European Journal of Human Genetics (2013) 21, 1320–1323

www.nature.com/ejhg

SHORT REPORT

Genetics of eye colours in different rural populations on the Silk Road

Sheila Ulivi*, Massimo Mezzavilla and Paolo Gasparini

Eye colour is a highly transmissible and discernible trait in humans. A genome-wide association scan for variants associated to eye pigmentation was carried out on a large group of individuals coming from the Silk Road. Significant associations were detected not only with *HERC2* (*P*-value $= 4.99 \times 10^{-37}$) and *OCA2* (*P*-value $= 4.51 \times 10^{-9}$) genes but also with *CTNNA2* gene (*P*-value $= 4.06 \times 10^{-8}$). Moreover, the multifactor dimensionality reduction analysis clearly showed the effect of *HERC2* haplotype over *OCA2* mostly associated with SNP, thus enabling a highly accurate eye-colour prediction. Finally, the regression tree analysis showed that individuals carrying a given combination of haplotypes have a significant probability to show a blue or green/grey iris colour as compared with brown, with a gradient from west to east.
European Journal of Human Genetics (2013) **21,** 1320–1323; doi:10.1038/ejhg.2013.41; published online 13 March 2013

Keywords: eye colours; GWAS; silk road

INTRODUCTION

Eye irises' colour is a common polygenic phenotype, which can vary between individuals.[1] The iris is a thin diaphragm composed mostly of connective tissue and smooth muscle fibres regulating the amount of light entering the pupil.[2] The colours, texture, and patterns of each person's eyes are unique and are due to different factors, such as the density and structure of the iris stroma, the pigment epithelium, the pigment content within the melanocytes of the iris stroma, and the light-absorption properties of the melanin colouring skin, hair, and eyes.

Furthermore, also genetic factors have a relevant role with well-known genes such as *OCA2* (OMIM: 203200), *HERC2* (OMIM: 605837), which regulates *OCA2* expression, and several other genes.[1]

In this paper, we analysed this polygenic trait in several communities scattered along the Silk Road, to fill a gap of knowledge on genes underlying iris colour among people living in this part of the world.

MATERIALS AND METHODS

An overall number of 1015 individuals randomly selected, ranging from 8 to 84 years of age, were recruited during the 'MARCO POLO' scientific expedition (www.marcopolo2010.it) and divided as follows: (a) Western Silk Road (WSR): Crimea (peninsula of Ukraine) (102), Georgia (147), Armenia (174), and Azerbaijan (73) and (b) Eastern Silk Road (ESR): Uzbekistan (122), Tajikistan (118), Kazakhstan (60), and Kyrgyzstan (219). Sociodemographic information, as well as data on professional activity, lifestyle, eating habits, and family history, was obtained. A colour eye chart was used to classify iris colour. Phenotypes were defined as previously described[1] (ie, blue = 0, intermediate (green, amber) = 1, and brown = 2).

Only 710 individuals out of 1015 gave a saliva sample. Before imputation, individuals with more than 5% missing genotypes and SNPs missing in more than 5% of samples were excluded. After doing the default filtering (call rate $\geq$ 97%, Rsq $\geq$ 0.3, Hardy–Weinberg Equilibrium *P*-value $\geq 1 \times 10^{-8}$, and minor allele frequency $\geq$ 0.01), 657 samples (WSR: 385; ESR: 272) and 2 157 485 directly genotyped and imputed SNPs passed quality controls. After genotyping (HumanOmniExpress; Illumina, Inc., San Diego, CA, USA), data were imputed using MACH (www.sph.umich.edu) to a common set of $\sim$2.5 million autosomal SNPs based on LD patterns observed in Hap Map release 22 CEU samples (http://www.hapmap.org). In order to avoid the presence of close relatives in our data set, statistical analyses were performed using a mixed model regression, in which the kinship matrix is the random effect, as implemented in GenABEL for genotyped SNPs and ProbABEL for imputed data. Haploview 4.2 (www.broadinstitute.org) and Plink v1.07 (http://pngu.mgh.harvard.edu) were used for haplotype construction, haplotype phases, and genetic-association studies. For studies on selection, a spatial ancestry analysis (SPA) was carried out to check for signal selection;[3] in addition, we performed BAYESCAN[4] to detect SNPs in the genomic region of *CTNNA2*, which show a higher level of population divergence than neutral loci; the decisive criterion that we adopted corresponds to a posterior probability $>$0.99, which is indicative that a locus is under selection. A recursive partitioning tree-based analysis, using the R party package, was also performed to predict eye colour on the basis of *HERC2* diplotypes and country of origin of each individual. For each node, we calculated the *P*-value associated after Bonferroni correction. Gene–gene interactions between analysed SNP and haplotype positions were tested using the multifactor dimensionality reduction (MDR) approach[5] (software version 2.0 beta 8.4, www.epistasis.org).

