Forensic Science International: Genetics 5 (2011) 170–180

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information

Susan Walsh, Fan Liu, Kaye N. Ballantyne, Mannis van Oven, Oscar Lao, Manfred Kayser *

Department of Forensic Molecular Biology, Erasmus University Medical Centre Rotterdam, Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received 23 October 2009
Received in revised form 8 February 2010
Accepted 11 February 2010

Keywords:
IrisPlex
Eye colour
Prediction
Forensic
Genotyping
HGDP-CEPH

ABSTRACT

A new era of 'DNA intelligence' is arriving in forensic biology, due to the impending ability to predict externally visible characteristics (EVCs) from biological material such as those found at crime scenes. EVC prediction from forensic samples, or from body parts, is expected to help concentrate police investigations towards finding unknown individuals, at times when conventional DNA profiling fails to provide informative leads. Here we present a robust and sensitive tool, termed IrisPlex, for the accurate prediction of blue and brown eye colour from DNA in future forensic applications. We used the six currently most eye colour-informative single nucleotide polymorphisms (SNPs) that previously revealed prevalence-adjusted prediction accuracies of over 90% for blue and brown eye colour in 6168 Dutch Europeans. The single multiplex assay, based on SNaPshot chemistry and capillary electrophoresis, both widely used in forensic laboratories, displays high levels of genotyping sensitivity with complete profiles generated from as little as 31 pg of DNA, approximately six human diploid cell equivalents. We also present a prediction model to correctly classify an individual's eye colour, via probability estimation solely based on DNA data, and illustrate the accuracy of the developed prediction test on 40 individuals from various geographic origins. Moreover, we obtained insights into the worldwide allele distribution of these six SNPs using the HGDP-CEPH samples of 51 populations. Eye colour prediction analyses from HGDP-CEPH samples provide evidence that the test and model presented here perform reliably without prior ancestry information, although future worldwide genotype and phenotype data shall confirm this notion. As our IrisPlex eye colour prediction test is capable of immediate implementation in forensic casework, it represents one of the first steps forward in the creation of a fully individualised EVC prediction system for future use in forensic DNA intelligence.

1. Introduction

Predicting externally visible characteristics (EVCs) using informative molecular markers, such as those from DNA, has started to become a rapidly developing area in forensic genetics [1]. With knowledge gleaned from this type of data, it could be viewed as a 'biological witness' tool in suitable forensic cases, leading to a new era of 'DNA intelligence' (sometimes referred to as forensic DNA phenotyping); an era in which the externally visible traits of an individual may be defined solely from a biological sample left at a crime scene or from a dismemberment of a missing person. The most relevant forensic cases for DNA-based EVC prediction would be those in which the evidence DNA sample does not match either a suspect's conventional short tandem repeat (STR) profile or any from a criminal DNA database, and also where no additional knowledge about the sample donor exists. DNA-based EVC prediction is also suitable in cases where eye witnesses are available, but their statements about the appearance of an unknown suspect may wish to be confirmed before use in intelligence work. Furthermore, in disaster victim identification or other cases of missing person identification, DNA-based EVC prediction would be useful whenever conventional STR profiles obtained do not match any putatively related individual. Unfortunately, at present, the molecular genetics of individual-specific EVCs remains largely unknown, with little expectation for immediate forensic application. However, a number of group-specific EVCs, such as eye colour, are being understood more and more in their genetic determination [2–9] and models for predicting phenotypes solely based on genotypes are being developed [10] with great promise for forensic applications. In certain cases, for example, if the police have no evidence on where/how to find a crime scene sample donor, or how to reveal the identity of a missing person, group-specific EVCs are already expected to be useful for tracing unknown individuals by focusing intelligence work on the most likely appearance group to which the individual in question belongs [1].

* Corresponding author. Tel.: +31 10 7038073; fax: +31 10 7044575.
E-mail address: m.kayser@erasmusmc.nl (M. Kayser).

1872-4973/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2010.02.004

Human eye (iris) colour is a highly polymorphic phenotype in people of European descent and, albeit less so in those from surrounding regions such as the Middle East or Western Asia [11], and is under strong genetic control [12]. Brown eye colour is assumed to reflect the ancestral human state [4] and is present everywhere in the world including Europe, although in lower frequencies, especially in its northern parts. Non-brown eye colours are assumed to be of European origin and to have been driven by positive selection starting in early European history, perhaps as a result of rare colour preferences in human mate choice [3,13]. Recent years have yielded intensive studies to increase the genetic understanding of human eye colour, via genome-wide association and linkage analysis or candidate gene studies [2–8,14–17]. The *OCA2* gene on chromosome 15 was originally thought to be the most informative human eye colour gene due to its association with the human P protein required for the processing of melanosomal proteins [9], and mutations in this gene do result in pigmentation disorders [18]. However, recent studies have shown that genetic variants in the neighbouring *HERC2* gene are more significantly associated with eye colour variation than those in *OCA2* [2–6]. Also, one of the most significant non-synonymous single nucleotide polymorphisms (SNPs) associated with eye colour, rs1800407 located in exon 12 of the *OCA2* gene, acts only as a penetrance modifier of rs12913832 in *HERC2* and is, to a lesser extent, independently associated with eye colour variation [5]. It is currently assumed that genetic variation in *HERC2* acts as a functional regulator of adjacent *OCA2* gene activity [3–5,19], although more work is needed to fully establish the functional relationship between these two genes. While the *HERC2*/*OCA2* region harbours most blue and brown eye colour information, other genes were also identified as contributing to eye colour variation, such as *SLC24A4, SLC45A2 (MATP), TYRP1, TYR, ASIP* and *IRF4,* although to a much lesser degree [2,6–8,17]. A recent study on 6168 Dutch Europeans demonstrated that with 15 eye-colour associated SNPs from eight genes, blue and brown eye colours can be predicted with >90% prevalence-adjusted accuracy and that most eye colour information is provided by a subset of just six SNPs from six genes [10].

Many of the currently known eye colour-associated SNPs, including those with high prediction value, are located in introns without functional evidence for causal trait involvement. They most likely provide eye colour information due to physical linkage with causal but currently unknown variants. This is because commercially available SNP microarrays, used in genome-wide association studies of complex traits including eye colour, are strongly biased towards non-coding markers. Due to the assumed positive selection history of non-brown eye colour in Europe, it can be expected that non-causal alleles, with association to non-brown eye colour in people of European (and neighbouring) ancestry, also exist in individuals of different ancestries that lack non-brown coloured eyes, which may result in wrong prediction outcomes. Indeed, an inspection of eye-colour associated SNPs in the limited non-European data of the International HapMap Project revealed small to considerable frequencies of blue-eye associated homozygote alleles, although blue-eyed individuals are very unlikely to occur in these East Asian and African populations. Examples are the CC/GG allele of rs916977 in the *HERC2* gene observed in 2/90 HapMap East Asians, or the TT/AA allele of rs4778138 and rs7495174, both in the *OCA2* gene, found in 5/90 and 7/90 HapMap East Asians, as well as in 3/60 and 43/60 HapMap Africans, respectively [4]. However, more detailed worldwide data are needed to assess whether DNA-based eye colour prediction only works reliably when the geographic origin of the person in question is known, e.g., from additional DNA-based ancestry testing.

Here, we have developed a single multiplex genotyping system, termed IrisPlex, for the six currently most eye-colour informative SNPs to accurately predict human blue and brown eye colour. To allow future forensic applications, we focussed on a technical platform widely applied by forensic laboratories, and investigated its degree of sensitivity. We include the prediction model which can correctly classify an individual's eye colour solely based on DNA data and illustrate the accuracy of the developed prediction system on individuals from various geographic origins. Furthermore, we applied the IrisPlex tool to the HGDP-CEPH samples representing 51 worldwide populations, and performed model-based eye colour prediction on a worldwide scale.

