Hum Genet (2007) 120:613–621
DOI 10.1007/s00439-006-0256-4

ORIGINAL INVESTIGATION

Identifying genes underlying skin pigmentation differences among human populations

Sean Myles · Mehmet Somel · Kun Tang · Janet Kelso · Mark Stoneking

Received: 6 July 2006 / Accepted: 28 August 2006 / Published online: 15 September 2006

Abstract Skin pigmentation is a human phenotype that varies greatly among human populations and it has long been speculated that this variation is adaptive. We therefore expect the genes that contribute to these large differences in phenotype to show large allele frequency differences among populations and to possibly harbor signatures of positive selection. To identify the loci that likely contribute to among-population human skin pigmentation differences, we measured allele frequency differentiation among Europeans, Chinese and Africans for 24 human pigmentation genes from 2 publicly available, large scale SNP data sets. Several skin pigmentation genes show unusually large allele frequency differences among these populations. To determine whether these allele frequency differences might be due to selection, we employed a within-population test based on long-range haplotype structure and identified several outliers that have not been previously identified as putatively adaptive. Most notably, we identify the *DCT* gene as a candidate for recent positive selection in the Chinese. Moreover, our analyses suggest that it is likely that different genes are responsible for the lighter skin pigmentation found in different non-African populations.

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00439-006-0256-4 and is accessible for authorized users.

S. Myles (✉) · M. Somel · K. Tang · J. Kelso · M. Stoneking
Department of Evolutionary Genetics,
Max Planck Institute for Evolutionary Anthropology,
Deutscher Platz 6, 04103 Leipzig, Germany
e-mail: myles@eva.mpg.de

Keywords Positive selection · Skin pigmentation · Fst · HapMap · Perlegen · SNP

Introduction

Uncovering the genetic causes of human phenotypic variation has been a main goal of human genetic research and is of great medical interest. Most commonly, within-population analyses (e.g. linkage and association studies) are employed to identify regions of the genome and/or genetic variants that associate with a particular phenotype. This approach is expected to be fruitful since roughly 85% of human genetic variation is found within populations and only 15% is found among populations (Lewontin 1972). For phenotypes that differ greatly among human populations, however, traditional within-population analyses may overlook many potential links between human genetic and phenotypic variation, as the genes that influence phenotypic variation within a population may differ from those that influence phenotypic variation among populations.

As humans recently expanded within and out of Africa, they encountered novel physical and cultural environments to which they likely had to adapt by local natural selection. Genetic variants that contribute to locally adaptive phenotypes are expected to show large between-population differentiation when compared to other loci (Beaumont and Balding 2004; Pollinger et al. 2005; Sabeti et al. 2006). It is often these loci, and the phenotypes associated with them, which are of most interest to evolutionary biologists and anthropologists and which may be overlooked by traditional association and linkage studies. In addition, it is possible, if

not likely, that many loci that have undergone local natural selection, and therefore show large among-population differentiation, are associated with disease [e.g. the thrifty gene hypothesis (Neel 1962)] or disease resistance, since disease is a potentially strong selective force (e.g. Tishkoff et al. 2001).

Skin pigmentation is a human phenotype that is known to differ greatly among human populations and it has long been speculated that skin pigmentation differences among human populations are adaptive (Darwin 1871). For example, while variation in human craniometric measures is similar to human genetic polymorphisms (i.e. 81% within-population variation), skin pigmentation shows the opposite pattern (i.e. 88% among-population variation), which suggests that local selection is responsible (Relethford 2002). Thus, we expect loci that contribute to these among-population differences in skin pigmentation to show large allele frequency differences among populations. Moreover, it is likely that such loci would harbor signatures of local selection.

To identify the genetic loci that likely underlie the among-population differences in human skin pigmentation, we calculated Fst values for several human pigmentation genes from the Perlegen (Hinds et al. 2005) and Hapmap (The International HapMap Consortium 2005) data sets. Our hypothesis is that skin pigmentation genes with large among-population variation (i.e. high Fst values) are strong candidate genes to account for among-population skin pigmentation differentiation. High Fst values alone may indicate which genes are responsible for the among-population differentiation in the phenotype, but only suggest that the change in allele frequency is adaptive. We therefore use a within-population statistic based on long-range haplotype structure to detect any signature of recent positive selection on candidate genes with high Fst values. We find patterns of variation consistent with positive selection at most of the previously identified adaptive skin pigmentation genes, and provide evidence of selection at some novel candidates as well.