RESULTS

The eye-colour distribution in the 1015 individuals (mean age 34.80, SD 17.01) turned out as follows: (a) ESR: 32 (6.2%) blue eyes, 35 (6.7%) green/grey, and 452 (87.1%) brown and (b) WSR: 46 (9.3%) had blue eyes, 105 (21.2%) green/grey, and 345 (69.5%) brown. The genomic control (GC) inflation factor (λ) to check for stratifications in our populations was calculated being λ of 1.010115 and indicating the absence of strong stratification in our samples.

The GC inflation factor was calculated considering all individuals from each population as a unique group.

GWAS results for a first analysis[1] are displayed in Table 1. The most significant association was obtained with two SNPs, known as being strongly correlated with both blue and brown eye colours, located

Institute for Maternal and Child Health—IRCCS 'Burlo Garofolo', Trieste, Italy
*Correspondence: Dr S Ulivi, Institute for Maternal and Child Health—IRCCS 'Burlo Garofolo', via dell'Istria 65/1, Trieste 34137, Italy; Tel: +39 040 3785539; Fax: +39 040 378540; E-mail: ulivi@burlo.trieste.it
Received 6 August 2012; revised 4 February 2013; accepted 13 February 2013; published online 13 March 2013

Table 1 Results of (a) GWAS for iris colour in silk road and (b) GWAS brown *versus* not brown

(a)

Gene	*Name*	*Chr*	*Position*	*A1*	*A2*	*Coded_allele_freq*	*P_value*	*Pgc_value*
OCA2	rs7495174	15	28344238	G	A	0.3	4.51×10^{-9}	4.99×10^{-9}
HERC2	rs1129038	15	28356859	A	G	0.28082192	4.99×10^{-37}	7.96×10^{-37}
	rs12593929	15	28359258	G	A	0.27549467	1.01×10^{-9}	1.13×10^{-9}
	rs12913832	15	28365618	G	A	0.28082192	4.99×10^{-37}	7.96×10^{-37}
	rs3935591	15	28374012	G	A	0.46183206	1.44×10^{-13}	1.69×10^{-13}
	rs11636232	15	28386626	T	C	0.07810107	1.28×10^{-10}	1.45×10^{-10}
	rs916977	15	28513364	G	A	0.48782344	3.29×10^{-16}	4.00×10^{-16}
	rs1667394	15	28530182	A	G	0.49162861	4.92×10^{-16}	5.96×10^{-16}
CTNNA2	rs13431891	2	80605790	A	G	0.938271	4.06×10^{-8}	4.44×10^{-8}

(b)

Gene	*Name*	*Chr*	*Position*	*A1*	*A2*	*Coded_allele_freq*	*P_value*	*Pgc_value*
OCA2	rs7495174	15	28344238	G	A	0.3	1.55×10^{-8}	1.45×10^{-8}
HERC2	rs1129038	15	28356859	A	G	0.28082	4.85×10^{-31}	3.61×10^{-31}
	rs12593929	15	28359258	G	A	0.27549	2.10×10^{-9}	1.94×10^{-9}
	rs12913832	15	28365618	G	A	0.28082	4.85×10^{-31}	3.61×10^{-31}
	rs3935591	15	28374012	A	G	0.46183	6.92×10^{-12}	6.23×10^{-12}
	rs11636232	15	28386626	T	C	0.0781	4.61×10^{-9}	4.27×10^{-9}
	rs916977	15	28513364	G	A	0.48782	7.19×10^{-14}	6.35×10^{-14}
	rs1667394	15	28530182	A	G	0.49163	8.99×10^{-14}	7.95×10^{-14}