2. Materials and methods

2.1. Sample collection and iris photography

Buccal swabs were taken from 40 volunteers with informed consent. A photographic image of their iris was taken concurrently with a macro lens, ensuring that similar distance and light conditions were used for each photo for normalisation. Information regarding the sex and country of birth for each individual was also collected (Supplementary Table 1). DNA was extracted using the QIAamp DNA Mini kit according to the manufacturer's protocol (Qiagen, Hagen, Germany). We also obtained the H952 subset of the HGDP-CEPH samples representing 952 individuals from 51 worldwide populations [20,21]. This subset excludes duplicates, mix-ups and relatives upto the level of first-degree cousins. Due to the lack of DNA, 18 samples could not be genotyped for all markers, leaving a total of 934 worldwide HGDP samples in this study (see Supplementary Table 2).

2.2. Multiplex design, genotyping and sensitivity testing

Six SNPs; rs12913832, rs1800407, rs12896399, rs16891982, rs1393350 and rs12203592 from the *HERC2*, *OCA2*, *SLC24A4*, *SLC45A2 (MATP)*, *TYR* and *IRF4* genes, respectively, were used in this study and marker details are available in Table 1. The six PCR primer pairs were designed using the free web-based design software Primer3Plus [22] using the default parameters of the program. Each PCR fragment size was limited to less than 150 bp to cater for degraded DNA samples, vital for future application on forensic samples. The sequences surrounding the relevant SNP were searched with BLAST [23] against dbSNP [24] for other SNP sites that may interfere with primer binding, and these sites were avoided. Also, to ensure there would be little interaction between all six forward and reverse primers, the software program AutoDimer [25] was used throughout the design. The PCR primer sequences can be found in Table 1. For the single multiplex PCR, a total of 1 μl (0.5–2 ng) genomic DNA extract from each individual was amplified in a 12 μl PCR reaction with 1× PCR buffer, 2.7 mM $MgCl_2$, 200 μM of each dNTP, primer concentrations of 0.416 μM each and 0.5 U AmpliTaq Gold DNA polymerase (Applied Biosystems Inc., Foster City, CA). Thermal cycling for PCR was performed on the gold-plated 96-well GeneAmp® PCR system 9700 (Applied Biosystems). The conditions for multiplex PCR were as follows: (1) 95 °C for 10 min, (2) 33 cycles of 95 °C for 30 s and 60 °C for 30 s, (3) 5 min at 60 °C. Both forward and reverse SBE primers were designed for each SNP and the six final primers chosen were based on their suitability for the multiplex and the genotype of the resultant product to allow complete multiplexing. The primer sequences and specifications can be found in Table 1. The design followed a similar protocol to the PCR primer design ensuring primer melting temperatures of approximately 55 °C for the SBE reaction and all possible primer interactions were screened. To ensure complete capillary separation between the

products, poly-T tails of varying sizes were added to the 5′ ends of the six SBE primers. Following PCR product purification to remove unincorporated primers and dNTPs, the multiplex SBE assay was performed using 1 μl of product with 1 μl SNaPshot reaction mix in a total reaction volume of 5 μl. Thermal cycling for SBE was performed on the gold-plated 96-well GeneAmp® PCR system 9700 (Applied Biosystems). The following thermocycling programme was used: 96 °C for 2 min and 25 cycles of 96 °C for 10 s, 50 °C for 5 s and 60 °C for 30 s. Excess fluorescently labelled ddNTPs were inactivated and 1 μl of cleaned multiplex extension products were then run on an ABI 3130*xl* Genetic Analyser (Applied Biosystems) following the ABI Prism® SNaPshot kit standard protocol (Applied Biosystems). Allele calling was performed using GeneMapper v. 3.7 software (Applied Biosystems). A custom designed bin set was implemented to allow automation of genotyping. For sensitivity testing, a threshold of 50 rfu for peak intensities was adopted to ensure accuracy of genotyping. Samples from three different individuals (brown, intermediate and blue eye colour) were measured and quantified in a dilution series using the Quantifiler™ Human DNA Quantification kit (Applied Biosystems). Template concentrations from 0.5 to 0.015 ng/μl were also run in duplicate to test the overall sensitivity of the multiplex.

2.3. Statistical analysis

Liu et al. [10] have previously published the formula used in this study for eye colour prediction. It is based on a multinomial logistic regression model. The probabilities of each individual being brown (π_1), blue (π_2), and otherwise (π_3) were calculated based on the sample genotypes,

$$\pi_1 = \frac{\exp(\alpha_1 + \sum \beta(\pi_1)_k x_k)}{1 + \exp(\alpha_1 + \sum \beta(\pi_1)_k x_k) + \exp(\alpha_2 + \sum \beta(\pi_2)_k x_k)}$$

$$\pi_2 = \frac{\exp(\alpha_2 + \sum \beta(\pi_2)_k x_k)}{1 + \exp(\alpha_1 + \sum \beta(\pi_1)_k x_k) + \exp(\alpha_2 + \sum \beta(\pi_2)_k x_k)}, \quad \text{and}$$

$$\pi_3 = 1 - \pi_1 - \pi_2.$$

where x_k is the number of minor alleles of the kth SNP. The model parameters, alpha and beta were derived based on 3804 Dutch individuals in the model-building set of the previous study [10] and can be found in Supplementary Table 3 of this paper. These probabilities can be calculated using the macro provided in the supplementary material (Supplementary Table 4). Each individual is classified as being brown, blue or intermediate based on the predicted probabilities derived from the above formula. For example, a phenotypic brown-eyed individual can give a probability value of 0.76 for brown, 0.09 for blue and 0.15 for intermediate. For the worldwide distribution, a threshold of 0.7 predicted eye colour probability was used for categorisation. For example, an individual is predicted as brown if $\pi_3 > 0.7$, otherwise they are predicted as undefined. This cut off was chosen based on the receiver operating characteristic (ROC) curve derived from the Dutch study [10], where after the false positive rate of 0.3 (corresponding to specificity of 0.7), the decrease of the true positive rate becomes costly, with the possibility of errors increasing, as seen in Supplementary Fig. 1 (see below for a discussion on the selection of the appropriate threshold). To evaluate the prediction accuracy on the worldwide samples, we assumed that all individuals outside of Europe and Western Russia are brown eyed, as phenotypic data are not available for the HGDP-CEPH individuals. MapViewer 7 (Golden Software, Inc., Golden, CO, USA) package was used to plot the distribution of SNP genotypes and the predicted eye colour, on the world map. A non-metric multidimensional scaling (MDS) plot was produced to illustrate the pairwise F_{ST} distances [26] of the six eye colour SNPs between

Table 1
SNP markers included in the IrisPlex system for eye colour prediction ordered according to prediction rank with their molecular details, and genotyping information.

SNP-ID	Prediction rank from [10]	Chr position	Gene	PCR Product (bp)	Forward PCR primer (5′-3′) and Reverse PCR primer (5′-3′)	Each primer concentration (μM)	Extension Primer (5′-3′) with t-tail for length differentiation	Primer direction	Primer length no tail (bp)	Total primer length (bp)	Concentration (μM)	Tm (°C)	Alleles detected
rs12913832	1	15—26039213	*HERC2*	87	TGGCTCTCTGTCTCTGATCC GGCCCCTGATGATGATAGC	0.416	ttttttttttttttttttttt- GCGTGCAGAACTTGACA	Reverse	17	41	0.2	55.0	T/C
rs1800407	2	15—25903913	*OCA2*	127	TGAAAGGCTGCCTCTGTTCT CGATGAGACAGAGCATGATGATGA	0.416	ttttttttCCCACACCCGTCCC	Reverse	14	23	1.0	57.3	C/T
rs12896399	3	14—91843416	*SLC24A4*	104	CTGGCGATCCAATTCTTTGT CTTAGCCCTGGGTCTTCGATG	0.416	ttttttttttttttttttttttttttttta- TCTTTAGGTCACGTATATATTTTCGGG	Forward	23	53	0.15	54.5	G/T
rs16891982	4	5—33987450	*SLC45A2* (*MATP*)	128	TCCAAGTTGTGCTAGACCAGA CGAAAGAGGAGTCGAGGTTG	0.416	ttttttttttAAACACGGAGTTGATGCA	Forward	18	29	0.22	55.9	C/G
rs1393350	5	11—88650694	*TYR*	80	TTCCTCAGTCCCTTCTCTGC GGGAAGGTGAATGATAACACG	0.416	ttttttttttttttttttta- TTTGTAAAAGACCACACAGATTT	Reverse	23	47	0.1	55.6	T/C
rs12203592	6	6—341321	*IRF4*	115	ACAGGGCAGCTGATCTCTTC GCTAAACCTGGCACCAAAAG	0.416	ttttttttttttt- AAAGTACCACAGGGGAATTT	Reverse	20	35	0.3	55.2	G/A

populations, using SPSS 15.0.1 for Windows (SPSS Inc., Chicago, USA). Analysis of molecular variance (AMOVA) [27] was performed using ARLEQUIN v3.11 [28].

