Simultaneous Purifying Selection on the Ancestral MC1R Allele and Positive Selection on the Melanoma-Risk Allele V60L in South Europeans

Conrado Martínez-Cadenas,1,2 Saioa López,1,2 Gloria Ribas,3 Carlos Flores,4,5 Oscar García,6 Arrate Sevilla,2 Isabel Smith-Zubiaga,7 Maider Ibarrola-Villaba,3 María del Mar Pino-Yanes,4,5,8 Jesús Gardeazabal,9 Dolores Boyano,10 Alicia García de Galdeano,10 Neskuts Izagirre,2 Concepción de la Rúa,2 and Santos Alonso*,2

1Department of Medicine, University of Castellón, Castellón, Spain
2Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country, Bizkaia, Spain
3INCLIVA Biomedical Research Institute, Valencia, Spain
4Research Unit, University Hospital N.S. de Candelaria, Tenerife, Spain
5CIBER on Respiratory Diseases, ISCIII, Madrid, Spain
6Forensic Science Unit, Forensic Genetics Section, Basque Country Police, Erandio, Bizkaia, Spain
7Department of Zoology and Animal Cell Biology, University of the Basque Country, Bizkaia, Spain
8Department of Medicine, University of California, San Francisco, CA, USA
9Dermatology Service, Cruces Hospital, Bizkaia, Spain
10Department of Cell Biology and Histology, University of the Basque Country, Bizkaia, Spain

†These authors contributed equally to this work.
*Corresponding author: E-mail: santos.alonso@ehu.es.

Abstract

In humans, the geographical apportionment of the coding diversity of the pigmentary locus melanocortin-1 receptor (MC1R) is, unusually, higher in Eurasians than in Africans. This atypical observation has been interpreted as the result of purifying selection due to functional constraint on MC1R in high UV-B radiation environments. By analyzing 3,142 human MC1R alleles from different regions of Spain in the context of additional haplotypic information from the 1000 Genomes (1000G) Project data, we show that purifying selection is also strong in southern Europe, but not so in northern Europe. Furthermore, we show that purifying and positive selection act simultaneously on MC1R. Thus, at least in Spain, regions at opposite ends of the incident UV-B radiation distribution show significantly different frequencies for the melanoma-risk allele V60L (a mutation also associated to red hair and fair skin and even blonde hair), with higher frequency of V60L at those regions of lower incident UV-B radiation. Besides, using the 1000G south European data, we show that the V60L haplogroup is also characterized by an extended haplotype homozygosity (EHH) pattern indicative of positive selection. We, thus, provide evidence for an adaptive value of human skin depigmentation in Europe and illustrate how an adaptive process can simultaneously help to maintain a disease-risk allele. In addition, our data support the hypothesis proposed by Jablonski and Chaplin (Human skin pigmentation as an adaptation to UVB radiation. Proc Natl Acad Sci U S A. 2010;107:8962–8968), which posits that habitation of middle latitudes involved the evolution of partially depigmented phenotypes that are still capable of suitable tanning.

Key words: skin pigmentation, MC1R, adaptive evolution, melanoma.

Introduction

The human melanocortin-1 receptor (MC1R) is an integral membrane G-protein-coupled receptor (GPCR) of skin and hair follicle melanocytes. GPCRs communicate the cells with their environment and upon receiving external stimuli trigger an adaptive response. In this regard, the binding of a small peptide, the α-melanocyte-stimulating hormone (α-MSH) to MC1R is coupled to a cAMP signaling pathway that stimulates eumelanogenesis (synthesis of brown/black melanins). In the absence of MC1R signaling, pheomelanogenesis (synthesis of yellow/red melanins) is the default pathway (Ito 2003). Therefore, MC1R is a regulator of the amount and type of pigment production, and has thus been referred to as a major determinant of skin phototype (García-Borrón et al. 2005). In this regard, some nonfunctional MC1R variants lead to phenotypes characterized by red hair, fair skin, freckles, and poor tanning ability (the red hair and fair skin phenotype, RHC) in a dominant manner. These have
been called R alleles, and include D84E, R151C, R160W, and D294H. Variants that do not lead to a total loss of function and that have weak or no association to the RHC phenotype (NRHC phenotype) are called r alleles, and include, for instance, V60L, V92M, and R163Q (Valverde et al. 1995; Box et al. 1997; Sturm et al. 2003; Duffy et al. 2004; Beaumont et al. 2007; Dessinioti et al. 2011).

As skin phototype is a risk factor for melanoma (Abdel-Malek and Ito 2013), considerable effort has been dedicated to find association between variants at MC1R (and other pigmentation genes) to melanoma risk (Valverde et al. 1996; Ichii-Jones et al. 1998; Palmer et al. 2000; Kennedy et al. 2001; Fernandez et al. 2007; Raimondi et al. 2008; Scherer et al. 2009; Williams et al. 2011; Davies et al. 2012). Less effort has been dedicated to understand pigmentation and melanoma risk from an evolutionary perspective. Interestingly, MC1R is highly diverse in humans: a recent literature review cataloged up to 57 nonsynonymous and 25 synonymous variants in approximately 1 kb of coding DNA (Gerstenblith et al. 2007). However, the geographical apportionment of this coding diversity shows that there is an increased diversity in Eurasian populations in comparison with African populations (Rana et al. 1999; Harding et al. 2000), which is an unusual observation, given that Africans typically exhibit greater genomic diversity than Eurasians. This is even more intriguing as the upstream region containing the promoter of MC1R does show the expected, opposite pattern, that is, a higher diversity in Africans compared with Asians and Europeans (Makova et al. 2001).

These patterns of diversity in the coding region of MC1R led both Rana et al. (1999) and Harding et al. (2000) to posit that MC1R is under strong functional constraint in Africa. However, their interpretation for the diversity patterns in Eurasia differed. Thus, while Rana et al. (1999) proposed local selection for mutations that result in lighter pigmentation (diversifying selection) in Eurasia, Harding et al. (2000) proposed instead that the level of MC1R polymorphism simply reflects the expectations of a neutral evolution for this locus in Eurasians.

In this work, we propose an explanation for the high diversity at MC1R and assess the role of selection in one large Spanish population sample in the European context. We conclude that purifying and positive selection can act simultaneously on the same locus. More interestingly, MC1R illustrates how an adaptive process can simultaneously help maintain a disease-risk allele.

