Discovery of copy number variants by multiplex amplifiable probe hybridization (MAPH) in candidate pigmentation genes

Saioa López1, Iker García1, Isabel Smith2, Arrate Sevilla1, Neskuts Izagirre1, Concepción de la Rúa1, and Santos Alonso1

1Genetics, Physical Anthropology and Animal Physiology and 2Department of Zoology and Animal Cell Biology, University of the Basque Country UPV/EHU, Barrio sarriena s/n, Leioa, Spain

Abstract

Background: Copy Number Variants (CNVs) contribute to a large fraction of genetic diversity and some of them have been reported to offer an evolutionary advantage.

Aim: To identify CNVs in pigmentary loci that could contribute to human skin pigmentation diversity.

Subjects and methods: This study assessed the existence of CNVs in every exon of candidate genes: TYR, TYRP1, DCT, MC1R and SLC24A5, using the Multiplex Amplifiable Probe Hybridization technique (MAPH). This study analysed a total of 99 DNA samples of unrelated individuals from different populations. Validation and further analysis in a larger Spanish sample were performed by RT-qPCR.

Results: Five CNVs were identified by MAPH: DCT exons 4 and 8, TYR exon 1 and SLC24A5 exons 1 and 4. Real-time quantitative PCR (RT-qPCR) confirmed the CNV in exon 1 of SLC24A5. This study further analysed the 5’ promoter region of SLC24A5 and found another CNV in this region. However, no association was found between the CNV and the degree of pigmentation.

Conclusion: Although the functional role of these structural variants in pigmentation should be the subject of future work, the results emphasize the need to consider all classes of variation (both SNPs and CNVs) when exploring the genetics of skin pigmentation.

Keywords

Evolution, human diversity, mutation, natural selection

Introduction

The colour of human skin is a complex trait that varies substantially across populations and correlates with the incident ultraviolet radiation (Chaplin, 2004; Jablonski & Chaplin, 2010). Although it is strongly assumed that pigmentation variation has been influenced by natural selection, it is currently being investigated how selection has affected the genetic architecture of pigmentation loci in different populations. In this regard, analysing the diversity patterns of pigmentation genes can offer some insights into the evolution of skin pigmentation (Alonso et al., 2008; de Gruijter et al., 2011; Hudjashov et al., 2013; Izagirre et al., 2006; Lao et al., 2007; Martínez-Cadenas et al., 2013).

A lot of work has been done on the identification of genes responsible for variation in human skin pigmentation; previous studies in mice revealed that pigmentation is presumed to be under the control of ~120 genes, which may act at different stages of melanogenesis (Fitch et al., 2003). Hence, it seems that the influence of gene interactions should also be considered an important factor contributing to the high variability of pigmentation among humans. Genetic studies have shown mutations in many genes that can cause pigmentation disorders, such as oculocutaneous albinism (OCA) caused by mutations in tyrosinase (TYR), oculocutaneous albinism II (OCA2), tyrosinase related protein 1 (TYRP1) or the solute carrier family 45 member 2 (SLC45A2) (reviewed in Simeonov et al., 2013); Xeroderma pigmentosum (XP), caused by a mutation in XPA gene that encodes a protein involved in DNA excision repair (Tanaka et al., 1990) or Waardenburg Syndrome type 2, associated with mutations in the microphthalmia-associated transcription factor (MITF) gene (Tassabehji et al., 1994). To date, however, only a few genes have been shown to have effects on normal variation in pigmentation. For instance the melanocortin receptor 1 (MC1R) (Naysmith et al., 2004); the solute carrier family 24 member 5 gene (SLC24A5), which encodes a cation exchanger that localizes to the melanosome or its precursor (Lamason et al., 2005); the agouti signalling protein gene (ASIP) (Bonilla et al., 2005); or those involved in the melanogenic pathway that catalyse the production of melanin from tyrosine: tyrosinase (TYR), tyrosinase-related protein 1 (TYRP1) and dopachrome tautomerase (DCT). Other studies have revealed that SLC45A2, implicated in processing intracellular trafficking of tyrosinase (Graf et al., 2005), interactions between MC1R and HERC2 (Branicki et al., 2009), IRF4 and SLC24A4 (Han et al., 2008) are also associated with normal human skin pigmentation variation.
However, apart from SNPs, additional layers of variability in the human genome also exist. Thus, stretches of DNA that are present at a variable copy number both within and among individuals (Copy Number Variants or CNVs) have been shown to contribute to a large fraction of genetic diversity (Feuk et al., 2006). The phenotypic effects at CNVs are likely to be due to modifications in expression patterns, because of a direct change in gene dosage, or indirectly through changing its location in the genome (Feuk et al., 2006). Many CNVs identified to date have been shown to be at the basis of many disorders, susceptibility to complex diseases or to bad prognosis in tumours (Hurles et al., 2008; Zhang et al., 2009). For instance, Charcot-Marie-Tooth type 1A (Lupski et al., 1991), Parkinson’s disease (Singleton et al., 2003) or predisposition to psoriasis (Hollox et al., 2008) are caused by a gain of genetic material. On the contrary, other diseases can be due to deletions in genetic fragments, such as autism (Weiss et al., 2008), schizophrenia (Stefansson et al., 2008; Xu et al., 2008) or systemic lupus erythematosus (Yang et al., 2007). However, while most of the extensive copy number variation is apparently non-adaptive or disadvantageous, it has also been shown that some copy number variants can offer evolutionary advantage and, thus, may have been under (recent) positive selection. For instance, an increased copy number of the salivary amylase gene, AMY1, has been suggested to be advantageous in populations with a diet rich in starch (Perry et al., 2007). Similarly, Popesco et al. (2006) found that the higher copy number of DUF1220 in humans, in comparison to other primates, was crucial to achieve higher cognitive functions.