3. Results and discussion

3.1. IrisPlex design and sensitivity

The design of the IrisPlex assay considered PCR fragment lengths of only 80–128 bp, allowing future application to forensic samples that often contain fragmented DNA due to degradation. It was also designed so that extension products were evenly separated by 6 bp in the region of 30–65 bp in length to ensure unequivocal marker differentiation. PCR and SBE multiplex optimisations aimed to balance all SNP alleles, generating similar peak intensities to ensure genotyping accuracy in a wide range of DNA quantities. However, despite extensive efforts, allele balance was not completely achieved, e.g., allele T of rs12896399 in its heterozygote state, or allele C of rs16891982 in its homozygote state were lower in comparison (Fig. 1). Nevertheless, this slight imbalance does not affect the genotyping accuracy, unless the DNA quantity falls below the sensitivity threshold, and thus appeared sufficient for practical applications. The assay works optimally between 0.25 and 0.5 ng of template DNA, but also reveals complete 6-SNP profiles down to a level of 31 pg representing approximately six human diploid cells. Only at 15 pg of DNA template were allelic drop-outs observed for some of the SNPs (Fig. 1).

Notably, the sensitivity achieved was considerably higher than those previously reported for autosomal SNP multiplexes introduced for human identification purposes (which may be influenced by SNP numbers included). For example, 500 pg of DNA was required for a full profile of 52 SNPs analysed in two SBE multiplexes after a single multiplex PCR [29], and also for a full 20 SNP (plus amelogenin) profile from a single tube PCR reaction [30]. The sensitivity of the IrisPlex is also considerably higher than that of the commercially available AmpF*l*STR Minifiler kit (Applied Biosystems) recommended for degraded DNA typing, which requires at least 125 pg of input DNA for full profiles of eight autosomal STRs (plus amelogenin) [31]. We therefore expect the sensitivity of our IrisPlex system to meet the requirements of routine forensic applications in most cases, with it expected to be more successful than multiplex SNP/STR systems currently used in forensic practice.

3.2. Prediction probability accuracy

We previously established in a large set of 6168 Dutch Europeans that the six SNPs from six genes included in this multiplex assay carry the most eye colour prediction information from all currently known eye-colour associated SNPs [10]. Considering the area under the receiver characteristic operating curves (AUC) as an overall measure for prevalence-adjusted prediction accuracy, whereby a completely accurate prediction is obtained at an AUC of one and random prediction at 0.5, very high values for brown eyes at 0.93 and for blue eyes at 0.91 were obtained [10]. To further illustrate the predictive performance of the IrisPlex and to demonstrate the systems reliability, we generated data on 40 individuals from various geographic origins (Supplementary Table 1). Fig. 2 represents how the prediction strength is correlated with each individual's eye colour phenotype in terms of the highest probability value, from blue to brown. The images were ordered based on eye colour prediction probabilities for blue starting in the upper left corner and for brown starting in the lower right corner of the picture, with the prediction probabilities for all three eye colour categories provided for each sample and eye image. Eye colour phenotypes were not considered in the ordering of the images. It is evident that there is a clear correlation between the predicted values and the visual representation of the eye colour phenotypes viewed from top to bottom, left to right, thus confirming the accuracy of the six SNP prediction model. For 37 (92.5%) of the individuals the genetic eye colour prediction perfectly agreed with the eye colour phenotype from visual inspection (Fisher's exact test P value = 9.78×10^{-9}). Only three individuals (see red box in Fig. 2) were incorrectly categorised into their brown/blue categories by the prediction model, or were inconclusive.

From this 40 person data set, the correct call rate (sensitivity) of the model when using an accuracy of above 0.7 was 91.6% for blue eye colour categorisation and 56% for brown. However, as can be seen from the examples (albeit limited) in Fig. 2, individuals with prediction probabilities for brown between 0.5 and 0.7 also have brown eyes. Lowering the eye colour probability threshold to 0.5 resulted in 87.5% correct brown eye colour categorisation, while the 91.6% for blue remained the same. The 0.5 probability level successfully illustrates the sensitivity of the model in comparison to established data from the previously published Dutch European cohort, where sensitivity values of 88.4% for brown and 93.4% for blue eye colour characterisation were achieved [10]. Altering the probability level can achieve higher specificity levels, although this will affect the overall sensitivity of the model. For example, probability levels of 0.9 and above will increase the specificity dramatically for true blue and true brown homozygotes, with 24 out of the 40 individuals showing 100% prediction accuracy. However, using such a high threshold, light and dark intermediates that could be visually viewed as slight variations of blue or brown, respectively, would then fail to be categorised into blue or brown. So far, these "intermediate" eye colours are more challenging to define using the present prediction model and the currently available SNPs. Notably, in our previous study involving several thousand Dutch Europeans, we observed that at the 0.5 threshold, the prediction accuracy for intermediate (i.e. non-blue/non-brown) eye colours was considerably lower at only 0.73 than that seen for blue and brown colours at >0.91 [10]. We hypothesised that the lower prediction accuracy reached for these intermediate colours may be explained by imprecise phenotype categorisation or the result of unidentified genetic determinants [10]. Hence, more work is needed to find genetic variants with high predictive value for the non-blue and non-brown eye colours. Finally, discrepancies between genetically predicted and true phenotypic eye colour may be caused by the fact that eye colour can change over one's lifetime. However, as this is a rare phenomenon [32], it is not expected to affect our prediction test significantly, but may be a contributing factor as to why we could not assign three test individuals in this study correctly, as well as deviations from 100% prediction accuracy in our previous study [10].

Each of the six SNPs included in the IrisPlex system provides mounting genetic information towards the overall prediction accuracy achievable with this DNA test system, although with different input. As previously established [10], rs12913832 in the *HERC2* gene alone carries most of the eye colour predictive information with an AUC of 0.899 for brown and 0.877 for blue achieved with this single SNP. This is in line with previous association studies showing that this SNP is the most strongly eye-colour associated SNP currently known [3,5,10]. The additional five SNPs from the additional five genes *OCA2*, *SLC24A4*, *SLC45A2 (MATP)*, *TYR* and *IRF4* included in the assay slightly increased the prediction accuracy as reflected in the prediction rank established previously [10] due to their lower (but still significant) eye colour association as established previously [2,6,7,10]. Notably, two of them, rs1393350 from the *TYR* gene and rs12896399 from the *SLC24A4* gene, reached much lower P values for association when

