in the Affymetrix 6.0 data and that were not in linkage disequilibrium (LD) were included. SNPs in *SLC24A5* (rs1426654) and *SLC45A2* (rs16891982) were also genotyped as they have been shown to be associated with skin pigmentation differences between European and non-European populations (Lamason et al. 2005; Norton et al. 2007). Therefore, they would be expected to influence pigmentation differences in this sample. In total, 29 SNPs passed quality control with a 97% call rate.

Five-hundred and fifteen individuals were successfully genotyped on the custom Sequenom array. These include individuals ascertained in a marketplace in Tlapa, Guerrero, Mexico ($N = 95$) (Bonilla et al. 2005), Hispanics participating in a diabetes study from the San Luis Valley, Colorado ($N = 180$) (Bonilla et al. 2004), self-identified Hispanic and Indigenous American students at the University of New Mexico in Albuquerque ($N = 67$) (Klimentidis et al. 2009) and individuals from Popayán, Colombia ($N = 173$). All individuals gave signed informed consent in their own language prior to participation and all sample collection and genotyping was performed under IRB approval. These populations differ in mean levels of Indigenous American and European admixture, but all reside predominately in regions with high ground-level exposure to UVR and show darker average skin pigmentation than populations in Northern Europe and East Asia (Jablonski and Chaplin 2000). Although current residence in high-UVR climates does not guarantee that their Indigenous American ancestors were exposed to the same conditions, it is the best available proxy.

For each individual, biogeographical ancestry (BGA) estimates were calculated using Maximum Likelihood Estimation (MLE) from panels of ancestry informative markers (AIMs). The AIMs varied among the populations and all available AIMs were used to calculate individual BGA. The samples from Tlapa and the San Luis Valley were genotyped previously for a panel of 22 AIMs (Bonilla et al. 2004, 2005) and the Albuquerque sample was genotyped for a panel of 76 AIMs (Klimentidis et al. 2009).

Table 1 SNPs genotyped in selection-nominated candidate genes

rsID	Gene	Chr	Location	European frequency	Indigenous American frequency
rs11895982	*ADAM17*	2	9484791	0.692	0.034
rs7596929	*ADAM17*	2	9635096	0.917	0.040
rs551573	*POMC*	2	25159008	0.900	0.068
rs12233134	*POMC*	2	25182520	0.042	0.875
rs10186053	*POMC*	2	25263694	0.792	0.313
rs6748403	*PAX3*	2	222589683	0.442	0.733
rs2726648	*KIT*	4	55379488	0.642	0.943
rs16891982	*SLC45A2*	5	33840180	1.000	0.000
rs2333857	*OPRM1*	6	154705260	0.559	1.000
rs6917661	*OPRM1*	6	154763249	0.450	0.756
rs12668421	*EGFR*	7	55076671	0.025	0.807
rs11238349	*EGFR*	7	55123565	0.008	0.972
rs4948023	*EGFR*	7	55421761	0.342	0.852
rs10736126	*HPS1*	10	100015914	0.642	0.102
rs705698	*SILV*	12	54670954	0.042	0.034
rs12439639	*PLDN*	15	43634372	0.150	0.693
rs1426654	*SLC24A5*	15	46142550	1.000	0.000
rs12396	*MYO5A*	15	50200608	0.058	0.051
rs8026828	*MYO5A*	15	50252879	0.667	0.845
rs4776017	*MYO5A*	15	50285718	0.083	0.205
rs11632529	*RAB27A*	15	53123546	0.300	0.136
rs16976220	*RAB27A*	15	53326152	0.325	0.585
rs472579	*RAB27A*	15	53455706	0.200	0.938
rs8105923	*AP3D1*	19	1860478	0.583	0.341
rs4807227	*AP3D1*	19	2269180	0.319	0.432
rs562926	*ATRN*	20	3573436	0.100	0.659
rs3859664	*ATRN*	20	3619891	0.775	0.256
rs1015361	*ASIP*	20	32202347	0.675	0.585
rs6087577	*ASIP*	20	32419084	0.550	0.057

The number of SNPs genotyped in each SNCG reflects both the size of the gene and the functional constraints of the array. SNPs were chosen with large allele frequency differences between Indigenous American and European populations for optimal performance in admixture linkage analysis. European and Indigenous American allele frequencies were taken from the Affymetrix dataset

The Colombian sample was genotyped for 112 AIMs (Fejerman et al. 2008) for this study. Individuals with more than 10% West African and/or East Asian genomic ancestry were removed from the sample to minimize the influence of alleles other than those commonly found in Indigenous American and European populations. Melanin Index (M) was calculated for all participants using reflectometry on the proximal medial portion of the arm with multiple measurements across both arms averaged to generate a single value. This location was chosen to best capture constitutive skin pigmentation—the basal amount of melanin produced by the body due to genes—not the

effect of sun exposure on the skin, known as facultative pigmentation.