**Results**

After resequencing MC1R in 1,217 healthy Spanish individuals, we detected 38 single-nucleotide polymorphism (SNPs), 6 of which have not been previously reported: V193M, A285T, R229H, Q233X, S302N, and W317C. We did not observe the R307G Neanderthal MC1R mutation described by Laloue-Fox et al. (2007), which adds support to their findings. We also found one new deletion, 520_522delGTC, and one new insertion, 539insC. However, for the subsequent diversity analyses, indel polymorphisms were not considered due to their different nature to SNPs. The most frequent variant was V60L, consistent with other studies on European populations (Flanagan et al. 2000; Fargnoli et al. 2006; Latreille et al. 2009). To this sample, we added those melanoma sequences previously published by some of us (Fernandez et al. 2007; Ibarrola-Villava et al. 2012), which included 354 individuals from Madrid (Spain). The global sample showed a total of 42 SNPs, with a total of 42 different haplotypes observed (see table 1 for a comparative diversity analysis).

We further resequenced MC1R in 127 melanoma patients and detected 15 SNPs. The most frequent variant was also V60L. To this sample, we added those melanoma sequences previously published by some of us (Fernandez et al. 2007; Ibarrola-Villava et al. 2012), which corresponded to 595 individuals. Then, the global melanoma sample showed a total of 37 different haplotypes formed from 28 SNPs, 7 of which did not show up in the control samples ($S41F$, $S41C$, $G89R$, $I180l$, $R213W$, $P268R$, and $N279K$).

In our healthy sample, the most likely root is the most common haplotype. Its likelihood was approximately 2 orders of magnitude higher than the one corresponding to the next most likely rooted tree. The most common haplotype in our sample is also the most common haplotype in the 1000G African sample. Mutation T314T, which lies in a CpG doublet, was removed because it seems to have arisen more than once (within different lineages). Its frequency in the healthy samples is 0.07 (total 2N = 3,142 alleles). Mutation R151C was also discarded because it seems to have arisen several times (within different lineages) in the melanoma samples (frequency in the healthy samples: 0.02).

**Diversity Analysis**

The coding region of MC1R shows a high nucleotide diversity for such a short coding region: We observed an estimated nucleotide diversity $π$ of 0.08%, whereas in a small set of European human exomes the average $π$ was 0.038% (Tennessen et al. 2010). This increased nucleotide diversity seems to be the result of a high mutation rate derived from a CpG content. Figure 1 shows the CpG content of the coding region of MC1R. For this region, but not for the flanking regions, the observed versus expected CpG ratio is higher than 0.65, which qualifies this region as a CpG island (Takai and Jones 2002). This high CpG content is also correlated with the highest $π$, but the lowest Tajima’s D values.

**Evidence for Purifying Selection at MC1R**

There are many mutations at coding MC1R, but these do not reach in general a substantial frequency in our sample (supplementary information S1, Supplementary Material online). An exception could be V60L, an r allele with a frequency of approximately 0.15, which will be dealt as discussed later. Neutrality tests indicate that the low frequency of new mutants can be due to the effect of purifying selection that tends to remove these new mutations from the population. Thus, table 2 shows that Tajima’s $D$ and Fu’s $D$ and $F$ tests are highly significant even under models that incorporate a recent Out-of-Africa demographic history (Gutenkunst et al. 2009; Laval et al. 2010), whereas the combined DHEW
Table 1. Diversity Parameters for the Coding Region of MC1R in Several Populations.

<table>
<thead>
<tr>
<th>Sample</th>
<th>2N^b</th>
<th>Length^c</th>
<th>S^d</th>
<th>No. of Haps^e</th>
<th>h^f</th>
<th>π^g</th>
<th>θw^h</th>
<th>k^i</th>
<th>Tajima's D (P Value)j</th>
</tr>
</thead>
<tbody>
<tr>
<td>This work</td>
<td>3,142</td>
<td>953</td>
<td>42</td>
<td>42</td>
<td>0.517</td>
<td>6.3</td>
<td>4.867</td>
<td>0.599</td>
<td>−2.143 (−0)</td>
</tr>
<tr>
<td>TSI</td>
<td>196</td>
<td>953</td>
<td>10</td>
<td>11</td>
<td>0.562</td>
<td>6.9</td>
<td>1.709</td>
<td>0.659</td>
<td>−1.437 (0.048)</td>
</tr>
<tr>
<td>ALL-AFR</td>
<td>492</td>
<td>950</td>
<td>11</td>
<td>12</td>
<td>0.184</td>
<td>2.1</td>
<td>1.624</td>
<td>0.192</td>
<td>−1.913 (0.001)</td>
</tr>
<tr>
<td>ASW</td>
<td>122</td>
<td>950</td>
<td>10</td>
<td>11</td>
<td>0.325</td>
<td>3.7</td>
<td>1.860</td>
<td>0.353</td>
<td>−2.014 (0.001)</td>
</tr>
<tr>
<td>LWK</td>
<td>194</td>
<td>950</td>
<td>4</td>
<td>5</td>
<td>0.071</td>
<td>0.8</td>
<td>0.685</td>
<td>0.072</td>
<td>−1.560 (0.010)</td>
</tr>
<tr>
<td>YRI</td>
<td>176</td>
<td>950</td>
<td>5</td>
<td>6</td>
<td>0.203</td>
<td>2.2</td>
<td>0.870</td>
<td>0.212</td>
<td>−1.477 (0.048)</td>
</tr>
<tr>
<td>ALL-NEU</td>
<td>536</td>
<td>951</td>
<td>10</td>
<td>11</td>
<td>0.674</td>
<td>9</td>
<td>1.458</td>
<td>0.854</td>
<td>−0.870 (0.210)</td>
</tr>
<tr>
<td>CEU</td>
<td>174</td>
<td>951</td>
<td>8</td>
<td>9</td>
<td>0.671</td>
<td>9</td>
<td>1.395</td>
<td>0.852</td>
<td>−0.873 (0.215)</td>
</tr>
<tr>
<td>FIN</td>
<td>186</td>
<td>951</td>
<td>7</td>
<td>8</td>
<td>0.663</td>
<td>8.7</td>
<td>1.207</td>
<td>0.827</td>
<td>−0.676 (0.282)</td>
</tr>
<tr>
<td>GBR</td>
<td>176</td>
<td>951</td>
<td>8</td>
<td>9</td>
<td>0.682</td>
<td>11</td>
<td>1.393</td>
<td>0.863</td>
<td>−0.852 (0.224)</td>
</tr>
<tr>
<td>ALL-ASI</td>
<td>572</td>
<td>953</td>
<td>6</td>
<td>8</td>
<td>0.525</td>
<td>9.1</td>
<td>0.866</td>
<td>0.913</td>
<td>0.097 (0.614)</td>
</tr>
<tr>
<td>CHB</td>
<td>194</td>
<td>953</td>
<td>4</td>
<td>5</td>
<td>0.555</td>
<td>10.2</td>
<td>0.685</td>
<td>0.967</td>
<td>0.736 (0.801)</td>
</tr>
<tr>
<td>CHS</td>
<td>200</td>
<td>953</td>
<td>4</td>
<td>5</td>
<td>0.538</td>
<td>10.6</td>
<td>0.681</td>
<td>1.009</td>
<td>0.856 (0.828)</td>
</tr>
<tr>
<td>JPT</td>
<td>178</td>
<td>953</td>
<td>4</td>
<td>6</td>
<td>0.435</td>
<td>7.3</td>
<td>0.695</td>
<td>0.664</td>
<td>0.081 (0.541)</td>
</tr>
</tbody>
</table>