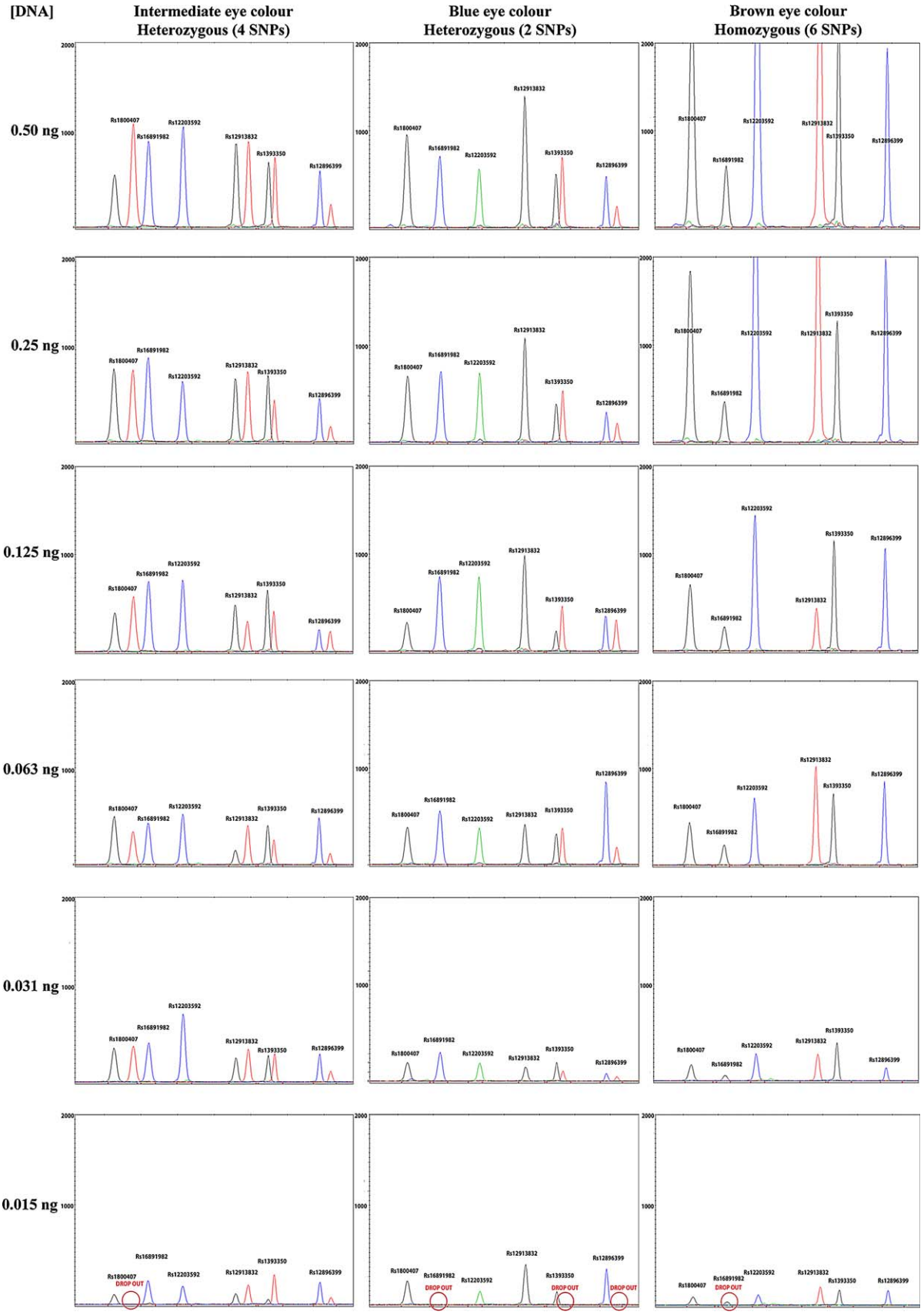

Fig. 1. Sensitivity testing of the IrisPlex genotyping assay for eye colour prediction. Multiplex single base extension (SBE) products from starting DNA amounts between 0.5 ng and 15 pg used in the prior single multiplex PCR are shown for three individuals with intermediate, blue and brown eye colour phenotypes and confirmatory IrisPlex eye colour prediction results. Allelic drop-outs are designated by red circles.

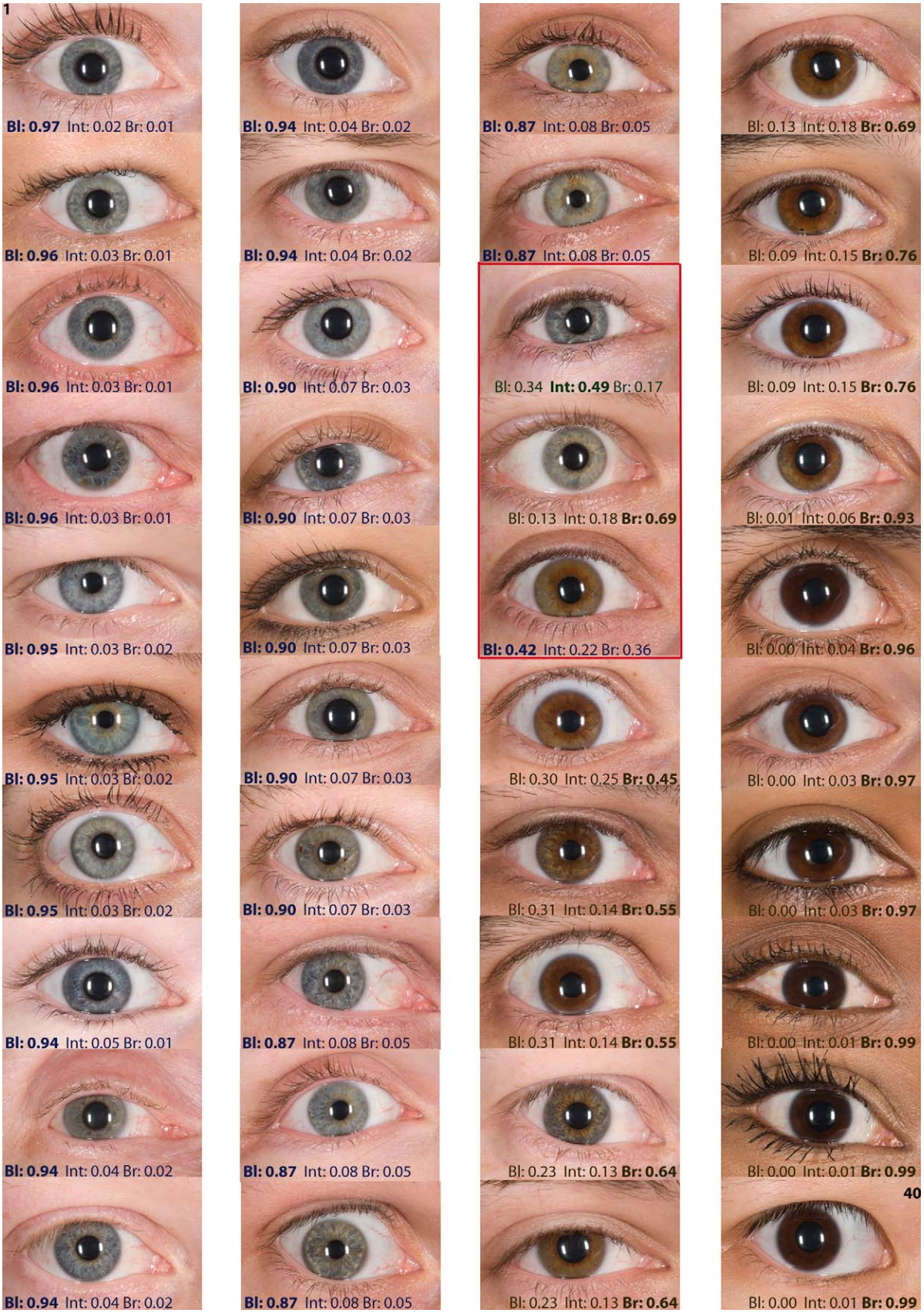

Fig. 2. Eye colour images of 40 individuals of various geographic origins and their estimated IrisPlex eye colour prediction probabilities for blue (Bl), intermediate (Int), and brown (Br) respectively. The highest of the three prediction probabilities per individual is highlighted in bold. Eye colour images are ordered based on maximal IrisPlex prediction probabilities for blue starting from the upper left corner and for brown starting from the lower right corner. Phenotype information was not considered in the ordering of the eye pictures. The area surrounded by the red line marks three cases with incorrect or non-conclusive IrisPlex eye colour prediction results. For IrisPlex genotype data and sample information see Supplementary Table 1.