Associations between the 29 SNPs and the *M* index were assessed using admixture linkage analysis implemented in ADMIXMAP (Hoggart et al. 2004) and analysis of variance (ANOVA). ADMIXMAP allows for finer scale analysis by imputing the likely ancestry for each allele at each locus of interest based on the allele frequencies of the AIMs in modern Indigenous American and European populations. To control for population stratification in the ANOVA, Indigenous American BGA was included as a covariate. The inclusion of BGA in these analyses is essential to avoid spurious associations between skin pigmentation and any marker with high *F*st between the parental populations. Because skin color and ancestry are strongly correlated in this sample ($R^2 = 0.37$, $p < 1e{-}15$), each of the 29 SNCG SNPs is associated with *M* when individual BGA is not included in the model. Admixture linkage analysis was performed on the set of 29 SNPs with individual BGA estimates supplied as prior to improve the convergence of locus-specific ancestry estimation. Age and sex were included as covariates for all analyses and a Bonferroni-corrected *p* value threshold of 0.0017 was considered for assessing statistical significance. This value is probably overly conservative as it treats SNPs within the same gene as independent. A less stringent correction, considering each of the 16 genes as an independent test, produces a *p* value threshold of 0.0042, which is considered here as a suggestive result.

Results

Previous research indicates that pigmentation genes tend to have higher *F*st values in pairwise comparisons between Europeans and Africans and between Europeans and East Asians than randomly selected regions of the genome (Pickrell et al. 2009). To determine if a similar pattern of selection was present for pigmentation genes in Indigenous Americans, 100 randomly selected sets of 76 genes were generated from the manually annotated GENCODE database (Harrow et al. 2006). The mean scores for normalized Tajima's *D* difference, ln*RH*, and LSBL were calculated for all genic SNPs in the pigmentation and random gene sets. In all cases, the mean scores for pigmentation genes fell in the middle of the distributions of mean scores for the randomly ascertained gene sets. This indicates that, as a class, pigmentation genes have not undergone greater selection than randomly selected genes in Indigenous Americans. However, it does not follow that individual pigmentation genes have not undergone substantial selection in these populations.

In order to determine the effects of individual pigmentation candidate genes on skin color, ADMIXMAP applies an allelic association test that calculates *z* scores equivalent to likelihood odds ratios (LOD scores) as well as *p* values. The highest statistically significant *z* scores are located on chromosome 15 in a region that includes the genes *PLDN*, *SLC24A5*, *MYO5A*, and *RAB27A* (Fig. 2). The effect of the *SLC24A5* SNP on skin pigmentation may be large enough and the region of LD sufficiently broad to cause the SNPs in *PLDN*, *MYO5A*, and *RAB27A* to show false positive associations. Although these SNPs are 2.58, 4.00, and 7.11 Mb away from *SLC24A5*, LD created by admixture can generate blocks of this length. If the adjacent SNPs are being swept along due to the strong effect of *SLC24A5*, the *z* scores should decrease steadily with increasing distance from that gene.

In order to assess the legitimacy of these associations, the two largest individual populations, those from Colombia and the San Luis Valley, were analyzed separately with all available AIMs included with the test SNPs. Although the *z* score for *SLC24A5* is the largest in the region for both populations, *MYO5A* and particularly