Genome-wide studies in individuals with different ancestry have revealed the existence of rare large structural variants comprising the pigmentation genes SLC24A5 (Wong et al., 2007), MC1R (Perry et al., 2008; Xu et al., 2011), TYR (Mills et al., 2006; Wang et al., 2008), TYRP1 (Redon et al., 2006; Simon-Sanchez et al., 2007) and DCT (Gusev et al., 2009). Therefore, it would not seem unrealistic to hypothesize that copy number variation in pigmentation loci could contribute to normal variation in skin pigmentation. In these genome-wide studies the detection of CNVs is often based on large genomic regions, rather than focusing on small regions such as single genes or exons. Thus, in this study we pursue the detection of CNVs in every exon of a set of genes related to human skin pigmentation (TYR, TYRP1, DCT, MC1R, SLC24A5) using the Multiplex Amplifiable Probe Hybridization technique (MAPH) (Armour et al., 2000) to assess copy number variation at each of the exons of the genes analysed (TYR, TYRP1, DCT, SLC24A5 and MC1R). We analysed a total of 99 DNA samples from unrelated individuals from five population groups: seven North-Saharan Africans (NSA), eight Sub-Saharan Africans (SSA), 10 Japanese (JPT), nine Chinese (CHN) and 65 Europeans. NSA, SSA, JPT and CHN samples were purchased from the Coriell Cell Repository (Camden, NJ). European DNAs consisted of a set of samples from individuals with several generations of Basque ancestry collected by us.

For the RT-qPCR experiments, after validation, we included an independent sample of 200 from a total of 650 unrelated individuals living in the Basque Country with self-reported Spanish ancestry and whose constitutive pigmentation had been measured by reflectance spectrometry. Pigmentation was measured on the inner surface of the upper arm, approximately mid-way between the axilla and the medial epicondyle of the humerus, using a reflectance spectrophotometer EEL DS29 Digital Unigalvo with filter 609, which is optimal for melanin quantification (Robins, 1991). From each subject we recorded sex, age, eye and hair colour, skin phototype according to the Fitzpatrick scale (Fitzpatrick, 1988) and presence of facial nevi. The correlation of the CNVs with pigmentation was statistically assessed by comparing the most and the least pigmented individuals from the reflectance distribution by means of a Fisher’s Exact Test (two tailed).

**Materials and methods**

**Ethics statement**

This study was approved by the Ethics Committee of The University of the Basque Country. Written informed consent was obtained from all subjects.

**DNA samples**

We used the multiplex amplifiable probe hybridization technique (MAPH) (Armour et al., 2000) to assess copy number variation at each of the exons of the genes analysed (TYR, TYRP1, DCT, SLC24A5 and MC1R). We analysed a total of 99 DNA samples from unrelated individuals from five population groups: seven North-Saharan Africans (NSA), eight Sub-Saharan Africans (SSA), 10 Japanese (JPT), nine Chinese (CHN) and 65 Europeans. NSA, SSA, JPT and CHN samples were purchased from the Coriell Cell Repository (Camden, NJ). European DNAs consisted of a set of samples from individuals with several generations of Basque ancestry collected by us.

For the RT-qPCR experiments, after validation, we included an independent sample of 200 from a total of 650 unrelated individuals living in the Basque Country with self-reported Spanish ancestry and whose constitutive pigmentation had been measured by reflectance spectrometry. Pigmentation was measured on the inner surface of the upper arm, approximately mid-way between the axilla and the medial epicondyle of the humerus, using a reflectance spectrophotometer EEL DS29 Digital Unigalvo with filter 609, which is optimal for melanin quantification (Robins, 1991). From each subject we recorded sex, age, eye and hair colour, skin phototype according to the Fitzpatrick scale (Fitzpatrick, 1988) and presence of facial nevi. The correlation of the CNVs with pigmentation was statistically assessed by comparing the most and the least pigmented individuals from the reflectance distribution by means of a Fisher’s Exact Test (two tailed).