Abbreviations: A1, coded allele; A2, non coded allele; Chr, chromosome; Coded_allele_freq, frequency of the coded allele; Gene, gene name; Name, SNP identifier; *P*_value, *P*-value obtained from GWAS analysis; Pgc_value, *P*-values corrected after GC; Position, physical position (dbSNP build 135).

within *HERC2* gene (rs1129038 and rs12913832; *P*-value = 4.99×10^{-37}). Other significant associations (*P*-value $< 5 \times 10^{-8}$) were found with both *CTNNA2* (OMIM: 114025)[6] and *OCA2* genes.

A second GWAS analysis was performed using the phenotype 'brown *versus* not brown' in a categorical way. In this case too, the most significant association was found with seven different SNPs located on *HERC2* gene and one on *OCA2* gene (Table 1).

According to SPA analysis, we found evidence of positive selection into the region of *HERC2* and *OCA2* genes (scores are over the top 2%) in agreement with previous data.[3] Interestingly, high SPA scores were also obtained for *CTNNA2* gene. BAYESCAN results confirm these outcomes, namely, that SNPs in the *CTNNA2* gene display evidence of positive selection.

Finally, looking at the most significant SNPs in *HERC2* gene, we were able to construct a seven-marker haplo-block, which detected 10 different haplotypes in our populations. Six of them show a frequency >5%: 5′-GGAACAG-3′ (27.2%), 5′-AAGGCGA-3′ (20.3%), 5′-GAAGCGA-3′ (20.1%), 5′-GAAACAG-3′ (17.7%), 5′-AAGGTGA-3′ (7.9%), and 5′-GAAGCAG-3′ (5.5%). All of them had correlation with eye colour ($0.036 > P > 2.05 \times 10^{-22}$), haplotype 5′-AAGGCGA-3′ being the most associated one ($P = 2.05 \times 10^{-22}$).

The 10 haplotypes originated by recombination and/or transversion from two ancestral haplotypes: 5′-AAGGCGA-3′ (H1) and 5′-GGAACAG-3′ (H2).

By using the combination of haplotypes H1 and H2 and the populations' origins, we carried out a partitioning analysis (Figure 1). Results indicate that homozygosity for H1 haplotype is strongly associated to blue and grey/green eye colour ($P < 0.001$). Subjects carrying H1 registered the following eye-colour percentages: 50.6% blue, 36.7% green/grey, and 12.7% brown. On the other hand, we found a large amount of brown pigmentation in subjects having at least one H2 (>80%). The effects of the other diplotypes depend on the country of origin (as shown in node 2, with $P = 0.038$). In particular, the populations of the WSR (Crimea, Georgia, Armenia, and Azerbaijan) have a significant increase of probability to have grey/green eye pigmentation (16.9%) compared with the ESR ones (Uzbekistan, Tajikistan, and Kazakhstan) (5%), thus highlighting a geographical gradient in eye pigmentation in these regions (Figure 1a).

Given the brown *versus* not-brown eye trait, we performed the algorithm for model-based recursive partitioning (Figure 1b). Interestingly, a person belonging to WSR and carrying H1/H1 genotype has a higher probability of having blue or green/grey eye (94%) as compared with a person from ESR (72%).

In the attempt to combine the roles of *OCA2* and *HERC2* genes, we used MDR method[5] (Figure 2), by using the two analysed haplotypes and rs7495174 SNP position. Our MDR analysis shows a synergistic interaction between them; the prediction accuracy of the tested model is 86.62%, with a sensitivity of 98.02%. The presence of *HERC2* H1/H1 plus the A/A *OCA2* genotype remarkably predicts blue eyes (ratio brown/not brown: 0.1324), whereas the presence of H2 compared with every alleles of *OCA2* SNP is enough to predict brown eye (Figure 2).