Materials and methods

Human pigmentation genes

A list of human pigmentation genes was obtained from Sturm et al. (2001) and Tomita and Suzuki (2004). Start positions, end positions and RefSeq IDs of the genes for build 34 and build 35 of the human genome were obtained by querying the kgXref table of the UCSC human genome browser. In cases where the gene name from the literature was linked to more than one entry in the kgXref table, the longest entry (in bp) was chosen. For the MITF and PAX3 genes, the query produced much larger genomic regions for build 35 than for build 34. To ensure that the regions remained of roughly equal size between builds, we assumed that the annotation was more accurate for build 35 and therefore used the liftover tool (http://www.genome.ucsc.edu/cgi-bin/hgLiftOver) to get the build 34 start and end positions from the build 35 positions for these two genes.

Polymorphism data

The raw genotype data from the Perlegen data set (Hinds et al. 2005) were downloaded from the Perlegen website (http://www.genome.perlegen.com/browser/download.html). The raw genotype data from the HapMap phase 2 (Rel #20, Jan 06, build 35) and phase 1 (Rel#16c.1, Oct 05, build 34) data sets (The International HapMap Consortium 2005) were downloaded from the HapMap website (http://www.hapmap.org/downloads/index.html.en). We excluded the X and Y-chromosomes because their Fst and riEHH values are not comparable to those of the autosomes. Only unrelated individuals from these data sets were included. Thus, the Perlegen data set includes 24 European Americans, 23 African Americans and 24 Han Chinese Americans. The HapMap data sets include 60 European Americans, 60 Yoruba from Nigeria and 45 Han Chinese; the Japanese HapMap data were not included in order to make the HapMap and Perlegen data more comparable. Here we use the following abbreviations: ''Eur'' for European Americans from both data sets; ''Chn'' for Han Chinese from both data sets; and ''Afr'' for African Americans and Yoruba. Analyses were performed with the statistical programming language R (http://www.r-project.org).

Fst calculation

Population pairwise Fst comparisons are denoted as follows: Afr–Eur, Afr–Chn, Eur–Chn. For each population pairwise Fst comparison, SNPs were removed that were fixed for the same allele in the two populations because they are noninformative (Weir and Cockerham 1984). Thus, the number of informative SNPs in the Perlegen data set for each Fst comparison is as follows: Afr–Eur = 1,521,010; Afr–Chn = 1,505,181; Eur–Chn = 1,337,863; threeway = 1,548,308. For the HapMap phase 2 data: Afr–Eur = 2,792,017; Afr–Chn = 2,736,308; Eur–Chn = 2,364,200; threeway = 2,862,817. A weighted average Fst was calculated for

each gene according to equation (10) in Weir and Cockerham (1984).

LRH test

We employ a long-range haplotype (LRH) statistic similar to the iHS statistic introduced by Voight et al. (2006) by using the extended haplotype homozygosity (EHH) measure proposed by Sabeti et al. (2002). The LRH test is performed on each population separately as follows. For both alleles at each SNP, the integral of the decay of the extended haplotype homozygosity (iEHH) is calculated. This is done in both directions from the SNP until EHH = 0.05. For each SNP, the relative iEHH is calculated as

$$\text{riEHH} = \frac{\text{iEHH}_a}{\text{iEHH}_b},$$

where the subscripts a and b are the two alleles at the SNP.

The phased SNP data from HapMap Phase 1 were used for the LRH test. Singletons and SNPs whose EHH decay plot contained gaps greater than 50 kb were removed from the analysis. The riEHH values were assigned to bins according to major allele frequency, ranked within each bin, and a corresponding percentile value was assigned to each SNP. The total number of SNPs analysed were: 775,110 (Afr), 730,213 (Eur), 669,212 (Chn).

Many of the genes extend over such short distances, and thus overlap with so few SNPs, that the riEHH statistic is unlikely to capture useful information. However, the signature of recent positive selection captured by LRH-based tests can extend over several 100 kb (Bersaglieri et al. 2004; Sabeti et al. 2002; Voight et al. 2006). Therefore, for the 19 genes that span less than 100 kb, we extended the boundaries equally in both directions until a region of 100 kb was reached. We define ''extreme'' riEHH values as those within the top and bottom 2.5% of the percentile distribution. For each gene region, the ''LRH statistic'' is defined as the proportion of SNPs with extreme riEHH values.

Recombination rates

Sex-averaged recombination rate estimates for the autosomes were obtained from the deCode pedigree-based genetic map (Kong et al. 2002). The recombination rate for a gene is given by the recombination rate of the 1-Mb window that overlaps with the middle position of the gene. The MC1R gene is not found within any 1-Mb window, and we therefore used the nearest recombination rate estimate.