"Samples labeled "This work" correspond to healthy individuals from Spain. The rest correspond to the 1000G Project populations: the south-European (TSI) population (the south-European IBS population from 1000G was not considered due to its low sample size; 2N=28); the African populations (individually: ASW, LWK, YRI, and all three pooled: ALL-AFR), the north-European populations (individually: CEU, FIN, GBR, the three pooled: ALL-NEU), and the Asian populations (individually: CHB, CHS, JPT, and all three pooled: ALL-ASI)."

"2N = Number of haplotypes.

"Nucleotides considered after removing coding synonymous polymorphisms. In the case of south European samples, we excluded T314T, caused by a polymorphism in nucleotide 942; in the case of the north European samples, we excluded I264I, T314T, and S316S, caused by polymorphisms at nucleotides 318, 531, and 942, respectively. In the African samples, we excluded L106L, T177T, F300F, and T314T, caused by polymorphisms at nucleotides 318, 531, 900, and 942, respectively. In the Asian samples, we excluded A103A, Y298Y, and T314T caused by polymorphisms at nucleotides 309, 894, and 942, respectively."

"Number of segregating sites.

"Number of different haplotypes observed in the sample.

"Haplotype diversity.

"Nucleotide diversity (×10^-4).

"Theta (per sequence) estimated from the number of segregating sites.

"Average number of nucleotide differences between haplotypes.

"Tajima's D value (significance inferred after 5,000 standard coalescent simulations with DnaSP 5.1)."

**FIG. 1.** Sliding window graph of polymorphism and CpG values around MC1R. The graph shows the values obtained for different parameters in sliding windows (of size 500 bp, shifted 100 bp) across the region. On top, the coding region (broad gray box) and UTRs (gray lines) of MC1R. Below this, the continuous line indicates that the GC content across the entire region is high (~65%). However, the observed to expected CpG ratio (continuous black line with regularly interspersed dots) substantially increases in the coding region (~0.75). A high content of CpGs, if methylated, could result in a high number of new variants (see Theta W, continuous line with regularly interspersed asterisks), which because of purifying selection result in highly negative Tajima's D values (approximately −2; discontinuous gray line).
test, which is specific of recent positive selection (due to the incorporation of the H test), is not. This is in agreement with the observation that many of the polymorphisms at this region are nonsynonymous and that the most common allele is also the ancestral one.

Using data from the 1000 Genomes (1000G) Project, we observe that this scenario of purifying selection in our Spanish population could be shared with other southern European countries (table 1). Thus, for instance, Tajima’s D test applied to the TSI sample shows a D value of $-1.437$ ($P$ value = 0.0438, under standard coalescent simulations). As expected, a similar pattern of purifying selection can be observed for the pooled 1000G African sample, for whom Tajima’s D is $-1.913$ ($P$ value = 0.001, under standard coalescent simulations), and also for the individual African populations (table 1). However, the pooled northern European sample (or the north European populations individually; table 1) does not share this pattern (Tajima’s D = $-0.870$, $P$ value = 0.21, under standard coalescent simulations). The south European–African purifying selection pattern is not observed in the pooled east Asian sample either (or in any of the individual populations).

Table 2. Neutrality Tests for the Observed MC1R Diversity in the Spanish Sample.

<table>
<thead>
<tr>
<th>Test</th>
<th>Tajima’s D[^a]</th>
<th>Fu and Li’s D[^b]</th>
<th>Fay and Wu’s h[^c]</th>
<th>Fu and Li’s F[^d]</th>
<th>DHEW[^e]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed values</td>
<td>$-2.143$</td>
<td>$0.161$</td>
<td>$-7.300$</td>
<td>$-6.000$</td>
<td>N/A[^f]</td>
</tr>
<tr>
<td>P values, under models</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single, constant-size pop[^g]</td>
<td>$-0.000^*$</td>
<td>$0.396$</td>
<td>$0.000^*$</td>
<td>$0.000^*$</td>
<td>$-0.112$</td>
</tr>
<tr>
<td>Guentekunst et al. (2009)</td>
<td>$-0.0002^*$</td>
<td>$-0.597$</td>
<td>$-0.0001^*$</td>
<td>$-0.0003^*$</td>
<td>$-0.325$</td>
</tr>
<tr>
<td>Laval et al. (2010)</td>
<td>$-0.001^*$</td>
<td>$-0.942$</td>
<td>$-0.0007^*$</td>
<td>$-0.002^*$</td>
<td>$-0.992$</td>
</tr>
</tbody>
</table>

[^a]Tajima’s D test.
[^b]Fay’s and Wu’s normalized H test.
[^c]Fu and Li’s D test.
[^d]Fu and Li’s F test.
[^e]Zeng et al’s combined DHEW test.
[^f]Not applicable.
[^g]No recombination assumed.
[^h]Significant after Bonferroni correction.

Evidence for Positive Selection at V60L in South Europeans

The strong pattern of purifying selection maintaining the ancestral MC1R allele contrasts with a substantial frequency of V60L in our sample. However, the relatively high frequency of V60L is not homogeneously distributed within Spain, as there seems to be some spatial heterogeneity in the frequency of this allele (table 3; supplementary information S2, Supplementary Material online). To explore whether this heterogeneity in the frequency of the V60L allele in Spanish populations could be explained by a general underlying demographic process in the region, we then assessed the existence of population stratification by means of principal component (PC) analyses based on data derived from 93 European ancestry informative markers (termed EuroAIMs), that are well known to recover the largest features of the north-northwest (NNW) to south-southeast (SSE) axis of genetic differentiation in Europe (Price et al. 2008). In fact, analyzing 1,187 individuals from different European populations including the Spanish, EIGENSOFT found significant only the first PC (Tracy–Widom test, $P$ value = $2.06 \times 10^{-35}$), differentiating NNW-Europeans from SSE-Europeans, and leaving the Spanish individuals at intermediate positions in the NNW–SSE axis of European differentiation. However, using this set of EuroAIMs, EIGENSOFT was unable to identify a significant PC (Tracy–Widom test, $P$ value = 0.646 for the first PC) when restricting the analysis to the Spanish populations (including all the Spanish provinces, the Spanish Basque, and the Canary Islanders). To check whether this heterogeneity could be reflecting the geographical UV-B irradiation pattern, we selected two population groups, one including those provinces for which UV-B irradiation values (fig. 2) lie in the lower 5th percentile in the Iberian Peninsula (group 1: Asturias, Bizkaia, Gipuzkoa; all of them northern populations), an another one including those provinces above the 95th percentile of UV-B in the Iberian Peninsula (the Canary Islands,

Table 3. Fisher Exact Test of V60L (rs1805005) Allele Frequency.

