UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Single Nucleotide Polymorphisms – from forensic identification to DNA phenotyping

Doutoramento em Biologia
Especialidade em Genética

Paulo Alexandre Paisana da Silva Dario

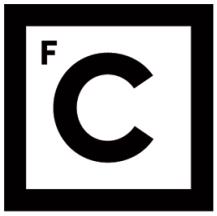
Tese orientada por:
Professor Doutor Francisco Corte-Real
Professora Doutora Deodália Dias

Documento especialmente elaborado para a obtenção do grau de doutor

2017

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS



Ciências
ULisboa

Single Nucleotide Polymorphisms – from forensic identification to DNA phenotyping

Doutoramento em Biologia
Especialidade em Genética

Paulo Alexandre Paisana da Silva Dario

Tese orientada por:
Professor Doutor Francisco Corte-Real
Professora Doutora Deodália Dias

Júri:
Presidente:

- Doutor Pedro Miguel Alfaia Barcia Ré, Professor Associado com Agregação
  Faculdade de Ciências da Universidade de Lisboa

Vogais:

- Doutor Francisco Manuel Andrade Corte-Real Gonçalves, Professor Associado com Agregação
  Faculdade de Medicina da Universidade de Coimbra e Presidente do Instituto Nacional de Medicina Legal e Ciências Forenses, I.P. (orientador)
- Doutora Ana Teresa Luís Lopes Maia, Professora Auxiliar
  Departamento de Ciências Biomédicas e Medicina da Universidade do Algarve
- Doutora Vanessa Raquel Branco Bogas, Especialista Superior de Medicina Legal
  Instituto Nacional de Medicina Legal e Ciências Forenses, I.P.
- Doutora Deborah Penque, Investigadora Auxiliar
  Departamento de Genética Humana do Instituto Nacional de Saúde Doutor Ricardo Jorge
- Doutora Maria Teresa Ferreira Ramos Nabais Oliveira Rebelo, Professora Auxiliar
  Faculdade de Ciências da Universidade de Lisboa
- Doutora Maria Helena Mouriño Silva Nunes, Professora Auxiliar
  Faculdade de Ciências da Universidade de Lisboa

Documento especialmente elaborado para a obtenção do grau de doutor

2017

Às mulheres que fazem o círculo da minha vida e que o preenchem com todas as cores da felicidade. À mulher que me deu vida, a Fina mas que chamo de Mãe, à mulher com quem criei vida, a Rita mas que chamo de Amor, e às nossas filhas, a Beatriz e a Leonor, obras-primas da nossa existência. Elas são o exemplo vivo daquilo que precisei para a realização desta tese e daquilo que preciso todos os dias da minha vida: a paciência, a alegria, o entusiasmo, a perseverança… resumindo, o Amor, pelos outros, por nós próprios e a dedicação aquilo a que nos propomos.

Agradecimentos

Ao finalizar este trabalho, não podia deixar de agradecer ao conjunto de pessoas que foram relevantes para elaboração deste projeto, por todo o apoio que deram para a sua realização. As palavras que lhes dirijo não são certamente todas as que merecem mas são o reflexo do meu sentimento de gratidão.

Ao Professor Doutor Francisco Corte-Real por ter acreditado nesta ideia, pela amizade, por ter aceitado a orientação deste Doutoramento e também por ter providenciado condições necessárias para a prossecução do mesmo.

À Professora Doutora Deodália Dias pela amizade, pela orientação desta Tese, pela sua ajuda, pelo seu entusiasmo e pelo estímulo que me transmitiu para que este projeto fosse em frente.

À Professora Doutora Helena Geada, pelo apoio e encorajamento para "embarcar nesta aventura" bem como pelos seus ensinamentos, pelos seus conselhos e pela amizade… sem o seu conselho este projeto nunca teria começado.

Ao Professor Doutor Jorge Costa Santos por ter confiado em mim e neste projeto tendo assegurado as condições necessárias para a realização do mesmo.

À Dra. Maria João Porto enquanto Diretora do Serviço de Genética e Biologia Forenses, bem como pelo seu valioso contributo científico na produção de artigos que deste trabalho resultaram.

À Dra. Teresa Ribeiro pelo apoio dado no Serviço, pela amizade e pela transmissão dos seus conhecimentos e experiência que muito serviram para a realização desta tese bem como para a minha preparação profissional.

À Doutora Mónica Carvalho, pela sua boa disposição e espírito positivo que por vezes permitiu um pensamento diferente em alturas em que os tempos eram mais difíceis.

À Professora Doutora Helena Mouriño pela amizade, pelo auxílio na análise estatística e pela bem-disposta ajuda que prestou e sem a qual esta tese seria certamente mais pobre.

À Dra. Manuela Marques, pelo apoio bibliográfico, por toda a investigação, por todas as dicas e ajuda dada para a produção deste trabalho bem como pela longa e grande amizade.

À Dra. Isabel Lucas e à Mestre Catarina Dourado, pelo carinho, pelo companheirismo, pela amizade e por todo o apoio que me prestaram nos bons momentos e naqueles que se revelaram menos bons, pelo seu auxílio no laboratório bem como em todo o muito trabalho com que temos que lidar no dia-a-dia.

Aos meus amigos, pelo seu apoio e incentivo, com um pensamento especial para aqueles que também estão "infetados com o bichinho forense", nomeadamente à Minda, à Filipa e à Dina.

À minha Mãe que sempre me apoiou em todos os momentos da minha vida, aos meus incríveis irmãos Zé-Tó e Nelita, ao Zé e à Joana bem como ao Pedro, à Zé, à Diana e à avó Ilda. Ao meu Pai que, já não estando entre nós, estará a tecer-me os maiores elogios aos que também já partiram. Obrigado por serem a melhor família do mundo e por todo o apoio que me deram e que me dão.

Finalmente, para a Rita pela sua ajuda, paciência, dedicação, companheirismo e acima de tudo, amor e para as nossas filhas, Beatriz e Leonor, que iluminam e aquecem o meu coração e às quais tive de privar muito do meu tempo para me poder dedicar a este projeto.

A todos o meu forte e sentido agradecimento!

Resumo

Os marcadores de nucleótido único ou SNPs (do inglês *single nucleotide polimorphisms*) apresentam baixas taxas de mutação e a sua biologia possibilita uma análise automatizada, com uma interpretação relativamente simples devido às técnicas de análise destes marcadores originarem poucos artefactos, levando à obtenção de informação genética adicional que, muitas vezes, pode ser uma mais-valia para o geneticista e para o esclarecimento dos acontecimentos, quer em casos do foro cível quer em casos do foro criminal. Por tudo isto, nos últimos anos tem-se assistido a um rápido desenvolvimento da metodologia de análise de SNPs como ferramenta usada em genética forense.

Do ponto de vista da análise forense, estes marcadores encontram-se agrupados em quatro classes. Uma destas classes é constituída pelos marcadores que podem ser úteis na identificação individual e uma outra por polimorfismos que podem ser usados na identificação de linhagens familiares, os quais facultam informação genética adicional à obtida aquando da análise de STRs (do inglês *short tandem repeats*), os marcadores mais utilizados nestas perícias. Assim, o estudo de SNPs, nestas situações, prende-se com a necessidade de obter informação genética adicional, derivada da existência de incompatibilidades nos sistemas STR estudados ou da falta de amostra pertencentes a indivíduos importantes para o estudo de relações de parentesco ou, ainda, devido à não amplificação ou amplificação incompleta dos *loci* STR, resultante de uma reduzida quantidade e/ou fraca qualidade do DNA existente para análise.

No entanto, as duas restantes classes de SNPs não permitem a concretização de uma identificação individual ou familiar mas não deixam, por isso, de conter um grande interesse forense. Estas classes possibilitam inferir sobre características do dador de uma amostra fornecendo informações que não são obtidas através da análise de STRs. Uma destas classes é constituída por SNPs que fornecem informação sobre a origem étnico-geográfica do dador da amostra e a restante classe é composta por marcadores que

fornecem indicações sobre características fenotípicas do indivíduo, características externamente visíveis que, na ausência de um suspeito, permitem fornecer informações valiosas aos investigadores policiais, auxiliando estas forças na descoberta do indivíduo que pretendem identificar. Este tipo de informação poderá ser principalmente importante quando o perfil genético de STRs do dador de uma determinada amostra, cuja análise seja importante para a resolução de um caso, não se encontre ainda inserido nas bases de dados de perfis de DNA e não existam fortes indícios ou suspeitos sobre os quais se possa dirigir a investigação em curso.

A aplicação destes marcadores está a ser estudada por vários grupos de investigação internacionais de modo a que o seu uso possa ser uma realidade futura nos laboratórios forenses. Contudo, e uma vez que a genética forense se prende com a descoberta jurídica da verdade dos factos, assente no conhecimento científico mas com base na variabilidade da espécie humana, importa fazer o estudo dos marcadores de interesse nas populações nas quais pode vir a ser utilizado este tipo de análise. Assim, também em Portugal interessa fazer o estudo dos marcadores forenses do tipo SNP para estes poderem vir a ser usados, no futuro, na realização de perícias envolvendo a nossa população. Com este trabalho pretendeu-se contribuir para este desenvolvimento, quer através da verificação na população Portuguesa das características populacionais de alguns dos multiplexes de SNPs mais investigados, quer através da modificação e aperfeiçoamento dessas mesmas ferramentas de modo a poder ser obtida mais e melhor informação com as mesmas.

No presente estudo, inicialmente com base na metodologia de análise simultânea de 52 SNPs desenvolvido pelo consórcio europeu SNP*for*ID, conjunto de marcadores denominado de 52-plex, começou por se desenvolver e otimizar um conjunto de 20 *loci* para identificação individual de modo a estes marcadores poderem ser utilizados, como ferramentas adicionais ao estudo de STRs, na identificação individual e na determinação de relações de parentesco. Posteriormente, avançou-se para o estudo do 52-plex na população Portuguesa e da sua eficácia na análise de amostras complexas, relacionadas com casos criminais e de investigação da identidade, tais como osso, dente ou vestígios provenientes de cenário forense e das quais não foi obtido um perfil completo de STRs quando da aplicação de um *kit* comercial.

Outro objetivo era estudar a aplicação de um multiplex de SNPs informativos de características fenotípicas, o IrisPlex, na mesma população. Adicionalmente, e uma vez que aquando da realização do projeto de tese foi verificado que os marcadores que compõem este multiplex para além de estarem relacionados com a determinação da cor dos olhos estão também envolvidos na determinação da cor do cabelo e da cor da pele, pelo facto dos principais genes envolvidos nos três mecanismos serem os mesmos, foi também definido suplementar este multiplex com marcadores adicionais de modo a verificar se a ferramenta resultante poderia fornecer informação adicional sobre estas características externamente visíveis.

Assim, esta tese de doutoramento apresenta resultados obtidos na investigação dos SNPs e da sua aplicabilidade à população Portuguesa, com enfoque nos marcadores de identificação individual e nos marcadores para fenotipagem forense de DNA, uma vez que estes foram os principais objetivos deste trabalho.

Deste modo, no Capítulo I é feita uma descrição sucinta do estado da arte sobre o conhecimento dos SNPs e da sua utilização em genética forense. Especial atenção foi dada à classe de SNPs utilizados para identificação individual (IISNPs) e à classe de SNPs informativos de características fenotípicas (PISNPs), uma vez que, como referido anteriormente, estas foram as classes sobre as quais esta tese de doutoramento se focou.

Posteriormente, são expostos os resultados obtidos nos capítulos II a V, na forma de artigos científicos, os quais foram publicados em revistas internacionais, indexadas e com revisão por pares. Para além destas publicações, é apresentado um manuscrito em fase de preparação.

A ordem de apresentação de cada um dos capítulos e subcapítulos desta tese não reflete necessariamente uma ordem cronológica, uma vez que alguns dos estudos descritos à frente decorreram simultaneamente e os resultados obtidos durante um trabalho particular viriam a influenciar o progresso de outro ou de outros trabalhos.

No Capítulo II é apresentada a validação e os resultados obtidos com um multiplex de 20 SNPs, derivado do 52-plex desenvolvido pelo consórcio SNP*for*ID, na investigação da paternidade e noutros casos de investigação de parentesco, alguns destes com um maior grau de complexidade.

No capítulo III, são apresentados os resultados obtidos com o uso de SNP*for*ID 52-plex no estudo dos parâmetros populacionais e forenses da população Portuguesa residente no sul de Portugal e também numa das maiores populações migratórias residentes no nosso país, os imigrantes provenientes da República da Guiné-Bissau. Neste capítulo são ainda apresentados os resultados obtidos com este multiplex na análise de amostras complexas relacionadas com casos criminais e de investigação da identidade.

O Capítulo IV, de menor dimensão, apresenta resultados obtidos com o desenvolvimento de um pequeno multiplex de dez SNPs informativos do haplótipo mitocondrial (LISNPs), o qual pode ser utilizado no estudo de linhagens maternas.

Já no Capítulo V são apresentados os resultados obtidos com o desenvolvimento de um novo multiplex de marcadores fenotípicos, baseado no IrisPlex, o qual foi desenvolvido para inferir acerca da cor dos olhos e da cor da pele e que mostrou ser aplicável à população Portuguesa. Neste capítulo é também feito o estudo da aplicação deste multiplex na investigação dum caso com interesse antropológico e forense.

No Capítulo VI é feita uma discussão geral dos resultados descritos nesta dissertação de doutoramento, bem como de algumas questões levantadas ao longo do desenvolvimento desta tese.

Por último, de forma resumida, são apresentadas as conclusões obtidas nesta tese. Entre estas, conclui-se que a análise de um pequeno número de IISNPs, como 20 marcadores, pode ser usada com sucesso na investigação da paternidade e na avaliação de casos de parentesco, como complemento à análise dos STRs. Todavia, para a análise de amostras complexas deverá ser utilizado um maior número de *loci*. Neste tipo de amostras deve ser analisado o 52-plex uma vez que a amplificação dos marcadores STR nestas amostras pode

ser ineficiente, revelando-se na obtenção de perfis incompletos ou na não obtenção de qualquer perfil genético. Foram estudados os parâmetros populacionais e forenses deste 52-plex na nossa população e na população de imigrantes provenientes da Guiné-Bissau, tendo sido obtidos resultados em linha com os descritos por outros autores. Foi, ainda, desenvolvido um pequeno multiplex de LISNPs, para determinação do haplogrupo mitocondrial, que poderá ser usado como teste rápido de triagem antes da sequenciação do DNA mitocondrial em amostras complexas.

Neste trabalho, foi também modificado o IrisPlex de modo a que o multiplex resultante pudesse fornecer informação sobre a cor da pele do indivíduo dador da amostra. O multiplex resultante demonstrou ser bastante preciso na determinação da cor dos olhos em indivíduos da população Portuguesa bem como a distinguir se o indivíduo em causa apresenta uma pele clara (fotótipos I a IV) ou se apresenta uma pele mais escura (fotótipos V e VI). Este multiplex demonstrou ser bastante sensível quando testada a sua aplicação numa amostra de músculo retirada de um indivíduo mumificado com data estimada de mumificação por volta dos anos 1930-40, apresentando, por isso, o DNA num estado degradado.

No futuro, a análise de SNPs será, certamente, usada com bastante frequência, na análise de casos criminais e de investigação da identidade, nomeadamente através da análise de IISNPs e de PISNPs.

Palavras-chave: Genética forense; identificação genética; fenotipagem forense de DNA – FDP; polimorfismos de um único nucleótido – SNP.

Abstract

Single nucleotide polymorphisms or SNPs present low mutation rates and its biology permits for an automated analysis, with a relatively simple interpretation due to the fact of the techniques used to analyze these markers originating few artifacts. This allows to attain additional genetic information that can offer great value to the geneticist and for the discovery of the truth in civil or criminal cases. For all these, in recent years we have witnessed the rapid development of SNP methodology as a tool in forensic genetics.

These markers can be characterized in four forensically relevant classes, one of which consists of SNPs useful in individual identification and other consists in polymorphisms used to identify family lineages, allowing the obtainment of additional genetic information to complement those usually achieved through STR analysis. Thus, SNP study in these situations relates to the need of additional genetic information due to the existence of STR incompatibilities in the studied systems or due to the lack of sample from individuals essential to the analysis of kinship casework. SNP study is also of great importance when incomplete amplification or no amplification of the studied STR systems is observed, resulting from the small quantity and / or poor quality of the analyzed DNA.

The two remaining SNP classes cannot provide individual or family identification but they present large forensic interest because they allow inferring about characteristics of the sample donor. The first of these classes is composed by SNPs that provide biogeografical ancestry information and the remainder class consists of markers that provide information about phenotypic characteristics. This type of information can be especially important when no information is known about the sample donor and not even his or her genetic profile has entered into DNA databases, in a way that these will provide valuable clues to the discovery of the individual yet to be identified.

The application of these markers to forensics is being studied by several research groups worldwide so that their use can become a reality in the near future. However, since forensic genetics is founded on scientific knowledge and based on human genetic diversity, it is important to study these in the populations in which this type of analysis can be used. Thus, it is important to study the SNP *loci* in the Portuguese population so that these markers can be used to carry out forensic testing in cases involving Portuguese individuals. The purpose of this work was to contribute to this development, both through the verification of SNP multiplexes in the Portuguese population and through the modification and improvement of these tools in order to obtain more and better Information with them. Thus, this work intended to study and optimize individual identification SNP multiplexes to be used on kinship testing and on individual identification and to study a phenotype informative SNP multiplex for application in the resolution of cases where there is no suspect's sample for comparison.

Initially, based on the 52-plex developed by the SNP*for*ID consortium, it was developed and optimized a set of 20 *loci* set for individual identification so that these could be used as an additional tool to STR investigation in the study of paternity testing and kinship analysis. Later, it was studied the population and forensic parameters of the 52-plex in the Portuguese population and the effectiveness of this multiplex in the analysis of complex samples such as bone, tooth or crime scene's samples from which it was not obtained a complete genetic profile when applying a commercial STR kit.

Another main objective was to study the application of IrisPlex in Portuguese population. This is a multiplex designed to infer about the eye color of a sample donor. In addition, since that when this thesis project was carried out it was verified that the markers that make up this multiplex, besides being related to eye color determination are also involved in hair and skin color, due to the main genes involved in the 3 mechanisms being the same, it was also decided to supplement this multiplex with additional markers in order to verify if the resulting tool could provide additional information on these other externally visible characteristics.

Thus, this doctoral thesis presents the results obtained through the analysis of SNPs and of its applicability to the Portuguese population, focusing on the individual identification *loci* and on the markers used for forensic DNA phenotyping. The results obtained are presented in chapters II to V, in the form of scientific articles which have been or will be published in indexed and peer-reviewed international scientific publications.

In the future, SNP analysis will certainly be used more quite often in criminal and identity investigation casework, namely through the analysis of identity investigation and of forensic DNA phenotyping SNPs.

Key-words: Forensic genetics; genetic identification; forensic DNA phenotyping – FDP; single nucleotide polymorphism – SNP.

Abbreviations

µl – microliter

A – adenine

AIM – ancestry informative marker

AISNP – ancestry informative SNP

AUC – area under the curve

bp – base pair

C – cytosine

CODIS – Combined DNA Index System

ddNTP – dideoxynucleotide triphosfate

DNA – deoxyribonucleic acid

EPG – electrophoretogram

EVC – external visible characteristic

FDP – forensic DNA phenotyping

F_{ST} – fixation index

G – guanine

GHEP-ISFG – Spanish and Portuguese-Speaking Working Group of the ISFG

GWAS – genome-wide association studies

He – expected heterozygosity

Ho – observed heterozygosity

HWE – Hardy-Weinberg equilibrium

IISNP – individual identification SNP

InDel – insertion deletion polymorphism

INMLCF – Portuguese National Institute of Legal Medicine and Forensic Sciences

ISFG – International Society for Forensic Genetics

LISNP – lineage informative SNP

LR – likelihood-ratio

MALDI-TOF - matrix-assisted laser desorption/ionization time-of-flight mass spectrometer

MP – matching probability

MPS – massive parallel sequencing

mtDNA – mitochondrial DNA

mtSNP – mitochondrial SNP

ng – nanogram

NGS – next generation sequencing

NPM – negative predictive value

PCR – polymerase Chain Reaction

PD – power of discrimination

PE – power of exclusion

PI – paternity index

PIC – polymorphism information content

PISNP – phenotype informative SNP

PP – posterior probability

PPV – positive predictive value

RFLP – restriction fragment length polymorphism

ROC – receiver operating characteristics

SBE – single base extension

SENS – sensitivity

SNP – single nucleotide polymorphism

SPEC – specificity

STR – short tandem repeat

SWGDAM – Scientific Working Group on DNA Analysis Methods

T – thymine

TPI – typical paternity index

VNTR – variable number tandem repeat

Table of Contents

Chapter I

General introduction

The Human Genome Project allowed the discovery of new genetic markers with high potential for many areas of genetics. Single Nucleotide Polymorphisms (SNPs) were perhaps the most important, at least for their abundance in the genome. It is difficult to say with certainty their frequency but it was estimated a frequency of 1 SNP per 1000 base pair (bp) (Venter *et al.*, 2001). However, if we consider the 84.7 million SNPs recently characterized by the 1000 Genomes Project, it is very likely that the frequency of SNPs on the human genome must be much higher than that (Auton *et al.*, 2015).

In recent years, we have witnessed a rapid development of SNPs methodologies in a variety of biomedical disciplines, and it is not different in forensic genetics. This fast development was due to these polymorphisms present low mutation rates and its biology allows for their automated analysis, namely through the recent and powerful massive parallel sequencing systems (MPS, also referred as Next Generation sequencing or NGS). The fact SNPs have few artifacts, make their interpretation relatively simple and allow obtention of additional genetic information, that can be a great value in different areas of biomedical sciences. In forensic genetics this is no different situation, SNPs are revealing to be an important tool for the analysis of complex cases where conventional methodologies, such as STRs or mitochondrial DNA analysis fail and cannot provide an adequate response or even to obtain information about external visible characteristics (EVC) or the ancestry of sample donor, when there is still no suspect.