**MAPH**

**Probe-set synthesis**

The probe-set consisted of four probes for TYR, seven for TYRP1, eight for DCT, nine for SLC24A5 and one for MC1R (Figure 1; Supplementary Material Table SI). Probes were synthesized so that they fulfilled the following conditions: they should map to unique regions of the genome, they should have a size range between 100–500 bp and a GC (guanines plus cytosines) content of 40–60%. In those cases in which the length of the exon was shorter than the length established for the probe, we extended the probe sequence into the flanking regions. In the case of TYR, exons 4 and 5 are products of a segmental duplication (Takeda et al., 1989), so we could not design a specific probe for each exon. Instead, we synthesized a single probe for a region of the intron 4, which is divergent enough so as to represent a unique target. Probe sequences were aligned against the human genome using BLAST to ensure that they were unique regions in the genome.

Amplicons from genomic DNA were obtained through PCR with the BIOTAQ DNA polymerase (Bioline) using specific primers (sequences shown in Supplementary Material Table SI) and were purified using Microspin S-400 HR columns (GE-Healthcare, UK) and cloned using the TOPO-XL PCR cloning kit (Life Technologies, Carlsbad, CA). Clones were amplified using the flanking vector primers M13, which add 242 bp to each fragment. Then, all fragments were re-amplified using the same inner primers TOPO-Fw.
(CCGCCAGTGTGCTGGAATTC) and TOPO-Rev (CCAGTGTGATGGATATCTGCAG) that only add 59 extra bp to each fragment. The size of the probes was checked by means of agarose gels. When the observed probe size did not seem to correspond to the expected one, sequencing was used to confirm that they actually corresponded to the expected products. From the agarose gel we estimated the concentration of the probes in order to homogenize approximately the quantity of each probe in a final probe mixture. The final concentration for each probe was 200 ng/ml.

Hybridization

Hybridization was done as described in Armour et al. (2000). Briefly, DNA samples (~1 µg) were denatured and fixed to positively charged nylon filters by UV irradiation (50 mJ). Filters were incubated in 1 ml pre-hybridization solution for 2 hours and then this solution was replaced with 200 µl of the pre-hybridization solution supplemented with Cot-1 DNA (10 µg/ml) (Life Technologies) for 1 hour at 65°C. These fixed samples were incubated at 65°C overnight with the probe mixture and the hybridization solution (described in Armour et al., 2000).

Post-hybridization washes and PCR

Next, filters were washed to remove non-hybridized probes following a stringent protocol with two different solutions: 1× SSC/1%SDS for 20 minutes and 0.2× SSC/2%SDS for 40 minutes. We recovered the hybridized probes from the filters by denaturation at 95°C in a solution of 1× PCR buffer (Thermo Scientific, Pittsburgh, PA). Probes were amplified with the primers TOPO-Fw and TOPO-Rev labelled with 6-FAM. These products were finally subject to capillary electrophoresis with fluorescence detection using GeneScan ROX 500 as the Size Standard in an ABI Prism 377 DNA Sequencer.

Data analysis

Using GeneScan software we recorded the area of each peak, which corresponds to one exon. The area of each peak should be proportional to the copy number of each exon. Normalization of peak areas was obtained by a global normalization: the peak area of each amplification product was divided by the combined area of all peaks in that individual, so that per individual the ratio of the peak area of each amplification product to the whole is known. Normalized peak area of a given probe amplification product was divided by the average normalized peak area of that product in a reference sample.

Thus, a value of ~1 indicates absence of variation, whereas values substantially higher or lower than 1 suggest possible duplication or deletion events, respectively. In order to avoid false positives, we made a conservative method of selection, in which only those values that were out of the range mean ±3 SD after data normalization were considered as potential CNVs.

RT-qPCR

In order to confirm the results obtained with the MAPH approach, we performed RT-qPCR reactions for all the samples showing an apparent CNV using SYBR Green Reagents (Life Technologies) with the StepOne Real-Time PCR System (Life Technologies). Primer concentrations and reaction profiles were previously optimized for each exonic region that was positive for MAPH. Criteria for primer optimization involved two parameters: reaction efficiency over 95% and Pearson correlation of each standard curve over 0.99. When primer sets did not work satisfactorily we performed TaqMan-based qPCR. Thus, for DCT exon 4 double-labelled probes were synthesized with 6-FAM at the 5′ end, LNA replacing three of the nucleotides in the sequence and BHQ-1 amidite at the 3′ end (Sigma-Aldrich, St. Louis, MO). The sequences of these primers and probes were 5′-GGAGGAACGAGTGTGATGTGT-3′, 5′-ACTAATCAGAGTCGGATCGTCTG-3′ and [6FAM]TGCCCCAA[+A]CAGC[+T]GG[+T]CTGTA[BHQ1].