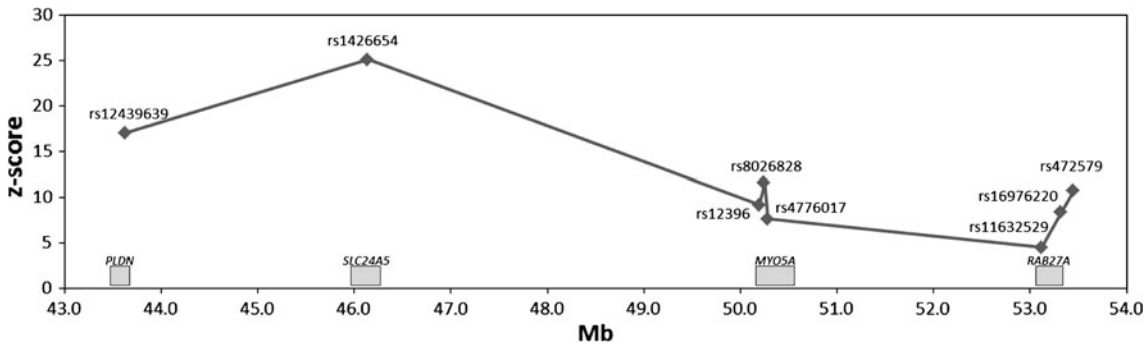

Fig. 2 ADMIXMAP *z* scores at 15q21.1. The region of the genome surrounding *SLC24A5* contains several other pigmentation candidate genes of which *PLDN, MYO5A*, and *RAB27A* were considered SNCG with one, three, and three SNPs genotyped in each gene, respectively. The approximate size and position of each gene can be seen along the *horizontal axis* and the *z* scores are plotted along the *vertical axis*. *Z* scores were significant after Bonferroni correction for all SNPs except rs11632529 in *RAB27A*

RAB27A have separate peaks as in the full sample analysis seen in Fig. 2. This may be an indication that these genes play an independent role in skin color variation in this sample. Notably, *PLDN*, *MYO5A*, and *RAB27A* show evidence of selection in Indigenous Americans, while *SLC24A5* does not. However, including the *SLC24A5* genotype as a covariate when analyzing the effects of the surrounding genes renders these genes' associations with skin color non-significant after Bonferroni correction. A larger sample size and denser marker panel may be necessary to independently assess the effects of these genes.

The second largest z score in the analysis was found at *SLC45A2* ($z = 20.66$, $p = 2e{-}31$), another gene previously associated with substantial European/non-European differences in skin pigmentation. *OPRM1* (rs6917661, $z = 8.613$, $p = 2.28e{-}4$) and *EGFR* (rs12668421, $z = 11.22$, $p = 6.25e{-}7$) were also significantly associated with skin color. When *SLC24A5* and *SLC45A2* genotypes are included as covariates, the z scores for *EGFR* and *OPRM1* increased. Treating these known contributing factors as covariates increases the proportion of observed variation that can potentially be explained by each candidate SNP and reinforces the independence of these genes' effects.

In order to confirm the ADMIXMAP results, ANOVA methods were employed using age and individual genome-wide BGA as covariates. Pigmentation is not significantly sexually dimorphic in this sample, so sex was not a significant covariate in the ANOVA. The results of the ANOVA are largely concordant with the ADMIXMAP results, with significant findings for the genes in the *SLC24A5* region and for *SLC45A2* as well as suggestive results for *OPRM1* and *EGFR*. The use of a genome-wide correction for ancestry in the ANOVA presents a challenge because we are seeking to identify genes contributing to differences in skin pigmentation between Europeans and Indigenous Americans. By including ancestry as a covariate, some of the variation potentially explained by these genes is removed and the power to identify genes associated with this variation is diminished. The degree to which this occurs is related to the effect size of the genes. For genes of larger effect like *SLC24A5* and *SLC45A2*, the associations appear weaker in the ANOVA analysis than in the ADMIXMAP analysis, but are still significant. For *OPRM1* and *EGFR*, the associations with ANOVA are suggestive but no longer significant. A second SNP in *EGFR*, rs11238349, was also suggestively associated with skin pigmentation in the ADMIXMAP results. When these two *EGFR* SNPs are combined into a haplotype, rs12668421*T–rs11238349*A, it is associated with darker skin pigmentation ($p = 0.022$).

This problem recurs when attempting to assess the magnitude of the effects of these genes on skin pigmentation. The residual values from an ANOVA including age and ancestry, referred to here as adjusted *M*, are graphed for each gene in Fig. 3, and the differences in adjusted *M* between the homozygotes were assessed using a *t* test. *SLC24A5* and *SLC45A2* were significant for mean differences in skin color with a magnitude of 3.06 and 3.07 adjusted melanin units between the homozygotes. *OPRM1* and *EGFR* also showed significant differences, but have smaller effects. For *OPRM1,* individuals who were homozygous for the T allele, which is more common in Indigenous Americans, were 0.982 adjusted melanin unit