1.1. SNP biology and its application

SNPs are the result of point mutations where one DNA bp is replaced by another. Figure 1 illustrates a point mutation in the DNA chain that may origin a new SNP *locus* in a given population.

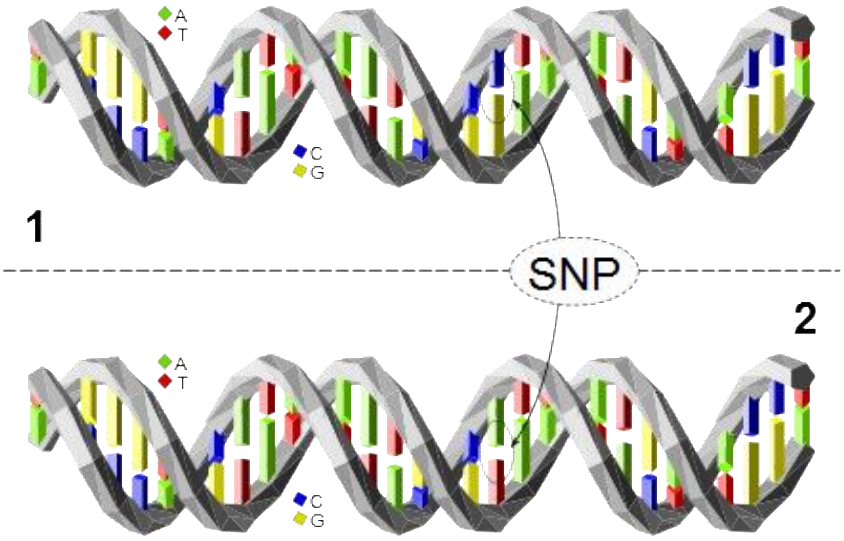


Figure 1 – Point mutation originating a SNP. The upper DNA molecule differs from the lower DNA molecule at a single base-pair location (C/A polymorphism). Source: SNP model by David Eccles (gringer), CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=2355125)

Over time and as result of other forces of evolution, namely natural selection and genetic drift, these mutations established in a given population by being transmitted to the progeny over the following generations. This process results in the entry of a new allele in that population and possibly in the establishment of a new polymorphism.

Because of this apparent simple generation mechanism, these polymorphisms are the most abundant polymorphisms class in the human genome and provide powerful tools for a variety of biomedical genetic studies (Wang *et al.*, 1998). These *loci* have been shown to be associated with different susceptibility to diverse forms of cancer (Engle *et al.*, 2006), heart disease (Aulchenko *et al.*, 2009) and neurodegenerative pathologies (Paulsen *et al.*, 2013), but also with different vulnerability to infection by several diseases, such as pulmonary tuberculosis meningococcal disease or AIDS (Chapman and Hill, 2012).

Recent advances in sequencing technology have drastically reduced the costs of DNA sequencing, making possible the analysis of a large number of individuals in order to provide a comprehensive resource on the human genetic variation, namely on SNPs. The

"1000 Genomes" was the first project to sequence the genome of a large number of samples having been sequenced 2504 genomes from individuals from 26 populations that led to the identification of 84.7 million SNPs (Auton *et al.*, 2015). As with other previously human genome sequencing projects, data from the "1000 Genomes" is available to the worldwide scientific community through a public open access database - The International Genome Sample Resource (IGSR, available at http://www.internationalgenome.org/). The initial goal of this project was to find genetic variants that have frequencies equal to or greater than 1% in the studied populations, i.e. find the different alleles of identified polymorphisms. These results are enabling and will enable geneticists to test the association between newly discovered SNPs and any genetic traits that pretends to be studied.

These studies can be used to better characterize the degree of association of a given polymorphism(s) with certain characteristic(s) and/or with certain populations, information that can be important in forensic genetics. Giving an example, the discovery of SNPs that are present in all or almost all populations in Hardy-Weinberg Equilibrium and that are not associated with any phenotypic characteristics makes these *loci* useful for human identification. On the other hand, if some markers are especially present in one or few populations, maybe its identification in an unknown sample may be an indication that the sample is from an individual from that specific population. More, taking as a phenotypic example the characteristic of the eye color, not all individuals present in northern Europe have blue eyes, although this is the area of the world where this characteristic is more present. A population study will tend to show a high degree of association between SNPs associated with this different physical characteristic with the different alleles displayed by these populations.

1.2. SNPs in forensic genetics

SNPs have a set of characteristics that makes them ideal molecular markers for use in forensics (Sanchez *et al.*, 2003, 2006; Gill *et al.*, 2004; Amorim and Pereira, 2005). First, SNPs present mutation rates lower than Short Tandem Repeats (STRs), *loci* typically used for analysis in genetic identification and in determination of biological kinship (10^{-8} compared to 10^{-3} to 10^{-5}, according to Huang *et al.*, 2002; Reich *et al.*, 2002). Second, SNPs can be analyzed through PCR amplification of small genomic regions (60-130 bp range), which can be extremely valuable when DNA samples are severely degraded especially when typing forensic or archaeological samples. This is reflected, among other cases, in the work of Vallone and coworkers (Vallone *et al.*, 2005), where they were able to identify some of the victims of the World Trade Center attack in samples where no results had been obtained through the amplification of STR *loci*. Third, SNPs can be rapidly analyzed through a large set of high throughput technologies, with accurate automated typing and allele calling. This is an important factor for the implementation of criminal DNA databases (Martin *et al.*, 2001; Schneider and Martin, 2001). This fact is becoming more relevant with the implementation of MPS sequencing systems in forensic labs due to the down coming costs and the advanced maturation stage of these technologies. Finally, SNPs, and binary polymorphisms in general, are relatively easy to validate because estimates of gene frequencies necessary for proper interpretation of forensic data can be obtained by analysis of a comparatively small number of samples, when compared with the number of samples needed to estimate allele frequencies of STRs. SNPs also offer a potentially cheaper, faster, and more automatable alternative to STRs in many applications (Pakstis *et al.*, 2010).

Ken Kidd at the 22nd Congress of the International Society of Forensic Genetics (ISFG) that was held on August 25, 2007 in Copenhagen, Denmark, during a panel discussion on SNPs and their application in forensic identity and relationship testing, classified SNPs in the following four categories: 1. Identity or Individual Identification SNPs (IISNPs) as “SNPs that collectively give very low probabilities of two individuals having the same multi-*locus* genotype”; 2. Lineage Informative SNPs (LISNPs) as “Sets of tightly linked SNPs that

function as multi-allelic markers that can serve to identify relatives with higher probabilities than simple bi-allelic SNPs"; 3. Ancestry Informative SNPs (AISNPs) as "SNPs that collectively give a high probability of an individual's ancestry being from one part of the world or being derived from two or more areas of the world"; and, 4. Phenotype Informative SNPs (PISNPs) as "SNPs that provide a high probability that the individual has particular phenotypes, such as a particular skin color, hair color, eye color, etc." (Butler *et al.*, 2008). This classification was commonly accepted by the forensic community and that is well described by Budowle and van Daal (Budowle and van Daal, 2008). Here it is reviewed the use of SNP for human forensic genetics and the recent advancements of the application of these markers.

In this manner, the first class of SNPs holds the *loci* that can be useful in individual identification and consequently in parental testing and the second one comprises the polymorphisms used to identify family lineages, mainly present in Y chromosome and in mitochondrial DNA. If we compare the analysis of these SNP markers to the analysis of standard methodologies typically used in forensic laboratories, the analysis of IISNPs can be fairly compared to the analysis of autosomal STRs and the analysis of the LISNPs compared to the analysis of Y-STRs and of mitochondrial DNA sequencing. As a consequence, the analysis of these types of SNPs is only executed in special situations as, for example: when there is the need to obtain additional genetic information due to the existence of STR incompatibilities in the studied systems, due to the lack of samples from individuals crucial to the analysis of the kinship relationships, or when there is incomplete PCR amplification or no amplification of the STR systems studied in complex samples, resulting from the small quantity or from the reduced quality of the DNA analyzed.

The two remaining SNP classes cannot provide individual or family identification but have great forensic interest because they allow inferring about characteristics of the sample donor. The AISNPs provide biogeographical ancestry information about sample donor and the PISNPs class consists of SNPs that provide information about phenotypic characteristics of sample donor. These can be especially important when there is no suspect to compare with the samples and no STR genetic profile has yet entered into DNA

databases, in a way that they provide valuable clues to police investigators, aiding the discovery of individual to be identified.

Still, is important to remember that this classification is not a tight one in a way that a marker that is included in one class may also provide some degree of other types of information, for example, LISNPs may provide information about the ancestry of an individual if his particular lineage is only observed in some part of the globe or PISNPs that provide a probability of an individual being blue eyed give also the information that the individual is probably European because blue-eyed color phenotypes are mainly of European ancestry. This type of information may even transpose to the clinical information, a situation that is recommended by the ISFG to be avoided for forensic autosomal STRs but that, in these cases, cannot be applicable. For example, identifying that an individual has a high probability of having a very light skin also informs scientists that there is an higher probability of that individual be more prone to melanoma development, but that comes from the fact of individuals with lighter tones of skin having more probabilities of develop this type of skin cancer.

1.2.1. Individual Identification SNPs

SNPs can be very efficient tools for human identity testing purposes or for distinguishing between close relatives, e.g. brothers, provided that a sufficient number of SNPs is used so that the average match probabilities (the probabilities of two unrelated individuals having the same multi-*locus* genotype) of the final panel at least be comparable to the standard CODIS STR markers (Budowle *et al.*, 2008). In order to obtain this, the SNP panel must be composed by somewhere in the range of 50 to 100 selected SNP *loci* so that alleles range in proportion between 0.2–0.8 and taking in consideration the assumption of independence across all *loci* (Chakraborty *et al.*, 1999; Gill, 2001; Amorim and Pereira, 2005). However, it was also suggested that 20 to 50 SNP *loci*, with no linkage and with a high degree of heterozygosity may be sufficient to adjust the population stratification in association studies based on populations (Pfaff *et al.*, 2004). Nevertheless, the more

recent MPS systems, using the newly available commercial forensic kits, can analyze more than one hundred SNP *loci* at once analyzing at the same time all the STRs more commonly used in routine forensic casework (Børsting and Morling, 2015; Churchill *et al.*, 2016).

Taking this in consideration, in recent years several autosomal SNP marker sets were developed for human identification (Dixon *et al.*, 2005; Kim *et al.*, 2010; Pakstis *et al.*, 2010; Freire-Aradas *et al.*, 2012; Cho *et al.*, 2014; Hou *et al.*, 2014; Wang *et al.*, 2016) and its use been widely debated by the forensic community. One of the best well-known of these is the 52-plex SNP assay that was developed by the SNP*for*ID consortium (Sanchez *et al.*, 2006).

The SNP*for*ID consortium was a project supported by the EU GROWTH program in 2002 which primary goal was to develop a SNP multiplex that could be used in forensic casework. This cooperation led to the creation of an assay which amplifies 52 autosomal SNPs in a single PCR reaction, with amplicons ranging in size from 59 to 115 bp, and that afterwards can be analyzed using several SNP detection methods (Musgrave-Brown *et al.*, 2007). This 52-plex assay revealed to be more successful in amplifying challenging samples than the assays based on STR analysis, indicating that forensic SNP typing can be very useful when STRs fail to present results for this kind of samples (Butler *et al.*, 2007; Musgrave-Brown *et al.*, 2007; Senge *et al.*, 2011). This was confirmed in some case reports involving real forensic samples such as the ones reported by Fondevila *et al.* or by Alimat *et al.* where full SNP profiles were generated from DNA extracts that yielded no or few STR *loci* (Fondevila *et al.*, 2008; Alimat *et al.*, 2013).

Another application where IISNPs can sometimes be important supplementary tool is in kinship testing. Kinship testing is perhaps the kind of forensic analysis with the greater number of exams done all over the world, performed by smaller and larger laboratories and by public and private entities in countries where legislation permits it. This can be observed when analyzing the number of laboratories participating in the kinship testing and forensic modules of proficiency testing programs of ISFG such as the GHEP-ISFG quality control proficiency testing.

Most of these laboratories perform kinship testing by comparing STR genotypes obtained analyzing one or two autosomal STR amplification kits produced by the following commercial companies: Applied Biosystems (kits AmpF*l*STR® Identifiler® / AmpF*l*STR® Identifiler® Plus PCR Amplification Kit and GlobalFiler™ PCR Amplification Kit), Promega (kits PowerPlex® 16 / 16 HS System and PowerPlex® Fusion) and Qiagen (Investigator ESSplex kits family). Using one or two of these kits is possible to analyze a great number of STRs including the CODIS and the ESS *loci* plus the amelogenin sex marker and, depending on the kits used, a few Y-STRs and InDels.

Although these kits are fully validated for forensic purposes and widely used all over the world, some particularly complex kinship situations exists where it is difficult to obtain a satisfactory answer to the case in question. This can be due to variable facts such as the existence of missing persons essential to the analysis, like the putative father (known as deficiency cases), or such as missing or misleading information like testing a close relative instead of the person of interest without knowing (like testing the brother or the father of the putative father). Other facts like the existence of genetic inconsistencies between the relatives in testing can lead to an inadequate answer, or even simply because a kit cannot resolve between two alleged parents can also present complex situations (Kersbergen *et al.*, 2008; González-Andrade *et al.*, 2009; Børsting and Morling, 2011).

Many times, to obtain additional information in these situations forensic scientists use other autosomal STR kits to achieve additional genetic information and complement the routine autosomal STR analysis previously achieved. However this methodology sometimes does not reveal abundant additional information due to the fact the majority of these kits markers being the ones already existing in the kits previously employed. Other option is to study the Y or the X chromosomal STR genotypes or even the mitochondrial haplotypes. However the use of these markers in forensic calculations is challenging since its segregation is not independent, therefore not following a Mendelian inheritance pattern, making these markers mainly used for the study of paternal and maternal lineages.

This way, other alternative for forensic scientists in these types of cases can be to complement it with different autosomal markers, like SNPs or even to do the kinship

testing using solely this type of markers. In 2008 Costa and coworkers (Costa *et al.*, 2008) begun to test the use of SNPs in paternity investigation using 10 SNPs. Still in 2008, Børsting and coworkers presented evidence that the SNP*for*ID 52-plex assay was a good alternative to standard STR investigations in relationship testing and suggested that the strength of STRs and SNPs could be combined in order to resolve complex cases. Børsting and coworkers compared the paternity indices obtained typing 15 STRs, 7 VNTRs and the SNP*for*ID multiplex in 124 paternity trios and concluded that due to SNPs mutating much less frequently, they were therefore ideal for additional marker testing in paternity cases where genetic inconsistencies were detected with the standard STR analysis, with the additional advantage of being both fast and cheaper to type. The typical paternity indices (PIs) obtained based on the 52 SNPs were 5–50 times lower than the typical PIs based on 15 STRs or 7 VNTRs. However, the mutation rate of SNPs was expected to be approximately 100.000 times lower than the mutation rate of tandem repeats, making Børsting *et al.* to consider the SNP*for*ID 52-plex assay as a valuable investigative tool in forensic casework where potential relatives are involved, especially in cases where various family scenarios are possible or in cases where a possible mutation has been detected in one of the tandem repeat *loci* and supplementary investigations are required. Furthermore, they considered the SNP*for*ID 52plex SNP assay as highly useful in disaster victim investigations because this assay can generate complete profiles from highly degraded DNA, where the STR kits can fail (Børsting *et al.*, 2008).

1.2.2. Lineage Informative SNPs

Similarly to what happens with mtDNA analysis or with Y-STR genotyping, SNPs that reside on the mtDNA genome or in the Y chromosome provide useful information about an individual's lineage. Because of a lack of recombination and a low mutation rate, these *loci* can be very informative for evolutionary studies and kinship analyses, especially in cases where the reference sample(s) and the evidence sample are several generations apart (Budowle and van Daal, 2008). These types of markers can be used in family testing for solving deficiency cases where putative fathers or mothers are not available for DNA

analysis and the unavailable male or female can be replaced by another relative of the same sex in the analysis of a descendent. Other applications of the use of these markers can include historic cases, as long as living descendants are available (Kayser, 2007).

Mitochondrial DNA is matrilineal inherited as haplotypes and all these haplotypes can ultimately be traced to a common ancestor that lived approximately 200,000 years ago in Africa (Mishmar *et al.*, 2003; Behar *et al.*, 2008). MtDNA sequence variation evolved as a result of the sequential accumulation of mutations along maternally inherited lineages, which can be represented in a tree reflecting the phylogenetic relationships of known mtDNA variants (van Oven and Kayser, 2009). For two individuals to belong to the same matrilineal lineage they have to share the same haplotype, an information that can present usefulness in forensic casework. Further, due to the nature of mtDNA and to the structure of the mitochondria itself the mtDNA is many times available in forensic samples where the autosomal DNA is highly limited or degraded such as bones, teeth or hairs. Consequently, mtDNA can provide key information in human identity testing from these types of samples (van Oven and Kayser, 2009).

This way, some SNP assays have been established for forensic casework like the ones developed by Brandstatter and coworkers (Brandstätter *et al.*, 2003), Quintáns and coworkers (Quintáns *et al.*, 2004), Vallone and coworkers (Vallone *et al.*, 2004), Coble and coworkers (Coble *et al.*, 2004, 2006), Nelson and coworkers (Nelson *et al.*, 2007) or Köhnemann and coworkers (Köhnemann *et al.*, 2009).

Similarly to what happens to mtDNA *loci*, SNPs on the non-recombining portion of the Y chromosome are inherited as haplotypes in the absence of recombination and with a low mutation rate but in this case in a patrilineal line (Jobling, 2001; Onofri *et al.*, 2006). These characteristics makes Y-SNPs particularly suitable for identifying stable paternal lineages (Onofri *et al.*, 2006) and reconstructing a common ancestor to help in the discovery of the history of our species (Hammer, 1995; Thomson *et al.*, 2000).

Several assays were also developed to genotype Y-SNPs that differ in the populations for which they were primarily aimed for, differ on the objective of the assay or on the

technologies used (Brion *et al.*, 2005; Krjutskov *et al.*, 2009; Geppert *et al.*, 2011; Park *et al.*, 2013).

Nevertheless, in forensics while it can be used for exclusion confirmation in criminal cases, Y inclusions are not possible to make because haplotypes are confined within lineages in which many male relatives of a suspect share the same Y chromosome (Jobling *et al.*, 1997).

Several autosomal haploblocks have also been identified by the HapMap project which for forensic purposes may collectively serve as autosomal lineage-based markers and improve the ability to identify missing persons when the number of family members for comparison is very limited (Budowle and van Daal, 2008).

The most likely use of lineage SNPs in forensics is for special instances such as deficiency cases in familial testing, missing person investigations or mass disaster identifications. Nevertheless, successful identification by genetic testing using kinship analysis is limited by the amount of DNA available for analysis, the number of family members for comparison, and the available genetic markers (Budowle and van Daal, 2008).

1.2.3. Ancestry Informative SNPs

Loci that can be used to infer about the genetic ancestry of a donor of a sample are commonly known as ancestry informative markers (AIMs). The analysis of these *loci* can present great importance in different medical areas but maybe even of greater importance in forensic applications, even if its use can present some ethical considerations.

All types of genetic markers can provide forensic investigators clues about the origin of an individual. Even STRs, markers commonly used in routine forensic practice present in some degree information about ancestry (Rosenberg, 2009) and is common for a well-

trained forensic geneticist who analyze a genotype that presents a rare allele in one population but that is normally found in another one to think that the donor of that sample should belong to that second population. That thought is just an educated guess but serves as the basis of the use of AIMs. If a certain sample presents a certain number of markers that are characteristic of a single population instead of others and does not present markers from other populations than there is a given probability of the donor of that sample belonging to that population. This thought is also important to understand that a single marker is not sufficient to address the question about ancestry information and caution should be taken when making individual inferences based on AIMs (Pardo-Seco *et al.*, 2014).

Although STRs can present some information about genetic ancestry, at least since 2003 that is known that using the most informative SNPs will be more efficient in this discovery (Rosenberg *et al.*, 2003). This was already evident back in 1994 when we revisit RFLP studies of Dean and coworkers where it was demonstrated that RFLPs differences were attributed to "racial differentiation" (Dean *et al.*, 1994) bearing in mind that RFLPs are the product of SNPs that alter DNA restriction sites altering molecule migration patterns.

In 2006 Lao and coworkers demonstrated that it was possible to use only 10 SNPs for detecting continental population structure using Affymetrix GeneChip Mapping 10K SNP array technology (Lao *et al.*, 2006). Since then, different AIM-SNP panels were published, using different technologies: the SNP*for*ID Consortium 34-plex, analyzed with SNaPshot (Phillips *et al.*, 2007); the 47 AIM-SNP Affymetrix GeneChip Mapping 10K SNP assay (Kersbergen *et al.*, 2009); the 128 AIM panel for American populations, developed by Kosoy *et al.* at Seldin lab (Kosoy *et al.*, 2009) and the Kidd lab 128 AISNP panel (Kidd *et al.*, 2011), both using TaqMan Assays; the LACE consortium 446 AIM panel, using Sequenome iPLEX system (Galanter *et al.*, 2012), and the Kidd lab ancestry informative 41-SNP panel to be analyzed with ABI SNPlex and Sequenome iPLEX system (Nievergelt *et al.*, 2013). In 2014, in an attempt to build a forensic ancestry panel from the ground up based on previously work from all the authors, the EUROFORGEN group proposed a Global AIM-SNP set composed of 128 SNPs to be used on different MPS forensic applications that would be developed by commercial companies (Phillips *et al.*, 2014). Nevertheless,

although some information is not already public, private companies developing forensic MPS applications are basing the choice of AIM SNP panel on the work developed in Seldin and in Kidd labs and in the 34-plex developed by the SNP*for*ID Consortium.