For normalization in SYBR Green based qPCR we took as reference genes those that in MAPH had shown no variability across the samples. RQ data were extracted with StepOne Software v2.0. In TaqMan based qPCR, we used RPPH1 as internal copy number control gene, the RNA component of the RNase P ribonucleoprotein. The sequences...
of these primers and probes were: 5′-GCTTGGAACAGA CTCACGG-3′, 5′-GTCTGAAATTGGGTTATGAGTCC-3′ and [HEX] TCACCTCAG[+C][+C]AT[+T]G[+A]ACTCA CTTCC [BHQ1].

The number of copies of the target sequence in each test sample was determined by relative quantitation (RQ) using the comparative CT (cycle threshold) method, which measures the CT difference between target and reference sequences and compares the ΔCT values of test samples to a calibrator sample known to have two copies of the target sequence.

CNV assays in 5′ promoter region were also performed by RT-qPCR reactions using SYBR Green Reagents. The sequences of the primers used were for the first region: 5′-AGGATGCGAGCTAGCCAAAG-3′ and 5′-TAGCTAT TTTCTCATGGCCTACAGTC-3′ and for the second region 5′AACGACCTTCTCTTAGTCTGTA-3′ and 5′-TGT CTCACACCTCGCAGGAAGCA-3′.

As a third validation technique for SLC24A5 exon 1 we also synthesized a LNA modified double-labelled probe (Sigma-Aldrich). The sequences of these primers and probes were: 5′-CTCACATGATCCAGTTTAATCCTC-3′, 5′-GAGCA ACAGAGCCTCTTCG-3′ and [6FAM] GAAGCT[+G] CACG[+C]TGC+[A]GTAAGAGCACAG [BHQ1].

Sequencing
The sequencing reactions were carried out on an ABI 310 Sequencer (Life Technologies) and chromatograms were analysed using GenalysWin 2.0 software.

Identification of transcription factor binding sites
The identification of putative transcription factor binding sites in the 5′ promoter region where a CNV was found was performed with PROMO (Messegue et al., 2002). Factors were predicted within a dissimilarity margin of 0%.

Results

MAPH
We initially identified five copy number variants by MAPH, corresponding to DCT exons 4 and 8, TYR exon 1 and SLC24A5 exons 1 and 4, all of them corresponding to duplications. The most frequent apparent CNV was in TYR exon 1, which was observed in eight individuals (two Chinese and six European), from a total of 99 samples. Apparent CNVs in DCT exon 8 and TYR exon 2 appeared in four individuals each (two African and two European; and four European, respectively). In the other cases, CNVs appeared just once: DCT exon 4 was duplicated in one Japanese sample, SLC24A5 exon 1 in one European and, finally, SLC24A5 exon 4 in one African from the North of the Sahara.

Despite the stringent criteria for CNV calling in MAPH, from the five variants that were identified, only one, that corresponding to exon 1 in SLC24A5, could be validated by RT-qPCR. It is possible that, despite the stringent protocol followed in post-hybridization washes in MAPH, probes were not totally washed in all the samples, leading to false positives. Every set of filters were washed together with a blank in order to account for this possibility; however, although these blanks were clean after the washes, we cannot exclude the possibility that in some isolated cases the non-hybridized probes were not completely washed. Thus, as these experiments confirmed the existence of copy number variation in SLC24A5 exon 1, we checked for association between this CNV and pigmentation phenotype.

Association between the CNV at SLC24A5 exon 1 and skin pigmentation

Samples
Among the 650 samples there were 415 females and 235 males, from whom we had recorded hair and eye colour, nevi presence and skin colour (reflectance). Hair colour was categorized as black, dark brown, light brown, blonde and red; categories for eye colour were dark brown, hazel, green, blue and grey. The number of nevi was categorized as: absence, 1–5, 6–10 and more than 10. We found significant differences for the frequencies of different hair colour categories among sexes (Fisher’s Exact test $p = 0$), the males having significantly higher frequencies of the darkest hair colours. Eye colour and number of nevi, however, did not differ significantly according to gender (Fisher’s Exact test $p > 0$ in all cases).

As regards skin pigmentation, it is still controversial how it is affected by sex. Thus, we also obtained and compared the means and the frequency distributions of skin reflectance of both sexes. In our sample there were no significant differences neither between the means of males and females (68.43 and 69.30, respectively; $t$-test for comparisons of two means: $p = 0.09$) nor between the distribution of the frequencies (Mann-Whitney’s $U$ for the comparison of two distributions: $p = 0.50$). This is at odds with traditional anthropological studies that have posited that females are significantly lighter than males (Robins, 1991) or the work by Candille et al. (2012), who observed that European males have a significantly lighter skin compared to females. Therefore, given the uncertainty in this regard and that it is thought by other authors that constitutive pigmentation can be influenced by sex, we decided to analyse males and females separately at this first stage.