P value calculation

To obtain *P* values for the Fst and LRH statistics, we resampled from the full SNP data sets while controlling for SNP density and recombination rate. To obtain a random SNP density distribution for a gene or gene region of length *X*, 1,000 regions of length *X* were sampled at random and the SNP density distribution from these 1,000 iterations was used as the random SNP density distribution. The recombination rate distribution consisted of all the recombination rate estimates from 1-Mb windows along the autosomes. We controlled for SNP density and recombination rate as follows. For a region of length *X*, 1,000 random regions of length *X* were sampled whose SNP densities fell within ±5% of the observed SNP density in the random SNP density distribution, and whose recombination rates fell within ±5% of the observed recombination rate in the recombination rate distribution. *P* values (p_cor) were calculated by comparing the observed Fst and LRH statistics to the distribution of values from the 1,000 random regions.

It has previously been demonstrated that genic SNPs have higher Fst values than nongenic SNPs (Hinds et al. 2005) and that genic regions contain an enrichment of SNPs with extreme iHS values (Voight et al. 2006), the equivalent of our riEHH values. Therefore, sampling at random from the full SNP data, which includes mostly nongenic regions, may artificially inflate *P* values. To control for this effect, genic regions were defined as follows. The ''refGene'' table was downloaded from the UCSC genome table browser to obtain Refseq IDs and positions for builds 34 and 35. Refseq genes were defined by the region between the transcription start and end positions. We defined 16,697 and 16,700 genic regions for builds 34 and 35, respectively, by identifying clusters of overlapping Refseq genes.

Average Fst values were calculated for all genic regions. A genic region size distribution was created by ordering the genic regions according to size. For each gene, *P* values (p_genic) were then obtained by comparing the observed Fst value to Fst values of genic regions whose size was within ±5% of the observed gene size and which contained at least one SNP. This resulted in at least 1,000 comparisons for each gene.

The LRH test was also performed on all genic regions. Genic regions <100 kb were extended in both directions until 100 kb was reached. For each gene, *P* values (p_genic) were obtained by comparing the

observed LRH statistic to genic regions ≥100 and ≤400 kb whose SNP density was within ±5% of the observed SNP density. These criteria resulted in at least 1,000 comparisons for most genes with a minimum of 924 comparisons.

Results

A list of the human pigmentation genes analyzed in the present study is presented in Table 1. The sizes of the genes range from 2.4 to 344 kb and thus the number of SNPs found within the genes differs dramatically among loci. The number of SNPs found within each gene is listed in Table S1. Two loci (HPS6 and POMC) did not overlap with any Perlegen SNPs and Fst values therefore could not be obtained for these two loci from the Perlegen data.

The average Perlegen and HapMap Fst values for the pigmentation genes are highly correlated ($r = 0.79$, $P < 1 \times 10^{-15}$). Several of the genes show unusually high Fst and LRH values when compared to the genome-wide distributions of these values. This information is captured in the two P values calculated for the Fst and LRH statistics, one controlling for SNP density and recombination rate (p_cor) and the other calculated only from genic regions (p_genic). Overall, the two P values are highly correlated for both the Fst values ($r = 0.98$, $P < 1 \times 10^{-15}$) and the LRH values ($r = 0.93$, $P < 1 \times 10^{-15}$). Figure 1 shows Fst values and LRH statistics for the 11 genes for which at least one of the two P values was significant at $P < 0.05$. The HPS6 gene did not overlap with any Perlegen SNPs and is therefore shown as NA in Fig. 1. Fst values, LRH values and their associated P values for each gene are provided in Table S1.

Several of the genes shown in Fig. 1 have been previously identified as showing a signature of local selection, as discussed below. However, several novel candidates appear in Fig. 1 as well. Most of these are significant for only Fst values or only for the LRH test, as discussed below; however, one new candidate, *DCT*, exhibits significant Fst values for the Afr–Chn and Eur–Chn comparisons in both the Perlegen and HapMap datasets, and moreover shows a significant LRH statistic in the Chinese sample. These results suggest that *DCT* may play a role in skin pigmentation differences between Chinese and other groups, and may have been subject to local selection in Chinese.

To further investigate the signature of selection at *DCT*, we investigated the long-range haplotype structure at this locus in more detail by generating a bidirectional haplotype branching diagram (Fig. 2) as described in Tang et al. (2004). We first chose a core