Lastly, even though there has been great advancement on the scientific knowledge and technological development in the investigation of ancestry informative SNPs there is the need of this matter to be addressed by society and policymakers (Budowle and van Daal 2008). Information about genetic ancestry, principally if undesirable and unexpected, can lead to the restructuring of personal, familial, or even group identity. Documented cases exist where anthropological and/or genetics research that casts doubt on ancestral relationships incited varying degrees of identity-related conflicts (Royal *et al.*, 2010). This way, balance must be done between the usefulness of the predicted physical appearance based indirectly on ancestry versus the potential for racial and ethnic discrimination (Budowle and van Daal, 2008). Nevertheless, it's not possible to forget that this type of markers were used in the past to aid in the resolution of important cases like the identification of the suspect in a serial murder case in the United States (Budowle and van Daal, 2008) or in the identification of the suspects involved in the Madrid train attack on March 11, 2004 that killed 191 people and injured 1755 (Phillips *et al.*, 2009).

1.2.4. Phenotype Informative SNPs

In recent years scientific community assisted an impressive discovery of *loci* associated with phenotypic characteristics. This progress was achieved, in the beginning, by evaluation of human orthologues of genes associated with EVCs in animal models (Kwon *et al.*, 1994; Lamason *et al.*, 2005; Lehner, 2013) but mostly, in recent years, through the analysis of genome-wide association studies (GWAS) in humans. In GWAS different characteristics are studied by comparison of several common genetic variants in different individuals to determine if some of those are associated with a specific trait, many times a human disease like eye degeneration (Klein *et al.*, 2005) or skin cancer (Gerstenblith *et al.*, 2010) for example.

Similarly, phenotypic characteristics GWAS have undisclosed *loci* associated with human EVCs such as eye color, skin color and hair color (Sulem *et al.*, 2007, 2008; Han *et al.*, 2008; Gerstenblith *et al.*, 2010).

These discoveries created great interest in forensic scientists as means of obtaining information about sample donors in cases where there are no suspects and doesn't exist any corresponding genetic profiles in DNA criminal databases (Sulem *et al.*, 2007; Kayser and Schneider, 2009; Kayser and de Knijff, 2011). This type of analysis known as Forensic DNA Phenotyping, also sometimes refereed as 'DNA intelligence', is still not widely used but can provide important information in criminal casework (Walsh *et al.*, 2011a, 2013; Keating *et al.*, 2013; Poetsch *et al.*, 2013). However, these traits are difficult to understand due to their high complexity due to their multifactorial nature, involving epistasis and gene–environment interaction phenomena (Pośpiech *et al.*, 2011, 2014).

Nowadays only eye color and hair color can be predicted with high accuracies using forensic DNA phenotyping (FDP) technologies although there are still a few assays developed to determine skin color. EVCs such as male baldness, hair morphology or facial format are under research to detect their DNA mechanisms and to estimate their phenotypes with a high predictive value (Kayser and de Knijff, 2011; Kayser, 2015).

One of these assays reported as having a high accuracy is IrisPlex, developed by Walsh *et al.* to determine eye color (Walsh *et al.*, 2011b). This assay consists of a single multiplex genotyping system for the six most informative eye color SNPs according to current knowledge: rs12913832 (HERC2), rs1800407 (OCA2), rs12896399 (SLC24A4), rs16891982 (SLC45A2 (MATP)), rs1393350 (TYR) and rs12203592 (IRF4). This multiplex in conjunction with a statistical prediction model (Liu *et al.*, 2009), was reported as having an average accuracy rate of 94% for predict eye color as being blue or brown in Europeans individuals. This analysis was done considering a probability threshold of 0.7 and ranged from 91% to 98% depending on the European population that was under consideration and despite the large variation in eye color frequencies between populations (Walsh *et al.*, 2012). Is important to refer that IrisPlex was validated for forensic use (Walsh *et al.*, 2011a). However, IrisPlex have limitations in accurately predicting intermediate eye colors such

as green because of the current unavailability of good DNA predictors for these non-blue and non-brown eye colors (Walsh *et al.*, 2011b).

Since IrisPlex publication, different studies were made to address its application on different populations namely in a Slovenian (Kastelic *et al.*, 2013), in different Eurasia populations (Yun *et al.*, 2014) and in a United States population (Dembinski and Picard, 2014). Although, using IrisPlex blue and brown eye color were reliably predicted from DNA samples in the Dutch, in the Slovenian and in the East Asia populations (Walsh *et al.*, 2012; Kastelic *et al.*, 2013; Yun *et al.*, 2014), what shows the interest of different authors in studying these matters.

Based on IrisPlex Walsh and coworkers developed a multiplex to determine eye color and hair color simultaneously – the HirisPlex (Walsh *et al.*, 2013). The HirisPlex assay consists in an assay of 24 SNPs, the 6 ones of IrisPlex multiplexed with INDEL polymorphism N29insA and 10 SNPs from the MC1R gene, rs11547464, rs885479, rs1805008, rs1805005, rs1805006, rs1805007, rs1805009, Y152OCH, rs2228479, and rs1110400, rs28777 (SLC45A2), rs12821256 (KITLG), rs4959270 (EXOC2), rs1042602 (TYR), rs2402130 (SLC24A4), rs2378249 (ASIP/ PIGU), and rs683 (TYRP1). IrisPlex 6 SNPs provide the information for the eye color prediction while the other 18 (together with IrisPlex SNPs 1, 2, 4 and 6) are used for hair color and shade investigation (Walsh *et al.*, 2013, 2014)

The authors report that this assay presents a hair color prediction model with average individual-based hair color prediction accuracies of 69.5% for blond, 78.5% for brown, 80% for red and 87.5% for black hair (Walsh *et al.*, 2013). However this assay have limitations in accurately predicting hair color in individuals who underwent age-dependent changes that influenced category shifts (such as blond to brown) because of the current unavailability of biomarkers to indicate such a color change so additional research is necessary (Walsh *et al.*, 2013). Nevertheless, HirisPlex was validated for forensic use (Walsh *et al.*, 2014).

As regards to an assay to determine skin color is important to refer the Maroñas *et al.* predictive test that was developed for analyzing skin color variation in Europeans, Africans

and admixed African-European individuals (Maroñas *et al.*, 2014). This test analyzes 10 SNPs located in 8 pigmentation genes in which one marker, rs16891982, is common to the IrisPlex eye color predictive test mentioned above. These authors identified the two most important markers amongst the 10 SNPs as being rs16891982 and rs1426654, respectively located in SLC45A2 and SLC24A5, with SNP rs16891982 accounting for 77.62% of the classification success for intermediate skin color and rs1426654 as being the major contributor to the differentiation of black vs. rest plus white vs. rest. In this study conclusions, it is also important to refer that the SNP rs6058017 in ASIP was identified as the second most important marker for classifying black individuals vs. rest, enhancing classification success by 2% and that SNP rs1448484 located in OCA2 was recognized as a major contributor to differentiate between black vs. white.

Despite all these studies, EVCs prediction tools are still not usually used in forensics because some phenotypic *loci* may provide ancestry or clinical information which can raise ethical concerns in the field. However the determination of these traits could be used in crime solving in such way as a conventional eyewitness testimony helping police in the determination of a suspect's profile, probably with lesser risk of exacerbating social pressure (Kayser and Schneider, 2009).

As regards the legal aspect, although the use of FDP is not permitted or is even forbidden in the legislation of most countries this technology is already allowed in legislation of Netherlands and some of states of the U.S. (Kayser and de Knijff, 2011; Murphy, 2013).

1.3. Objectives

The present PhD thesis aimed to achieve two main global objectives: the first was the study and optimization of autosomal IISNP multiplexes for human identification for use primarily in Portuguese population but also in others for which there is great casuistic in the Portuguese National Institute of Legal Medicine and Forensic Sciences, where the practical part of this work was conducted. The idea was the development of methodologies based on the SNP*for*ID 52-plex (Schneider, SNP*for*ID Consortium; Sanchez

et al., 2006) and later, the implementation of the 52-plex to casuistic, if necessary. As the application of lineage markers can sometimes be important for the resolution of forensic cases, the study of LISNPs for aid in this resolution could also be investigated; the second global objective was the development and optimization of a PISNP multiplex designed for phenotyping eye color and skin color that could be used in the same population(s) described above and, avoiding the use of purely ancestral markers. The main idea was based on the fact that the same *loci* that are responsible for eye color determination are also involved in skin color and hair color variation then if we based on the literature would supplement IrisPlex (Walsh *et al.*, 2011b) with other important *loci* then it would be possible to obtain some information about these other physical traits. As mentioned above, this same idea was also trailed by Walsh *et al.* (2011b) in the development of HirisPlex multiplex where they supplemented their previously developed IrisPlex with additional *loci* what allowed for the determination of hair color in addition to the eye color (Walsh *et al.*, 2013). One important fact was that for the development and implementation of all the developed methodologies it was necessary to use Applied Biosystems SNaPshot™ technology (*ABI PRISM® SNaPshot™ Multiplex Kit Protocol*, 2000) because it was the technological platform available at the laboratory, a situation that is very common for most forensic laboratories all over the world.

1.4. General Methodology

SNP genotyping can be achieved through a variety of methods that differ in the type of allelic discrimination reaction used to discriminate between the different alleles and in the method of detection of the products of those reactions. Allelic discrimination reaction can be based on allelic-specific hybridization, primer extension, oligonucleotide ligation or invasive cleavage and detection methods can be based on fluorescence (which can be analyzed using different technologies such as electrophoresis, arrays or some sequencing by synthesis systems), FRET, fluorescence polarization mass spectroscopy or luminescence. These different SNP typing methodologies that are applied to forensic genetics are well reviewed by Sobrino and coworkers (Sobrino *et al.*, 2005) exception for MPS technologies that are more recent than that publication.

In this PhD thesis, samples were analyzed by minisequencing reaction with single base extension (SBE) using SNaPshot™ (Applied Biosystems), following electrophoresis and fluorescence detection. This is probably the most common commercial minisequencing reaction technology followed by electrophoresis and fluorescence detection and is based on the high accuracy of nucleotide incorporation by DNA polymerases. It is also important that this technology is available in most forensic laboratories.

In SNaPshot™ minisequencing reactions, an unlabelled primer is positioned with the 3' end at the base immediately upstream to the SNP site. This annealed primer is then extended with a single ddNTP complementary to the base interrogated by the DNA polymerase that only introduces those nucleotides if the primers have a perfect match with the template (Fig. 2). Important for this assay is that in the reaction, each ddNTP is labeled with a different fluorescent dye. The products are then separated electrophoretically in an automated capillary DNA sequencer and the products of single base expansion detected by fluorescence detection (Sobrino *et al.*, 2005).

To construct multiplexes for analyzing multiple *loci* the design of the primers have to be done taking advantage of the spatial separation of the minisequencing products that can be obtained using ''tails'' at the 5' end of the SNaPshot™ with primers of varying lengths of non-human sequence (Sobrino *et al.*, 2005).

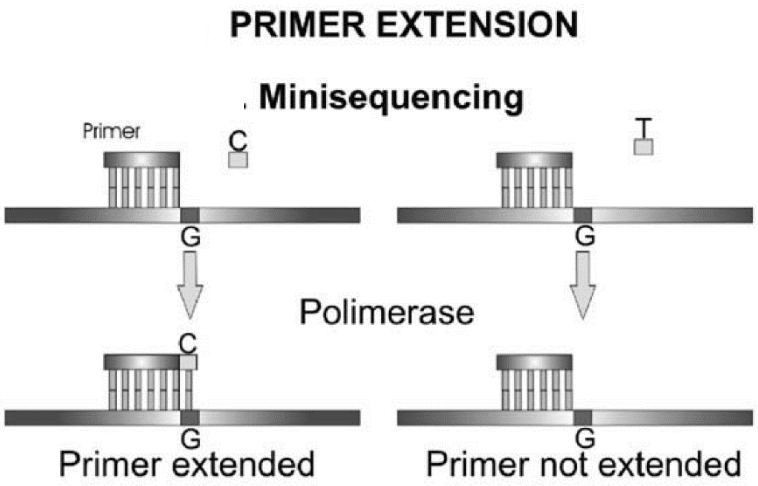


Figure 2 – Illustration of the allelic discrimination reactions by minisequencing: a primer anneals to its target DNA immediately upstream to the SNP and is extended with a single nucleotide complementary to the polymorphic base (from Sobrino *et al.*, 2005, reprinted with permission from the publisher © Elsevier).

Although SNaPshot™ minisequencing methodology was used in all studies developed during this PhD thesis, different PCR conditions were used in order to achieve the specific purposes of each of these studies. This way, these detailed PCR conditions can be consulted on the scientific papers that resulted from this thesis and that are presented on the following sections.

Comprising the different studies about SNP analysis in forensic analyses carried out during this PhD thesis, a total of 721 samples of human origin were analyzed. These belong to the Portuguese National Institute of Legal Medicine and Forensic Sciences and were studied in casuistic of the Forensic Genetics and Biology Department of the Institute. Of these samples, 662 were reference ones, the majority of peripheral blood collected on Whatman cards although some samples of saliva had also been used. All samples were collected under informed consent in healthy individuals, involved in biological kinship investigation, being subsequently anonymized for research purposes. The remaining 59 samples were forensic casework ones (bone tissue, tooth material, hairs and other samples of diverse nature). Additionally, some samples of vertebrate animals were also analyzed in order to validate the specificity of the methodologies developed. DNA from

reference samples was extracted using Chelex® 100 method (Walsh *et al.*, 1991) but for forensic casework samples diverse extraction methods were used: Phenol-Chloroform, Machery Nagel Nucleospin, Qiagen Mini, Micro and Investigator kits depending on type, availability and quality of samples presented for analysis. When necessary, DNA quantification was performed using Real-Time PCR and STR typing was performed with standard commercial kits commonly used in routine forensic analysis. Detail about the samples analyzed and the methodological conditions used in the different studies carried out during this PhD can also be consulted on the scientific papers that are presented on the following sections.

1.5. Thesis Structure

This PhD dissertation reports the outcomes obtained in the investigation of SNP *loci* and its applicability to Portuguese population with special focus on individual identification and phenotyping that were the main objectives of this work. This way, obtained results are exposed in chapters II to IV in the form of scientific papers that were published in peer reviewed international publications or that are being prepared to be submitted for publication.

The order of presentation of each chapter and sub-chapter in the present PhD dissertation does not necessarily reflect a chronological order, since some of the works described below were done simultaneously and the results obtained during one particular work would influence the progress of the other and vice-versa.

Chapter I describes the current knowledge of SNPs and its use in forensic genetics, framing the reader within the thematic and preparing it for the next chapters where are presented the results obtained during this thesis. Special attention was given to the classes of IISNPs and of PISNPs since these were the classes this PhD thesis focused.

Chapter II presents the validation and results obtained with a 20-plex derived from the SNP*for*ID multiplex in paternity investigation, in kinship testing and in complex casework.

The work developed under the Chapter II is presented on the following papers:

1. Dario, P., Ribeiro, T., Oliveira, A.R., Dias, D., Geada, H., Corte Real, F. and Costa Santos, J. (2012) Internal Validation of 20 SNP Multiplex for Forensic Genetics. *International Journal of Legal Medicine*, 126(S1), pp. 357. doi:10.1007/s00414-012-0711-9 (Only abstract published);

2. Dario, P., Ribeiro, T., Espinheira, R. and Geada, H. (2009) SNPs in paternity investigation: The simple future. *Forensic Science International: Genetics Supplement Series*, 2(1), pp. 127–128. doi:10.1016/j.fsigss.2009.08.136;

3. Dario, P., Ribeiro, T., Espinheira, R., Dias, D., Geada, H. and Corte-Real, F. (2011). 20 SNPs as supplementary markers in kinship testing. *Forensic Science International: Genetics Supplement Series*, 3(1), pp. e508–e509. doi:10.1016/j.fsigss.2011.09.111.

4. Dario, P., Ribeiro, T., Dias, D., Corte-Real, F. and Geada, H. (2011) Complex casework using single nucleotide polymorphisms. *Forensic Science International: Genetics Supplement Series*, 3(1), pp. e379–e380. doi:10.1016/j.fsigss.2011.09.051;

Chapter III reports the results obtained with the SNP*for*ID 52-plex in the study of population and forensic parameters of the Southern Portuguese population and from immigrants from Guinea-Bissau, one of the largest migratory populations residing in Portugal. In addition, this chapter reports the results obtained with this assay when examining problematic samples from criminal or identification casework.

The work developed under the Chapter III is presented on the following papers:

5. Dario, P., Oliveira, A. R., Ribeiro, T., Porto, M. J., Dias, D. and Corte Real, F. (2017) Autosomal SNPs study of a population sample from Southern Portugal and from a sample of immigrants from Guinea-Bissau residing in Portugal. *Legal Medicine*, 24, pp. 32-35. doi:10.1016/j.legalmed.2016.11.004;

6. Dario, P., Oliveira, A. R., Ribeiro, T., Porto, M. J., Santos, J. C., Dias, D. and Corte Real, F. (2015) SNP*for*ID 52-plex in casework samples: "Cracking" bones and other difficult samples. *Forensic Science International: Genetics Supplement Series*, 5, pp. 118–120. doi:10.1016/j.fsigss.2015.09.048.

Chapter IV is a smaller section where are presented the results obtained with the development of a small mitochondrial LISNPs multiplex that can be used as a lineage marker tool. This assay proved to be very useful on the investigation of haplogroups in the analysis of criminal DNA samples prior to its DNA sequencing.

The work developed under the Chapter IV is presented on the following paper:

7. Dario, P., Bom, J., Ribeiro, T. and Geada, H. (2009) MtSNP typing before mtDNA sequencing: Why do it? *Forensic Science International: Genetics Supplement Series*, 2(1), pp. 187–188. doi:10.1016/j.fsigss.2009.08.137.

Chapter V exhibits the results obtained with the development of a new PISNP multiplex based on IrisPlex, developed for the determination of eye color and skin color in Portuguese individuals and the application of this on a case with anthropological and forensic interest.

The work developed under the Chapter V is presented on the following papers:

8. Dario, P., Mouriño, H., Oliveira, A. R., Lucas, I., Ribeiro, T., Porto, M. J., Costa Santos, J., Dias, D. and Corte Real, F. (2015) Assessment of IrisPlex-based multiplex for eye and skin color prediction with application to a Portuguese population. *International Journal of Legal Medicine*, 129(6), pp. 1191–1200. doi:10.1007/s00414-015-1248-5;

9. Dario, P., Oliveira, A. R., Marques, M., Ribeiro, T., Porto, M. J., Dias, D., Costa Santos, J. and Corte Real, F. (in preparation). Forensic phenotyping of a mummified body – genetics contribution to a glimpse from the past. *To be submitted*.

Chapter VI presents an overall discussion of the results described in this PhD dissertation, as well as some questions raised along the development of this thesis.

At the end, it is possible to verify the conclusions achieved with this PhD thesis.

References

ABI PRISM® SNaPshot™ Multiplex Kit Protocol (2000). Applied Biosystems.

Alimat, S., Hadi, S. and Goodwin, W. (2013) SNP genotyping of forensic casework samples using the 52 SNPforID markers. *Forensic Science International: Genetics Supplement Series*, 4(1), pp. e178–e179. doi: 10.1016/j.fsigss.2013.10.092.

Amorim, A. and Pereira, L. (2005) Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Science International*, 150(1), pp. 17–21. doi: 10.1016/j.forsciint.2004.06.018.

Aulchenko, Y. S., Ripatti, S., Lindqvist, I., Boomsma, D., IM, H., PP, P., BW, P., AC, J., JF, W., Spector, T., NG, M., NL, P., KO, K., Kaprio, J., Hofman, A., NB, F., MR, J., Gyllensten, U., Campbell, H., Rudan, I., Johansson, A., Marroni, F., Hayward, C., Vitart, V., Jonasson, I., Pattaro, C., Wright, A., Hastie, N., Pichler, I., AA, H., Falchi, M., Willemsen, G., JJ, H., EJ, de G., GW, M., Whitfield, J., Magnusson, P., Saharinen, J., Perola, M., Silander, K., Isaacs, A., EJ, S., AG, U., JC, W., BA, O., Elliott, P., Ruokonen, A., Sabatti, C., Gieger, C., Meitinger, T., Kronenberg, F., Döring, A., HE, W., JH, S., MI, M., CM, van D. and Peltonen, L. (2009) Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nature Genetics*, 41(1), pp. 47–55. doi: 10.1038/ng.269.

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., Mardis, E. R., Marth, G. T., McVean, G. A., Nickerson, D. A., Schmidt, J. P., Sherry, S. T., Wang, J., Wilson, R. K., Gibbs, R. A., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., Kovar, C., Lee, S., Muzny, D., Reid, J. G., Zhu, Y., Wang, J., Chang, Y., Feng, Q., Fang, X., Guo, X., Jian, M., Jiang, H., Jin, X., Lan, T., Li, G., Li, J., Li, Y., Liu, S., Liu, X., Lu, Y., Ma, X., Tang, M., Wang, B., Wang, G., Wu, H., Wu, R., Xu, X., Yin, Y., Zhang, D., Zhang, W., Zhao, J., Zhao, M., Zheng, X., Lander, E. S., Altshuler, D. M., Gabriel, S. B., Gupta, N., Gharani, N., Toji, L. H., Gerry, N. P., Resch, A. M., Flicek, P., Barker, J., Clarke, L., Gil, L., Hunt, S. E., Kelman, G., Kulesha, E., Leinonen, R., McLaren, W. M., Radhakrishnan, R., Roa, A., Smirnov, D., Smith, R.