CNV analysis
First, we chose the 60 women with the highest reflectance values and the 60 with the lowest. These showed different phenotypes as regards eye and hair colour and presence of facial nevi and/or freckles. On this sub-set of samples, we performed subsequent RT-qPCRs to identify the CNV in SLC24A5 exon 1. We found that four out of the 120 women analysed shared this novel CNV. Next, we analysed another sub-set of 80 DNA samples (the 40 most pigmented males and the 40 less pigmented males). We found one male with a CNV in this exon. The copy number of the CNV was not the same in all the individuals. As seen in Figure 2, the RQ (relative quantitation, see Methods) was 1.5 (three copies) in one case, 2 (four copies) in two cases and 2.5 (five copies) in two cases. There were no statistically significant differences between the groups of the most and the least pigmented individuals as regards the frequency of the CNV (two-tailed Fisher’s Exact Test, $p = 1$). Table 1 summarizes the
characteristics of these individuals. These results were replicated and confirmed by the LNA-modified probe.

**Extent of the CNV at exon 1 of SLC24A5**

In order to find out the actual extent of the fragment we looked for CNVs in the second exon of SLC24A5 in the individual with a CNV in the first exon. We performed RT-qPCR, using the primers previously designed for MAPH and optimized for quantitative experiments and we obtained that the exon 2 was not variable in copy number, consistent with the results obtained by MAPH technique.

In view of the fact that no association was found between this structural variant and pigmentation variability and prompted by previous observations that associate a polymorphism in this gene with pigmentation variability (Lamason et al., 2005; Stokowski et al., 2007), further work was done in the upstream region of SLC24A5, on the individuals with a CNV in exon 1, to determine if the CNV also spanned the upstream region of the gene. A loss of genetic material was found in one individual in the 5′ promoter (5′ PR_1), around 300 bp from the start of exon 1. Another region was also analysed, ∼1 Kb upstream from the first one (5′ PR_2), showing no CNVs. We searched for transcription factor binding sites (TFBS) in this region (dissimilarity margin of 0%) and found seven putative TFBS: FOXP3, C/EBPbeta, YY1, STAT4, GR-beta, Pax-5 and TFII-I.

Prompted by the discovery of a deletion in this 5′ promoter region, we re-analysed 60 random individuals from the subset by RT-qPCR. We found that seven of them shared a CNV in the upstream region, five of which corresponded to a loss of genetic material and two a genetic gain (Figure 3). Table 2 summarizes the characteristics of these individuals.

![Figure 2](image2.png)

**Table 1. Characteristics of the individuals with the CNV in exon 1 of SLC24A5.**

| ID   | Reflectance (filter 609) | Sex | Age | Eye colour | Hair colour | Facial nevi | Freckles | Phototype | Copy number |
|------|--------------------------|-----|-----|------------|-------------|-------------|----------|-----------|-------------|
| Bu24 | 61.47                    | F   | 51  | Blue       | Blonde      | 0           | No       | III       | 4           |
| Bu17 | 62.23                    | F   | 24  | Black      | Black       | 0           | After sun exposure | IV | 3           |
| Li15 | 75.40                    | F   | 18  | Black      | Dark brown  | 0           | No       | III       | 5           |
| Ip1  | 68.03                    | M   | 35  | Green      | Blonde      | 0           | No       | IV        | 5           |
| Sa24 | 67.60                    | F   | 21  | Brown      | Brown       | 0           | No       | III       | 4           |

![Figure 3](image3.png)
Diversity at SLC24A5 exon 1 and 5' promoter region

We speculated that the presence of CNVs could leave an imprint of the patterns of SNP diversity of SLC24A5, given that, on the one hand, cryptic deletions would result in apparent homozygosity and, on the other, recent cryptic duplications of one allele would mask possible variant alleles in sequencing or genotyping experiments. Thus, we assessed the diversity at this locus (where the CNV was found) in the European samples from the 1000 Genomes Project (1KGP) and we compared it to that corresponding to a region around SLC24A5 exon 3, where SNP rs1426654 locates. This SNP is particularly relevant as it has been strongly associated with human pigmentation (Lamason et al., 2005) and has also been reported to be under the action of positive selection (Basu Mallick et al., 2013; Norton et al., 2010; Sabeti et al., 2010). As positive selection also results in a drop in diversity (actually SNP rs1426654 is almost fixed in Europeans) we wanted to compare the profiles of diversity between these two loci in order to assess if it would be possible to identify a specific diversity signature for unaccounted (cryptic) copy number variation. Alternatively, a similar signature could indicate that, when declaring a locus under positive selection, caution should be taken that CNV is actually not the agent responsible for the observed drop in diversity.