<table>
<thead>
<tr>
<th>Group 1 vs. Group 2[^i]</th>
<th>2N Alleles</th>
<th>Frequency V60L Allele</th>
<th>Fisher Exact Test (P Value)[^j]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 1</td>
<td>Group 2</td>
<td>Group 1</td>
</tr>
<tr>
<td>a) UV-B $\leq$ 5th vs. UV-B $\geq$ 95th</td>
<td>188</td>
<td>134</td>
<td>32</td>
</tr>
<tr>
<td>b) UV-B $\leq$ 10th vs. UV-B $\geq$ 90th</td>
<td>266</td>
<td>170</td>
<td>41</td>
</tr>
<tr>
<td>c) UV-B $\leq$ 15th vs. UV-B $\geq$ 85th</td>
<td>348</td>
<td>270</td>
<td>55</td>
</tr>
<tr>
<td>d) UV-B $\leq$ 20th vs. UV-B $\geq$ 80th</td>
<td>466</td>
<td>314</td>
<td>72</td>
</tr>
</tbody>
</table>

[^i]Population groups compared: in a) group 1 includes those provinces whose daily (annual mean) received UV-B radiation is lower than or equal to the 5th percentile of the distribution of UV-B radiation in the Iberian Peninsula. Group 1: Asturias (1), Bizkaia (2), and Gipuzkoa (3) (all of them northern populations); group 2 includes those provinces whose daily (annual mean) received UV-B radiation is greater than or equal to the 95th percentile of the distribution of UV-B radiation in the Iberian Peninsula: the Canary Islands (12), Cadiz (13), and Granada (14). Similarly for b), c), and d) but considering the percentiles indicated. In b) group 1, as in a) plus Cantabria (4), Araba (5), and Lugo (6); group 2 as in a) plus Almeria (15), Huelva (16), and Caceres (17). In c) Group 1, as in b) plus A Coruna (7) and Navarra (8); Group 2 as in b) plus Sevilla (18) and Jalon (19). In d) Group 1 as in c) plus Pontevedra (9), Ourense (10), and La Rioja (11); Group 2 as in c) plus Ciudad Real (17), Malaga (20), and Melilla (21). Only individuals whose grandparents were also born in these provinces were considered. See figure 2 to localize provinces or regions numeric codes on the map.

[^j]Fisher’s exact test one-tailed $P$ values.
Ca´ diz, and Granada; all of them southern populations) plus (group 2), Fisher’s exact test shows that the frequency of V60L in group 1 (0.17) is significantly higher than that in group 2 (0.09) (one-tailed P value = 0.042). This is confirmed if we consider other percentiles (table 3) also. Nevertheless, the distribution of the V60L allele frequency is far from closely following a strictly linear latitudinal pattern. Thus, when we group samples according to their UV-radiation percentile (at noninclusive 0.05 quantiles from the lower and upper ends, respectively), there is actually a negative correlation between the V60L allele frequency and UV radiation (Pearson’s correlation coefficient = −0.165), but this is not significant (t test P value = 0.25). This suggests that other evolutionary processes, including demographic history, may interfere. However, we also have to take into account the low sample sizes for many of the bins, which result in imprecise estimates of allele frequencies (supplementary information S2, Supplementary Material online).

Notwithstanding, the relative high frequency of V60L in the Spanish population could be indicative of some selective advantage. Therefore, we then investigated whether recent positive selection could explain this pattern. For that, we applied the extended haplotype homozygosity (EHH) test to the 1000G southern European sample, the best proxy available, as the TSI population has similar latitude to north Spain. Figure 3 shows that the extent of homozygosity from V60L is in fact longer than expected under the neutral demographic models considered. It is also longer than that observed for the 1000G north European sample set (see supplementary information S3, Supplementary Material online for EHH analysis in single populations). We performed an additional test of selection based on a haplogroup-specific Tajima’s D test on a region of 10 kb S’ of rs1805005 for the set of European rs1805005-T (V60L) haplotypes (see Materials and Methods). Tajima’s D for the rs1805005-T haplogroup in this population was −2.045, whereas for the alternative rs1805005-G haplogroup Tajima’s D was 1.693. The P value under the Gutenkunst et al. (2009) demographic model for the T haplogroup is 0.021 (0.032 under the Laval et al. [2010] model). This indicates that positive selection can be held responsible for the substantial frequency of V60L at least in some areas of Europe.

The Collateral Damage of Positive Selection
It is interesting to note that individuals with melanoma show a significantly higher frequency of V60L than their corresponding controls (Fernandez et al. 2007; Ibarrola-Villava et al. 2012; see also http://www.melgene.org/, last accessed October 3, 2013). We have improved this data set with new melanoma cases and compared it with our Spanish reference set as the control sample. Cochran–Armitage test also shows an association between V60L and melanoma risk in an additive model (table 4) only when considering as controls the set of populations belonging to the provinces with highest UV-B irradiation (above the 95th, 90th, 85th, or 80th percentiles). This suggests that the adaptive evolution of skin depigmentation is concurrently responsible for an increased health risk in, most typically, postreproductive age.
On the other hand, there were no significant differences when using as controls either the set of populations with the lowest UV-B radiation or the whole set of Spanish provinces. This is illustrative of how population structure in allele frequencies can be an issue in genetic association tests, which can lead to false negatives if not properly accounted for.