E., Streeter, I., Thormann, A., Toneva, I., Vaughan, B., Zheng-Bradley, X., Bentley, D. R., Grocock, R., Humphray, S., James, T., Kingsbury, Z., Lehrach, H., Sudbrak, R., Albrecht, M. W., Amstislavskiy, V. S., Borodina, T. A., Lienhard, M., Mertes, F., Sultan, M., Timmermann, B., Yaspo, M.-L., Mardis, E. R., Wilson, R. K., Fulton, L., Fulton, R., Sherry, S. T., Ananiev, V., Belaia, Z., Beloslyudtsev, D., Bouk, N., Chen, C., Church, D., Cohen, R., Cook, C., Garner, J., Hefferon, T., Kimelman, M., Liu, C., Lopez, J., Meric, P., O.Sullivan, C., Ostapchuk, Y., Phan, L., Ponomarov, S., Schneider, V., Shekhtman, E., Sirotkin, K., Slotta, D., Zhang, H., McVean, G. A., Durbin, R. M., Balasubramaniam, S., Burton, J., Danecek, P., Keane, T. M., Kolb-Kokocinski, A., McCarthy, S., Stalker, J., Quail, M., Schmidt, J. P., Davies, C. J., Gollub, J., Webster, T., Wong, B., Zhan, Y., Auton, A., Campbell, C. L., Kong, Y., Marcketta, A., Gibbs, R. A., Yu, F., Antunes, L., Bainbridge, M., Muzny, D., Sabo, A., Huang, Z., Wang, J., Coin, L. J. M., Fang, L., Guo, X., Jin, X., Li, G., Li, Q., Li, Y., Li, Z., Lin, H., Liu, B., Luo, R., Shao, H., Xie, Y., Ye, C., Yu, C., Zhang, F., Zheng, H., Zhu, H., Alkan, C., Dal, E., Kahveci, F., Marth, G. T., Garrison, E. P., Kural, D., Lee, W.-P., Fung Leong, W., Stromberg, M., Ward, A. N., Wu, J., Zhang, M., Daly, M. J., DePristo, M. A., Handsaker, R. E., Altshuler, D. M., Banks, E., Bhatia, G., del Angel, G., Gabriel, S. B., Genovese, G., Gupta, N., Li, H., Kashin, S., Lander, E. S., McCarroll, S. A., Nemesh, J. C., Poplin, R. E., Yoon, S. C., Lihm, J., Makarov, V., Clark, A. G., Gottipati, S., Keinan, A., Rodriguez-Flores, J. L., Korbel, J. O., Rausch, T., Fritz, M. H., Stütz, A. M., Flicek, P., Beal, K., Clarke, L., Datta, A., Herrero, J., McLaren, W. M., Ritchie, G. R. S., Smith, R. E., Zerbino, D., Zheng-Bradley, X., Sabeti, P. C., Shlyakhter, I., Schaffner, S. F., Vitti, J., Cooper, D. N., Ball, E. V., Stenson, P. D., Bentley, D. R., Barnes, B., Bauer, M., Keira Cheetham, R., Cox, A., Eberle, M., Humphray, S., Kahn, S., Murray, L., Peden, J., Shaw, R., Kenny, E. E., Batzer, M. A., Konkel, M. K., Walker, J. A., MacArthur, D. G., Lek, M., Sudbrak, R., Amstislavskiy, V. S., Herwig, R., Mardis, E. R., Ding, L., Koboldt, D. C., Larson, D., Ye, K., Gravel, S., Swaroop, A., Chew, E., Lappalainen, T., Erlich, Y., Gymrek, M., Frederick Willems, T., Simpson, J. T., Shriver, M. D., Rosenfeld, J. A., Bustamante, C. D., Montgomery, S. B., De La Vega, F. M., Byrnes, J. K., Carroll, A. W., DeGorter, M. K., Lacroute, P., Maples, B. K., Martin, A. R., Moreno-Estrada, A., Shringarpure, S. S., Zakharia, F., Halperin, E., Baran, Y., Lee, C., Cerveira, E., Hwang, J., Malhotra, A., Plewczynski, D., Radew, K., Romanovitch, M., Zhang, C., Hyland, F.

C. L., Craig, D. W., Christoforides, A., Homer, N., Izatt, T., Kurdoglu, A. A., Sinari, S. A., Squire, K., Sherry, S. T., Xiao, C., Sebat, J., Antaki, D., Gujral, M., Noor, A., Ye, K., Burchard, E. G., Hernandez, R. D., Gignoux, C. R., Haussler, D., Katzman, S. J., James Kent, W., Howie, B., Ruiz-Linares, A., Dermitzakis, E. T., Devine, S. E., Abecasis, G. R., Min Kang, H., Kidd, J. M., Blackwell, T., Caron, S., Chen, W., Emery, S., Fritsche, L., Fuchsberger, C., Jun, G., Li, B., Lyons, R., Scheller, C., Sidore, C., Song, S., Sliwerska, E., Taliun, D., Tan, A., Welch, R., Kate Wing, M., Zhan, X., Awadalla, P., Hodgkinson, A., Li, Y., Shi, X., Quitadamo, A., Lunter, G., McVean, G. A., Marchini, J. L., Myers, S., Churchhouse, C., Delaneau, O., Gupta-Hinch, A., Kretzschmar, W., Iqbal, Z., Mathieson, I., Menelaou, A., Rimmer, A., Xifara, D. K., Oleksyk, T. K., Fu, Y., Liu, X., Xiong, M., Jorde, L., Witherspoon, D., Xing, J., Eichler, E. E., Browning, B. L., Browning, S. R., Hormozdiari, F., Sudmant, P. H., Khurana, E., Durbin, R. M., Hurles, M. E., Tyler-Smith, C., Albers, C. A., Ayub, Q., Balasubramaniam, S., Chen, Y., Colonna, V., Danecek, P., Jostins, L., Keane, T. M., McCarthy, S., Walter, K., Xue, Y., Gerstein, M. B., Abyzov, A., Balasubramanian, S., Chen, J., Clarke, D., Fu, Y., Harmanci, A. O., Jin, M., Lee, D., Liu, J., Jasmine Mu, X., Zhang, J., Zhang, Y., Li, Y., Luo, R., Zhu, H., Alkan, C., Dal, E., Kahveci, F., Marth, G. T., Garrison, E. P., Kural, D., Lee, W.-P., Ward, A. N., Wu, J., Zhang, M., McCarroll, S. A., Handsaker, R. E., Altshuler, D. M., Banks, E., del Angel, G., Genovese, G., Hartl, C., Li, H., Kashin, S., Nemesh, J. C., Shakir, K., Yoon, S. C., Lihm, J., Makarov, V., Degenhardt, J., Korbel, J. O., Fritz, M. H., Meiers, S., Raeder, B., Rausch, T., Stütz, A. M., Flicek, P., Paolo Casale, F., Clarke, L., Smith, R. E., Stegle, O., Zheng-Bradley, X., Bentley, D. R., Barnes, B., Keira Cheetham, R., Eberle, M., Humphray, S., Kahn, S., Murray, L., Shaw, R., Lameijer, E.-W., Batzer, M. A., Konkel, M. K., Walker, J. A., Ding, L., Hall, I., Ye, K., Lacroute, P., Lee, C., Cerveira, E., Malhotra, A., Hwang, J., Plewczynski, D., Radew, K., Romanovitch, M., Zhang, C., Craig, D. W., Homer, N., Church, D., Xiao, C., Sebat, J., Antaki, D., Bafna, V., Michaelson, J., Ye, K., Devine, S. E., Gardner, E. J., Abecasis, G. R., Kidd, J. M., Mills, R. E., Dayama, G., Emery, S., Jun, G., Shi, X., Quitadamo, A., Lunter, G., McVean, G. A., Chen, K., Fan, X., Chong, Z., Chen, T., Witherspoon, D., Xing, J., Eichler, E. E., Chaisson, M. J., Hormozdiari, F., Huddleston, J., Malig, M., Nelson, B. J., Sudmant, P. H., Parrish, N. F., Khurana, E., Hurles, M. E., Blackburne, B., Lindsay, S. J., Ning, Z., Walter, K., Zhang, Y., Gerstein, M. B., Abyzov,

A., Chen, J., Clarke, D., Lam, H., Jasmine Mu, X., Sisu, C., Zhang, J., Zhang, Y., Gibbs, R. A., Yu, F., Bainbridge, M., Challis, D., Evani, U. S., Kovar, C., Lu, J., Muzny, D., Nagaswamy, U., Reid, J. G., Sabo, A., Yu, J., Guo, X., Li, W., Li, Y., Wu, R., Marth, G. T., Garrison, E. P., Fung Leong, W., Ward, A. N., del Angel, G., DePristo, M. A., Gabriel, S. B., Gupta, N., Hartl, C., Poplin, R. E., Clark, A. G., Rodriguez-Flores, J. L., Flicek, P., Clarke, L., Smith, R. E., Zheng-Bradley, X., MacArthur, D. G., Mardis, E. R., Fulton, R., Koboldt, D. C., Gravel, S., Bustamante, C. D., Craig, D. W., Christoforides, A., Homer, N., Izatt, T., Sherry, S. T., Xiao, C., Dermitzakis, E. T., Abecasis, G. R., Min Kang, H., McVean, G. A., Gerstein, M. B., Balasubramanian, S., Habegger, L., Yu, H., Flicek, P., Clarke, L., Cunningham, F., Dunham, I., Zerbino, D., Zheng-Bradley, X., Lage, K., Berg Jespersen, J., Horn, H., Montgomery, S. B., DeGorter, M. K., Khurana, E., Tyler-Smith, C., Chen, Y., Colonna, V., Xue, Y., Gerstein, M. B., Balasubramanian, S., Fu, Y., Kim, D., Auton, A., Marcketta, A., Desalle, R., Narechania, A., Wilson Sayres, M. A., Garrison, E. P., Handsaker, R. E., Kashin, S., McCarroll, S. A., Rodriguez-Flores, J. L., Flicek, P., Clarke, L., Zheng-Bradley, X., Erlich, Y., Gymrek, M., Frederick Willems, T., Bustamante, C. D., Mendez, F. L., David Poznik, G., Underhill, P. A., Lee, C., Cerveira, E., Malhotra, A., Romanovitch, M., Zhang, C., Abecasis, G. R., Coin, L., Shao, H., Mittelman, D., Tyler-Smith, C., Ayub, Q., Banerjee, R., Cerezo, M., Chen, Y., Fitzgerald, T. W., Louzada, S., Massaia, A., McCarthy, S., Ritchie, G. R., Xue, Y., Yang, F., Gibbs, R. A., Kovar, C., Kalra, D., Hale, W., Muzny, D., Reid, J. G., Wang, J., Dan, X., Guo, X., Li, G., Li, Y., Ye, C., Zheng, X., Altshuler, D. M., Flicek, P., Clarke, L., Zheng-Bradley, X., Bentley, D. R., Cox, A., Humphray, S., Kahn, S., Sudbrak, R., Albrecht, M. W., Lienhard, M., Larson, D., Craig, D. W., Izatt, T., Kurdoglu, A. A., Sherry, S. T., Xiao, C., Haussler, D., Abecasis, G. R., McVean, G. A., Durbin, R. M., Balasubramaniam, S., Keane, T. M., McCarthy, S., Stalker, J., Chakravarti, A., Knoppers, B. M., Abecasis, G. R., Barnes, K. C., Beiswanger, C., Burchard, E. G., Bustamante, C. D., Cai, H., Cao, H., Durbin, R. M., Gerry, N. P., Gharani, N., Gibbs, R. A., Gignoux, C. R., Gravel, S., Henn, B., Jones, D., Jorde, L., Kaye, J. S., Keinan, A., Kent, A., Kerasidou, A., Li, Y., Mathias, R., McVean, G. A., Moreno-Estrada, A., Ossorio, P. N., Parker, M., Resch, A. M., Rotimi, C. N., Royal, C. D., Sandoval, K., Su, Y., Sudbrak, R., Tian, Z., Tishkoff, S., Toji, L. H., Tyler-Smith, C., Via, M., Wang, Y., Yang, H., Yang, L., Zhu, J., Bodmer, W., Bedoya, G., Ruiz-Linares, A., Cai, Z., Gao, Y.,

Chu, J., Peltonen, L., Garcia-Montero, A., Orfao, A., Dutil, J., Martinez-Cruzado, J. C., Oleksyk, T. K., Barnes, K. C., Mathias, R. A., Hennis, A., Watson, H., McKenzie, C., Qadri, F., LaRocque, R., Sabeti, P. C., Zhu, J., Deng, X., Sabeti, P. C., Asogun, D., Folarin, O., Happi, C., Omoniwa, O., Stremlau, M., Tariyal, R., Jallow, M., Sisay Joof, F., Corrah, T., Rockett, K., Kwiatkowski, D., Kooner, J., Tịnh Hiê`n, T., Dunstan, S. J., Thuy Hang, N., Fonnie, R., Garry, R., Kanneh, L., Moses, L., Sabeti, P. C., Schieffelin, J., Grant, D. S., Gallo, C., Poletti, G., Saleheen, D., Rasheed, A., Brooks, L. D., Felsenfeld, A. L., McEwen, J. E., Vaydylevich, Y., Green, E. D., Duncanson, A., Dunn, M., Schloss, J. A., Wang, J., Yang, H., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Min Kang, H., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A. and Abecasis, G. R. (2015) A global reference for human genetic variation. *Nature*, 526(7571), pp. 68–74. doi: 10.1038/nature15393.

Behar, D. M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan, H., Tzur, S., Comas, D., Bertranpetit, J., Quintana-Murci, L., Tyler-Smith, C., Wells, R. S. and Rosset, S. (2008) The Dawn of Human Matrilineal Diversity. *American Journal of Human Genetics*, 82(5), pp. 1130–1140. doi: 10.1016/j.ajhg.2008.04.002.

Børsting, C. and Morling, N. (2011) Mutations and/or close relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers. *Forensic Science International: Genetics*, 5(3), pp. 236–41. doi: 10.1016/j.fsigen.2010.02.007.

Børsting, C. and Morling, N. (2015) Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics*, 18, pp 78-79. doi: 10.1016/j.fsigen.2015.02.002.

Børsting, C., Sanchez, J. J., Hansen, H. E., Hansen, A. J., Bruun, H. Q., Morling, N. and Borsting, C. (2008) Performance of the SNPforID 52 SNP-plex assay in paternity testing. *Forensic Science International: Genetics*, 2(4), pp. 292–300. doi: 10.1016/j.fsigen.2008.03.007.

Brandstätter, A., Parsons, T. J. and Parson, W. (2003) Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *International Journal of Legal Medicine*, 117(5), pp. 291–298. doi: 10.1007/s00414-003-0395-2.

Brion, M., Sobrino, B., Blanco-Verea, A., Lareu, M. V. and Carracedo, A. (2005)

Hierarchical analysis of 30 Y-chromosome SNPs in European populations. *International Journal of Legal Medicine*, 119(1), pp. 10–15. doi: 10.1007/s00414-004-0439-2.

Budowle, B. and van Daal, A. (2008) Forensically relevant SNP classes. *Biotechniques*, 44(5), p. 603–608,610. doi: 10.2144/000112806.

Butler, J. M., Budowle, B., Gill, P., Kidd, K. K., Phillips, C., Schneider, P. M., Vallone, P. M. and Morling, N. (2008) Report on ISFG SNP Panel Discussion. *Forensic Science International: Genetics Supplement Series*, 1(1), pp. 471–472. doi: 10.1016/j.fsigss.2007.10.159.

Butler, J. M., Coble, M. D. and Vallone P. M. (2007) STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Science, Medicine, and Pathology*, 3(3), pp. 200–205. doi: 10.1007/s12024-007-0018-1.

Chakraborty, R., Stivers, D. N., Su, B., Zhong, Y. and Budowle, B. (1999) The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems. *Electrophoresis*, 20(8), pp. 1682–1696. doi: 10.1002/(SICI)1522-2683(19990101)20:8<1682::AID-ELPS1682>3.0.CO;2-Z.

Chapman, S. J. and Hill, A. V. S. (2012) Human genetic susceptibility to infectious disease. *Nature Reviews Genetics*, 13(3), pp. 175–188. doi: 10.1038/nrg3114.

Cho, S., Yu, H. J., Han, J., Kim, Y., Lee, J. and Lee, S. D. (2014) Forensic application of SNP-based resequencing array for individual identification. *Forensic Science International: Genetics*, 13, pp. 45–52. doi: 10.1016/j.fsigen.2014.07.005.

Churchill, J. D., Schmedes, S. E., King, J. L. and Budowle, B. (2016) Evaluation of the Illumina Beta Version ForenSeq DNA Signature Prep Kit for use in genetic profiling. *Forensic Science International: Genetics*, 20, pp. 20–29. doi: 10.1016/j.fsigen.2015.09.009.

Coble, M. D., Just, R. S., O.Callaghan, J. E., Letmanyi, I. H., Peterson, C. T., Irwin, J. A. and Parsons, T. J. (2004) Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. *International Journal of Legal Medicine*, 118(3), pp. 137–146. doi: 10.1007/s00414-004-0427-6.

Coble, M. D., Vallone, P. M., Just, R. S., Diegoli, T. M., Smith, B. C. and Parsons, T. J. (2006) Effective strategies for forensic analysis in the mitochondrial DNA coding region. *International Journal of Legal Medicine*, 120(1), pp. 27–32. doi:

10.1007/s00414-005-0044-z.

Costa, G., Dario, P., Lucas, I., Ribeiro, T., Espinheira, R. and Geada, H. (2008) Autosomal SNPs in paternity investigation. *Forensic Science International: Genetics Supplement Series*, 1(1), pp. 507–509. doi: 10.1016/j.fsigss.2007.10.198.

Dean, M., Stephens, J. C., Winkler, C., Lomb, D. A., Ramsburg, M., Boaze, R., Stewart, C., Charbonneau, L., Goldman, D. and Albaugh, B. J. (1994) Polymorphic admixture typing in human ethnic populations. *American Journal of Human Genetics*, 55(4), pp. 788–808.

Dembinski, G. M. and Picard, C. J. (2014) Evaluation of the IrisPlex DNA-based eye color prediction assay in a United States population. *Forensic Science International: Genetics*, 9(0), pp. 111–117. doi: 10.1016/j.fsigen.2013.12.003.

Dixon, L. A., Murray, C. M., Archer, E. J., Dobbins, A. E., Koumi, P. and Gill, P. (2005) Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. *Forensic Science International*, 154(1), pp. 62–77. doi: 10.1016/j.forsciint.2004.12.011.

Engle, L. J., Simpson, C. L. and Landers, J. E. (2006) Using high-throughput SNP technologies to study cancer. *Oncogene*, 25(11), pp. 1594–1601. doi: 10.1038/sj.onc.1209368.

Fondevila, M., Phillips, C., Naveran, N., Fernandez, L., Cerezo, M., Salas, A., Carracedo, A. and Lareu, M. V (2008) Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. *Forensic Science International: Genetics*, 2(3), pp. 212–8. doi: 10.1016/j.fsigen.2008.02.005.

Freire-Aradas, A., Fondevila, M., Kriegel, A.-K., Phillips, C., Gill, P., Prieto, L., Schneider, P. M., Carracedo, A. and Lareu, M. V (2012) A new SNP assay for identification of highly degraded human DNA. *Forensic Science International: Genetics*, 6(3), pp. 341–9. doi: 10.1016/j.fsigen.2011.07.010.

Galanter, J. M., Fernandez-Lopez, J. C., Gignoux, C. R., Barnholtz-Sloan, J., Fernandez-Rozadilla, C., Via, M., Hidalgo-Miranda, A., Contreras, A. V., Figueroa, L. U., Raska, P., Jimenez-Sanchez, G., Zolezzi, I. S., Torres, M., Ponte, C. R., Ruiz, Y., Salas, A., Nguyen, E., Eng, C., Borjas, L., Zabala, W., Barreto, G., González, F. R., Ibarra, A., Taboada, P., Porras, L., Moreno, F., Bigham, A., Gutierrez, G., Brutsaert, T., León-

Velarde, F., Moore, L. G., Vargas, E., Cruz, M., Escobedo, J., Rodriguez-Santana, J., Rodriguez-Cintrón, W., Chapela, R., Ford, J. G., Bustamante, C., Seminara, D., Shriver, M., Ziv, E., Burchard, E. G., Haile, R., Parra, E. and Carracedo, A. (2012) Development of a panel of genome-wide ancestry informative markers to study admixture throughout the americas. *PLoS Genetics*, 8(3). doi: 10.1371/journal.pgen.1002554.

Geppert, M., Baeta, M., Núñez, C., Martínez-Jarreta, B., Zweynert, S., Cruz, O. W. V., González-Andrade, F., González-Solorzano, J., Nagy, M. and Roewer, L. (2011) Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America. *Forensic Science International: Genetics*, 5(2), pp. 100–104. doi: 10.1016/j.fsigen.2010.08.016.

Gerstenblith, M. R., Shi, J. and Landi, M. T. (2010) Genome-wide association studies of pigmentation and skin cancer: a review and meta-analysis. *Pigment Cell Melanoma Res*, 23(5), pp. 587–606. doi: 10.1111/j.1755-148X.2010.00730.x.

Gill, P. (2001) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *International Journal of Legal Medicine*, 114(4–5), pp. 204–210. doi: 10.1007/s004149900117

González-Andrade, F., Sánchez, D., Penacino, G. and Martínez Jarreta, B. (2009) Two fathers for the same child: a deficient paternity case of false inclusion with autosomic STRs. *Forensic Science International: Genetics*, 3(2), pp. 138–40. doi: 10.1016/j.fsigen.2008.09.008.

Hammer, M. F. (1995) A recent common ancestry for human Y chromosomes. *Nature*, 378(6555), pp. 376–378. doi: 10.1038/378376a0.

Han, J., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., Hankinson, S. E., Hu, F. B., Duffy, D. L., Zhao, Z. Z., Martin, N. G., Montgomery, G. W., Hayward, N. K., Thomas, G., Hoover, R. N., Chanock, S. and Hunter, D. J. (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS genetics*, 4(5), p. e1000074. doi: 10.1371/journal.pgen.1000074.