Thus, we downloaded genotypes for a region of 30 Kb containing the SLC24A5 locus, for the European populations from 1000 Genomes Project (1KGP) (Phase 1 data from May 2011), consisting of 380 individuals, by means of SPSmart (Amigo et al., 2008). The construction of haplotypes and the estimation of Theta per nucleotide and Tajima's D values were performed using a customized Perl script. Theta and Tajima’s D values were calculated in overlapping sliding windows (window length 4000 bp; step size 400 bp). We observed, as expected, a loss of diversity defined by Theta ($\theta = 4 N_e \mu$; where $N_e$ = effective population size and $\mu$ = mutation rate per generation) (see Figure 4) and an increased Tajima’s D value in the region encompassing SNP rs1426654 in exon 3. Tajima’s D is one of the most used tests to assess deviations from neutrality in a given locus. It measures the normalized difference between two estimates of $\theta$: one based on the mean pairwise differences between sequences (n) and the other, on the number of segregating sites (S). In this regard, as the derived allele of SNP rs1426654 is almost fixed in Europeans, diversity decays in that locus, and consequently, the power of Tajima’s D to detect selection is substantially reduced. Interestingly, we found similar patterns of diversity and selection in the region surrounding the CNVs discovered herein, especially in the 5' promoter region.

Table 2. Characteristics of the individuals with the CNV in the 5' promoter region of SLC24A5.

| ID   | Reflectance (filter 609) | Sex | Age | Eye colour | Hair colour | Facial nevi | Freckles | Phototype | Copy number |
|------|--------------------------|-----|-----|------------|-------------|-------------|----------|-----------|-------------|
| Bu24 | 61.47                    | F   | 51  | Blue       | Blonde      | 0           | No       | III       | 1           |
| Ll21 | 73.33                    | M   | 37  | Blue       | Brown       | 4           | No       | III       | 1           |
| Esc19| 74.51                    | F   | 61  | Green      | Brown       | 0           | In childhood | III       | 4           |
| Sa4  | 63.50                    | M   | 22  | Brown      | Brown       | 0           | No       | III       | 1           |
| Ba6  | 76.00                    | M   | 27  | Brown      | Brown       | 1           | After sun exposure | III       | 4           |
| Ni29 | 72.15                    | M   | 24  | Brown      | Brown       | 3           | No       | III       | 1           |
| Esc11| 66.39                    | M   | 57  | Green      | Brown       | 0           | In childhood | III       | 1           |

Figure 4. Nucleotide diversity (pi) and Theta for SLC24A5 exons 1 and 3. The regions analysed corresponded to a region of 10 Kb surrounding exon 1 (left) and surrounding exon 3 (right), in the European populations from 1 KGP.
Discussion

Our results provide robust evidence for the existence of a CNV in the first exon of SLC24A5. Structural variation can contribute to phenotypic differences, but in many cases it is difficult to estimate how relevant these changes are, both from an evolutionary and a functional point of view. SLC24A5 encodes the NCX5 protein, a member of the potassium-dependent sodium calcium exchanger protein family that locates in normal human epidermal melanocytes. SLC24A5 is one of the major human pigmentation genes, as it is crucial for melanin synthesis (Ginger et al., 2008). In particular, Lamason et al. (2005) showed a polymorphism in SLC24A5 exon 3, rs1426654 (Ala111Thr), which accounts for a high percentage of normal variation in pigmentation among human populations. Whereas the ancestral allele (alanine) is found in Africans and East Asians at a substantial frequency (0.92 and 0.99, respectively), in Europeans the derived allele (threonine) is nearly fixed. Besides, this derived allele has shown strong evidence of recent positive selection in Europeans (Basu Mallick et al., 2013; Norton et al., 2010; Sabeti et al., 2010).

However, we have not found any statistical association between the CNV in exon 1 and the degree of pigmentation when comparing the most and less pigmented individuals from our population sample. The individuals in which this CNV has been detected form a heterogeneous group as regards skin pigmentation, hair and eye colour and presence of facial nevi and/or freckles (Table 1). So, the occurrence of this mutation cannot be associated with a specific phenotypic trait and, therefore, it does not have an evident role in normal variation of skin pigmentation in our sample. This lack of association could be real or could be due to the fact that the range of phenotypic variability between the groups of most and least skin-pigmented individuals might not be marked enough to find that association. Thus, an effect on pigmentation may exist, but its effect may be too small to be highlighted with the modest sample size we have used. As skin pigmentation is a trait that varies among populations according to the incidence of ultraviolet irradiation in an evolutionary process driven by natural selection, a key strategy to be adopted would be the analysis and comparison of populations with marked differences in skin pigmentation. In this regard, knowing the demographic processes that could have influenced the biological composition of populations under study is also crucial (population bottlenecks, population growth or admixture, for example) as they may hinder the interpretation of the variability detected or the selective pressures acting on a given variant.