Table 1 List of 24 human pigmentation genes and their chromosomal locations

Locus name	Refseq_ID	Chromosome	Start positions		End positions		Length (bp)	
			Build 34	Build 35	Build 34	Build 35	Build 34	Build 35
AP3B1	NM_003664	5	77,382,223	77,333,906	77,674,601	77,626,284	292,378	292,378
ASIP	NM_001672	20	33,563,846	32,311,831	33,572,824	32,320,809	8,978	8,978
DCT	NM_001922	13	92,789,843	93,889,841	92,829,924	93,929,924	40,081	40,083
DTNBP1	NM_032122	6	15,631,019	15,631,017	15,771,250	15,771,250	140,231	140,233
F2RL1	NM_005242	5	76,198,926	76,150,609	76,215,212	76,166,895	16,286	16,286
HPS1	NM_000195	10	99,840,542	100,165,945	99,871,291	100,196,694	30,749	30,749
HPS3	NM_032383	3	150,168,279	150,330,068	150,212,214	150,374,003	43,935	43,935
HPS4	NM_022081	22	25,171,999	25,172,000	25,204,374	25,204,374	32,375	32,374
HPS5	NM_007216	11	18,264,525	18,256,792	18,308,030	18,300,297	43,505	43,505
HPS6	NM_024747	10	103,489,733	103,815,136	103,492,379	103,817,782	2,646	2,646
MATP	NM_016180	5	33,990,222	33,980,477	34,030,281	34,020,537	40,059	40,060
MC1R	NM_002386	16	89,727,256	88,512,526	89,729,615	88,514,885	2,359	2,359
MITF	NM_198159	3	69,721,702	69,871,322	69,950,556	70,100,176	228,854	228,854
MLPH	NM_024101	2	238,682,679	238,177,929	238,749,386	238,244,636	66,707	66,707
MYO5A	NM_000259	15	50,321,365	50,392,602	50,537,303	50,608,539	215,938	215,937
OCA2	NM_000275	15	25,602,391	25,673,627	25,946,825	26,018,061	344,434	344,434
PAX3	NM_181457	2	223,267,145	222,890,111	223,366,239	222,989,205	99,094	99,094
POMC	NM_000939	2	25,358,256	25,295,372	25,366,094	25,303,210	7,838	7,838
RAB27A	NM_004580	15	53,211,857	53,283,093	53,278,641	53,349,877	66,784	66,784
SILV	NM_006928	12	54,634,156	54,634,156	54,646,093	54,646,093	11,937	11,937
SLC24A5	NM_205850	15	46,129,224	46,200,460	46,150,644	46,221,880	21,420	21,420
SOX10	NM_006941	22	36,611,358	36,692,819	36,623,578	36,705,039	12,220	12,220
TYR	NM_000372	11	88,599,192	88,550,687	88,716,979	88,668,474	117,787	117,787
TYRP1	NM_000550	9	12,683,448	12,683,448	12,700,258	12,700,249	16,810	16,801

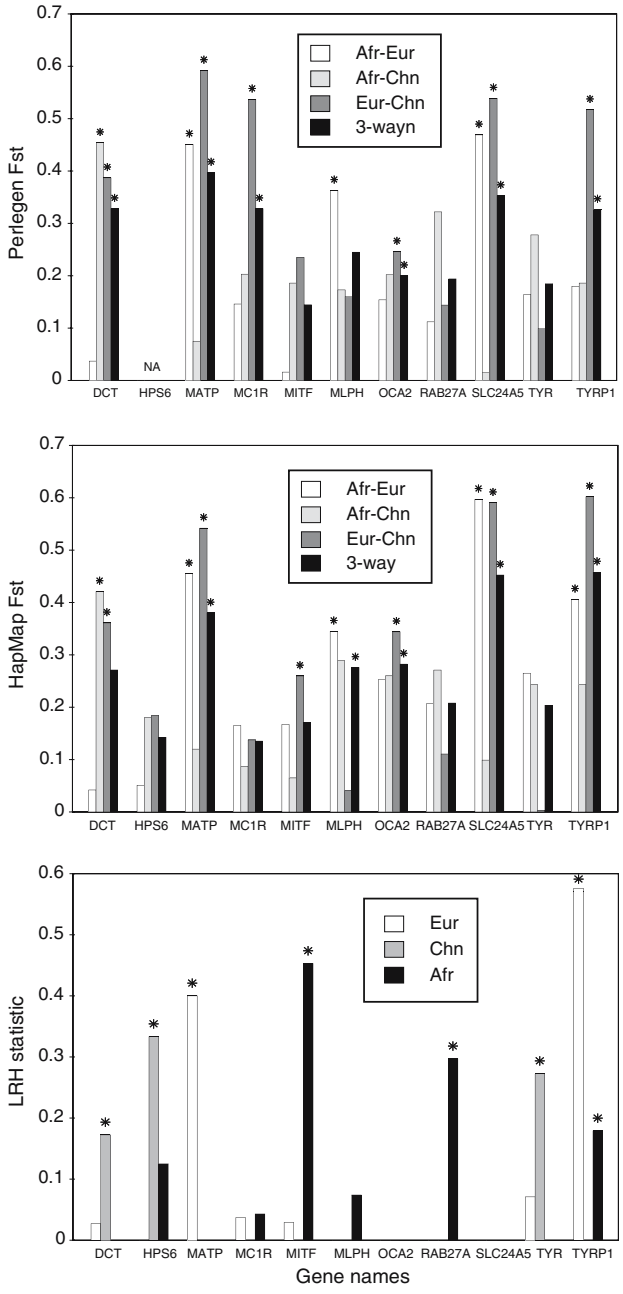

Fig. 1 Results of the Fst and LRH tests for the 11 skin pigmentation genes for which at least one Fst value was significant at $P < 0.05$ (see Materials and methods). Significance at $P < 0.05$ is indicated by an *asterisk*