Hou, G., Jiang, X., Yang, Y., Jia, F., Li, Q., Zhao, J., Guo, F. and Liu, L. (2014) A 21-locus autosomal SNP multiplex and its application in forensic science. *Journal of Forensic Sciences*, 59(1), pp. 5–14. doi: 10.1111/1556-4029.12259.

Huang, Q.-Y., Xu, F.-H., Shen, H., Deng, H.-Y., Liu, Y.-J., Liu, Y.-Z., Li, J.-L., Recker, R. R. and

Deng, H.-W. (2002) Mutation patterns at dinucleotide microsatellite loci in humans. *American Journal of Human Genetics*, 70(3), pp. 625–634. doi: 10.1086/338997.

Jobling, M. A. (2001) Y-chromosomal SNP haplotype diversity in forensic analysis. *Forensic Science International*, 118(2–3), pp. 158–62. doi: 10.1016/S0379-0738(01)00385-1.

Jobling, M. A., Pandya, A. and Tyler-Smith, C. (1997) The Y chromosome in forensic analysis and paternity testing. *International Journal of Legal Medicine*, 110(3), pp. 118–124. doi: 10.1007/s004140050050.

Kastelic, V., Pospiech, E., Draus-Barini, J., Branicki, W. and Drobnic, K. (2013) Prediction of eye color in the Slovenian population using the IrisPlex SNPs. *Croatian Medical Journal*, 54(4), pp. 381–386. doi: 10.3325/cmj.2013.54.381.

Kayser, M. (2007) Uni-parental markers in human identity testing including forensic DNA analysis. *BioTechniques*, 43(6), p. Sxv-Sxxi. doi: 10.2144/000112580.

Kayser, M. (2015) Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Science International: Genetics*, 18, pp. 33–48. doi: 10.1016/j.fsigen.2015.02.003.

Kayser, M. and de Knijff, P. (2011) Improving human forensics through advances in genetics, genomics and molecular biology. *Nature Reviews Genetics*, 12(3), pp. 179–192. doi: 10.1038/nrg2952.

Kayser, M. and Schneider, P. M. (2009) DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. *Forensic science international. Genetics*, 3(3), pp. 154–61. doi: 10.1016/j.fsigen.2009.01.012.

Keating, B., Bansal, A. T., Walsh, S., Millman, J., Newman, J., Kidd, K., Budowle, B., Eisenberg, A., Donfack, J., Gasparini, P., Budimlija, Z., Henders, A. K., Chandrupatla, H., Duffy, D. L., Gordon, S. D., Hysi, P., Liu, F., Medland, S. E., Rubin, L., Martin, N. G., Spector, T. D., Kayser, M. and International Visible Trait Genetics, C. (2013) First all-in-one diagnostic tool for DNA intelligence: genome-wide inference of biogeographic ancestry, appearance, relatedness, and sex with the Identitas v1 Forensic Chip. *International Journal of Legal Medicine*, 127(3), pp. 559–572. doi: 10.1007/s00414-012-0788-1.

Kersbergen, P., van Duijn, K., Kloosterman, A. D., den Dunnen, J. T., Kayser, M. and de

Knijff, P. (2009) Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans. *BMC genetics*, 10(1), p. 69. doi: 10.1186/1471-2156-10-69.

Kersbergen, P., Vaneede, P., Kraaijenbrink, T., Lardy, N. M. M., Sijen, T., Bakker, E., Deknijff, P., van Eede, P. H. H. and de Knijff, P. (2008) "False positive" or true paternity: Investigating one or two STR mismatches by detailed SNP analyses. *Forensic Science International: Genetics Supplement Series*, 1(1), pp. 518–519. doi: 10.1016/j.fsigss.2007.10.106.

Kidd, J. R., Friedlaender, F. R., Speed, W. C., Pakstis, A. J., De La Vega, F. M. and Kidd, K. K. (2011) Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples. *Investigative genetics*, 2(1), p. 1. doi: 10.1186/2041-2223-2-1.

Kim, J.-J., Han, B.-G., Lee, H.-I., Yoo, H.-W. and Lee, J.-K. (2010) Development of SNP-based human identification system. *International Journal of Legal Medicine*, 124(2), pp. 125–131. doi: 10.1007/s00414-009-0389-9.

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J.-Y., Sackler, R. S., Haynes, C., Henning, A. K., SanGiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C. and Hoh, J. (2005) Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science*, 308(5720), pp. 385–389. doi: 10.1126/science.1109557.

Köhnemann, S., Hohoff, C., Pfeiffer, H., Kohnemann, S., Hohoff, C. and Pfeiffer, H. (2009) An economical mtDNA SNP assay detecting different mitochondrial haplogroups in identical HVR 1 samples of Caucasian ancestry. *Mitochondrion*, 9(5), pp. 370–375. doi: 10.1016/j.mito.2009.08.002.

Kosoy, R., Nassir, R., Tian, C., White, P. a, Butler, L. M., Silva, G., Kittles, R., Alarcon-Riquelme, M. E., Gregersen, P. K., Belmont, J. W., De La Vega, F. M. and Seldin, M. F. (2009) Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Human Mutation*, 30(1), pp. 69–78. doi: 10.1002/humu.20822.

Krjutskov, K., Viltrop, T., Palta, P., Metspalu, E., Tamm, E., Suvi, S., Sak, K., Merilo, A., Sork, H., Teek, R., Nikopensius, T., Kivisild, T. and Metspalu, A. (2009) Evaluation of the 124-plex SNP typing microarray for forensic testing. *Forensic Science*

International: Genetics, 4(1), pp. 43–48. doi: 10.1016/j.fsigen.2009.04.007.

Kwon, H. Y., Bultman, S. J., Löffler, C., Chen, W. J., Furdon, P. J., Powell, J. G., Usala, A. L., Wilkison, W., Hansmann, I. and Woychik, R. P. (1994) Molecular structure and chromosomal mapping of the human homolog of the agouti gene. *Proceedings of the National Academy of Sciences of the United States of America*, 91(21), pp. 9760–9764.

Lamason, R. L., Mohideen, M. A., Mest, J. R., Wong, A. C., Norton, H. L., Aros, M. C., Jurynec, M. J., Mao, X., Humphreville, V. R., Humbert, J. E., Sinha, S., Moore, J. L., Jagadeeswaran, P., Zhao, W., Ning, G., Makalowska, I., McKeigue, P. M., O.Donnell, D., Kittles, R., Parra, E. J., Mangini, N. J., Grunwald, D. J., Shriver, M. D., Canfield, V. A. and Cheng, K. C. (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science*, 310(5755), pp. 1782–1786. doi: 10.1126/science.1116238.

Lao, O., van Duijn, K., Kersbergen, P., de Knijff, P. and Kayser, M. (2006) Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. *American Journal of Human Genetics*, 78(4), pp. 680–690. doi: 10.1086/501531.

Lehner, B. (2013) Genotype to phenotype: lessons from model organisms for human genetics. *Nature Reviews Genetics*, 14(3), pp. 168–178. doi: 10.1038/nrg3404.

Liu, F., van Duijn, K., Vingerling, J. R., Hofman, A., Uitterlinden, A. G., Janssens, a C. J. W. and Kayser, M. (2009) Eye color and the prediction of complex phenotypes from genotypes. *Current Biology : CB*, 19(5), pp. R192-3. doi: 10.1016/j.cub.2009.01.027.

Maroñas, O., Phillips, C., Söchtig, J., Gomez-Tato, A., Cruz, R., Alvarez-Dios, J., De Cal, M. C., Ruiz, Y., Fondevila, M., Carracedo, Á. and Lareu, M. V. (2014) Development of a forensic skin colour predictive test. *Forensic Science International: Genetics*, 13, pp. 34–44. doi: 10.1016/j.fsigen.2014.06.017.

Martin, P. D. D., Schmitter, H. and Schneider, P. M. M. (2001) A brief history of the formation of DNA databases in forensic science within Europe. *Forensic Science International*, 119(2), pp. 225–231. doi: 10.1016/S0379-0738(00)00436-9.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M. D., Sukernik, R. I., Olckers, A. and Wallace, D. C. (2003) Natural selection shaped regional mtDNA variation in humans. *Proceedings*

of the National Academy of Sciences of the United States of America, 100(1), pp. 171–6. doi: 10.1073/pnas.0136972100.

Murphy, E. (2013) Legal and Ethical Issues in Forensic DNA Phenotyping.

Musgrave-Brown, E., Ballard, D., Balogh, K., Bender, K., Berger, B., Bogus, M., Borsting, C., Brion, M., Fondevila, M., Harrison, C., Oguzturun, C., Parson, W., Phillips, C., Proff, C., Ramos-Luis, E., Sanchez, J. J., Sanchez Diz, P., Sobrino Rey, B., Stradmann-Bellinghausen, B., Thacker, C., Carracedo, A., Morling, N., Scheithauer, R., Schneider, P. M. and Syndercombe Court, D. (2007) Forensic validation of the SNPforID 52-plex assay. *Forensic Science International: Genetics*, 1(2), pp. 186–190. doi: 10.1016/j.fsigen.2007.01.004.

Nelson, T. M., Just, R. S., Loreille, O., Schanfield, M. S. and Podini, D. (2007) Development of a multiplex single base extension assay for mitochondrial DNA haplogroup typing. *Croatian Medical Journal*, 48, pp. 460–472.

Nievergelt, C. M., Maihofer, A. X., Shekhtman, T., Libiger, O., Wang, X., Kidd, K. K. and Kidd, J. R. (2013) Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel. *Investigative Genetics*, 4(1), p. 13. doi: 10.1186/2041-2223-4-13.

Onofri, V., Alessandrini, F., Turchi, C., Pesaresi, M., Buscemi, L. and Tagliabracci, A. (2006) Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. *Forensic Science International*, 157(1), pp. 23–35. doi: 10.1016/j.forsciint.2005.03.014.

van Oven, M. and Kayser, M. (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, 30(2), pp. 386–394. doi: 10.1002/humu.20921.

Pakstis, A. J., Speed, W. C., Fang, R., Hyland, F. C. L., Furtado, M. R., Kidd, J. R. and Kidd, K. K. (2010) SNPs for a universal individual identification panel. *Human Genetics*, 127(3), pp. 315–24. doi: 10.1007/s00439-009-0771-1.

Pardo-Seco, J., Martinón-Torres, F. and Salas, A. (2014) Evaluating the accuracy of AIM panels at quantifying genome ancestry. *BMC Genomics*, 15(1), p. 543. doi: 10.1186/1471-2164-15-543.

Park, M. J., Lee, H. Y., Kim, N. Y., Lee, E. Y., Yang, W. I. and Shin, K. J. (2013) Y-SNP miniplexes for East Asian Y-chromosomal haplogroup determination in degraded

DNA. *Forensic Science International: Genetics*, 7(1), pp. 75–81. doi: 10.1016/j.fsigen.2012.06.014.

Paulsen, J. S., Nance, M., Kim, J. I., Carlozzi, N. E., Panegyres, P. K., Erwin, C., Goh, A., McCusker, E. and Williams, J. K. (2013) A review of quality of life after predictive testing for and earlier identification of neurodegenerative diseases. *Progress in Neurobiology*, 110, pp. 2–28. doi: 10.1016/j.pneurobio.2013.08.003.

Gill, P., Werrett, D. J., Budowle, B. and Guerrieri, R. (2004) An assessment of whether SNPs will replace STRs in national DNA databases--joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM). *Science and Justice*, 44(1), pp. 51–53.

Pfaff, C. L., Barnholtz-Sloan, J., Wagner, J. K. and Long, J. C. (2004) Information on Ancestry from Genetic Markers. *Genetic Epidemiology*, 26(4), pp. 305–315. doi: 10.1002/gepi.10319.

Phillips, C., Parson, W., Lundsberg, B., Santos, C., Freire-Aradas, A., Torres, M., Eduardoff, M., Børsting, C., Johansen, P., Fondevila, M., Morling, N., Schneider, P., Carracedo, A. and Lareu, M. V (2014) Building a forensic ancestry panel from the ground up: The EUROFORGEN Global AIM-SNP set. *Forensic Science International: Genetics*, 11C, pp. 13–25. doi: 10.1016/j.fsigen.2014.02.012.

Phillips, C., Prieto, L., Fondevila, M., Salas, A., Gómez-Tato, A., Alvarez-Dios, J., Alonso, A., Blanco-Verea, A., Brión, M., Montesino, M., Carracedo, A. and Lareu, M. V. (2009) Ancestry analysis in the 11-M Madrid bomb attack investigation. *PloS one*, 4(8), p. e6583. doi: 10.1371/journal.pone.0006583.

Phillips, C., Salas, A., Sánchez, J. J., Fondevila, M., Gómez-Tato, a, Alvarez-Dios, J., Calaza, M., de Cal, M. C., Ballard, D., Lareu, M. V, Carracedo, A., Sanchez, J. J. and Gomez-Tato, A. (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Science International: Genetics*, 1(3–4), pp. 273–80. doi: 10.1016/j.fsigen.2007.06.008.

Poetsch, M., Blohm, R., Harder, M., Inoue, H., von Wurmb-Schwark, N. and Freitag-Wolf, S. (2013) Prediction of people.s origin from degraded DNA - presentation of SNP assays and calculation of probability. *International Journal of Legal Medicine*, 127(2), pp. 347–357. doi: 10.1007/s00414-012-0728-0.

Pośpiech, E., Draus-Barini, J., Kupiec, T., Wojas-Pelc, A. and Branicki, W. (2011) Gene-gene interactions contribute to eye colour variation in humans. *Journal of Human Genetics*, 56(6), pp. 447–455. doi: 10.1038/jhg.2011.38.

Pośpiech, E., Wojas-Pelc, A., Walsh, S., Liu, F., Maeda, H., Ishikawa, T., Skowron, M., Kayser, M. and Branicki, W. (2014) The common occurrence of epistasis in the determination of human pigmentation and its impact on DNA-based pigmentation phenotype prediction. *Forensic Science International: Genetics*, 11C, pp. 64–72. doi: 10.1016/j.fsigen.2014.01.012.

Quintáns, B., Alvarez-Iglesias, V., Salas, A., Phillips, C., Lareu, M. V and Carracedo, A. (2004) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Science International*, 140(2–3), pp. 251–7. doi: 10.1016/j.forsciint.2003.12.005.

Reich, D. E., Schaffner, S. F., Daly, M. J., McVean, G., Mullikin, J. C., Higgins, J. M., Richter, D. J., Lander, E. S. and Altshuler, D. (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. *Nature Genetics*, 32(1), pp. 135–142. doi: 10.1038/ng947.

Rosenberg, N. A. (2009) Genetic Structure of Human Populations. *Science*, 2381(2002), pp. 2381–2386. doi: 10.1126/science.1078311.

Rosenberg, N. A., Li, L. M., Ward, R. and Pritchard, J. K. (2003) Informativeness of Genetic Markers for Inference of Ancestry. *American Journal of Human Genetics*, 73, pp. 1402–1422. doi: 10.1086/380416.

Royal, C. D., Novembre, J., Fullerton, S. M., Goldstein, D. B., Long, J. C., Bamshad, M. J. and Clark, A. G. (2010) Inferring Genetic Ancestry: Opportunities, Challenges, and Implications., *American Journal of Human Genetics*. The American Society of Human Genetics, 86(5), pp. 661–673. doi: 10.1016/j.ajhg.2010.03.011.

Sanchez, J. J., Børsting, C., Hallenberg, C., Buchard, A., Hernandez, A. and Morling, N. (2003) Multiplex PCR and minisequencing of SNPs—a model with 35 Y chromosome SNPs. *Forensic Science International*, 137(1), pp. 74–84. doi: 10.1016/S0379-0738(03)00299-8.

Sanchez, J. J., Phillips, C., Borsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P. M., Carracedo, A., Morling, N. and Børsting, C. (2006) A multiplex assay with 52 single

nucleotide polymorphisms for human identification. *Electrophoresis*, 27(9), pp. 1713–1724. doi: 10.1002/elps.200500671.

Schneider, P. M. *SNPforID.org*. SNPforID Consortium. Available at: http://www.snpforid.org/.

Schneider, P. M. and Martin, P. D. (2001) Criminal DNA databases: the European situation. *Forensic Science International*, 119(2), pp. 232–8.

Senge, T., Madea, B., Junge, A., Rothschild, M. A. and Schneider, P. M. (2011) STRs, mini STRs and SNPs--a comparative study for typing degraded DNA. *Legal Medicine*, 13(2), pp. 68–74. doi: 10.1016/j.legalmed.2010.12.001.

Sobrino, B., Brión, M. and Carracedo, A. (2005) SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International*, 154(2–3), pp. 181–94. doi: 10.1016/j.forsciint.2004.10.020.

Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Jakobsdottir, M., Steinberg, S., Gudjonsson, S. a, Palsson, A., Thorleifsson, G., Pálsson, S., Sigurgeirsson, B., Thorisdottir, K., Ragnarsson, R., Benediktsdottir, K. R., Aben, K. K., Vermeulen, S. H., Goldstein, A. M., Tucker, M. a, Kiemeney, L. a, Olafsson, J. H., Gulcher, J., Kong, A., Thorsteinsdottir, U. and Stefansson, K. (2008) Two newly identified genetic determinants of pigmentation in Europeans. *Nature Genetics*, 40(7), pp. 835–7. doi: 10.1038/ng.160.

Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Magnusson, K. P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G., Jakobsdottir, M., Steinberg, S., Pálsson, S., Jonasson, F., Sigurgeirsson, B., Thorisdottir, K., Ragnarsson, R., Benediktsdottir, K. R., Aben, K. K., Kiemeney, L. a, Olafsson, J. H., Gulcher, J., Kong, A., Thorsteinsdottir, U., Stefansson, K. and Palsson, S. (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature Genetics*, 39(12), pp. 1443–1452. doi: 10.1038/ng.2007.13.

Thomson, R., Pritchard, J. K., Shen, P., Oefner, P. J. and Feldman, M. W. (2000) Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proceedings of the National Academy of Sciences*, 97(13), pp. 7360–7365. doi: 10.1073/pnas.97.13.7360.

Vallone, P. M., Decker, A. E. and Butler, J. M. (2005) Allele frequencies for 70 autosomal SNP loci with U.S. Caucasian, African-American, and Hispanic samples. *Forensic*

Science International, 149(2–3), pp. 279–286. doi: 10.1016/j.forsciint.2004.07.014.

Vallone, P. M., Just, R. S., Coble, M. D., Butler, J. M. and Parsons, T. J. (2004) A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. *International Journal of Legal Medicine*, 118(3), pp. 147–157. doi: 10.1007/s00414-004-0428-5.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A, Holt, R. A, Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V, Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K.,

Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Sjolander, K. V, Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. and Zhu, X. (2001) 'The sequence of the human genome.', Science, 291(5507), pp. 1304–51. doi: 10.1126/science.1058040.

Walsh, P. S., Metzger, D. A. and Higuchi, R. (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, *Biotechniques*, 10(4), pp. 506–513.

Walsh, S., Chaitanya, L., Clarisse, L., Wirken, L., Draus-Barini, J., Kovatsi, L., Maeda, H., Ishikawa, T., Sijen, T., de Knijff, P., Branicki, W., Liu, F. and Kayser, M. (2014) Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Science International: Genetics*, 9, pp. 150–61. doi: 10.1016/j.fsigen.2013.12.006.

Walsh, S., Lindenbergh, A., Zuniga, S. B., Sijen, T., de Knijff, P., Kayser, M. and Ballantyne, K. N. (2011a) Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence. *Forensic Science International: Genetics*, 5(5), pp. 464–71. doi: 10.1016/j.fsigen.2010.09.008.

Walsh, S., Liu, F., Ballantyne, K. N., van Oven, M., Lao, O., Kayser, M. and Oven, M. Van (2011b) IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International*, 5(3), pp. 170–80. doi: 10.1016/j.fsigen.2010.02.004.

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W. and Kayser, M. (2013) The HIrisPlex system for simultaneous prediction of hair and eye

colour from DNA. *Forensic Science International: Genetics*, 7(1), pp. 98–115. doi: 10.1016/j.fsigen.2012.07.005.

Walsh, S., Wollstein, A., Liu, F., Chakravarthy, U., Rahu, M., Seland, J. H., Soubrane, G., Tomazzoli, L., Topouzis, F., Vingerling, J. R., Vioque, J., Fletcher, A. E., Ballantyne, K. N. and Kayser, M. (2012) DNA-based eye colour prediction across Europe with the IrisPlex system. *Forensic Science International: Genetics*, 6(3), pp. 330–40. doi: 10.1016/j.fsigen.2011.07.009.

Wang, D. G., Fan, J.-B., Siao, C.-J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lipshutz, R., Chee, M. and Lander, E. S. (1998) Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. *Science*, 280(5366), pp. 1077–1082. doi: 10.1126/science.280.5366.1077.

Wang, Q., Fu, L., Zhang, X., Dai, X., Bai, M., Fu, G., Cong, B. and Li, S. (2016) Expansion of a SNaPshot assay to a 55-SNP multiplex: Assay enhancements, validation, and power in forensic science. *Electrophoresis*, 37(10), pp. 1310–7. doi: 10.1002/elps.201500353.

Yun, L., Gu, Y., Rajeevan, H. and Kidd, K. K. (2014) Application of six IrisPlex SNPs and comparison of two eye color prediction systems in diverse Eurasia populations. *International Journal of Legal Medicine*, 128(3), pp. 447-53. doi: 10.1007/s00414-013-0953-1.

Chapter II

Validation and application of an Individual Identification 20 SNP multiplex in kinship investigation

Paper 1.

Abstract published in:

Dario, P., Ribeiro, T., Oliveira, A.R., Dias, D., Geada, H., Corte Real, F. and Costa Santos, J. (2012) Internal Validation of 20 autosomal SNP Multiplex for Forensic Genetics. *International Journal of Legal Medicine* 126, pp. S1: 357. doi: 10.1007/s00414-012-0711-9

Paper 2.

Dario, P., Ribeiro, T., Espinheira, R. and Geada, H. (2009). SNPs in paternity investigation: The simple future. *Forensic Science International: Genetics Supplement Series*, 2(1), pp. 127–128. doi:10.1016/j.fsigss.2009.08.136

Paper 3.