**Discussion**

We have shown that MC1R shows high coding nucleotide diversity in the (south European) population analyzed. We have proposed that methylation-mediated deamination of 5-methylcytosine is ultimately responsible for the high nucleotide diversity observed at MC1R. The following observations add support to the hypothesis that: 1) CpGs within genes, contrary to regulatory CpG islands, tend to be highly methylated (Jones 1999); 2) methylome analysis of H1 human embryonic stem cells and IMR90 fetal lung fibroblasts shows that MC1R is highly methylated (Lister et al. 2009) (http://neomorph.salk.edu/human_methylome/browser.html, last accessed October 3, 2013); 3) our own bisulfite sequencing data indicate that MC1R is methylated in sperm; 4) in mammalian genomes, CpG sites are hotspots for transition mutations because of methylation-mediated deamination of 5-methylcytosine (Cooper and Krawczak 1993); 5) 88% (44 out of 50) of the SNPs observed in our sample (considering both controls and melanomas) are

**Table 4.** Cochrane–Armitage Association Test for V60L (rs1805005) and Melanoma Risk.

<table>
<thead>
<tr>
<th>Sample</th>
<th>N (inds)</th>
<th>Genotypes</th>
<th>Exact HW Test (P Value)</th>
<th>Cochran–Armitage Trend Test (P Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melanoma</td>
<td>722</td>
<td>GG</td>
<td>494</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GT</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>UV-B (upper 95%)</td>
<td>72</td>
<td>52</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>UV-B (upper 90%)</td>
<td>85</td>
<td>62</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>UV-B (upper 85%)</td>
<td>139</td>
<td>101</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>UV-B (upper 80%)</td>
<td>157</td>
<td>115</td>
<td>28</td>
<td>3</td>
</tr>
</tbody>
</table>

*The T allele is responsible for the V60L amino acid change.

*Tested against the melanoma sample under an additive model.

*Includes individuals from our own sequenced melanoma samples from Castellón (22) plus those of Fernández et al. (2007) and Ibarrola-Villava et al. (2012).

*As in the footnote “d” plus Almería (15), Huelva (16), and Cáceres (17).

*As in the footnote “e” plus Sevilla (18) and Jaén (19).

*As in the footnote “f” plus Málaga (20), Melilla (21), and Ciudad Real (22). See figure 2 for localizing the provinces or regions on the map by using the corresponding numeric codes between parentheses.
transitions; and 6) CpG sites show a mutation rate 10 times higher than non-CpGs (Nachman and Crowell 2000).

This high variability is, however, shaped by contrasting evolutionary forces. It has been argued that environmental pressure is responsible for maintaining the ancestral MC1R allele at high frequency in Africa, whereas in Eurasia instead, this functional constraint is absent (Rana et al. 1999; Harding et al. 2000). However, we show that strong purifying selection detected in our sample, we show that strong functional constraint on the ancestral MC1R haplotype also takes place at least in some areas of southern Europe. This observation might subsequently lead us to think that MC1R has played no role in the adaptive lightening of the human skin in Eurasia. Actually, MC1R variability has been found not to contribute to light pigmentation in mice from the Florida’s Atlantic Coast (Steiner et al. 2009), although it had been highlighted as an adaptive locus in beach mice from Florida’s Gulf Coast (Hoekstra et al. 2006). Similarly, in nonhuman primates, MC1R variability does not seem either to show a simple relationship to coat color (Mundy and Kelly 2003; Haitina et al. 2007). Besides, the presence of a functional, MSH-responsive MC1R in orangutan shows that alternative mechanisms of coat color generation can exist (Haitina et al. 2007).

However, our results indicate that two selective forces shape human MC1R diversity in Southern Europe. Similarly to Africa, purifying selection is strong maintaining the ancestral allele in Southern Europe. But simultaneously, positive selection is responsible for an increased frequency of the V60L allele, a low-penetrance “red hair and fair skin” allele (also associated to blonde hair; Box et al. 1997). The fact that V60L is under positive selection constitutes evidence that MC1R is also playing an active adaptive role in human skin depigmentation, at least for some areas of south Europe. Actually, the world frequency distribution of V60L shows that this allele is circumscribed to Europe and the near east (Gerstenblith et al. 2007; Galore-Haskel et al. 2009; Eskandani et al. 2010; Acar et al. 2012). However, we show that this increase in fitness for the population has collateral damaging consequences for the individual’s health. In this regard, the same variant (V60L) that is under positive selection is also a melanoma-risk allele. As melanoma is typically a postreproductive disease (Sáenz et al. 2005), we can consider that we are postponing the payment due to our adaptation to a reduced-sunlight environment. Does this mean that melanoma, or more generally, skin cancer risk has no fitness consequences itself? We cannot answer that question with our data, because although the ancestral MC1R allele is under strong purifying selection (not only in our sample but also in more demanding regions like Africa), we would also need to assess other protective roles that have been proposed for dark skin, most notably photoprotection of circulating folate (Jablonski and Chaplin 2010). However, the role of MC1R in DNA repair and apoptosis suggests that MC1R is not simply encoding a passive filter (melanin) (García-Borrón et al. 2005). In any case, whatever the fitness forces at play, the final fitness balance seems to favor depigmentation in Europe.

On the other hand, despite proving evidence that depigmentation can be adaptive in relation to scarce sunlight, we however, cannot explicitly say which is the underlying physiological mechanism involved. The vitamin D hypothesis (Loomis 1967) would constitute a straightforward connection. Nevertheless, there is still a gap that needs to be further explored in order for the earlier mentioned hypothesis to be proven.

Finally, it also remains to be known why other R/r mutations are not so common in our population. One hypothesis is that r mutations result in a lighter phenotype but in a way that does not compromise tanning during seasonal increases in UV-B, and thus enabling protection against skin cancer (Robins 1991) or against folate destruction (Jablonski and Chaplin 2010). In this regard, V60L has been shown to achieve significantly lower CAMP levels after agonist stimulation than the wild type MC1R protein, which indicates substantial impaired function, similar to that of the R variants (Schiotz et al. 1999; Beaumont et al. 2007, 2009). This result is supported by genetic data, as the skin color of individuals homozygous for V60L is significantly fairer than wild-type individuals, to the same extent, for instance, as homozygotes for the R allele R160W (Beaumont et al. 2007). However, in heterodimers with the wild-type receptor, V60L does not seem to show (but most R mutants do) dominant-negative effect on wild-type receptor’s cell-surface expression. A dominant-negative effect would result in a decreased ability of the wild-type receptor (the ancestral allele) to elevate intracellular cAMP levels (Beaumont et al. 2007, 2009), and would add to the decreased ability of the V60L allele itself. Thus, generally speaking, V60L would preserve the ability to increase eumelanin production (tanning) during seasonal increases in UV-B better than the R alleles. This would be useful particularly in southern European regions, where summer UV-B radiation is higher than in northern Europe. Thus, in conclusion, our results would support the hypothesis proposed by Jablonski and Chaplin (2010) that “habitation of middle latitudes between approximately 23° and 46° involved the evolution of partially depigmented phenotypes capable of tanning.”