SNP within the *DCT* gene (indicated by the black circles in Fig. 2) from which the breakdown of haplotype homozygosity is portrayed bidirectionally, both distal and proximal to the core SNP. The core SNP was chosen as the SNP within the *DCT* gene with the most extreme riEHH value. This SNP, rs2031526, has a riEHH value within the top 1.7% of the riEHH distribution and has Fst values within the top 0.6 and 0.3% of the Afr–Chn and Eur–Chn HapMap Fst distributions respectively (AC Fst = 0.718; EC Fst = 0.607). Figure 2 demonstrates that the derived A allele in the Chinese is found on a high frequency haplotype with long-range homozygosity, while the Europeans and Africans show substantial breakdown of homozygosity over the same physical distance on the high frequency, ancestral G allele haplotype. Similar results were obtained when other core SNPs with extreme riEHH values from the *DCT* gene were used (data not shown). Thus, the *DCT* gene harbors a signature of local positive selection in Chinese using both Fst and LRH-based tests, and is therefore a potential candidate to account for the differences in skin pigmentation between the Chinese and other human populations.

Discussion

While association and linkage studies often fail to consider among-population genetic and phenotypic variation, Fst-based approaches may be useful in identifying loci that contribute to phenotypes that differ greatly among human populations. Skin pigmentation is a human phenotype that shows much more among-group variation than most phenotypic traits and thus we expect large allele frequency differences in the genes that contribute to the among-group variation in skin pigmentation. Furthermore, skin pigmentation may be locally adaptive and thus the large among-population differentiation may be the result of positive local selection. Here we identify many genes that likely explain among-population skin pigmentation variation and show signatures of local positive selection.

It is clear from Fig. 1 that the Fst values from the Perlegen and Hapmap data sets are highly similar. Only the MC1R locus stands out as an exception (see Fig. 1). This discrepancy is largely due to the difference in frequency between the Perlegen and Hapmap Chinese samples at SNP rs3212363 (Perl–Chn = 0.095; Hap–Chn = 0.367). The low frequency in the Perlegen Chinese sample (freq = 0.095) compared to the Perlegen European sample (freq = 0.761) results in a high Fst value for the Eur–Chn (Fst = 0.61) and the three-way (Fst = 0.403) comparisons at this SNP, while the intermediate frequency in the Hapmap Chinese sample (freq = 0.367) produces near average Fst values at this SNP (Eur–Chn Fst = 0.169; threeway Fst = 0.109). If this difference between the Perlegen and Hapmap data sets is not due to genotyping error, then it highlights the possible influence of within-population heterogeneity of allele frequencies on population-based studies.

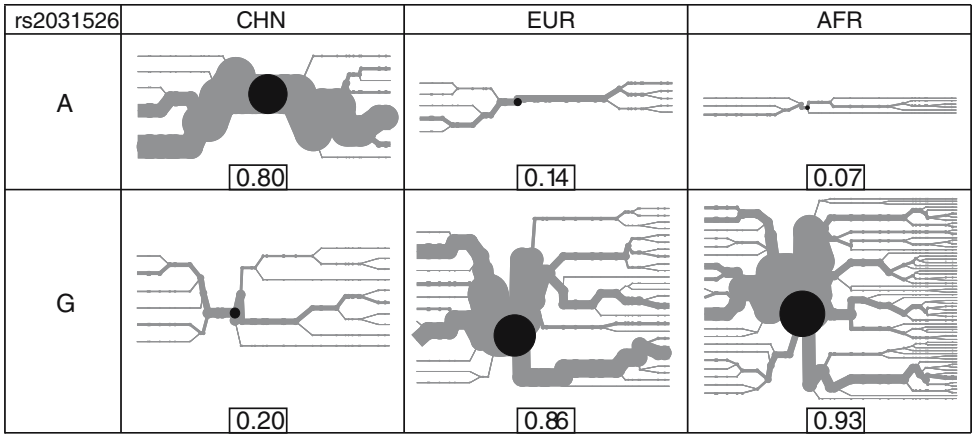

Fig. 2 Haplotype branching diagram of the *DCT* gene for Chinese, European and Africans. The "core" of the diagram is the SNP rs2031526 and is indicated by the *black circle*. The allele frequencies at this SNP are indicated in the *box* below each branching diagram. The bi-directional breakdown of haplotype homozygosity of the derived A allele and the ancestral G allele are shown in the first and second rows, respectively. The thickness of the branches is directly proportional to the frequency of the long-range haplotype