Dario, P., Ribeiro, T., Espinheira, R., Dias, D., Geada, H. and Corte-Real, F. (2011) 20 SNPs as supplementary markers in kinship testing. *Forensic Science International: Genetics Supplement Series*, 3(1), pp. e508–e509. doi:10.1016/j.fsigss.2011.09.111

Paper 4.

Dario, P., Ribeiro, T., Dias, D., Corte-Real, F. and Geada, H. (2011) Complex casework using single nucleotide polymorphisms. *Forensic Science International: Genetics Supplement Series*, 3(1), pp. e379–e380. doi:10.1016/j.fsigss.2011.09.051

Int J Legal Med (2012) 126 (Suppl 1):S121–S384 S357

PP-759

Internal Validation of 20 SNP Multiplex for Forensic Genetics

Paulo Dario[1], Teresa Ribeiro[1], Ana Rita Oliveira[2], Deodália Dias[3], Helena Geada[4], Francisco Corte Real[1], Jorge Costa Santos[1]
[1]National Institute of Legal Medicine, Portugal
[2]CESAM – Centre for Environmental and Marine Studies, Portugal
[3]Faculty of Science, University of Lisbon, Lisbon, Portugal
[4]CENCIFOR – Forensic Sciences Centre, Portugal

Background: There has been a growing interest in Single Nucleotide Polymorphisms (SNPs) analysis in Forensic Genetics. In 2006, the SNPforID Consortium published the development of a multiplex PCR assay for detection of 52 autosomal SNPs for human identification, validated for forensic use in the following year. Even though the analysis of 52 SNPs may be ideal, a smaller number of loci can be sufficient to provide additional data for complex casework. The use of a 20 SNP multiplex has already been developed for paternity and kinship analysis. Here we present the internal validation study of that multiplex.

Method: DNA was extracted using Chelex®100 resin method from buccal swabs and/or blood samples of 113 unrelated individuals involved in paternity testing casework. DNA from animal samples was also extracted. PCR amplification of autosomal SNPs was performed in two 25 μl 10-plex reactions designed to amplify SNP loci rs1490413, rs1029047, rs763869, rs735155, rs2107612, rs1454361, rs2111980, rs1005533, rs8037429, rs891700 in the first multiplex and SNP loci rs2046361, rs717302, rs1886510, rs729172, rs1024116, rs1463729, rs2076848, rs1355366, rs907100, rs737681 in the second one. PCR conditions and SBE reactions were performed as described by Sanchez et al. with slight modifications and SBE products were analyzed by capillary electrophoresis on a 3130 Genetic Analyzer.

Results: An optimum PCR annealing temperature of 58°C was obtained although Sanchez et al. described it at 60°C. This assay revealed sensibility to obtain full profiles from 2.5 ng, 2.0 ng, 1.0 ng and 0.5 ng of human DNA. No SNP loci amplification was detected from animal samples, which included various vertebrates such as cats, dogs, horses and others.

Population studies were performed with 113 individual samples. All studied loci revealed allelic frequencies higher than 0.310, sixteen above 0.400. Discrimination Power of 99.9999995 % and Power of Exclusion (PE) of 97.88 % were obtained with the 20 SNP studied. Although PE is relatively low, these 20 SNP multiplex can be very informative when analyzed together with one or two routine STR multiplex loci.

Conclusion: A sensible and specific 20 SNP multiplex for forensic testing, based on SNPforID 52-plex, was validated. This multiplex shows human specificity, full profiles can be obtained from very low DNA quantity and is very informative when analyzed together with routine STR multiplex loci. Autosomal SNP analysis can be a valuable tool in Forensic Genetics, especially in kinship analysis and human identification.

(Original document modified for easy reading)

1. Introduction

There has been a growing interest in single nucleotide polymorphisms (SNPs) analysis in forensic genetics. In 2006, the SNP*for*ID Consortium published the development of a multiplex PCR assay for detection of 52 autosomal SNPs for human identification [1], validated for forensic use in the following year [2]. Other development topics such as genetic marker characterization, inheritance, chromosomal location, polymorphism or population allelic frequencies are fully studied and accessible in Musgrave-Brown work [2] or online in SNP*for*ID Consortium webpage [3].

Even though the analysis of 52 SNPs may be ideal, a smaller number of *loci* can be sufficient to provide additional data for complex casework. The use of a 20 SNP multiplex has already been developed for paternity and kinship analysis [4-6]. Here we present an validation study of that multiplex, based on SWGDAM and ISO/IEC 17025:2005 guidelines [7,8], with proper adaptations due to SNP nature.

2. Material and methods

DNA was extracted using Chelex®100 resin method from buccal swabs and/or blood samples of 113 unrelated individuals involved in paternity testing casework. DNA from animal samples was also extracted.

PCR amplification of autosomal SNPs was performed in two 25µl 10-plex reactions designed to amplify SNP *loci* rs1490413, rs1029047, rs763869, rs735155, rs2107612, rs1454361, rs2111980, rs1005533, rs8037429, rs891700 in the first multiplex and SNP *loci* rs2046361, rs717302, rs1886510, rs729172, rs1024116, rs1463729, rs2076848, rs1355366, rs907100, rs737681 in the second one.

PCR conditions and SBE reactions were performed as described by Sanchez et al. [1] with slight modifications and SBE products were analyzed by capillary electrophoresis on a 3130 Genetic Analyzer.

3. Results

Various experiments were performed to obtain SNPs electrophoregrams with good peak heights and shapes, as presented in figure 1.

For this multiplex an optimum PCR annealing temperature of 58ºC was obtained (figure 2), although Sanchez et al. described it at 60ºC for the 52 SNP*for*ID multiplex [1] from which this was derived.

As shown in figure 3, this assay revealed sensibility to obtain full profiles from 2.5 ng, 2.0 ng, 1.0 ng and 0.5 ng of human DNA, although partial profiles were obtained with less concentrated samples.

Known samples, including 9947A, 9948 and K562, were genotyped to evaluate correct genotyping and reproducibility and to evaluate peak size average and standard deviations (figure 4 and Table I). Allele matching was noticed to be affected by room temperature (data not shown).

SNP *loci* amplification was not detected from animal samples, which included various vertebrates such as cats, dogs, horses and others and DNA contamination was not detected using SNP procedures (data not shown).

Population parameters were examined studying samples from 113 individuals. All studied *loci* revealed allelic frequencies higher than 0.310, sixteen above 0.400. Discrimination Power of 99.9999995% and Power of Exclusion (PE) of 97.88% were obtained for the studied multiplex.

For comparison to established methods, paternity testing results obtained with standard STR results were compared to the ones obtained with the 20 SNP multiplex. As exemplified in figure 5 results are coincident between the two methodologies.

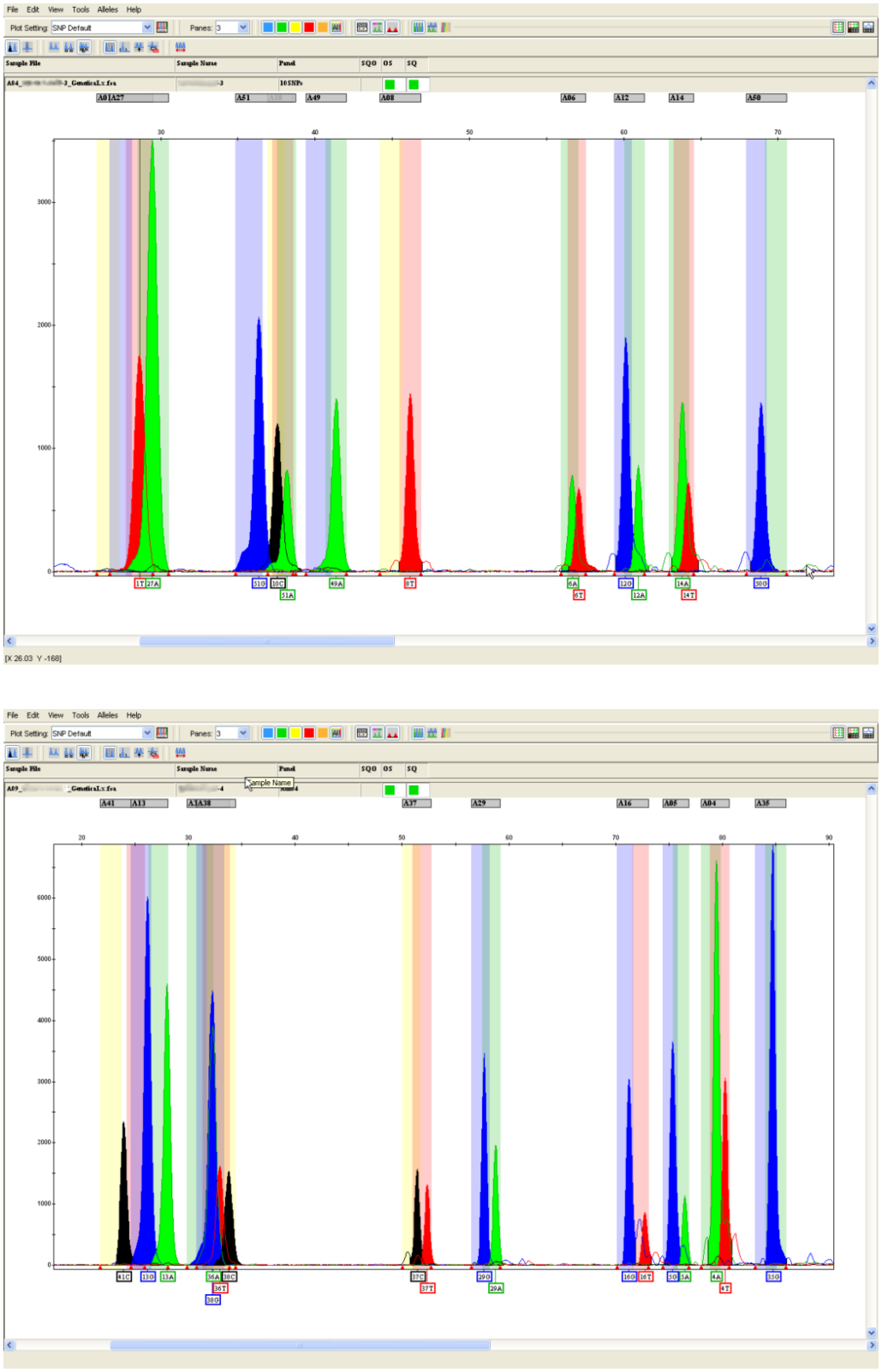


Fig. 1 - Electropherogram showing the amplification of the 20 SNPs after optimization. *As shown in the electropherogram, the good peak heights and shapes were attained after optimizatio*n. Upper panel includes *loci* A01, A27, A51, A10, A49, A08, A06, A12, A14 and A50 and lower panel includes *loci* A41, A13, A36, A38, A37, A29, A16, A05, A04 and A35.

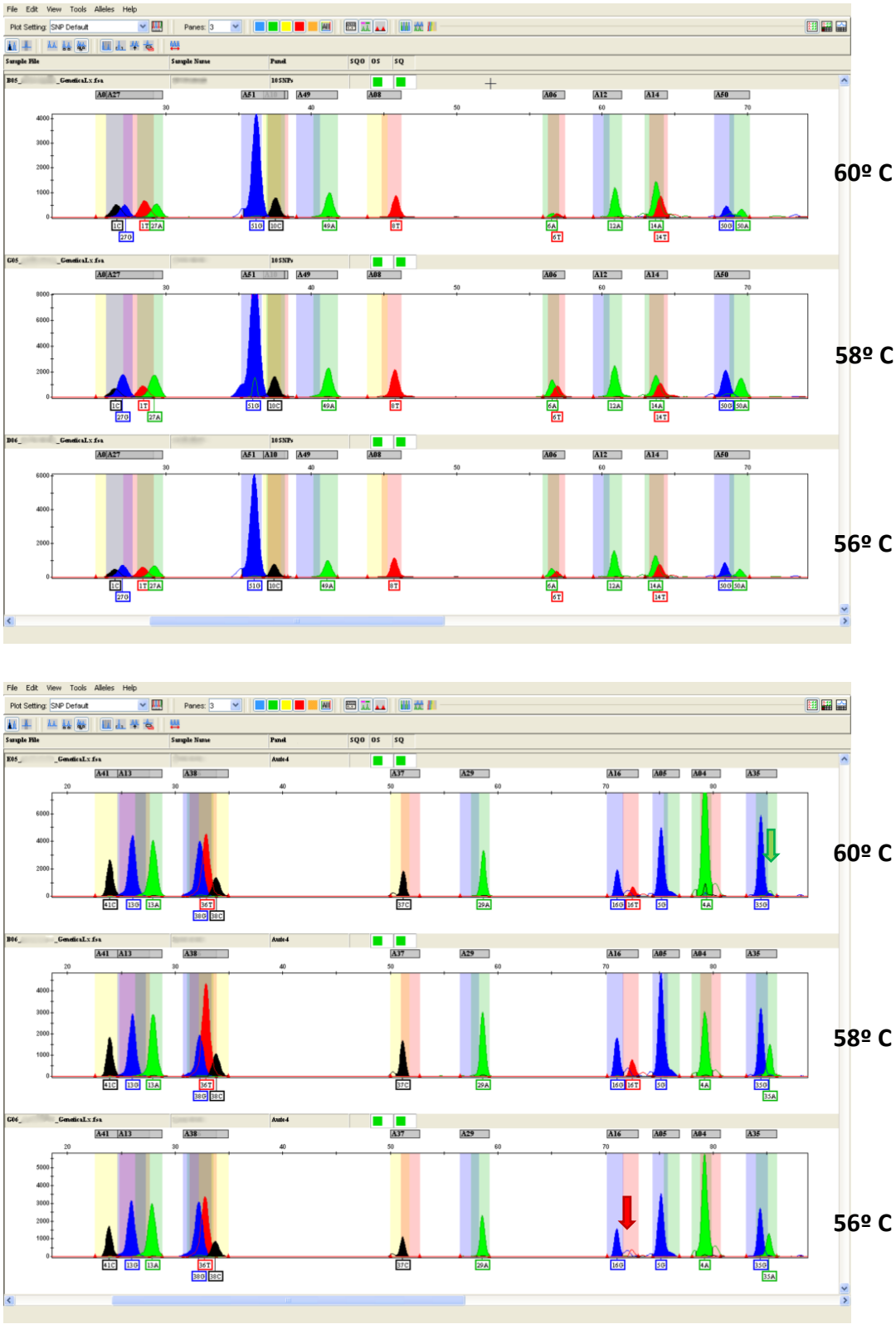


Fig. 2 - PCR annealing temperature optimization. As can be seen in this figure, internal PCR optimization was achieved with PCR annealing temperature of 58ºC. Minor peaks dropouts, using other annealing temperatures, are indicated with arrows of same color and place of the lost allele. Upper panel includes *loci* A01, A27, A51, A10, A49, A08, A06, A12, A14 and A50 and lower panel includes *loci* A41, A13, A36, A38, A37, A29, A16, A05, A04 and A35 r each temperature of annealing tested.

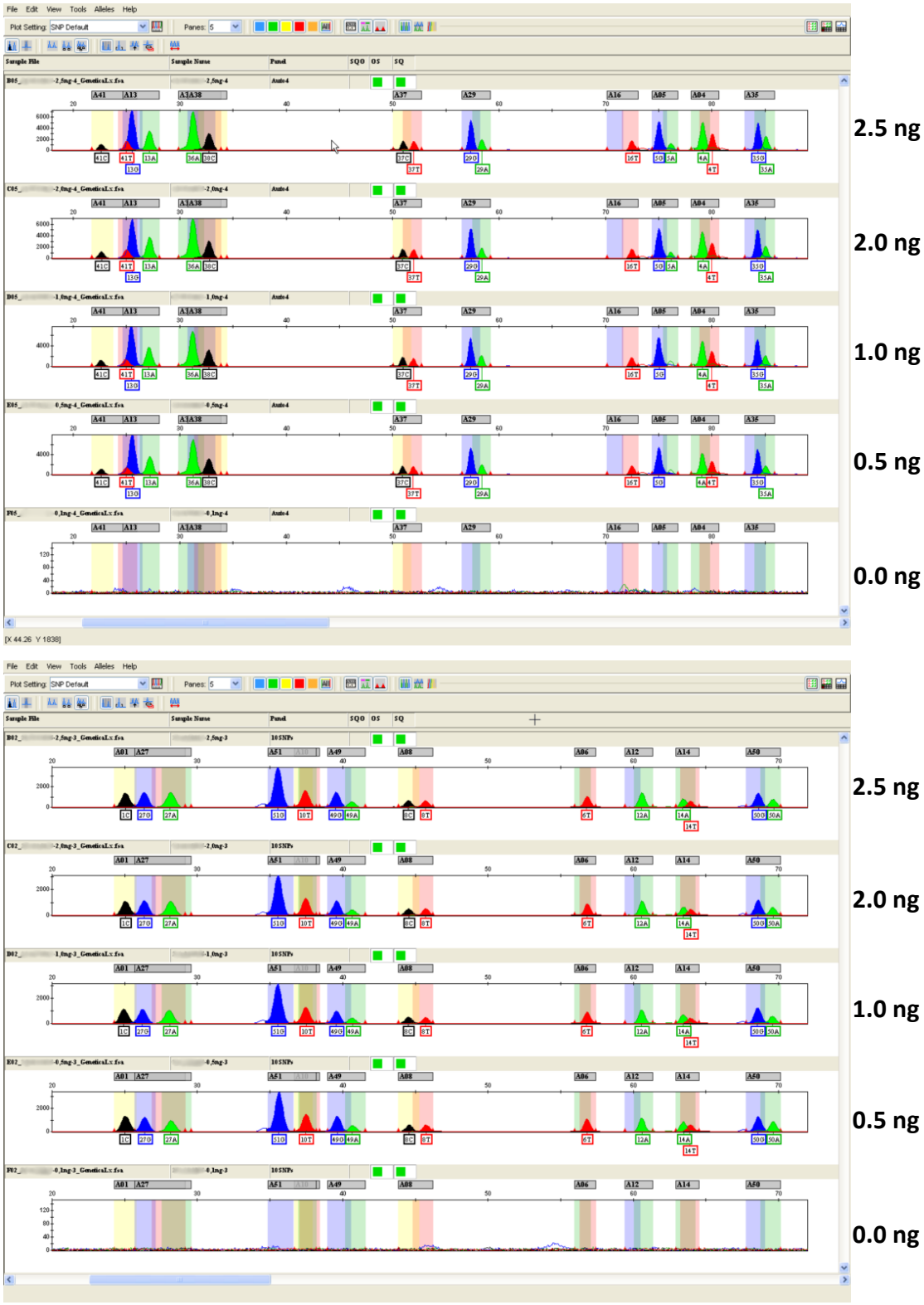


Fig. 3 - Sensitivity studies. Sensitivity of the proposed multiplex was investigated using different samples concentration: 2.5 ng, 2.0 ng, 1.0 ng, 0.5 ng, 0.1 ng (not presented) and 0.0 ng. Full genetic profiles were obtained from samples containing 0.5 ng, although partial profiles were obtained with less concentrated samples. Upper panel includes *loci* A01, A27, A51, A10, A49, A08, A06, A12, A14 and A50 and lower panel includes *loci* A41, A13, A36, A38, A37, A29, A16, A05, A04 and A35 for each concentration tested.

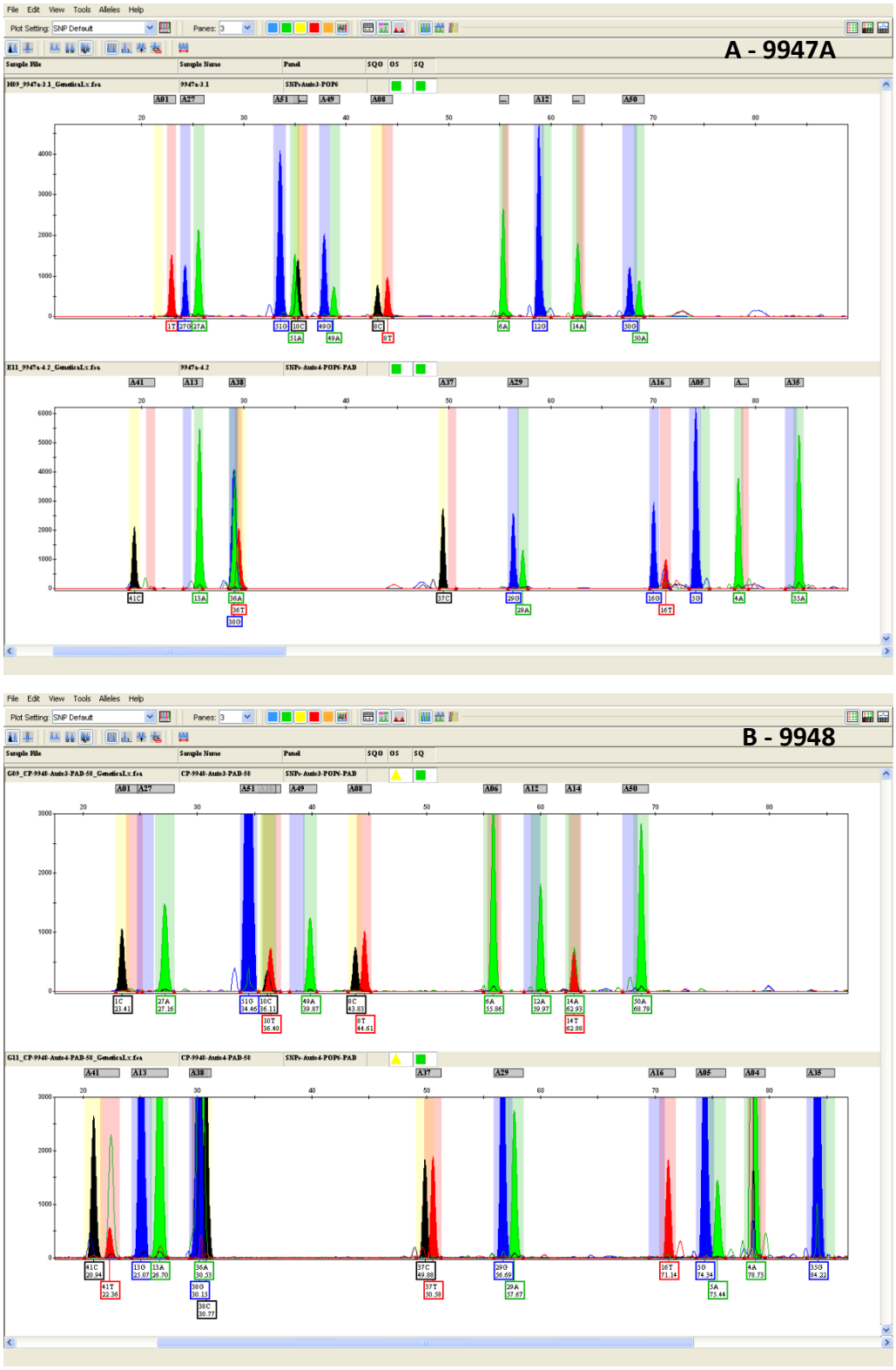


Fig. 4 - Known samples and reproducibility. To evaluate this multiplex correct genotyping, several injections of references samples 9947A, 9948 and K562 were genotyped at different times, presenting good results from samples with concentrations above 0.5 ng.