Materials and Methods

Population Samples Sequenced

We analyzed 3,142 human MC1R alleles (1,571 individuals) from different regions of Spain through direct DNA sequencing of the coding region. Of the 1,571 individuals, 962 came from the Spanish National DNA Bank (http://www.bancoadn.org/en/home.htm, last accessed October 3, 2013) and represent different regions in Spain. Of these, 670 individuals were known to have all grandparents from a single specific province of Spain. In addition, we sequenced a further sample of 172 alleles from the Basque Country (86 individuals), 60 alleles (30 individuals) from the Canary Islands (a subset of those described in Pino-Yanes et al. [2011], selected based on the lower ancestry to northern Africa as deduced from 93 EuroAIMs), and 278 alleles (139 individuals) from the province of Castellón. Sequences have been deposited in GenBank with accession numbers KC981760–KC984193. Finally, 708
already sequenced $MC1R$ alleles (354 individuals) from Madrid were recruited from previously published work (Fernandez et al. 2007; Ibarrola-Villava et al. 2012).

Data Compiled from Other Sources

We obtained full $MC1R$ sequence information as strings of SNP genotypes from the 1000G project data (http://www.1000genomes.org/, last accessed October 3, 2013) by means of SPSmart v.5.1.1 (http://spsmart.cesga.es/, last accessed October 3, 2013) (Amigo et al. 2008), which uses the information 1000 Genomes Phase I data (data freeze as of May 2011) for the following populations:

European populations: 268 individuals from northern Europe, comprising Utah residents with northern and western European ancestry from the CEPH collection (CEU, $N = 87$), Finnish from Finland (FIN, $N = 93$), and British from England and Scotland (GBR, $N = 88$); and 112 individuals from southern Europe, comprising Tuscan from Italy (TSI, $N = 98$) and Iberian Spanish (IBS, $N = 14$).

African populations: 246 individuals of African ancestry from southwest USA (ASW, $N = 61$), Luhya in Webuye, Kenya (LWK, $N = 97$), and Yoruba in Ibadan, Nigeria (YRI, $N = 88$).

East Asian populations: 268 individuals from China (CHB), Han Chinese South (CHS), and Japanese in Tokyo, Japan (JPT).

Resequencing

All DNA samples were analyzed for the coding sequence of the $MC1R$ gene by direct automated DNA sequencing. Polymerase chain reaction (PCR) primer sequences used to frame a 1,150-bp fragment containing the entire sequence of the $MC1R$ gene single exon were as follows: 5'-CAGCACCATGAACTAAGCAGCACTT-3' (sense) and 5'-AAGGCTCCGCCTTTACAACATTTACTCACGC-3' (antisense). PCR reactions were carried out using 100 ng of DNA as template, and the PCR reaction mixture also included 2 μl of 10 × PCR buffer, 25 mM MgCl2, 5 μM each of dGTP, dATP, dCTP, and dTTP, 25 pmol of each PCR primer, and 1 unit of AmpliTaq Gold DNA polymerase (Life Technologies). Thermocycling conditions included a denaturation step at 95 °C for 8 min, followed by 35 cycles of 95 °C for 20 s, 56 °C for 45 s, and 72 °C for 1 min, plus a final extension cycle of 72 °C for 10 min. Amplified products were purified using ExoSAP-IT (USB Corporation) according to the manufacturer's instructions. In order to obtain both forward and reverse sequences of the entire $MC1R$ exon, purified PCR products were subjected to four sequencing reactions using forward and reverse sequencing primers: $MC1R$-F1 5'-CAGCACCATGAACTAAGCAGCACTT-3' and $MC1R$-R1 5'-AAGGCTCCGCCTTTACAACATTTACTCACGC-3', $MC1R$-F2 5'-TGGATGAGCTGCACTTTCTAC-3' and $MC1R$-R2 5'-CAGCACCATGAACTAAGCAGCACTT-3', $MC1R$-F3 5'-ATGCCAGAGGTGGTCTCTGACCTGATC-3' and $MC1R$-R3 5'-AAGGCTCCGCCTTTACAACATTTACTCACGC-3' and $MC1R$-R4 5'-AAGGCTCCGCCTTTACAACATTTACTCACGC-3'. Sequencing reactions were performed using 10 ng of PCR purified product (MicroSpin S-400 HR Columns, GE Healthcare) and BigDye Terminator Cycle Sequencing kits (Life Technologies). Sequencing reaction products were then washed with Performa DTR V3 96-well plates (EdgeBio), and they were finally run on an ABI 3130 Genetic Analyzer (Life Technologies). Sequences were edited with Genalyx v2.8 (Takahashi et al. 2003) and further checked manually afterward.

Bisulfite Methylation Analysis

Methylation was assessed using a bisulfate treatment protocol based on Kaneda et al. (2004). In brief, 3 μg of DNA from sperm was digested with Hinfl and incubated with 6 N NaOH for 15 min at 37 °C. To this solution, we added 120 μl of the conversion solution: sodium bisulfite 4.04 M (Sigma), Hydroquinone 10 mM (Sigma), and NaOH 6 N (pH 5), and performed 20 cycles of denaturing for 30 s at 95 °C and incubation for 15 min at 50 °C. DNA was then purified using the QIAquick PCR purification kit (Qiagen), desulfonated by incubation with NaOH 6 N for 5 min at room temperature, and precipitated with ethanol 100% and ammonium acetate 6 M (pH 7) (Sigma).

Converted DNA was amplified by PCR with the following primers: Fw 5'-TGTGGGTCGGGTGTGTTA-3' and Rev 5'-ACCTCCTAAGGTCTTACGC-3', which were designed to anneal specifically to the bisulfite-converted sequence (considering methylation in CpG). PCR conditions were as follows: 3 min at 96 °C; 45 cycles of 30 s at 96 °C, 20 s at 59 °C, and 1 min at 70 °C; and a final step of 10 min at 70 °C. Sequencing reactions were performed using the BigDye Terminator kit 3.1 (Life Technologies) and run on an ABI 310 using the same primers that were used for the PCR.

Probabilistic and Experimental Phasing

Haplotypes were estimated with PHASE (available from: http://stephenslab.uchicago.edu/software.html#phase, last accessed October 3, 2013). For those pairs of SNPs whose phase could not be statistically inferred with a $P$ value > 0.9, we proceeded to infer their phase experimentally by cloning (TOPO-TA kit, Invitrogen) and resequencing again.