Although several variants within the MC1R locus are clearly associated with skin pigmentation variation within populations (for reviews see Makova and Norton 2005; Rees 2003; Sturm et al. 2001), it is unlikely that they account for among-population pigmentation variation: extensive sequencing in several human populations have failed to find high levels of among-population differentiation or strong evidence of positive local selection at MC1R (Harding et al. 2000; Makova et al. 2001; Rana et al. 1999). Thus, it is possible, if not likely, that many genes that contribute to phenotypic variation within populations are not the same genes that contribute to among-population phenotypic differences.

Several of the genes presented in Fig. 1 have previously been shown to harbor signatures of local positive selection. For example, the *MATP* and *SLC24A5* genes both harbor derived nonsynonymous alleles at high frequency in Europeans, show unusually large allele frequency differences between Europeans and other populations, and have reduced haplotype diversity: patterns consistent with the action of recent natural selection on these genes in Europeans (Lamason et al. 2005; Soejima et al. 2006). The results from Fig. 1 for *MATP* are consistent with this scenario: Fst values are significantly high for all comparisons involving the Europeans and the LRH statistic is also significantly high within Europeans. For *SLC24A5*, although the Fst results are consistent with positive selection acting in Europeans, the LRH statistic is not. The reason for this is the lack of variation in *SLC24A5* in Europeans: there are only nine SNPs with minor allele frequencies >0.05 in the Europeans in the Hapmap data set for the 100 kb region surrounding this locus. Because the LRH test used here compares the breakdown of haplotype homozygosity of one allele against that of the other allele within a population, the power of this approach is severely limited in regions of low variation and selective sweeps causing extremely low diversity within a population are thus unlikely to be detected (Voight et al. 2006). However, a novel between-population LRH-based test in which the breakdown of haplotype homozygosity at a SNP in one population is compared to the same SNP in a different population does detect *SLC24A5* as an outlier (K. Tang and M. Stoneking, in preparation).

Voight et al. (2006) used an LRH-based test of selection on the Hapmap data set and identified four additional skin pigmentation genes (*OCA2, MYO5A, DTNBP1*, and *TYRP1*) that exhibit signals of positive selection in Europeans. Contrary to their results, *DTNBP1* does not contain any SNPs with extreme riEHH values in our analysis. In addition, *DTNBP1* shows little differentiation among populations according to Fst (see Table S1). *MYO5A* also does not appear among the significant genes in our analysis, but the LRH statistic is nearly significant in Europeans (p_cor = 0.053, p_genic = 0.070). However, Fst values at *MYO5A* are not unusually high (see Table S1). Thus, *DTNBP1* and *MYO5A* are unlikely to account for among-population skin pigmentation differences. We find no unusual long-range haplotype structure at *OCA2* but find Fst values that are consistent with selection acting in Europeans: the Afr–Eur Fst values for HapMap and Perlegen both approach significance (HapMap p_cor = 0.067; Perlegen p_cor = 0.073) and

the Eur–Chn and threeway comparisons are both significant in both data sets (Fig. 1). Finally, our data are consistent with *TYRP1* having undergone recent positive selection in Europeans: Fst is high among comparisons involving Europeans and the LRH statistic in Europeans is also significant for this locus. An unusually high degree of European admixture in the African Americans at this locus in the Perlegen data set may be the reason why the Perlegen Afr–Eur Fst is substantially lower than the HapMap Fst value for *TYRP1*. Also, the significant LRH statistic in Africans indicates the possibility that selection has acted at this locus in Africa.

Recently, Izagirre et al. (2006) used several publicly available SNP data sets including European, Asian and African samples to identify signatures of selection from a set of 81 pigmentation loci. From their Fst-based analysis on genes that were examined in the present study, they identified as candidates for positive selection *SLC24A5* in the Eur–Afr comparison and *MATP*, *OCA2*, *TYRP1* and *DCT* in the Eur–Chn comparison. Genes showing a significant signature according to their LRH test include *TYRP1* and *SCL24A5* in Europeans. A large proportion of the SNP data used by Izagirre et al. (2006) came from the Perlegen and Hapmap databases and it is therefore noteworthy that they failed to find significant Fst values for *MATP* in the Afr–Eur comparison and *DCT* in the Afr–Chn comparison. A possible explanation for this discrepancy is the heterogeneity in their sampling strategy: frequency data from different sources were pooled so that, for example, the frequency of the allele in the ''African'' population was sometimes derived from the frequency in African Americans, in some cases from the Yoruba from Nigeria, and other cases from both. The frequency of an allele can differ considerably among groups of individuals within these ''populations''. Thus, the pooling of allele frequencies across data sets makes comparisons within the data set difficult to interpret and may be the reason for the differences observed between our results and those of Izagirre et al. (2006).