The CNV found in the 5′ promoter region is also of interest. In this sense, phenotypes could be affected by CNVs in different ways. The copy number variant could contain regulatory elements, which if deleted could alter the expression of the regulated exons. In this study we have found five individuals with a loss of genetic material in the 5′ promoter region of the gene SLC24A5, which is very likely to contain such regulatory elements. We searched for putative TFBS and observed that there are seven in this region. Interestingly, we identified five binding sites for the transcription factor C/EBPbeta, which is involved in the motility of primary human melanocytes (Damm et al., 2010), which are skin cells that synthesize melanin, the pigment responsible for skin colour. We have also found two individuals carrying additional copies of the region, which could provide additional copies of regulatory regions. Moreover, this region is identified as a DNaseI hypersensitivity (DNase HS) site in ENCODE (Encyclopedia of DNA Elements, https://genome.ucsc.edu/ENCODE/), a region prone to cleavage by the DNase I enzyme due to the loss of the condensed structure of the chromatin. This chromatin state favours the binding of transcription factors, thus playing a major role in the transcriptional regulation of nearby genes (Thurman et al., 2012). It has been reported that the number of DNase HS sites is associated with the expression levels of nearby genes (Crawford et al., 2006), thus reinforcing the hypothesis that the CNV at the 5′ promoter region might affect the expression of SLC24A5. Even so, there is no evidence among our samples that this gain of genetic material leads to any obvious phenotypic effect. Research into the Database of Genomics Variants (http://dgv.tcag.ca/dgv/app/home) revealed that other CNVs in the upstream region of SLC24A5 also exist. The closest, corresponding to a deletion and found in four individuals, is ~70 kb upstream from the CNV discovered by us (Conrad et al., 2010). Although there are still no functional assays and they have not been associated with a specific phenotype, the existence of various structural variants in the promoter region of SLC24A5 suggests that it has evolved under neutral conditions with no significant purifying selection acting on CNVs. Whether these structural variants have a functional role in the regulation of SLC24A5 expression should be the subject of future work.

We compared the genetic diversity patterns in Europeans in a region of 10 kb surrounding exon 1 and 10 kb surrounding exon 3, where SNP rs1426654 locates, and observed a decrease in diversity in both exons. In exon 3, this drop in diversity is a consequence of selection acting on SNP rs1426654. Although this might lead one to think that positive selection is also acting on exon 1 or the 5′ promoter region, we suggest that the decrease in diversity reported here is due to the presence of the CNVs, which leads to a detection of an excess of homozygotes. Thus, this observation opens a new issue related to the detection of signatures of selection on the genome, underlining the difficulties and misunderstandings that might arise if the drop in diversity provoked by structural variation is not taken into account.

Skin pigmentation is a complex trait regulated by the interaction of several genes and the discovery of CNVs in SLC24A5 in this work emphasizes the need to account for all classes of variation (both SNPs and CNVs) when exploring the genetics of skin pigmentation. It should be remarked that the novel CNVs have been found in this work exclusively in South European individuals by MAPH or RT-qPCR. Thus, in this regard, it would be interesting to make bigger efforts in the identification of these or other putative structural variants among other populations worldwide. Apart from the evolutionary perspective, the identification of CNVs associated to variation in pigmentation may have further applications, such as phenotypic characterization of individuals or susceptibility to melanoma.
Acknowledgements

We thank Prof. John Armour for his invaluable help in the design and performance of the experiments.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

This work was supported by the former Spanish Ministerio de Ciencia e Innovacion, project CGL2008-04066/BOS to S.A.; by the Dpt. Educacion, Universidades e Investigacion del Basque Government, project IT542-10 to C.R.; and a predoctoral fellowship from the Dpt. Educacion, Universidades e Investigacion del Basque Government to S.L.

References

Alonso S, Izagirre N, Smith-Zubiaga I, Gardeazabal J, Diaz-Ramon JL, Diaz-Perez JL, Zelenika D, et al. 2008. Complex signatures of selection for the melanogenic loci TYR, TYRP1 and DCT in humans. BMC Evol Biol 8:74.

Amigo J, Salas A, Phillips C, Carracedo A. 2008. SPSmart: adapting population-based SNP genotype databases for fast and comprehensive web access. BMC Bioinformatics 9:428.

Armour JA, Sismani C, Patsalis PC, Cross G. 2000. Measurement of locus copy number by hybridization with amplifiable probes. Nucleic Acids Res 28:605–609.

Basu Mallick C, Iliescu FM, Mills M, Hill S, Tamang R, Chaubey G, Goto R, et al. 2013. The light skin allele of SLC24A5 in South Asians and Europeans shares identity by descent. PLoS Genet 9: e1003912.

Bonilla C, Boxill LA, Donald SA, Williams T, Sismani C, Patsalis PC, Cross G. 2000. Measurement of copy number differences in the human genome. Nature 404:444–454.

Candille SI, Absher DM, Beleza S, Bauchet M, McEvoy B, Garrison B, Brudnik U, Wojas-Pelc A. 2009. Interactions between selection and neuronal expression of DUF1220 domains. Science 313:219–232.

Fitch KR, Carson AR, Scherer SW. 2006. Structural variation in the human genome. Nature 438:85–97.

Graf J, Hodgson R, van Daal A. 2005. Single Nucleotide Polymorphisms in the MATP gene are associated with normal human pigmentation variation. Hum Mutat 25:278–284.

Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, Friedman JM, Pe’er I. 2009. Whole population, genome-wide mapping of hidden relatedness. Genome Res 19:318–326.

Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, Hankinson SE, et al. 2008. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet 4:e1000074.