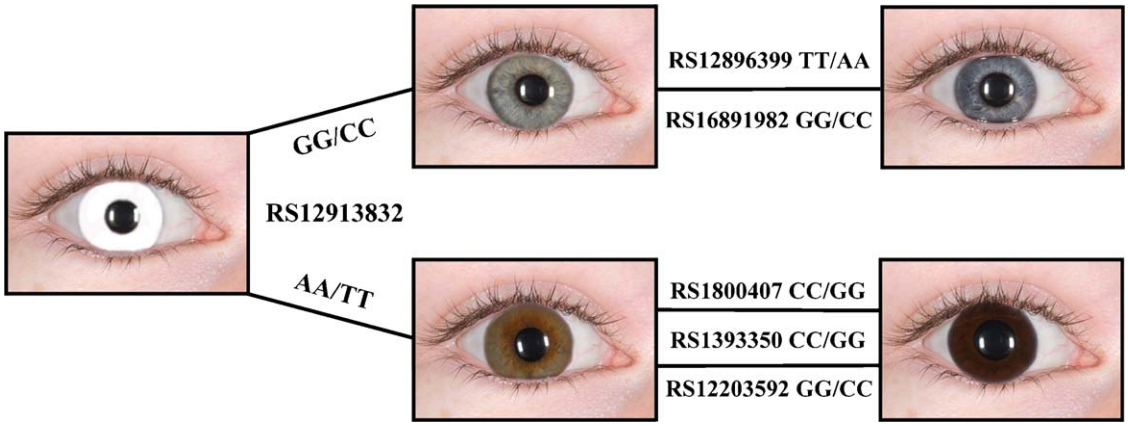

Fig. 3. Hypothesised scenario for genetic determination of brown and blue eye colours showing the impact of the most influential SNP genotypes from the 6-SNP model.

comparing individuals with blue versus green eyes relative to blue versus brown eyes previously [2]. This may indicate that they contribute more to the blue and intermediate prediction and less to the brown prediction. The *P* values and adjusted rs12913832 beta values for the six SNPs involved in this prediction model can be found in the supplementary material of Liu et al. [10] as the six highest ranking SNPs. In general, to understand the impact of each SNP on the prediction model, two scenarios have been displayed in Fig. 3 for the genetic variation in eye colour based on the six SNPs presented. As depicted, the major impact in determining whether the eye colour will be brown versus non-brown comes from rs12913832 (*HERC2*) with its AA/TT versus GG/CC homozygote genotypes. Further determination of non-brown is provided by rs12896399 (*SLC24A4*) and its TT/AA, as well as by rs16891982 (*SLC45A2 (MATP)*) and its GG/CC homozygote genotype. On the other hand, further darkening of brown is determined by the homozygote genotype CC/GG of rs1800407 (*OCA2*) and rs1393350 (*TYR*), respectively, as well as the GG/CC homozygote genotype of rs12203592 (*IRF4*).

3.3. Genetic diversity and eye colour prediction on a worldwide scale

Fig. 4 displays the genotypes of 934 individuals from 51 HGDP-CEPH populations for each of the six SNPs included in the multiplex prediction test (see Supplementary Table 2 for the genotype data). Fig. 4a presents the most eye-colour associated and the highest prediction-ranking rs12913832 (*HERC2*) SNP. It is apparent that the blue-eye associated homozygote genotype (CC/GG) as well as the heterozygote genotype, are both almost exclusively restricted to Europe and the surrounding areas such as the Middle East and West Asia, where blue and intermediate eye colours are expected. On the other hand, the brown eye associated homozygote genotype (TT/AA) of this SNP exists everywhere in the world and is (almost) the only genotype found in areas such as East Asia, Oceania and Sub-Saharan Africa where only brown eye colour is expected. Our HGDP data on the Japanese population confirm a recent study, which demonstrated that all the Japanese individuals involved had brown eyes and carried the TT/AA genotype [33]. The geographic pattern observed for rs12913832 (*HERC2*) is also evident for rs16891982 (*SLC45A2 (MATP)*), ranked number 4 in the Dutch cohort prediction [10]. However, unlike rs12913832 (HERC2), the blue-eye associated homozygote genotype of rs16891982 (*SLC45A2 (MATP)*) is of much higher frequency in Europe, Middle East and West Asia. rs1800407 (*OCA2*), ranked number 2 [10] (Fig. 4b), is postulated to act as a penetrance modifier on rs12913832 (*HERC2*), and is less defined in its homozygote blue eye genotype, which is very rare, but does display the heterozygous genotype to a considerable degree within Europe, with minor frequency in the Middle East and West Asia and quite uncommon in the rest of the world. Similar findings are obtained for rs1393350 (*TYR*), ranked number 5 [10] (Fig. 4e) and rs12203592 (*IRF4*) ranked number 6 in the large Dutch population study [10] (Fig. 4f). rs12896399 (*SLC24A4*) displays no recognisable trend in the geographic distribution of the two alleles (Fig. 4c), which is remarkable as this SNP was ranked third best in the Dutch cohort [10]. Hence, apart from rs12896399, there is an increase in frequency of the blue eye associated homozygote and the heterozygote genotypes towards Europe and, albeit less so in the Middle East and West Asia, which corroborates the degree of expected eye colour variation in these regions. Conversely, brown-eye associated homozygote genotypes are predominant in East Asia, Sub-Saharan Africa, Native America and Oceania in agreement with the expected monomorphy of eye colour (brown) in these areas.

The variation in worldwide allele distributions between these six SNPs underlines the importance of using a combined SNP model to accurately predict eye colour. Fig. 5 is an illustration of the predicted eye colours of the HGDP-CEPH worldwide panel using this model, in which a probability threshold of 0.7 was applied. The results clearly demonstrate that blue eye colour is only predicted in Europe, and, albeit more rarely so in the Middle East and West Asia, but never elsewhere in the world. In particular, blue eye colour was predicted in Europeans (including Western Russians) with an average probability of 0.86. Similarly, individuals with predicted non-blue and non-brown colours, which are included in the prediction group below the probability threshold, are mostly observed in Europe and, albeit less so in the Middle East and West Asia, but never elsewhere in the world (with the exception of a single individual from Brazil with a brown probability of 0.48 and two from Algeria with brown probabilities just short of the 0.7 threshold at 0.69). Moreover, brown eye colour is predicted everywhere in the world but is the only predicted eye colour in East Asia, Oceania, Sub-Saharan Africa and Native America (with the noted single exception). In particular, brown eye colour in the HGDP samples from outside Europe, Middle East and West Asia, i.e. in regions where only brown eyes are expected, was predicted with an average probability of 0.997. Unfortunately, there are no individual eye colour phenotypes available for the HGDP-CEPH samples; but our DNA-predicted eye colour results are in agreement with general knowledge and reported data [11] on the distribution of eye colour phenotypic variation around the world. However, without eye colour phenotype information we