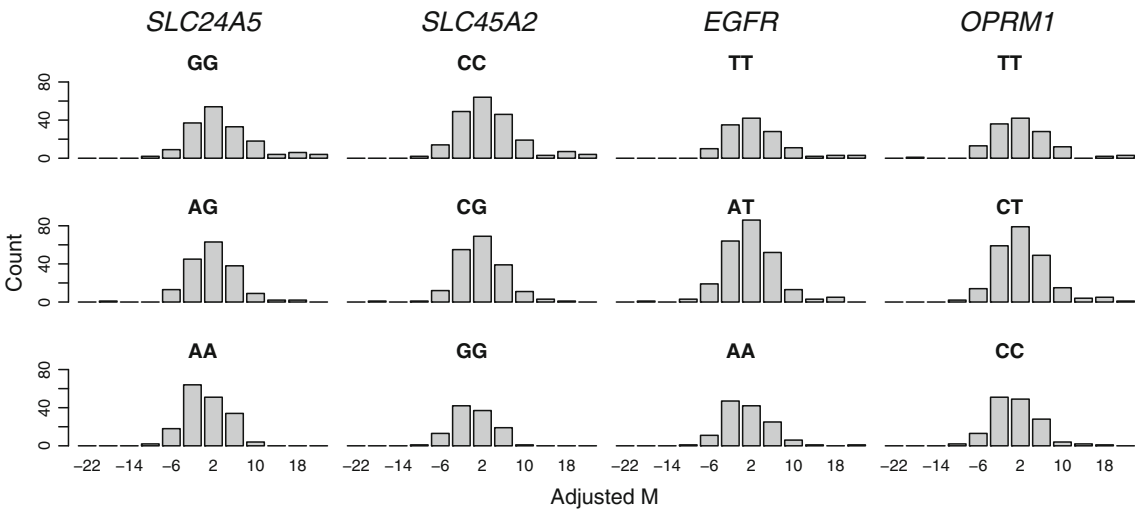

Fig. 3 Distribution of adjusted melanin index (*M*) values by genotype. Adjusted *M* was calculated as the residual values of *M* in an ANOVA including age and BGA. Higher values of adjusted *M* indicate darker skin color. In each case, the allele more common in Indigenous Americans is shown at the *top* of the graph and the allele more common in Western Europeans is shown at the *bottom* of the graph

darker than individuals homozygous for the C allele. *EGFR* showed the same pattern with TT homozygotes, more common in Indigenous Americans, on average 1.73 adjusted melanin units darker than AA homozygotes. These differences are for adjusted melanin units which have been corrected for the effect of ancestry and normalized. It is imperative to bear in mind that adjusting for ancestry will decrease the apparent effect size of the genes, while failing to adjust for ancestry will falsely inflate the apparent effect. For example, when *M* is considered directly, the difference in homozygotes for *OPRM1* and *EGFR* is 4.93 and 6.11 *M* units, respectively.

Discussion

Four genes were found to contribute to skin pigmentation variation in the admixed sample and reflect the differences that have arisen since the split between the ancestors of modern European and Indigenous American populations. Two of these genes, *SLC24A5* and *SLC45A2,* have previously been investigated for their roles in differences between European and West African skin pigmentation. Both show evidence of recent selection in Europe contributing to lighter skin pigmentation found on that continent and in parts of North Africa and the Middle East but not in East Asia (Norton et al. 2007). Based on this distribution, these genes are expected to contribute to variation in skin color between western Europeans and any other group, including Indigenous Americans. However, this association has not previously been described for Hispanic populations.

Although not previously associated with pigmentation variation among human populations, *EGFR* is a type-1 tyrosine kinase receptor that plays a major role in many cellular functions including keratinocyte proliferation, differentiation, migration, and death (reviewed in Jost et al. 2000). The keratinocytes, in turn, regulate the proliferation and differentiation of melanocytes (Hirobe 2011). This pathway is the hypothesized mechanism for the Dsk5 mutation in the homologous mouse gene that causes both the proliferation of basal keratinocytes and an increased number of melanocytes (Fitch et al. 2003).