Population Genetics Analysis of the $MC1R$ Sequences

Population diversity and neutrality tests parameters were obtained using DnaSP v5.10 (Librado and Rozas 2009) (http://www.ub.edu/dnasp, last accessed October 3, 2013). For the neutrality tests, $P$ values were initially obtained from simulations run under a neutral, constant-sized population model using the program ms (Hudson 2002). Twenty thousand simulations were obtained fixing on $\hat{d}_T$. To take into consideration the possible effect of demography on the neutrality tests, we carried out additional simulations incorporating optimized parameters that attempt to reflect the human evolutionary history. These parameters were obtained
from the models of Gutenkunst et al. (2009) and that of Laval et al. (2010). The ms command lines for the Gutenkunst et al. (2009) model were as follows:

```
/ms 3142 20000 -t 1.922 -l 3 0 3142 0 -n 1 1.68202
-n 2 3.73683 -n 3 7.29205 -g 2 116.010723 -g 3 160.264047 -ma x 0.881098 0.561966 0.881098 x 2.79746 0.561966 2.79746 x -ej 0.028985 3 2 -en 0.028985 2 0.287184 -ema 0.028985 3 x 7.29314 x 7.29314 x x x x x -ej 0.197963 2 1 -en 0.303501 1 1
```

The ms command lines for the Laval et al. (2010) parameters were as follows:

```
/ms 3142 20000 -t 1.922 -l 3 0 3142 0 -n 1 1.666667
-n 2 2.261 -n 3 1.051 -g 1 372.278 -g 2 380.217
-g 3 259.366 -ma x 0.7176 0.7176 0.7176 x 0.7176 0.7176 x -ej 0.0163 3 2 -en 0.0435 2 0.2029
-ej 0.0435 2 1
```

Using these simulations, the P values for the Tajima’s D, Fay’s and Wu’s H (normalized), and Fu’s and Li’s D and F tests were obtained using the software msstats (K. Thornton’s lab, http://www.molpopgen.org/software/msstats/, last accessed October 3, 2013). The DHEW combined test, which includes Tajima’s D, Fay’s and Wu’s H (normalized), and the Ewens and Watterson test were performed with software kindly provided by K. Zeng and modified by us to allow for larger sample sizes. This compound test is more robust against the presence of recombination, is insensitive to background selection and demography, and is thus more powerful in detecting positive selection (Zeng et al. 2007).

The genealogical relationships among MC1R haplotypes were inferred by means of Network v4.6 (http://www.fluxus-engineering.com, last accessed October 3, 2013). We used genetree v9.0 (by R.C. Griffiths, available from: http://www.stats.ox.ac.uk/~griff/software.html, last accessed October 3, 2013) to infer the root of the tree.

Structure Analysis

To explore the genetic structure in the Spanish populations, data from 93 EuroAIMs recovering the largest features of the European NNW to SSE axis of differentiation (Price et al. 2008), from a total of 1,187 individuals, was utilized to assess a PC analysis by means of EIGENSOFT 4.2 (Price et al. 2006). Data finally included in this analysis were derived from individuals known to have all grandparents from a single specific province of Spain, and from reference populations, satisfying a genotyping completion rate more than 90%. We restricted the genotyping and analyses to those 93 SNPs that overlapped with samples used as reference (see Pino-Yanes et al. [2011] for a full description of methods).

Briefly, genotyping was conducted utilizing the iPLEX Gold assay on MassARRAY system (Sequenom, San Diego, CA) by the Spanish National Genotyping Center, Santiago de Compostela Node (CeGen, http://www.cegen.org, last accessed October 3, 2013), and was performed in the following:

- Samples from 553 individuals from 14 Spanish provinces from the Spanish National DNA Bank (www.bancoadn.org, last accessed October 3, 2013): Andalusia (n = 105), Murcia (n = 27), Extremadura (n = 25), Castile-La Mancha (n = 50), Valencia (n = 38), Castile and Leon (n = 93), Madrid (n = 6), Catalonia (n = 55), Galicia (n = 69), Cantabria (n = 18), Navarre (n = 16), La Rioja (n = 21), Asturias (n = 14), and Aragon (n = 16).
- Samples from 24 Spanish Basques satisfying that all their four parental surnames were of Basque origin.

In addition, data from 80 Canary Islanders taken from previous studies (Pino-Yanes et al. 2011), selected based on the lower ancestry to northern Africa, were used for comparisons, along with data from 144 Swedish, 52 Polish, 69 English, 117 Italians, 64 Greeks, and 28 French from Price et al. (2008), as well as from 56 Utah residents with ancestry from northern and western Europe (CEU), all of them also satisfying a genotyping completion rate more than 90%.

EHH Analysis of the 1000G Data

To assess the possibility of positive selection acting on single alleles, we used our own implementation of the EHH method (Sabeti et al. 2002), which uses simulations obtained by means of Hudson’s ms program (Hudson 2002) to estimate the P values. This process was implemented in the following way:

We selected two 100 kb regions, one on each side of the SNP we wanted to test for recent positive selection, and, for each region, we estimated divergence to the orthologous region in the chimpanzee. From divergence (assuming a 6 My divergence time and the Jukes and Cantor model; Li 1997), we estimated the mutation rate μ, which in its turn was used to estimate the parameter θ for both the Gutenkunst et al. (2009) and Laval et al. (2010) models. The recombination parameter ρ was estimated from the genomewide recombination rate from Phase 2 HapMap estimated from phased haplotypes in HapMap Release 22 (NCBI 36) (International HapMap Consortium 2007). In each simulation, the specific θ and ρ values were randomly sampled from a normal distribution with a standard deviation equal to the mean. The demographic models used for the simulations were the same as mentioned in Gutenkunst et al. (2009) and Laval et al. (2010). For each region, the first SNP in the set of simulations was assumed to represent the SNP of interest. Only those simulations for which the frequency pattern of this SNP, taking into account the ancestral or derived state, was equal (±10% allowance) to the observed values in the 1000G European population, were considered for P value estimations. In addition, only those simulations showing a total number of SNPs per 100 kb equal to the observed value (±10% allowance) were considered for EHH estimations.

EHH profiles from the set of simulated haplotypes carrying the (derived) allele under consideration (at the first SNP) were then used to infer P values. Typically, for each region several tens of thousands simulations were needed to obtain a set of 400–500 simulations satisfying these conditions. After this, the EHH values at points distant from the first
SNP 0.1, 0.2, 0.3, and 0.5 units (in a scale of 0 to 1, being 1 equivalent to 100 kb) were recorded and the individual simulations showing the EHH value at the 95th percentile at each point were selected. All simulation selected were lumped together. Next, duplicated points were removed, and then we forced all points to satisfy the condition $EHH_i + 1 \leq EHH_i$ (where $i$ increases from SNP number 1 in any direction). The resulting distribution of EHH values was used as the 95% cut-off line.

**Tajima’s D Test for Specific Haplogroups**

As an additional test for positive selection on V60L, we carried out Tajima’s D test on that subset of haplotypes defined by the derived state T at rs1805005 (V60L), using the information from the 1000G project. We initially tried to carry out this test on the south European 1000G samples, but within the haplogroup defined by allele rs1805005-T, there is no variability at least 10 kb on both sides of rs1805005. Thus, we decided to test the whole 1000G European sample (GBR, CEU, FIN, TSI, and IBS pooled). Note that it is unlikely that pooling will result in more extreme negative Tajima’s D values, so this is a priori, a conservative approach. Besides, there is some evidence (although weaker) from the EHH test that the V60L allele is also under some positive selection in the 1000G north European populations (fig. 3). Note that evidence for selection on the pooled 1000G European population is only apparent when we analyze the internal diversity of the rs1805005-T haplogroup and not when we analyze the global diversity, that is, including both rs1805005-T and -G haplogroups.