DISCUSSION

The Silk Road, a well-known trading route of the ancient Chinese civilisation, has been an important way for cultural and commercial exchange between subjects living in East, South, and Western Asia with those settled in the Mediterranean and European regions for almost 3000 years. The collapse of the Mongol empire and the late fifteenth-century discovery of the sea route from Europe to Asia led to a progressive decline of the Silk Road's trade in Central Asia. The geography of the Silk Road is a complex interaction between the

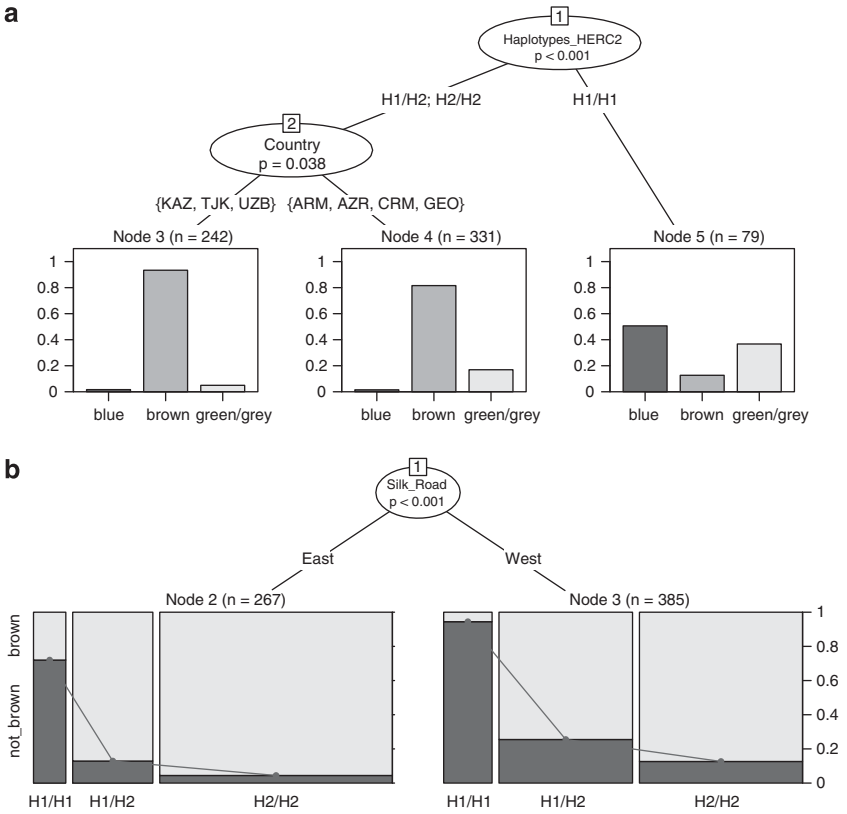

Figure 1 Results obtained from the partitioning analysis. (**a**) A regression tree with two nodes based on *HERC2* haplotypes and country (or region) of origin of each individual (ARM, Armenia; AZR, Azerbaijan; CRM, Crimea (peninsula of Ukraine); GEO, Georgia; KAZ, Kazakhstan; TJK, Tajikistan; UZB, Uzbekistan), most striking is the major node where we have the higher percentage of blue eyes (50.6%) in individuals carrying H1/H1. In the first node, the origin of population of the individuals carrying H1/H1 is not reported because the *P*-value is not significant. (**b**) Tree obtained from model-based recursive partitioning; East refers to populations from ESR, and West refers to those from WSR.