While selection at *EGFR* has been identified in both East Asians and Africans (Akey et al. 2002; Lao et al. 2007; Johansson and Gyllesnsten 2008), this is the first association of *EGFR* with selection in the Americas and with pigmentation variation in a phenotyped sample. To assess shared selective pressures among the populations, haplotype blocks were generated from the phased Affymetrix data and characterized by solid spines of LD with blocks combined if the multiallelic D' between adjacent blocks was greater than 0.95 using Haploview (Barrett et al. 2005). As expected, the haplotype blocks were smallest and most variable in the Yorubans with longer blocks found in the East Asian populations and the longest blocks in the Indigenous Americans. Lao et al. identified statistically significant EHH values 0.067 cM upstream in African populations and both 0.027 cM downstream and 0.03 cM upstream in East Asians. This is reflected in the HapMap East Asians by a large haplotype block approximately 233 kbp upstream of rs12668421, the SNP associated with pigmentation variation in Indigenous American populations. The second large LD block in the East Asian populations is located approximately 35 kbp downstream of the haplotype block defined by rs12668421 and rs11238349. In the Yorubans, the haploblock identified by Lao et al. is also more than 220 kbp upstream of rs12668421. Although the haploblocks found in the East Asians are apparent in the Indigenous Americans, neither these nor the Yoruban haploblocks show strong LD with rs12668421 or overlap with the 102 kb haploblock defined by rs12668421 and rs11238349. The core haplotype associated with darker skin color in the admixed populations is the most common in both the Indigenous Americans (40.2%) and the East Asians (47.3%), but absent in the Yorubans. Although it does not overlap with the region Lao et al. identified, this shared haploblock raises the possibility of a shared role for EGFR in pigmentation in these populations. However, association between pigmentation variation and *EGFR* in East Asians and Africans remains to be investigated.

OPRM1 is an unusual candidate for influencing pigmentation because it is largely known as an opioid receptor. However, its function as a neuronal regulator suggests a possible association with skin color based on the common origins of the skin and nervous system (Simonin et al. 1995). Opioid receptors are active in both keratinocytes and melanocytes and may influence the differentiation of these cells (Bigliardi et al. 2009; Ho et al. 2010). Although the mechanism remains unknown, this association suggests that further investigations would be worthwhile.

This research has demonstrated, for the first time, evidence of selection at multiple skin pigmentation genes in New World populations. Although there is substantial statistical evidence for selection in these skin pigmentation genes, this evidence alone should not be used to conclude that these genes necessarily influence skin pigmentation. Selection driving these non-neutral patterns could be acting on some trait other than skin pigmentation. While a relatively large proportion of the genes show evidence of selection but no association with skin pigmentation variation in this sample, it remains possible that these genes contribute to skin color but have very small effects that are below the power of this study to detect.

The broad, continuous nature of skin pigmentation indicates that many genes are involved in both within- and among-population variation. Genomic signatures of

selection provide a means to prioritize candidate genes for among-population differences; but additional associations based on phenotypic observations are essential to bolster the functional significance of SNCGs in pigmentation. The utility of this approach is demonstrated in the novel associations of *EGFR* and *OPRM1* with pigmentation differences between Indigenous Americans and Europeans.

Acknowledgments The authors would like to thank the individuals who volunteered their time to participate in this research and the anonymous reviewers for their comments. This project was funded by National Science Foundation Doctoral Dissertation Improvement Grant #0925976 and intramural Research and Graduate Studies Office and Hill Fellowships from Pennsylvania State University to EEQ. MB and MSt were funded by the Max Planck Society. YCK was funded by a Graduate Research Development grant from the University of New Mexico. The authors have no conflicts of interest to disclose.

References

Akey JM, Zhang G, Zhang K et al (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome Res 12:1805–1814. doi:10.1101/gr.631202

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics (Oxford, England) 21:263–265. doi:10.1093/bioinformatics/bth457

Bigham A, Bauchet M, Pinto D et al (2010) Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genetics 6:14

Bigliardi PL, Tobin DJ, Gaveriaux-Ruff C, Bigliardi-Qi M (2009) Opioids and the skin—where do we stand? Exp Dermatol 18:424–430. doi:10.1111/j.1600-0625.2009.00844.x

Bonilla C, Parra EJ, Pfaff CL et al (2004) Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. Ann Hum Genet 68:139–153. doi:10.1046/j.1529-8817.2003.00084.x

Bonilla C, Gutiérrez G, Parra EJ et al (2005) Admixture analysis of a rural population of the state of Guerrero, Mexico. Am J Phys Anthropol 128:861–869. doi:10.1002/ajpa.20227

Fejerman L, John EM, Huntsman S et al (2008) Genetic ancestry and risk of breast cancer among US Latinas. Cancer Res 68:9723. doi:10.1158/0008-5472.CAN-08-2039

Fitch KR, McGowan KA, van Raamsdonk CD et al (2003) Genetics of dark skin in mice. Genes Dev 17:214–228. doi:10.1101/gad.1023703