Several other genes from Fig. 1 show significant results in at least one comparison. For example, *RAB27A* and *MITF* both have significant LRH statistics in Africans, while *HPS6* and *TYR* have significant LRH statistics in Chinese. Also, *MLPH* and *MITF* have significant Fst values, but the latter only for the Eur–Chn comparison, whereas the LRH statistic is significant in Africans for this locus. Inconsistent signals (i.e., Fst values significant but not one of the relevant LRH statistics, or LRH values significant but not Fst values) may reflect noise in the data, or may reflect the fact that Fst and LRH capture different aspects of the data. In general, if one of the three populations has experienced positive selection at a particular locus, then we may expect a high LRH statistic in this population and high Fst values for comparisons between this population and the other two populations. However, it has been shown that LRH-based tests have low power to detect sweeps near or at fixation and that power is highest for very recent selective events, <10,000-years old (Sabeti et al. 2002; Voight et al. 2006). Fst measures allele frequency differences among populations, which may persist at a locus for longer than the signal detected by the LRH statistic but must post-date the major human migrations out of Africa, some 50,000–75,000 years ago (Sabeti et al. 2006). In a scenario of recent local selection, however, there may not have been enough time to produce extreme allele frequency differentiation at a locus and thus the signal of selection may go unnoticed using Fst-based approaches.

We identify one gene, *DCT*, from which the Fst and LRH statistics are consistent with the action of positive selection in the Chinese (see Fig. 1) and which has not previously been identified as a potentially adaptive skin pigmentation gene in the Chinese. The haplotype branching diagram (Fig. 2) shows extended haplotype homozygosity on a high frequency allele in the Chinese at the *DCT* gene, a signature indicative of positive selection. *DCT*, dopachrome tautomerase, is an enzyme involved in melanin synthesis and is closely related to *TYR* and *TYRP1*, genes in which mutations are associated with pigmentary diseases in humans (del Marmol and Beermann 1996; Toyofuku et al. 2001a, b). *DCT* has not been associated with human pigmentary disease, but *DCT* knockout mice and mice with nonsynonymous mutations at this locus show a lighter coat color phenotype (Costin et al. 2005; Guyonneau et al. 2004). There are no nonsynonymous SNPs in the Perlegen and HapMap data sets within the *DCT* gene and it is therefore difficult to identify a putatively causal allele that may have been the direct target of selection. Additional work, including full resequencing, will be required in order to identify a putatively causal site at this locus.

Numerous adaptive hypotheses have been put forth to explain the variation in skin pigmentation among human populations including resistance to sunburn and skin cancer, vitamin D synthesis, resistance to cold injury, camouflage, protection from parasites and disease, regulation of UV radiation into the skin, sexual selection and relaxation of purifying selection (for a review see Jablonski 2004). As a first step in evaluating these hypotheses, we identify genes that likely contribute to

among-population skin pigmentation differences by searching for genes associated with pigmentation that show unusually large allele frequency differentiation among populations as measured by Fst. Loci that have been under local selection also tend to have unusually large Fst values and thus the genes that we identify here are strong candidates for having experienced local positive selection. Furthermore, we employed a long-range haplotype test to verify the signature of selection at these loci. It is worth noting that Fst and LRH are likely not independent even under neutrality: alleles that have risen quickly to high frequency by drift in one population will also carry linked variants to high frequency by genetic hitchhiking and thus also show long-range haplotype homozygosity. Future studies are required to clarify the relationship between these two statistics.

In conclusion, we identify the *DCT* gene as likely to contain variant(s) that have experienced positive selection in Chinese. Also, we confirm the signature of selection at *MATP*, *SLC24A5*, *OCA2*, and *TYRP1* in Europeans and provide suggestive evidence of selection at several other loci. It is notable that no gene shows a shared signature of selection in Europeans and Chinese, relative to Africans. These results suggest that the lighter skin pigmentation observed in non-African populations is the result of positive selection on different loci in different human populations. The identification and analysis of additional genes involved in human skin pigmentation and the functional characterization of the allelic variants at the candidate loci presented here will help clarify the nature and extent of skin pigmentation adaptation in human populations.

Acknowledgments We thank Ed Green for technical assistance; and David Hughes, Naim Matasci, Susan Ptak and Michael Lachmann for useful discussion. Supported by the Max Planck Society.