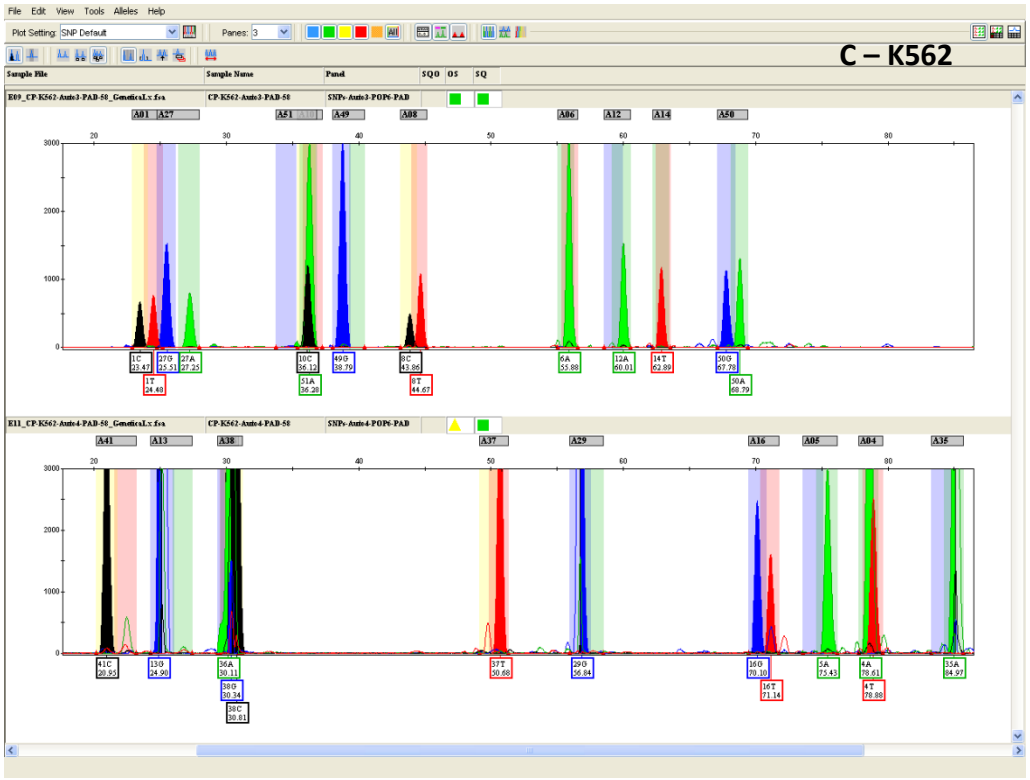


Fig. 4 (cont.) - Known samples and reproducibility. To evaluate this multiplex correct genotyping, several injections of references samples 9947A (A), 9948 (B) and K562 (C) were genotyped at different times, presenting good results from samples with concentrations above 0.5 ng.

Locus		**A01**		**A27**		**A51**		**A10**		**A49**		***BinSet* Auto3**
	Allele	01C	01T	27G	27A	51G	51A	10C	10T	49G	49A	
	Average	26,14	28,11	26,85	28,95	35,97	37,72	37,35	37,81	39,91	41,02	
	SD	0,34	0,30	0,23	0,27	0,16	0,18	0,14	0,10	0,14	0,15	
Locus		**A08**		**A06**		**A12**		**A14**		**A50**		
	Allele	08C	08T	06A	06T	12G	12A	14A	14T	50G	50A	
	Average	44,61	45,69	56,49	56,87	59,97	60,78	63,62	63,97	68,50	69,52	
	SD	0,16	0,16	0,08	0,08	0,07	0,07	0,10	0,13	0,16	0,18	
Locus		**A41**		**A13**		**A36**		**A38**		**A37**		***BinSet* Auto4**
	Allele	41C	41T	13G	13A	36A	36T	38G	38C	37C	37T	
	Average	23,38	25,71	25,73	27,59	31,78	32,48	31,81	33,34	50,96	51,95	
	SD	0,37	0,52	0,27	0,32	0,38	0,34	0,29	0,34	0,27	0,24	
Locus		**A29**		**A16**		**A05**		**A04**		**A35**		
	Allele	29G	29A	16G	16T	05G	05A	04A	04T	35G	35A	
	Average	57,39	58,47	70,98	72,45	75,08	76,22	79,15	79,90	84,28	85,08	
	SD	0,15	0,14	0,18	0,21	0,15	0,16	0,20	0,28	0,21	0,20	

Table I - Match criteria, calculation and comparison of obtained peak size average and standard deviations. In this table is presented the values, in bp, of obtained peak size average and standard deviations.

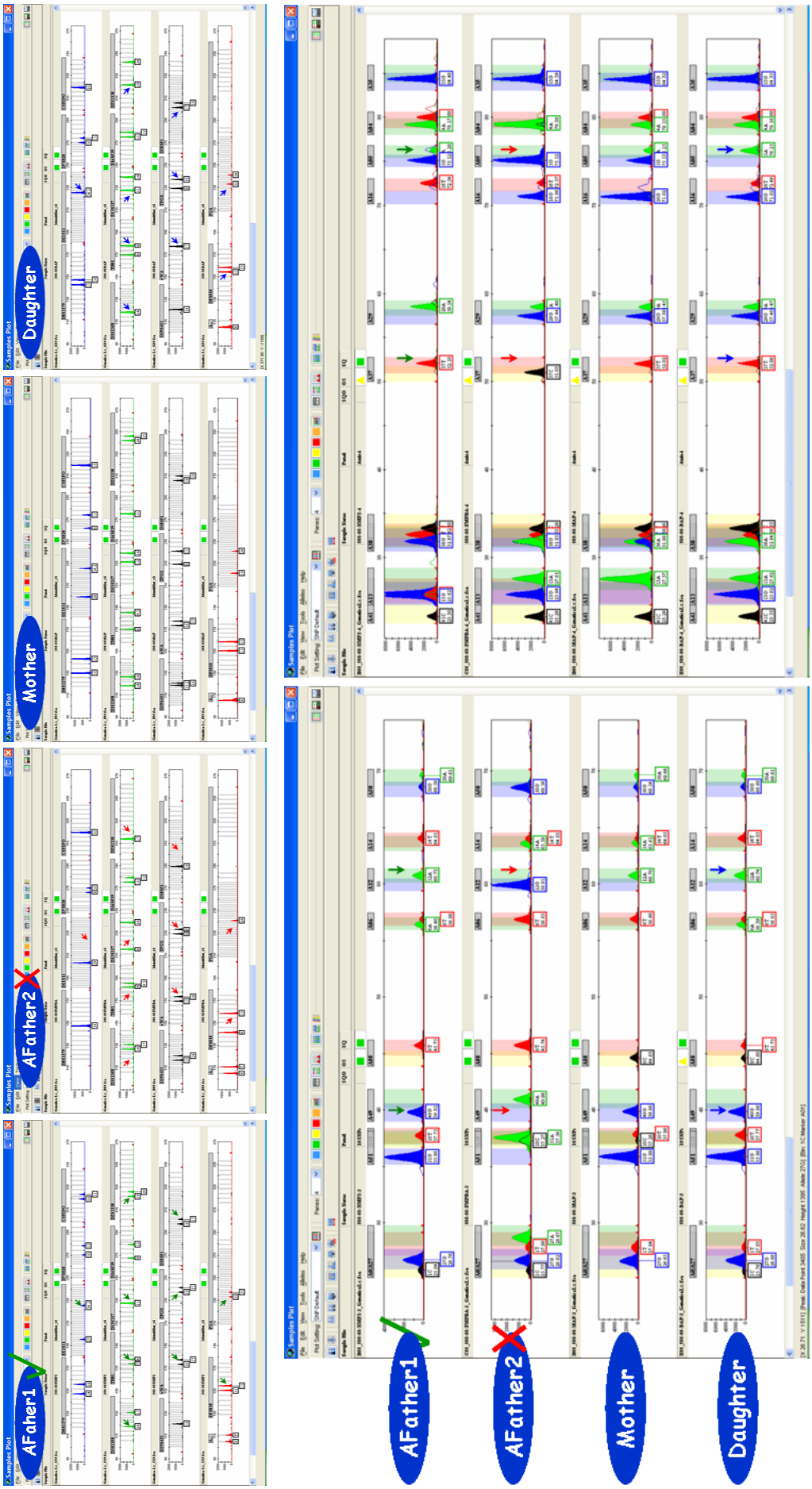


Fig. 5 – Comparison to established methods. Example of a paternity testing results obtained through STR (upper panel) and SNP analysis (downer panel) AFather1 - Putative Father 1; AFather2 - Putative Father 2 (excluded); Mother - Mother; Daughter - Child.

4. Discussion

Following SWGDAM guidelines [7] for the internal validation of this multiplex, the method presented the correct genotypes of known and non-probative evidence samples, with good reproducibility and match criteria, although affected by room temperature, probably due to small SNP amplicons.

Although this multiplex is based in SNP*for*ID one [1] which is validated for forensic use [2], there are differences, namely in the number of tested *loci*, and for that reason some development validation topics were also evaluated: species specificity and PCR reaction conditions.

This small multiplex revealed to be human specific; with no cross contamination just as current procedures for STR analysis when strictly followed.

More important, this is a very sensible method and, although giving a PE relatively low, this 20 SNP multiplex can be very informative specially when analyzed together with routine standard STR *loci*.

5. Conclusions

SNPs are genetic *loci* that are been investigated for forensic use in recent years and SNP*for*ID multiplex [1] is a milestone of that effort. Since the validation of this multiplex [2] this proved to be very useful in paternity testing [9] and even in providing supplementary genetic information for resolution of relationship tests that show ambiguous STR results [10].

Another multiplex derived from SNP*for*ID 52-plex was also developed - Børsting 49 SNPs multiplex, with slight modifications relatively to the original [11]. This multiplex is also validated for forensic genetic testing [11] which demonstrates the importance of SNP*for*ID multiplex, and its variations, for forensic genetic testing.

Here we present the validation of a 20 SNP multiplex for use in forensic genetics, also based in SNP*for*ID multiplex. This has the objective of providing complement genetic information for assisting resolution of deficient or complex paternity and relationship testing or for individual genetic identification.

This relatively small SNP multiplex shows human specificity and full profiles can be generated from very low DNA quantities. This can be very informative, especially when analyzed together with routine STR multiplex *loci*, as presented in previous studies [4-6]. Autosomal SNP analysis can be a valuable tool in genetic analysis, especially in kinship investigation and human identification. These *loci* will probably have a bigger importance in future forensic genetics with the current development of new genotyping technologies outcome.

Role of funding

National Institute of Legal Medicine, Portugal, financially supported this work.

Conflict of interest

None.

References

1. Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27 (9):1713-1724.

2. Musgrave-Brown E, Ballard D, Balogh K, Bender K, Berger B, Bogus M, Børsting C, Brion M, Fondevila M, Harrison C, Oguzturun C, Parson W, Phillips C, Proff C, Ramos-Luis E, Sanchez JJ, Sanchez Diz P, Sobrino Rey B, Stradmann-Bellinghausen B, Thacker C, Carracedo A, Morling N, Scheithauer R, Schneider PM, Syndercombe Court D (2007) Forensic validation of the SNP*for*ID 52-plex assay. Forensic Sci Int Genet 1 (2) 186-190.

3. Schneider PM SNP*for*ID.org. SNPforID Consortium. <http://www.SNPforID.org>.

4. Dario P, Ribeiro T, Espinheira R, Geada H (2009) SNPs in paternity investigation: The simple future. Forensic Science International: Genetics Supplement Series 2 (1):127-128.

5. Dario P, Ribeiro T, Dias D, Corte-Real F, Geada H (2011) Complex casework using single nucleotide polymorphisms. Forensic Science International: Genetics Supplement Series 3 (1):e379-e380.

6. Dario P, Ribeiro T, Espinheira R, Dias D, Geada H, Corte-Real F (2011) 20 SNPs as supplementary markers in kinship testing. Forensic Science International: Genetics Supplement Series 3 (1):e508-e509.

7. SWGDAM (2004). Revised validation guidelines. Forensic Science Communications, 6(3).

8. ISO/IEC 17025:2005 (2005). General requirements for the competence of testing and calibration laboratories. Available from <http://www.iso.org>.

9. Børsting C, Sanchez JJ, Hansen HE, Hansen AJ, Bruun HQ, Morling N (2008) Performance of the SNP*for*ID 52 SNP-plex assay in paternity testing. Forensic Sci Int Genet 2 (4):292-300.

10. Phillips C, Fondevila M, Garcia-Magarinos M, Rodriguez A, Salas A, Carracedo A, Lareu MV (2008) Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. Forensic Sci Int Genet 2 (3):198-204.

11. Børsting C, Rockenbauer E, Morling N (2009) Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard. Forensic Sci Int Genet 4 (1):34-42.

Forensic Science International: Genetics Supplement Series 2 (2009) 127–128

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

Research article

SNPs in paternity investigation: The simple future

Paulo Dario [a,b], Teresa Ribeiro [a], Rosa Espinheira [a], Helena Geada [a,b,*]

[a] Genetics and Forensic Biology Department, National Institute of Legal Medicine, South Branch, 1150-219 Lisbon, Portugal
[b] Faculty of Medicine, University of Lisbon, Portugal

ARTICLE INFO

Article history:
Received 15 August 2009
Accepted 25 August 2009

Keywords:
SNPs
Paternity Investigation
SNaPshot

ABSTRACT

Based on the 52 SNP-plex developed by the SNP*for*ID Consortium, we designed two 10-plex to study single nucleotide polymorphisms (SNPs) for human identification and to establish its usefulness in paternity casework. This 20 autosomal SNP set was studied in 56 paternity investigation cases from South Portuguese resident population, also analyzed with 17 Short Tandem Repeats (STRs). Results obtained with both methodologies were consistent with each other, except for one case where the alleged father could not be excluded by SNPs. No mutation was found in the SNP *loci*, whereas a mismatch in STRs was detected. The use of SNPs as a complement to the analysis of autosomal STRs in paternity casework can result in paternity index and paternity probability values equivalent or higher than those obtained with more STR *loci*, but with lower costs. This study shows that instead of using additional STR *loci*, the analysis of 20 autosomal SNPs, as a complement technique to standard methodologies, is an appealing alternative in paternity investigation cases.

1. Introduction

Nowadays, Forensic Genetic Laboratories perform paternity investigations with one or two commercial Short Tandem Repeat (STR) multiplex kits, one kit containing most of the markers previously studied in the other one. This strategy, although giving more information about the case and assuring quality control, can be an expensive way of getting such small extra information from two or three markers. More, due to the relatively high mutation rates of these markers, the use of STR *loci* in paternity cases sometimes leads to genetic inconsistencies, which oblige a higher number of STR analyses to present a reliable conclusion.

SNPs are receiving enormous attention in the field of Forensic Genetics. This is due, in part, to their low mutation rates which make them very suitable for human identification [1], and especially, for paternity casework as demonstrated by Børsting et al. [2]. With the aim of using SNP analysis as a complement method in paternity investigation, we have chosen a set of 20 autosomal SNPs from the 52-plex developed by the SNP*for*ID Consortium [3] to establish an alternative method to the use of more STR *loci*, especially for difficult or complex paternity investigation cases, continuing previous work [4].

2. Materials and methods

Selected from the 52-plex SNP*for*ID for human identification [3], two 10-plex were designed to analyze a total of 20 SNPs using SNP*for*ID browser [5], choosing the most polymorphic *loci* in the nearby populations, namely North Portuguese and Galicia Spanish populations, populations that share great similarity with ours. Care was also taken to ensure that *loci* chosen were also polymorphic in African populations, due to the number of African origin residents in our country by historical reasons.

Using standard SNaPshot® methodology (Applied Biosystems), the 20 SNP set was investigated in 56 paternity investigation cases from different ethnic backgrounds. All cases have been studied previously with standard STR methods, using AmpFℓSTR® Identifiler® PCR Amplification Kit (Applied Biosystems) and PowerPlex® 16 System (Promega), the two kits usually used in our routine paternity casework. Forty-four non-exclusion cases were obtained, one of which with a genetic inconsistency in D2S1338. For both methodologies, samples were run in 3130/3130xl Genetic Analyzers using POP4™ polymer and analyzed with GeneMapper® ID v3.2 software (Applied Biosystems).

3. Results and discussion

Results obtained with both methodologies were consistent with each other, except for one case where the alleged father

* Corresponding author. Tel.: +351 218 811 800; fax: +351 218 864 493.
E-mail address: hgeada@dlinml.mj.pt (H. Geada).

1875-1768/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigss.2009.08.136

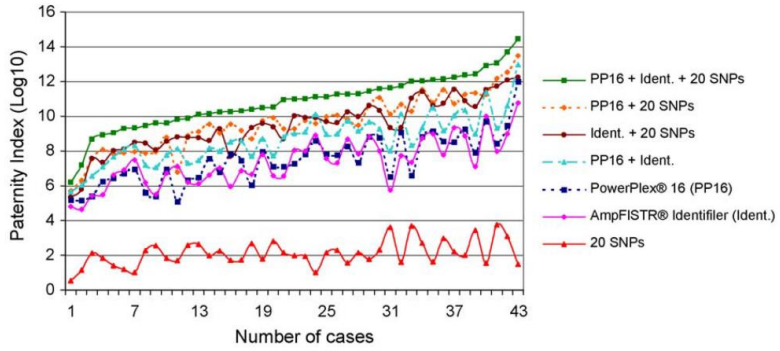


Fig. 1. Paternity index comparison of 43 non-exclusion cases obtained with different methodologies—STRs with one or both kits, SNPs and the combined use of STRs and SNPs.

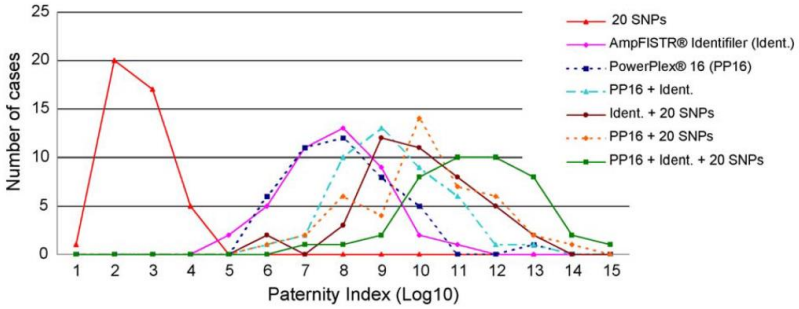


Fig. 2. Distribution of paternity indices according to different methodologies, which shows that the use of Identifiler® or PowerPlex® 16 plus 20 SNPs gave similar or better results than the use of both kits together. As expected both kits plus 20 SNPs provide the best results.

could not be excluded, probably due to the small number of SNP *loci* analyzed. No mutation was found in the SNP *loci* studied. Paternity investigation cases are usually studied with AmpFℓSTR® Identifiler® PCR Amplification Kit or/and PowerPlex® 16 System, giving paternity indices and paternity probabilities around 10^8 to 10^9 and 99,999 999(9)%, respectively when both kits are used. SNP methodology studies gave very low forensic parameters as it would be expected. However, when these 20 SNPs are analyzed together with STR Identifiler or Promega kits, paternity index and paternity probability have equivalent or higher values than those obtained with the combination of both kits. For most non-exclusion cases, the use of one commercial kit in combination with SNPs is almost always more informative than the use of both commercial kits together, as can be seen in Figs. 1 and 2. Moreover, the use of both kits and 20 SNPs altogether results in very high forensic parameters, around 10^{11} to 10^{12} and 99,999 999 999(9)%, as emphasized in Fig. 2 for paternity indices.

This study demonstrates the usefulness of only 20 SNPs for paternity investigation as a complement to standard methodology, being an alternative to the use of more STR *loci* in simple and complex cases. Furthermore, this is a simple and less expensive methodology using genetic markers that are the near future for Forensic Laboratories.

Conflict of interest

None.

Funding Source

This work was financially supported by Instituto Nacional de Medicina Legal, I.P. and Faculty of Medicine, University of Lisbon, Portugal.