For this haplogroup-specific Tajima’s D test, we decided to analyze a region comprising 10 kb of the upstream flanking region to V60L. Using just one flanking region of V60L in which one SNP at the beginning or the end is the V60L, SNP is more efficient when calculating the $P$ values (discussed later). Even after pooling 1000G data from north and south European samples, there is just one polymorphism within the T haplogroup in the downstream 10 kb region to V60L, which does not give us power to perform Tajima’s D test. However, there are 6 SNPs in the upstream 10 kb region within the T haplogroup.

To estimate the $P$ values for the rs1805005-T haplogroup, we proceeded as follows:

a) First, run coalescent simulations using parameters from the whole European population

For the Gutenkunst et al. (2009) model:

\[
/m_5 760 252 -t tbs -r tbs 10000 \cdot l 3 0 760 0 -n 1 1666666 -n 2 2.261 -n 3 1.051 -g 1 372.278 -g 2 380.217 -g 3 259.366 - ma x 0.7176 0.7176 0.7176 0.7176 x 0.7176 0.7176 0.7176 x -eg 0.0063 2 0 -eg 0.0063 2 0 -ej 0.0163 3 2 -eg 0.0163 0.0199 1 0 -ej 0.0435 2 1 -en 0.0435 2 0 -2029
\]

where the -t (theta) option is a random number from a normal distribution with mean 18.4 and variance equal to the mean, and the -r (recombination parameter) option is a random number from a normal distribution with mean 3.25 and variance equal to the mean.

For the Laval et al. (2010) model:

\[
./ms 760 252 -t tbs -r tbs 10000 \cdot l 3 0 760 0 -n 1 1666666 -n 2 2.261 -n 3 1.051 -g 1 372.278 -g 2 380.217 -g 3 259.366 -ma x 0.881098 0.881098 0.881098 0.881098 x 0.881098 0.881098 0.881098 0.881098 x -eg 0.0063 2 0 -eg 0.0063 2 0 -ej 0.0163 3 2 -eg 0.0163 0.0199 1 0 -ej 0.0435 2 1 -en 0.0435 2 0 -2029
\]

b) Then, for each simulation, we check that at a certain predefined position allele frequencies are close (allowance of 5%) to those observed at rs1805005 at that population. This predefined position is normally the first or the last SNP, because this facilitates steps b), c), and d). Similarly, we check that the number of segregating sites is close to the observed value (allowance of 10%) for the whole sample. If these conditions are satisfied, we keep that simulation, otherwise we discard it.

c) Then, for each accepted simulation, we extract the set of haplotypes that are of a desired allelic state at the predefined position (0 = ancestral, 1 = derived), emulating the rs1805005-T (derived) allele in that population.

d) We check that the number of segregating sites within this extracted set of haplotypes is similar (allowance 10%) to that observed for that haplogroup in the 1000G population. For 10 kb upstream, rs1905005-T alleles 6 segregating sites are observed, so the allowed range was between 5 and 7 segregating sites. Note that several million simulations are needed to end up with 100–200 simulations satisfying our conditions. For each accepted simulation, we calculate its corresponding Tajima’s D value (for the specified set of haplotypes forming a haplogroup). The set of the so-calculated Tajima’s D values forms our distribution from which to infer the $P$ value for the observed Tajima’s D of the rs1805005-T haplogroup of the 1000G European sample.

**Association Analysis**

To detect the association of some MC1R variants to melanoma susceptibility, we sequenced MC1R from 127 DNA samples from melanoma patients attending the outpatient clinic at the Castellón Province Hospital (Castellón, Spain) for a 5-year period, between 2007 and 2011. In addition, already published MC1R sequence data on 595 melanoma patients from Madrid were also recruited (Fernandez 2017).
et al. 2007; Ibarrola-Villava et al. 2012). Both sets of samples are registered at the National Health Institute Carlos III with reference numbers C0000274 and C0000293. For the association analyses, we used the R libraries MaXact and HardyWeinberg (http://cran.r-project.org/web/packages/last_accessed_October_3, 2013).

**UV Radiation Maps**

We elaborated a UV-radiation map of Spain to correlate the daily annual average of incident UV-B radiation (J/m²) with MC1R haplotype frequencies. Surface UV-B data (from 1995 to 2010, although not for all the provinces) were supplied to us by AEMET (the Spanish National Meteorological Agency), who used 16 meteorological stations spread over the Spanish territory. We also obtained data on the satellite-inferred global solar irradiation (annual mean; irradiation in the range from 0.2 to 4 μm) per province from the period 1983–2005, obtained from CM-SAF (Climate Satellite Application Facilities), EUMETSAT, and elaborated by AEMET (http://www.aemet.es/documentos/es/serviciosclimaticos/datosclimatologicos/atlas_radiacion_solar/atlas_de_radiacion_24042012.pdf, last accessed October 3, 2013). Because UV-B data and global data across the Iberian Peninsula (i.e., excluding the Canary Islands) correlated well (Pearson’s r > 0.9) for those provinces lacking a meteorological station recording UV-B data, UV-B values were inferred by linear regression from the global irradiation. Irradiation maps were elaborated with QGIS v1.8.0 (http://www.qgis.org/, last accessed October 3, 2013).

**Supplementary Material**

Supplementary information S1–S3 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

This work was supported by the former Spanish Ministerio de Ciencia e Innovación, project CGL2008-04066/BOS to S.A.; the Dpt. Educacion, Universidades e Investigación of the Basque Government, project IT542-10; the University of the Basque Country program UFI11/09; a predoctoral fellowship from the Dept. Educación, Universidades e Investigación of the Basque Government to S.L. (BI09.248); “Programa de Investigación Científica de la Universidad de La Laguna” (boc-a-2010-255-7177); the Health Institute “Carlos III” (FIS PI08/1383, FIS PI11/00623) to C.F.; the European Regional Development Funds “A way of making Europe”; and a postdoctoral fellowship from Fundación Ramón Areces to M.P.Y. The authors thank the Spanish Banco Nacional de ADN (BNADN) (http://www.bancoadn.org/) for providing us with DNA samples from all over Spain. They also thank the Spanish Agencia Estatal de Meteorología (AEMET) (http://www.aemet.es/es/portada) for kindly providing us with the UV-B radiation data.

**References**