Hedjazov G, Vilkens R, Kivisild T. 2013. Global patterns of diversity and selection in human tyrosinase gene. PLoS One 8: e74307.

Hulskes M, Dermitzakis E, Tyler-Smith C. 2008. The functional impact of structural variation in humans. Trends Genet 24:238–245.

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 African Americans. Mol Biol Evol 23:1697–1706.

Jablonski NG, Chaplin G. 2010. Human skin pigmentation as an adaptation to UV radiation. Proc Natl Acad Sci USA 107: 8962–8967.

Lamason RL, Mohideen MA, Mestre JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, et al. 2005. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science 310: 1782–1786.

Lao O, de Grujiter JM, van Duijn K, Navarro A, Kayser M. 2007. Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. Ann Hum Genet 71:354–369.

Lupski JR. 2010. De Oca-Luna RM, Slagtenhaupt S, Pentao L, Guzzetta V, Trask BJ, Saucedo-Cardenas O, et al. 2011. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. Cell 66: 219–232.

Martinez-Cadenas C, Lopez S, Ribas G, Flores C, Garcia O, Sevilla A, Smith-Zubiaga I, Ibarrola-Villaba M, et al. 2013. Simultaneous purifying selection on the ancestral MC1R allele and positive selection on the melanoma-risk allele V60L in South Europeans. Mol Biol Evol 30:2654–2665.

Messeguer X, Escudero R, Farré D, Nuñez O, Martínez J, Albá MM. 2002. PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics 18:333–334.

Mills RE, Luttingt CT, Larkins CE, Beauchamp A, Tsu C, Pittard WS, Devine SE. 2006. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res 16: 1182–1190.

Naymushin L, Waterson K, Ha T, Flanagan N, Bisset Y, Ray A, Wakamatsu K, et al. 2004. Quantitative measures of the effect of melanocortin 1 receptor on human pigmentary status. J Invest Dermatol 122:423–428.

Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, et al. 2007. Diet and the evolution of human amylase gene copy number variation. Nature Genet 39:1256–1260.

Popescu MC, Malcaren JD, Hopkins J, Dumas L, Cox M, Meltesen L, McGavran L, et al. 2006. Human lineage-specific amplification, selection and neuronal expression of DUF1220 domains. Science 313: 1304–1307.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, et al. 2006. Global variation in copy number in the human genome. Nature 444:444–454.

Robins AH. 1991. Biological perspectives on human pigmentation. Cambridge: Cambridge University Press.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie Y, et al. 2010. Genome-wide detection and characterization of positive selection in human populations. Nature 449:913–918.
Simeonov DR, Wang X, Wang C, Sergeev Y, Dolinska M, Bower M, Fischer R, et al. 2013. DNA variations in oculocutaneous albinism: an updated mutation list and current outstanding issues in molecular diagnostics. Hum Mutat 34:827–835.

Simon-Sanchez J, Scholz S, Fung HC, Matarin M, Hernandez D, Gibbs JR, Britton A, et al. 2007. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. Hum Mol Genet 16:1–14.

Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, et al. 2003. alpha-Synuclein locus triplication causes Parkinson’s disease. Science 302:841.

Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S, Fossdal R, et al. 2008. Large recurrent microdeletions associated with schizophrenia. Nature 455:237–244.

Stokowski RP, Pant PV, Dadd T, Fereday A, Hinds DA, Jarman C, Filsell W, et al. 2007. A genomewide association study of skin pigmentation in a South Asian population. Am J Hum Genet 81:1119–1132.

Tanaka K, Miura N, Satokata I, Miyamoto I, Yoshida MC, Satoh Y, Kondo S, et al. 1990. Analysis of a human DNA excision repair gene involved in group A xeroderma pigmentosum and containing a zinc-finger domain. Nature 348:73–76.

Tassabehji M, Newton VE, Read AP. 1994. Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. Nature Genet 8:251–255.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, et al. 2012. The accessible chromatin landscape of the human genome. Nature 489:75–82.

Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, et al. 2008. The diploid genome sequence of an Asian individual. Nature 456:60–65.

Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, et al; Autism Consortium. 2008. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med 358:667–675.

Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, et al. 2007. A comprehensive analysis of common copy-number variations in the human genome. Am J Hum Genet 80:91–104.

Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. 2008. Strong association of the novo copy number mutations with sporadic schizophrenia. Nat Genet 40:880–885.

Xu H, Poh WT, Sim X, Ong RT, Suo C, Tay WT, Khor CC, et al. 2011. SgD-CNV, a database for common and rare copy number variants in three Asian populations. Hum Mutat 32:1341–1349.

Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, Hebert M, et al. 2007. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. Am J Hum Genet 80:1037–1054.

Zhang F, Gu W, Hurles ME, Lupski JR. 2009. Copy number variation in human health, disease and evolution. Annu Rev Genomics Hum Genet 10:451–481.

Supplementary material available online
Supplementary Material Table SI