References

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol 13:969–980

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN (2004) Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74:1111–1120

Costin GE, Valencia JC, Wakamatsu K, Ito S, Solano F, Milac AL, Vieira WD, Yamaguchi Y, Rouzaud F, Petrescu AJ, Lamoreux ML, Hearing VJ (2005) Mutations in dopachrome tautomerase (Dct) affect eumelanin/pheomelanin synthesis, but do not affect intracellular trafficking of the mutant protein. Biochem J 391:249–259

Darwin C (1871) The descent of man. Princeton University Press, Princeton

del Marmol V, Beermann F (1996) Tyrosinase and related proteins in mammalian pigmentation. FEBS Lett 381:165–168

Guyonneau L, Murisier F, Rossier A, Moulin A, Beermann F (2004) Melanocytes and pigmentation are affected in dopachrome tautomerase knockout mice. Mol Cell Biol 24:3396–3403

Harding RM, Healy E, Ray AJ, Ellis NS, Flanagan N, Todd C, Dixon C, Sajantila A, Jackson IJ, Birch-Machin MA, Rees JL (2000) Evidence for variable selective pressures at MC1R. Am J Hum Genet 66:1351–1361

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR (2005) Whole-genome patterns of common DNA variation in three human populations. Science 307:1072–1079

Izagirre N, Garcia I, Junquera C, de la Rua C, Alonso S (2006) A scan for signatures of positive selection in candidate loci for skin pigmentation in humans. Mol Biol Evol 23:1697–1706

Jablonski NG (2004) The evolution of human skin and skin color. Ann Rev Anthro 33:585–623

Kong A, Gudbjartsson D, Sainz J, Jonsdottir G, Gudjonsson S, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson S, Frigge M, Thorgeirsson T, Gulcher J, Stefansson K (2002) A high-resolution recombination map of the human genome. Nat Genet 31:241–247

Lamason RL, Mohideen M-APK, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE, Sinha S, Moore JL, Jagadeeswaran P, Zhao W, Ning G, Makalowska I, McKeigue PM, O'Donnell D, Kittles R, Parra EJ, Mangini NJ, Grunwald DJ, Shriver MD, Canfield VA, Cheng KC (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science 310:1782–1786

Lewontin RC (1972) The apportionment of human diversity. Evol Biol 6:391–398

Makova K, Norton H (2005) Worldwide polymorphism at the MC1R locus and normal pigmentation variation in humans. Peptides 26:1901–1908

Makova KD, Ramsay M, Jenkins T, Li W-H (2001) Human DNA sequence variation in a 6.6-kb region containing the melanocortin 1 receptor promoter. Genetics 158:1253–1268

Neel JV (1962) Diabetes mellitus: a ''thrifty'' genotype rendered detrimental by ''progress''? Bull WHO 77:694–703

Pollinger JP, Bustamante CD, Fledel-Alon A, Schmutz S, Gray MM, Wayne RK (2005) Selective sweep mapping of genes with large phenotypic effects. Genome Res 15:1809–1819

Rana BK, Hewett-Emmett D, Jin L, Chang BHJ, Sambuughin N, Lin M, Watkins S, Bamshad M, Jorde LB, Ramsay M, Jenkins T, Li W-H (1999) High Polymorphism at the human melanocortin 1 receptor locus. Genetics 151:1547–1557

Rees JL (2003) Genetics of hair and skin color. Annu Rev Genet 37:67–90

Relethford JH (2002) Apportionment of global human genetic diversity based on craniometrics and skin color. Am J Phys Anthropol 118:393–398

Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES (2006) Positive natural selection in the human lineage. Science 312:1614–1620

Soejima M, Tachida H, Ishida T, Sano A, Koda Y (2006) Evidence for recent positive selection at the human AIM1 locus in a European population. Mol Biol Evol 23:179–188

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

Tang K, Wong LP, Lee EJD, Chong SS, Lee CGL (2004) Genomic evidence for recent positive selection at the human MDR1 gene locus. Hum Mol Genet 13:783–797

The International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437:1299

Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drousiotou A, Dangerfield B, Lefranc G, Loiselet J, Piro A, Stoneking M, Tagarelli A, Tagarelli G, Touma EH, Williams SM, Clark AG (2001) Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science 293:455–462

Tomita Y, Suzuki T (2004) Genetics of pigmentary disorders. Am J Med Genet 131C:75–81

Toyofuku K, Wada I, Spritz RA, Hearing VJ (2001a) The molecular basis of oculocutaneous albinism type 1 (OCA1): sorting failure and degradation of mutant tyrosinases results in a lack of pigmentation. Biochem J 355:259–269

Toyofuku K, Wada I, Valencia JC, Kushimoto T, Ferrans VJ, Hearing VJ (2001b) Oculocutaneous albinism types 1 and 3 are ER retention diseases: mutation of tyrosinase or Tyrp1 can affect the processing of both mutant and wild-type proteins. FASEB J 15:2149–2161

Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biol 4:e72

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370