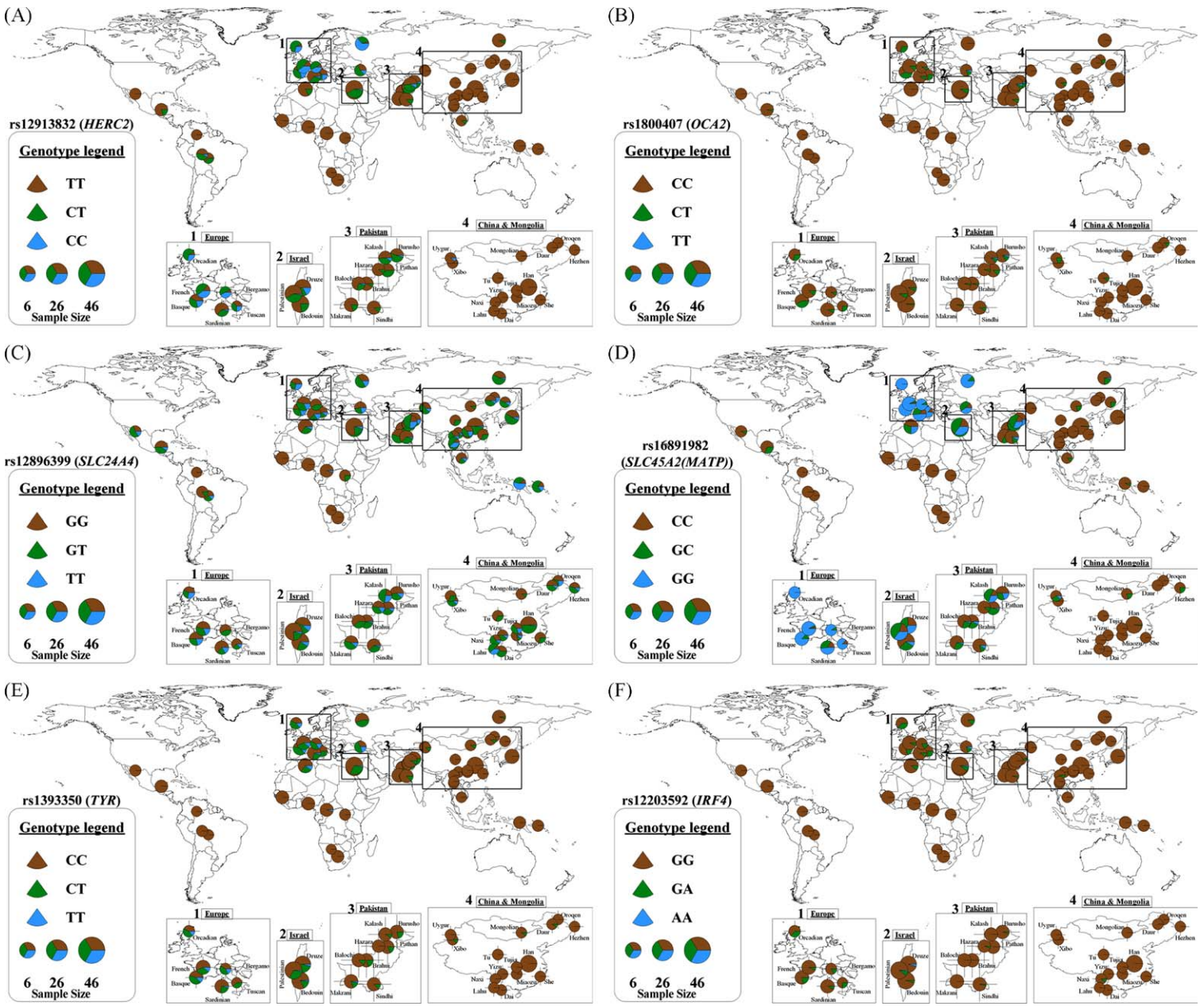

Fig. 4. Worldwide genotype distribution of the six IrisPlex SNPs in 934 individuals of the H952 HGDP-CEPH set from 51 worldwide population groups, in order of prediction rank revealed from a large Dutch cohort [10]: (a) rs12913832 (*HERC2*), (b) rs1800407 (*OCA2*), (c) rs12896399 (*SLC24A4*), (d) rs16891982 (*SLC45A2(MATP)*), (e) rs1393350 (*TYR*), (f) rs12203592 (*IRF4*). Blue indicates the proportion of individuals with blue-eye-associated homozygote genotypes as revealed from previous European studies, brown indicates the proportion of individuals with brown-eye-associated homozygote genotypes from previous European studies, and green indicates the proportion of individuals with heterozygote genotypes. For genotype data and sample information see Supplementary Table 2.

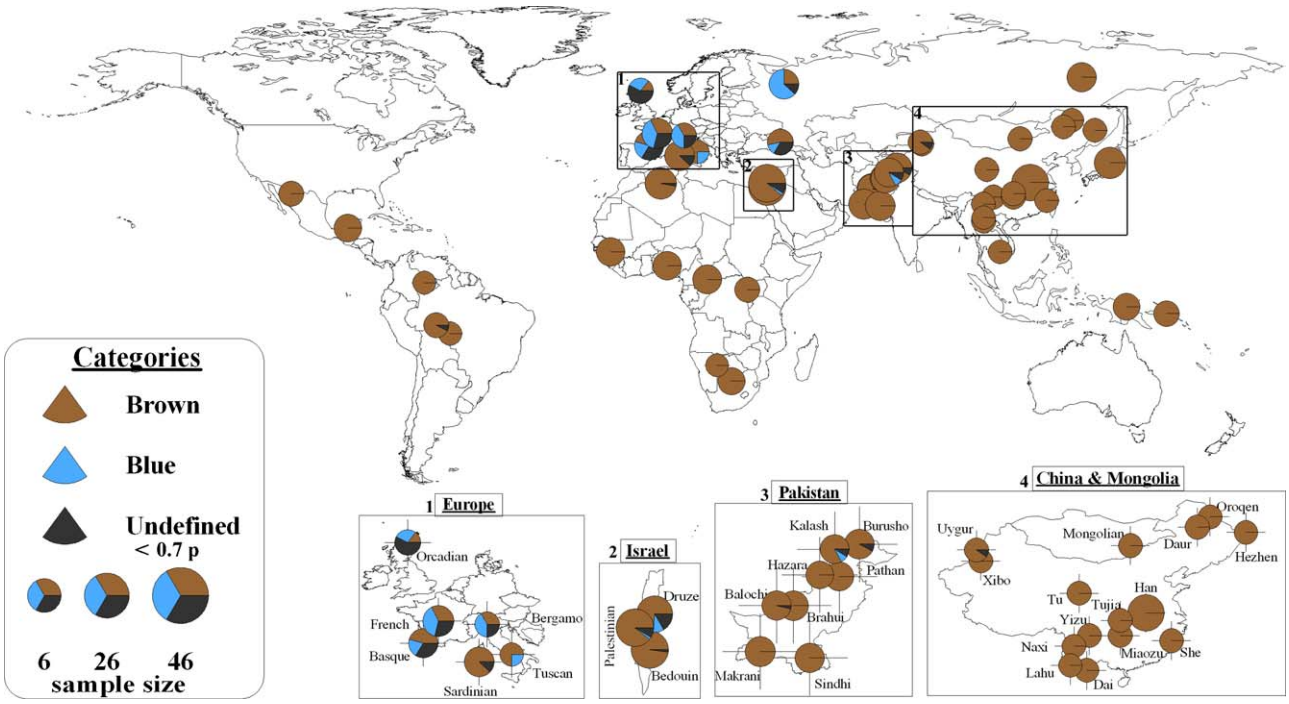

Fig. 5. IrisPlex eye colour prediction on a worldwide scale, using 934 individuals of the H952 HGDP-CEPH set from 51 worldwide populations and applying a prediction probability threshold of 0.7. Blue indicates the proportion of individuals with predicted blue eye colour, brown indicates the proportion of individuals with predicted brown eye colour, and black indicates undefined individuals given the prediction probability threshold applied. For IrisPlex eye colour prediction probability values see Supplementary Table 2.

cannot exclude for certain the existence of non-brown eye colour outside Europe that remains undetectable via the SNPs used here (which were all identified in previous association studies on European populations). Nonetheless, we regard this scenario as highly unlikely given the assumed European origin of non-brown eye colour variation [13]. For additional confirmation of the reliable worldwide use of our eye colour prediction test without prior ancestry information, we would also like to emphasise that the brown eye colour of all of the individuals from the 40-person illustration dataset whose country of origin is outside of Europe were predicted correctly (Supplementary Table 1).