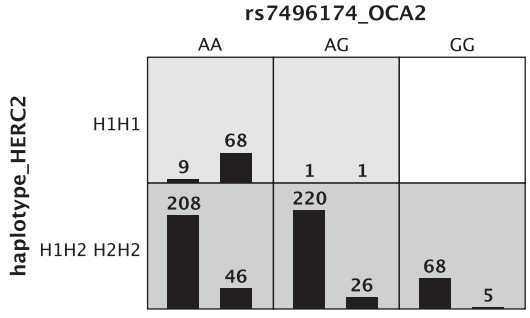

Figure 2 MDR analysis. Representation of haplotype–SNP combinations among attributes considered in interaction model for brown *versus* not brown; the dark grey cells represent the genotype combinations associated with 'high risk' of having a specific eye colour, light grey cells are associated with 'low risk', and white cells mean lack of data. Each cell is a representation of the number of individuals with brown eyes (left bar) and not brown eyes (right bar).

different physical and climate areas featuring mountains, moorlands or grasslands, and river valleys and oases, often neighbouring uninhabitable deserts. Thus, populations could be dispersed (in the grasslands) or concentrated in the oases and river valleys. The majority of the population is of mixed Turkish descent. The Uighurs are the largest ethnic group, whereas Kyrgyz, Kazaks, Uzbeks, and Tartars are further strongly represented populations. From a linguistic point of view, different varieties of old Turkish are spoken.

The colour of the eye's iris can vary dramatically between individuals, due to genetic differences. Present results confirm, also in these populations, the role of genes such as *HERC2* and *OCA2*, already known to be involved in defining iris colour. Moreover, our findings demonstrate that not only *HERC2* and *OCA2* genes but also *CTNNA2* gene could be under selection pressure. As reported by Sturm and Duffy,[7] eye colour is a feature that may fall under multiple selection pressures, including sexual, cultural, and environmental factors (ie, the level of sunlight). In this light, the presence of brown eyes, especially in populations living in the ESR, might be probably due to the combined action of both environmental and cultural factors.

One of the interesting outcomes of the present data is the demonstration on how genetic information can be used to predict eye colours. For example, homozygous individuals for a specific *HERC2* haplotype (H2) lead to a higher probability of carrying brown eyes in all populations, whereas carriers of the other haplotype (H1) have a significant probability to show a blue or green/grey iris colour. Furthermore, homozygous individuals for the same haplotype have a different probability to develop green/grey iris colour depending on the region in which they live (ie, a person belonging to the Caucasus region has an higher probability as compared with individuals living in Central Asia). An explanation for these findings is the possible presence of population-specific polymorphisms that might interact with *HERC2* and *OCA2* genes and thus contribute to the phenotypes. These polymorphisms could have a higher frequency in the Caucasus region because of a different history of gene flow between the various

populations. Apart from a significant relevance at population level, present findings might also be extremely useful in forensic medicine.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Laura Esposito, Angela D'Eustacchio, Federico Fornasier, and Emmanouil Athanasakis for their technical support. We also thank Terra Madre organization and the Terra Madre communities who participated in the project. This work is part of the scientific activities carried out within the expedition Marcopolo2010–2012 and it has largely been benefited from funds coming from the sponsors of the scientific expedition.

1 Kayser M, Liu F, Janssens AC *et al*: Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am J Hum Genet* 2008; **82**: 411–423.
2 Rennie IG: Don't it make my blue eyes brown: heterochromia and other abnormalities of the iris. *Eye (Lond)* 2012; **26**: 29–50.
3 Yang WY, Novembre J, Eskin E, Halperin E: A model-based approach for analysis of spatial structure in genetic data. *Nat Genet* 2012; **44**: 725–731.
4 Foll M, Gaggiotti OE: A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* 2008; **180**: 977–993.
5 Moore JH, Gilbert JC, Tsai CT *et al*: A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *J Theor Biol* 2006; **241**: 252–261.
6 Liu J, Hutchison K, Perrone-Bizzozero N, Morgan M, Sui J, Calhoun V: Identification of genetic and epigenetic marks involved in population structure. *PLoS One* 2010; **5**: e13209.
7 Sturm RA, Duffy DL: Human Pigmentation genes under environmental selection. *Genome Biol* 2012; **13**: 248.