Harrow J, Denoeud F, Frankish A et al (2006) GENCODE: producing a reference annotation for ENCODE. Genome Biol 7(Suppl 1):S4.1–S4.9. doi:10.1186/gb-2006-7-s1-s4

Hirobe T (2011) How are proliferation and differentiation of melanocytes regulated? Pigment Cell Melanoma Res 24:462–478. doi:10.1111/j.1755-148X.2011.00845.x

Ho H, Milenković T, Memisević V et al (2010) Protein interaction network topology uncovers melanogenesis regulatory network components within functional genomics datasets. BMC Syst Biol 4:84. doi:10.1186/1752-0509-4-84

Hoggart CJ, Shriver MD, Kittles RA et al (2004) Design and analysis of admixture mapping studies. American J Hum Genet 74:965–978. doi:10.1086/420855

Jablonski NG, Chaplin G (2000) The evolution of human skin coloration. J Hum Evol 39:57–106. doi:10.1006/jhev.2000.0403

Jablonski NG, Chaplin G (2010) Human skin pigmentation as an adaptation to UV radiation. Proc Natl Acad Sci 107(Suppl 2):8962–8968. doi:10.1073/pnas.0914628107

Johansson A, Gyllesnsten U (2008) Identification of local selective sweeps in human populations since the exodus from Africa. Hereditas 145:126–137. doi:10.1111/j.2008.0018-0661.02054.x

Jost M, Kari C, Rodeck U (2000) The EGF receptor–an essential regulator of multiple epidermal functions. Eur J Dermatol 10:505–510

Kent WJ, Sugnet CW, Furey TS et al (2002) The Human Genome Browser at UCSC. Genome Res 12:996–1006

Klimentidis YC, Miller GF, Shriver MD (2009) Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among Hispanics and Native Americans. Am J Phys Anthropol 138:375–383. doi:10.1002/ajpa.20945

Lamason RL, Mohideen M-APK, Mest JR et al (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science 310:1782–1786. doi:10.1126/science.1116238

Lao O, de Gruijter JM, van Duijn K et al (2007) Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. Ann Hum Genet 71:354–369. doi:10.1111/j.1469-1809.2006.00341.x

Li JZ, Absher DM, Tang H et al (2008) Worldwide human relationships inferred from genome-wide patterns of variation. Science (New York) 319:1100–1104. doi:10.1126/science.1153717

Myles S, Tang K, Somel M et al (2008) Identification and analysis of genomic regions with large between-population differentiation in humans. Ann Hum Genet 72:99–110. doi:10.1111/j.1469-1809.2007.00390.x

Norton HL, Kittles RA, Parra E et al (2007) Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. Molecular Biol Evol 24:710–722. doi:10.1093/molbev/msl203

Pickrell JK, Coop G, Novembre J et al (2009) Signals of recent positive selection in a worldwide sample of human populations. Genome Res 19:826–837. doi:10.1101/gr.087577.108

Relethford JH (1997) Hemispheric difference in human skin color. Am J Phys Anthropol 104:449–457. doi:10.1002/(SICI)1096-8644(199712)104:4<449:AID-AJPA2>3.0.CO;2-N

Sabeti PC, Reich DE, Higgins JM et al (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837. doi:10.1038/nature01140

Sabeti PC, Varilly P, Fry B et al (2007) Genome-wide detection and characterization of positive selection in human populations. Nature 449:913–918. doi:10.1038/nature06250

Schlötterer C, Dieringer D (2005) A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In: Nurminsky D (ed) Selective Sweep. Kluwer Academic/Plenum Publishers, New york, pp 55–64

Shriver MD, Kennedy GC, Parra EJ et al (2004) The genomic distribution of population substructure in four populations using 8, 525 autosomal SNPs. Human Genomics 1:274–286

Simonin F, Gavériaux-Ruff C, Befort K et al (1995) kappa-Opioid receptor in humans: cDNA and genomic cloning, chromosomal assignment, functional expression, pharmacology, and expression pattern in the central nervous system. Proc Natl Acad Sci USA 92:7006–7010

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

The International HapMap Consortium (2003) The International HapMap Project. Nature 426:789–796. doi:10.1038/nature02168

Thong HY, Jee SH, Sun CC, Boissy R (2003) The patterns of melanosome distribution in keratinocytes of human skin as one determining factor of skin colour. Br J Dermatol 149:498–505

Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS biology 4:e72. doi:10.1371/journal.pbio.0040072