3.4. Ancestry inference with eye colour SNPs

A notion that has been advocated in the past, is inferring biogeographic ancestry (or genetic origins) from DNA markers derived from pigmentation genes [34]. We tested the power of the six eye colour predictive SNPs for differentiating worldwide human populations and individuals. First, we asked by means of an AMOVA test how much of the total genetic variation provided by the six SNPs is explained by geography with assigning the 51 HGDP-CEPH populations to seven regional geographic groups. A variance proportion of 24.1% was estimated from 10100 permutations, which was statistically significant ($P < 0.000005$). However, AMOVA based on predicted eye colour grouping (brown, blue and undefined using a probability threshold of 0.7) resulted in an increased variance proportion of 48.7% ($P < 0.000005$). Hence, although a considerable and significant proportion of genetic eye colour variance is indeed explained by geography, about twice as much is explained when considering predicted eye colour, as may be expected. Since the eye colour prediction was solely based on the genetic variation (not using phenotype information), the AMOVA results highlight the presence of genetic homogeneity within each predicted eye colour category.

To understand the geographic information content provided by the six eye colour SNPs in more detail, we performed a non-metric multidimensional scaling (MDS) analysis of F_{ST} values estimated between pairs of all 51 HGDP-CEPH populations. As seen from the plot which was performed using $k = 2$ dimensions (Fig. 6, *S*-stress value 0.05998), all central, eastern and western European populations, which carry considerable amounts of predicted non-brown eyed individuals, cluster together and separately from all African, East Asian, Native American, Oceania populations as

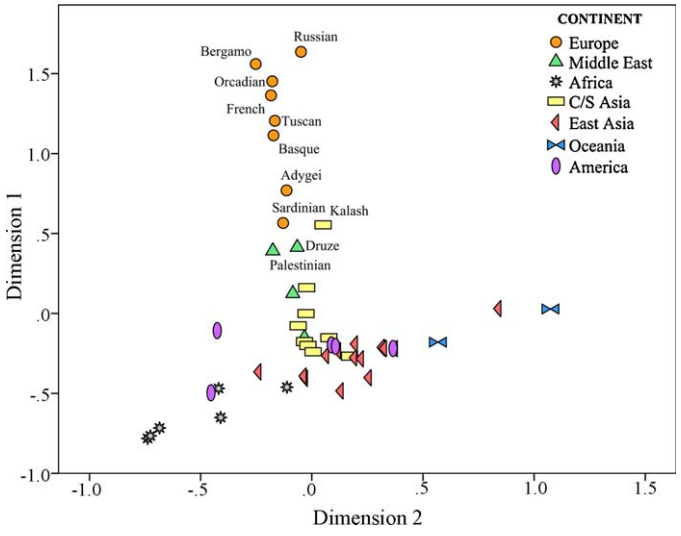

Fig. 6. Non-metric multidimensional scaling (MDS) plot of the pairwise F_{ST} distances between HGDP-CEPH populations using the six IrisPlex SNPs, colour code is according to geographic regions as provided in the legend. All populations with variation in IrisPlex predicted eye colour are given with names.

well as most of the Central South Asian and some Middle Eastern groups, i.e. all populations where brown was the only predicted eye colour. The two southern European populations cluster together with the particular Middle Eastern and Pakistani groups who included low numbers of predicted non-brown eyed individuals; they all appeared somewhat between the non-southern Europeans on one side, and the remaining worldwide populations on the other side. Hence, on the population level, European geographic information can be inferred from the eye colour SNPs used here (perhaps with the exception of southern Europe). However, on the individual level, which is of a greater concern in forensic applications, the situation appears different. No clear geographic clustering of individuals was evident in a MDS plot of identical-by-state distances obtained from the genotypes of the six eye colour SNPs (data not shown). Noteworthy, European individuals are indeed differentiable from non-Europeans via hundreds of thousands of "random" SNPs [35,36]. Also, European individuals, together with their neighbours from the Middle East and West Asia, can be differentiated from other worldwide individuals using small numbers of carefully ascertained ancestry-sensitive SNPs either obtained from regions outside pigmentation genes [37,38], or applying a combination of markers from both pigmentation genes and other genomic regions [39,40].

3.5. Conclusions

Here we present a robust and sensitive DNA tool, termed IrisPlex, for the accurate prediction of blue and brown eye colour. The developed multiplex genotyping system based on the six currently most eye-colour informative SNPs (i) allows prediction of blue and brown eye colour with high levels of accuracy, (ii) is extremely sensitive allowing successful analyses of picogram amounts of DNA, (iii) is designed to cater for degraded DNA, and (iv) is based on a genotyping technology that relies on equipment widely used by the forensic community. Hence, the IrisPlex system is highly suitable for application to forensic casework, including those with limited DNA quantity and quality. Our data from applying this system to eye colour prediction on a worldwide scale provided supporting evidence that correct interpretation of blue and brown eye colour prediction does not require additional ancestry information when this test and model are used. However, even considering this supporting evidence, it would still be of interest to perform a worldwide study on eye colour prediction where phenotypic data is available. Also, future research into the genetic basis of non-blue and non-brown eye colours will need to show if such colours can be predicted with similarly high levels of accuracy as already possible for blue and brown eye colour representing the two extremes of the continuous eye colour distribution. As EVC prediction can create many new avenues of investigation combined with other means of intelligence, the IrisPlex eye colour prediction system presented here is expected to become of great benefit to the forensic community in the coming years.

Acknowledgements

We are grateful to all volunteers who provided eye pictures as well as DNA samples for this study and to Ruud Koppenol for the eye photography. We thank Peter de Knijff and Marjan Sjerps for useful discussion on this matter. We additionally thank the HGDP-CEPH donors for their contribution to this panel, and Howard Cann for providing the DNA samples. This work was funded by the Netherlands Forensic Institute (NFI), and received additional support by a grant from the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands (FGCN).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2010.02.004.

References

[1] M. Kayser, P.M. Schneider, DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations, Forensic Sci. Int. Genet. 3 (2009) 154–161.
[2] P. Sulem, D.F. Gudbjartsson, S.N. Stacey, A. Helgason, T. Rafnar, K.P. Magnusson, A. Manolescu, A. Karason, A. Palsson, G. Thorleifsson, M. Jakobsdottir, S. Steinberg, S. Palsson, F. Jonasson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K.R. Benediktsdottir, K.K. Aben, L.A. Kiemeney, J.H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir, K. Stefansson, Genetic determinants of hair, eye and skin pigmentation in Europeans, Nat. Genet. 39 (2007) 1443–1452.
[3] H. Eiberg, J. Troelsen, M. Nielsen, A. Mikkelsen, J. Mengel-From, K. Kjaer, L. Hansen, Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression, Hum. Genet. 123 (2008) 177–187.
[4] M. Kayser, F. Liu, A.C.J.W. Janssens, F. Rivadeneira, O. Lao, K. van Duijn, M. Vermeulen, P. Arp, M.M. Jhamai, W.F.J. van Ijcken, J.T. den Dunnen, S. Heath, D. Zelenika, D.D.G. Despriet, C.C.W. Klaver, J.R. Vingerling, P.T.V.M. de Jong, A. Hofman, Y.S. Aulchenko, A.G. Uitterlinden, B.A. Oostra, C.M. van Duijn, Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene, Am. J. Hum. Genet. 82 (2008) 411–423.
[5] R.A. Sturm, D.L. Duffy, Z.Z. Zhao, F.P.N. Leite, M.S. Stark, N.K. Hayward, N.G. Martin, G.W. Montgomery, A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color, Am. J. Hum. Genet. 82 (2008) 424–431.
[6] J. Han, P. Kraft, H. Nan, Q. Guo, C. Chen, A. Qureshi, S.E. Hankinson, F.B. Hu, D.L. Duffy, Z.Z. Zhao, N.G. Martin, G.W. Montgomery, N.K. Hayward, G. Thomas, R.N. Hoover, S. Chanock, D.J. Hunter, A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation, PLoS Genet. 4 (2008) e1000074.
[7] P. Sulem, D.F. Gudbjartsson, S.N. Stacey, A. Helgason, T. Rafnar, M. Jakobsdottir, S. Steinberg, S.A. Gudjonsson, A. Palsson, G. Thorleifsson, S. Palsson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K.R. Benediktsdottir, K.K. Aben, S.H. Vermeulen, A.M. Goldstein, M.A. Tucker, L.A. Kiemeney, J.H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir, K. Stefansson, Two newly identified genetic determinants of pigmentation in Europeans, Nat. Genet. 40 (2008) 835–837.
[8] P.A. Kanetsky, J. Swoyer, S. Panossian, R. Holmes, D. Guerry, T.R. Rebbeck, A polymorphism in the Agouti signaling protein gene Is associated with human pigmentation, Am. J. Hum. Genet. 70 (2002) 770–775.
[9] T.R. Rebbeck, P.A. Kanetsky, A.H. Walker, R. Holmes, A.C. Halpern, L.M. Schuchter, D.E. Elder, D. Guerry, P gene as an inherited biomarker of human eye color, Cancer Epidemiol. Biomarkers Prev. 11 (2002) 782–784.
[10] F. Liu, K. van Duijn, J.R. Vingerling, A. Hofman, A.G. Uitterlinden, A.C.J.W. Janssens, M. Kayser, Eye color and the prediction of complex phenotypes from genotypes, Curr. Biol. 19 (2009) R192–R193.
[11] R.L. Beals, H. Hoijer, An Introduction to Anthropology, Macmillan, New York, 1965
[12] R.A. Sturm, T.N. Frudakis, Eye colour: portals into pigmentation genes and ancestry, Trends Genet. 20 (2004) 327–332.
[13] P. Frost, European hair and eye color: a case of frequency-dependent sexual selection? Evol. Hum. Behav. 27 (2006) 85–103.
[14] D.L. Duffy, G.W. Montgomery, W. Chen, Z.Z. Zhao, L. Le, M.R. James, N.K. Hayward, N.G. Martin, R.A. Sturm, A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation, Am. J. Hum. Genet. 80 (2007) 241–252.
[15] G. Zhu, D.M. Evans, D.L. Duffy, G.W. Montgomery, S.E. Medland, N.A. Gillespie, K.R. Ewen, M. Jewell, Y.W. Liew, N.K. Hayward, R.A. Sturm, J.M. Trent, N.G. Martin, A genome scan for eye color in 502 twin families: most variation is due to a QTL on chromosome 15q, Twin Res. 7 (2004) 197–210.
[16] D. Posthuma, P.M. Visscher, G. Willemsen, G. Zhu, N.G. Martin, P.E. Slagboom, E.J. de Geus, D.I. Boomsma, Replicated linkage for eye color on 15q using comparative ratings of sibling pairs, Behav. Genet. 36 (2006) 12–17.
[17] T.N. Frudakis, M. Thomas, Z. Gaskin, K. Venkateswarlu, K.S. Chandra, S. Ginjupalli, S. Gunturi, S. Natrajan, V.K. Ponnuswamy, K.N. Ponnuswamy, Sequences associated with human iris pigmentation, Genetics 165 (2003) 2071–2083.
[18] M.H. Brilliant, The mouse p (pink-eyed dilution) and human P genes, oculocutaneous albinism type 2 (OCA2), and melanosomal pH, Pigment Cell Res. 14 (2001) 86–93.
[19] R.A. Sturm, Molecular genetics of human pigmentation diversity, Hum. Mol. Genet. 18 (2009) R9–17.
[20] N.A. Rosenberg, Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives, Ann. Hum. Genet. 70 (2006) 841–847.
[21] N.A. Rosenberg, J.K. Pritchard, J.L. Weber, H.M. Cann, K.K. Kidd, L.A. Zhivotovsky, M.W. Feldman, Genetic structure of human populations, Science 298 (2002) 2381–2385.
[22] A. Untergasser, H. Nijveen, X. Rao, T. Bisseling, R. Geurts, J.A.M. Leunissen, Primer3Plus, an enhanced web interface to Primer3, Nucleic Acids Res. 35 (2007) W71–74.

[23] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[24] S.T. Sherry, M.-H. Ward, M. Kholodov, J. Baker, L. Phan, E.M. Smigielski, K. Sirotkin, dbSNP: the NCBI database of genetic variation, Nucleic Acids Res. 29 (2001) 308–311.
[25] P.M. Vallone, J.M. Butler, AutoDimer: a screening tool for primer-dimer and hairpin structures, Biotechniques 37 (2004) 226–231.
[26] B.S. Weir, C. Cockerham, Estimating F-statistics for the analysis of population structure, Evolution 38 (1984) 1358–1370.
[27] L. Excoffier, P.E. Smouse, J.M. Quattro, Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data, Genetics 131 (1992) 479–491.
[28] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis, Evol. Bioinform. Online 1 (2005) 47–50.
[29] J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, C.D. Harrison, E. Musgrave-Brown, A. Salas, D. Syndercombe-Court, P.M. Schneider, A. Carracedo, N. Morling, A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724.
[30] L.A. Dixon, C.M. Murray, E.J. Archer, A.E. Dobbins, P. Koumi, P. Gill, Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes, Forensic Sci. Int. 154 (2005) 62–77.
[31] J.J. Mulero, C.W. Chang, R.E. Lagacé, D.Y. Wang, J.L. Bas, T.P. McMahon, L.K. Hennessy, Development and validation of the AmpFlSTR MiniFiler PCR Amplification Kit: a MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA, J. Forensic Sci. 53 (2008) 838–852.
[32] L.Z. Bito, A. Matheny, K.J. Cruickshanks, D.M. Nondahl, O.B. Carino, Eye color changes past early childhood: the Louisville twin study, Arch. Ophthalmol. 115 (1997) 659–663.
[33] R. Iida, M. Ueki, H. Takeshita, J. Fujihara, T. Nakajima, Y. Kominato, M. Nagao, T. Yasuda, Genotyping of five single nucleotide polymorphisms in the OCA2 and HERC2 genes associated with blue-brown eye color in the Japanese population, Cell Biochem. Funct. 27 (2009) 323–327.
[34] H. Pulker, M.V. Lareu, C. Phillips, A. Carracedo, Finding genes that underlie physical traits of forensic interest using genetic tools, Forensic Sci. Int. Genet. 1 (2007) 100–104.
[35] J.Z. Li, D.M. Absher, H. Tang, A.M. Southwick, A.M. Casto, S. Ramachandran, H.M. Cann, G.S. Barsh, M.W. Feldman, L.L. Cavalli-Sforza, R.M. Myers, Worldwide human relationships inferred from genome-wide patterns of variation, Science 319 (2008) 1100–1104.
[36] M. Jakobsson, S.W. Scholz, P. Scheet, J.R. Gibbs, J.M. VanLiere, H.-C. Fung, Z.A. Szpiech, J.H. Degnan, K. Wang, R. Guerreiro, J.M. Bras, J.C. Schymick, D.G. Hernandez, B.J. Traynor, J. Simon-Sanchez, M. Matarin, A. Britton, J. van de Leemput, I. Rafferty, M. Bucan, H.M. Cann, J.A. Hardy, N.A. Rosenberg, A.B. Singleton, Genotype, haplotype and copy-number variation in worldwide human populations, Nature 451 (2008) 998–1003.
[37] O. Lao, K. van Duijn, P. Kersbergen, P. de Knijff, M. Kayser, Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry, Am. J. Hum. Genet. 78 (2006) 680–690.
[38] P. Kersbergen, K. van Duijn, A.D. Kloosterman, J.T. den Dunnen, M. Kayser, P. de Knijff, Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans, BMC Genet. 10 (2009) 69.
[39] C. Phillips, A. Salas, J.J. Sanchez, M. Fondevila, A. Gomez-Tato, J. Alvarez-Dios, M. Calaza, M.C. de Cal, D. Ballard, M.V. Lareu, A. Carracedo, Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs, Forensic Sci. Int. Genet. 1 (2007) 273–280.
[40] D. Corach, O. Lao, C. Bobillo, K. van der Gaag, S. Zuniga, M. Vermeulen, K. van Duijn, M. Goedbloed, P.M. Vallone, W. Parson, P. de Knijff, M. Kayser, Inferring continental ancestry of Argentineans from autosomal, Y-chromosomal and mitochondrial DNA, Ann. Hum. Genet. 74 (2010) 65–76.