References

[1] K.K. Kidd, A.J. Pakstis, W.C. Speed, E.L. Grigorenko, S.L. Kajuna, N.J. Karoma, S. Kungulilo, J.J. Kim, R.B. Lu, A. Odunsi, F. Okonofua, J. Parnas, L.O. Schulz, O.V. Zhukova, J.R. Kidd, Developing a SNP panel for forensic identification of individuals, Forensic Sci. Int. 164 (2006) 20–32.
[2] C. Børsting, J.J. Sanchez, H.E. Hansen, A.J. Hansen, H.Q. Bruun, N. Morling, Performance of the SNPforID 52 SNP-plex assay in paternity testing, Forensic Sci. Int. Genet. 2 (2008) 292–300.
[3] J.J. Sanchez, C. Phillips, C. Borsting, 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.
[4] G. Costa, P. Dario, I. Lucas, T. Ribeiro, R. Espinheira, H. Geada, Autosomal SNPs in paternity investigation, Forensic Sci. Int. Genet. Suppl. Sér. 1 (2008) 507–509.
[5] J. Amigo, C. Phillips, M. Laréu, A. Carracedo, The SNP*for*ID browser: an online tool for query and display of frequency data from the SNP*for*ID project, Int. J. Legal Med. 122 (2008) 435–440.

Forensic Science International: Genetics Supplement Series 3 (2011) e508–e509

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

20 SNPs as supplementary markers in kinship testing

Paulo Dario [a,b,c,d,*], Teresa Ribeiro [a,b], Rosa Espinheira [a,b], Deodália Dias [c,d], Helena Geada [b,e], Francisco Corte-Real [a,b,f]

[a] National Institute of Legal Medicine, Portugal
[b] CENCIFOR – Forensic Sciences Centre, Portugal
[c] CESAM – Centre for Environmental and Marine Studies, Portugal
[d] University of Lisbon, Faculty of Science, Portugal
[e] University of Lisbon, Faculty of Medicine, Portugal
[f] University of Coimbra, Faculty of Medicine, Portugal

ARTICLE INFO

Article history:
Received 15 September 2011
Accepted 30 September 2011

Keywords:
Autosomal SNPs
Kinship testing
SNaPshot®

ABSTRACT

Single Nucleotide Polymorphisms (SNPs) are having an increasingly role in Forensic Genetics due to very low SNP mutation rates and the possibility to multiplex a great number of loci. The purpose of this study was to evaluate the use of 20 autosomal SNPs as additional markers in the resolution of kinship casework where the alleged father was not available for testing and close relatives were used instead. A total of six caseworks which included alleged paternal grandparents, alleged uncles or alleged brothers were studied. All individuals studied in these cases were typed before with 17 autosomal STRs using Identifiler(Plus)® and Powerplex 16® systems. Twenty SNPs were typed using SNaPshot® methodology with two 10-plex, previously shown useful in paternity testing. LRs were calculated with "*Familias*" using South Portugal STR and SNP frequency databases. This study confirms that even as few as 20 autosomal SNP loci can be very useful in kinship analysis as a complement to standard methodologies. Moreover, SNaPshot® methodology can be easily implemented in any Forensic Laboratory.

1. Introduction

In routine casework, forensic scientists can be faced with difficult kinship studies due to lack of crucial persons for reconstruction of family trees. In these cases, all genetic markers performed in the Laboratory are used to get a better knowledge of the cases. This is even more important when genetic incompatibilities are detected and/or forensic statistics do not permit strong confidence in one of the possible kinship hypothesis.

As in normal paternity testing, kinship study is usually done using routine Short Tandem Repeats (STRs) kits, Identifiler(Plus)® and/or Powerplex 16® System, many times not being sufficient enough to obtain results to understand the familiar connections involved. For that, other forensic tools are needed as additional autosomal STRs, sexual chromosome markers or mitochondrial DNA. Nevertheless, one type of genetic markers that can be used in these situations is autosomal Single Nucleotide Polymorphisms (SNPs). These have been used for human identification since the development of SNP panels like the 52-plex SNP*for*ID [1] and its use validated [2]. Because of the characteristics of these genetic markers, especially its low mutation rates, they have been successfully used in paternity testing [3,4] and proved to be very useful in challenging cases and in complex kinship casework situations [5,6].

The aim of this work is to demonstrate that the use of merely 20 autosomal SNPs as additional genetic markers can be useful in challenging kinship casework, as previously shown for paternity testing [4].

2. Material and methods

Six kinship testing caseworks were studied, as can be seen in Fig. 1, where H1 is the claimed hypothesis, H2 is the exclusion hypothesis and in filled light blue are visualized the individuals in dispute. These six cases consisted in two situations involving putative paternal grandparents in the absence of putative father (Casework 1 and 2), two cases where the putative paternal uncle (Casework 3) or the putative maternal aunt (Casework 4) were tested in the absence of other paternal or maternal relatives, respectively, in paternity and maternity tests. Two other forensic caseworks were studied in which putative brothers were used to infer about paternity of individuals in dispute (Casework 5 and 6). For this, a total of 20 individuals were typed for 17 autosomal STRs and 20 autosomal SNPs.

STRs were typed using Identifiler(Plus)® PCR Amplification Kit (Applied Biosystems) and PowerPlex® 16 System (Promega). SNPs were typed with SNaPshot® (Applied Biosystems) with two 10-

* Corresponding author at: National Institute of Legal Medicine, Portugal. Tel.: +351 218 811 800; fax: +351 218 809 958.
E-mail addresses: pdario@dlinml.mj.pt, paulo.dario@gmail.com (P. Dario).

1875-1768/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigss.2011.09.111

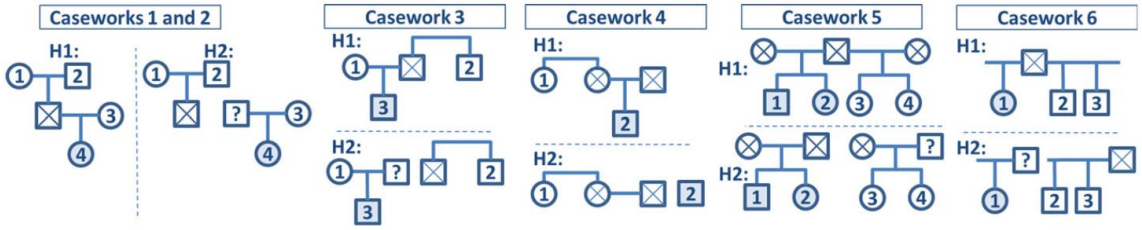


Fig. 1. Six caseworks presenting tested alternative hypothesis (studied individuals are indicated by numbers and in filled light blue are visualized the individuals in dispute). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

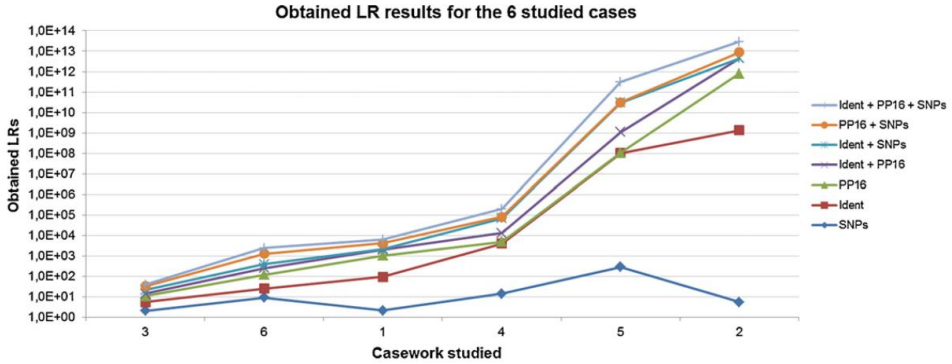


Fig. 2. Six studied caseworks (1–6) presented according to the LR values obtained using different STR and/or SNP multiplex.

plex, using previously described methodology [4] based on SNP*for*ID 52-plex [1]. Obtained data was analyzed with GeneMapper® ID v3.2.1 (Applied Biosystems).

Likelihood-ratios and posterior probabilities were calculated with *Familias* software v.1,97 using allele frequencies for the South Portuguese population and assuming 0.5 prior probabilities for tested hypothesis depicted in Fig. 1.

3. Results and discussion

Posterior probabilities (PPs) and likelihood-ratios (LRs) from these six caseworks were calculated (data not shown). LRs comparison when using the different marker systems are shown in Fig. 2, where the cases studied are presented according to the LR values obtained. The single use of SNPs in these caseworks gave weak results as expected. However, the combined use of one autosomal STR multiplex kit with the analysis of 20 SNPs presented better results than the use of both autosomal STR amplification kits – Identifiler (Ident) and Powerplex16 (PP16). The best results were obtained when SNPs were analyzed together with both kits, being this enough to increase the LR value more than 1 log, in five out of six cases, when compared with the use of both STR kits alone.

4. Conclusions

This study demonstrates the utility of merely 20 autosomal SNPs as a complement to standard methodologies, namely to STR analysis, in kinship testing, although it would be an advantage to increase the number of studied SNP loci. Nevertheless, this is a simple and inexpensive methodology that can be easily used in the resolution of this kind of complex casework.

Conflict of interest statement

None.

Funding source

The National Institute of Legal Medicine, Portugal, financially supported this work.

References

[1] J.J. Sanchez, C. Phillips, C. Borsting, 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.
[2] C. Borsting, E. Rockenbauer, N. Morling, Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard, Forensic Sci. Int. Genet. 4 (2009) 34–42.
[3] C. Borsting, J.J. Sanchez, H.E. Hansen, A.J. Hansen, H.Q. Bruun, N. Morling, Performance of the SNPforID 52 SNP-plex assay in paternity testing, Forensic Sci. Int. Genet. 2 (2008) 292–300.
[4] P. Dario, T. Ribeiro, R. Espinheira, H. Geada, SNPs in paternity investigation: the simple future, Forensic Sci. Int.: Genet. Suppl. Ser. 2 (2009) 127–128.
[5] C. Phillips, M. Fondevila, M. Garcia-Magarinos, A. Rodriguez, A. Salas, A. Carracedo, M.V. Lareu, Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers, Forensic Sci. Int. Genet. 2 (2008) 198–204.
[6] C. Borsting, N. Morling, Mutations and/or close relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers, Forensic Sci. Int. Genet. 5 (2011) 236–241.

Forensic Science International: Genetics Supplement Series 3 (2011) e379–e380

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

Complex casework using single nucleotide polymorphisms

Paulo Dario [a,b,c,d,*], Teresa Ribeiro [a,b], Deodália Dias [c,d], Francisco Corte-Real [a,b,e], Helena Geada [b,f]

[a] National Institute of Legal Medicine, Portugal
[b] CENCIFOR – Forensic Sciences Centre, Portugal
[c] CESAM – Centre for Environmental and Marine Studies, Portugal
[d] University of Lisbon, Faculty of Science, Portugal
[e] University of Coimbra, Faculty of Medicine, Portugal
[f] University of Lisbon, Faculty of Medicine, Portugal

ARTICLE INFO

Article history:
Received 26 August 2011
Accepted 15 September 2011

Keywords:
Autosomal SNPs
Complex kinship testing
SNaPshot®

ABSTRACT

Complex kinship analyses are normally a challenge for forensic laboratories, especially in cases in which the individuals involved can have criminal responsibilities. This paper presents two complex relationship caseworks studied with routine STRs and autosomal SNPs as supplementary markers. In the first case, to exclude trafficking of children, maternity investigation of a child was requested involving two alleged mothers – a 39-year-old woman, the alleged grandmother, and her absent daughter. The second one was a possible incest case with a young girl with Trisomy 21 where her father was also the alleged child's father. The individuals of these cases were typed for 17 autosomal STRs with AmpF*l*STR Identifiler or IdentifilerPlus and Powerplex 16. Twenty autosomal SNPs were also typed using SNaPshot® methodology, with two 10-plex previously revealed useful in paternity testing. Both cases gave low likelihood ratio values with STRs and a genetic incompatibility was also detected in the first case. SNP studies strongly indicated that the alleged grandmother was not the child's mother but indeed the grandmother in a real complex immigrant kinship case, while in the second casework reinforced the incest relationship. Therefore, SNPs revealed useful as additional markers in complex kinship testing.

1. Introduction

Forensic geneticists come across casework difficult to analyze due to the complexity of family relationships. Some cases also have serious criminal consequences for the individuals studied. These caseworks can involve incest leading to pregnancy of very young adolescents. Routine STRs are not sufficient to give an answer in some of these cases and therefore more loci need to be studied. The two caseworks presented here were brought to the laboratory by police forces to investigate possible criminal situations that involved complex family relationships. Their study with 17 routine Short Tandem Repeats (STRs) rose up some doubts about the conclusions obtained, so there was a need to study more genetic markers.

The first casework involved an immigrant family composed of 7 persons: the oldest one, a 39-year-old female, the alleged grandmother, one of her daughters and a son, her alleged grandson in dispute and three other very young possible grandchildren of two absent daughters. Police wanted to determine if those children belonged indeed to that woman's family and the possible relationship existing among them, to exclude a situation of trafficking of children. The second case consisted in a possible incest by the father of a young girl with Trisomy 21 that presumably abused his daughter leading to her pregnancy, which ended in voluntary abortion.

Due to the relevance of the presented situations, these two cases were also studied with 20 autosomal Single Nucleotide Polymorphisms (SNPs) [1], with a fast methodology, which have performed good results in the resolution of complex family cases [2], also involving STR mutations [3].

2. Material and methods

In both cases, a total 9 individual were typed from DNA extracted from buccal swabs with Chelex®, except for the mother of the second case and her abortion product in which the extracted DNAs were sent to the laboratory by the police. STRs were typed using AmpF*l*STR® Identifiler® (first case) or IdentifilerPlus® PCR Amplification Kit (second case) (Applied Biosystems) and Power-Plex® 16 System (Promega) in both cases. For these systems, manufacturer's instructions were modified for reduction of amplification reaction mix to half of the recommended volume.

Twenty SNPs were genotyped with SNaPshot® methodology (Applied Biosystems) with two 10-plex using previously described

* Corresponding author at: National Institute of Legal Medicine, Portugal. Tel.: +351 218 811 800; fax: +351 218 809 958.
E-mail address: pdario@dlinml.mj.pt (P. Dario).

1875-1768/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigss.2011.09.051

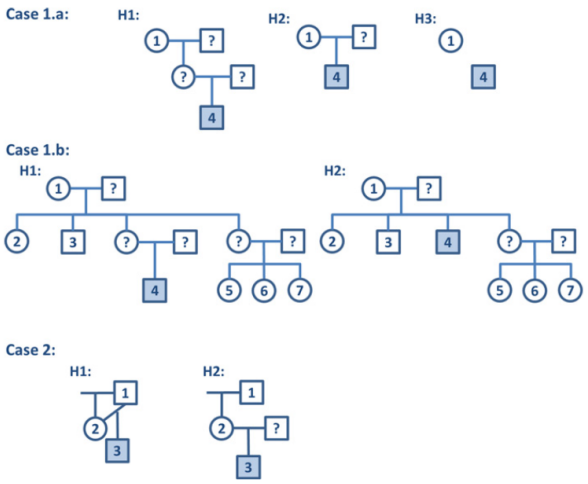


Fig. 1. Genetic pedigrees of the two studied cases evidencing tested hypothesis: Case 1.a – first step simple analyzed pedigrees; Case 1.b – final pedigrees hypothesis analyzed; Case 2 – incest case.

methodology [1] based on SNP*for*ID 52-plex [4]. Samples were run in 3130/3130xl Genetic Analyzers equipped with 36 capillary arrays,using POP4™ polymer for STRs and POP6™ polymer for SNPs. Obtained data was analyzed with GeneMapper® ID v3.2.1.

Likelihood-ratios (LR) and posterior probabilities (PP) were calculated with *Familias* software v.1,97 [5] using allele frequencies for the South Portuguese population [6,7].

3. Results and discussion

Case 1 – According to the simple pedigrees shown in Fig. 1 Case 1.a, a genetic incompatibility in the D18S51 locus was detected with 17 routine autosomal STRs, between the alleged grandmother and the tested child.

PPs and LRs of that woman being grandmother *vs* being mother *vs* having no relation with the child were calculated assuming 1/3 prior probabilities (Fig. 1 Case 1.a). Studying the case with Identifiler and Powerplex was obtained a PP of 92.34% (LR = 949) of being the grandmother. But, when 20 SNP analysis was done altogether with the 17 previously studied STRs, PP was raised to 99.97% (LR = 4292). The other possible relationships were tested in a similar way leading to a complex pedigree as shown in Figure 1Case 1.b. When the hypothesis of the women presented being the grandmother was tested against that of being the mother of the child, the final LR was of 2.11×10^{17} favoring the 1st hypothesis using all genetic markers previously referred. These immigrant situations with difficulties in communication, not only due to language barriers, but also mainly because of poor education background, make them possible targets for social discrimination.

Case 2 – In this case, two simple tested hypotheses were obtained as presented in Figure 1Case2. Posterior probabilities of 99.9999996% and 99.9999991% corresponding to LRs of 2.94×10^{8} and 1.18×10^{8} were, respectively, obtained with Identifiler or PowerPlex. When analyzed with both kits, PP raised to 99.99999997% and LRto3.91 $\times 10^{9}$. Whoever, when SNPs were analyzed with each one of these kits better results were obtained: PP 99.9999999992% and LR1.39 $\times 10^{11}$ for SNPs/Identifiler and PP 99.999999998% and LR 5.61 $\times 10^{10}$ for SNPs/Powerplex. When all systems studied were analyzed altogether the best results were obtained: PP 99.999999999994% and LR of 1.85×10^{12} – a very strong result favoring incest hypothesis, 4 logs above the original results obtained with a single STR amplification kit and three above the conjugation of both kits. These results raised credibility for the hypothesis of child abuse.

4. Conclusions

This study demonstrates the usefulness of 20 autosomal SNPs, performed with two 10-plex, in complex kinship casework as complement to standard STR methodologies, as was also performed before in paternity investigation cases [1] and kinship analysis [8]. Besides being a very fast technique, which enables a better and a quick response of the Forensic Laboratory, the SNP methodology is simple and inexpensive and can be easily used in the resolution of this kind of challenging work.

Conflict of interest statement

None.

Funding source

The National Institute of Legal Medicine, Portugal, financially supported this work.

References

[1] P. Dario, T. Ribeiro, R. Espinheira, H. Geada, SNPs in paternity investigation: the simple future, Forensic Sci. Int. Genet.: Suppl. Ser. 2 (2009) 127–128.
[2] C. Borsting, N. Morling, Mutations and/or close relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers, Forensic Sci. Int. Genet. 5 (2011) 236–241.
[3] C. Phillips, M. Fondevila, M. Garcia-Magarinos, A. Rodriguez, A. Salas, A. Carracedo, M.V. Lareu, Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers, Forensic Sci. Int. Genet. 2 (2008) 198–204.
[4] J.J. Sanchez, C. Phillips, C. Borsting, 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.
[5] T. Egeland, P.F. Mostad, B. Mevag, M. Stenersen, Beyond traditional paternity and identification cases. Selecting the most probable pedigree, Forensic Sci. Int. 110 (2000) 47–59.
[6] C. Vieira-Silva, C. Cruz, T. Ribeiro, R. Espinheira, South Portugal population genetic analysis with 17 loci STRs, Int. Congr. Ser. 1288 (2006) 367–368.
[7] R. Pereira, M. Fondevila, C. Phillips, A. Amorim, A. Carracedo, L. Gusmão, Genetic characterization of 52 autosomal SNPs in the Portuguese population, Forensic Sci. Int. Genet.: Suppl. Ser. 1 (2008) 358–360.
[8] P. Dario, T. Ribeiro, R. Espinheira, D. Dias, H. Geada, F. Corte-Real, 20 SNPs as supplementary markers in kinship testing, Forensic Sci. Int. Genet.: Suppl. Ser. 3 (2011) e508–e509.

Chapter III

Study and application of SNP*for*ID 52-plex in Portuguese population and casuistic

Paper 5.

Dario, P., Oliveira, A. R., Ribeiro, T., Porto, M. J., Dias, D. and Corte Real, F. (2017). Autosomal SNPs study of a population sample from Southern Portugal and from a sample of immigrants from Guinea-Bissau residing in Portugal. *Legal Medicine*, 24, pp. 32-35. doi:10.1016/j.legalmed.2016.11.004.

Paper 6.

Dario, P., Oliveira, A. R., Ribeiro, T., Porto, M. J., Santos, J. C., Dias, D. and Corte Real, F. (2015) SNP*for*ID 52-plex in casework samples: "Cracking" bones and other difficult samples, *Forensic Science International: Genetics Supplement Series*, 5, pp. 118–120. doi:10.1016/j.fsigss.2015.09.048.

Legal Medicine 24 (2017) 32–35

Contents lists available at ScienceDirect

Legal Medicine

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

Announcement of Population Data

Autosomal SNPs study of a population sample from Southern Portugal and from a sample of immigrants from Guinea-Bissau residing in Portugal

Paulo Dario [a,b,c,d,*], Ana Rita Oliveira [a,b,d], Teresa Ribeiro [a,c], Maria João Porto [a,c], Deodália Dias [b,d], Francisco Corte Real [a,c,e]

[a] INMLCF – National Institute of Legal Medicine and Forensic Sciences, Largo da Sé Nova, s/n, 3000-213 Coimbra, Portugal
[b] University of Lisbon, Faculty of Sciences, Campo Grande, 1749-016 Lisboa, Portugal
[c] CENCIFOR – Forensic Sciences Centre, Largo da Sé Nova, s/n, 3000-213 Coimbra, Portugal
[d] CESAM – Centre for Environmental and Marine Studies, Pólo de Lisboa, Edifício C2, Campo Grande, 1749-016 Lisboa, Portugal
[e] University of Coimbra, Faculty of Medicine, Rua Larga, 3004-504 Coimbra, Portugal

ARTICLE INFO

Article history:
Received 14 June 2016
Received in revised form 30 October 2016
Accepted 22 November 2016
Available online 23 November 2016

Keywords:
SNP
South of Portugal
Guinea-Bissau
52plex
DNA frequency estimate
Genetic distance

ABSTRACT

In recent years, autosomal single nucleotide polymorphisms (SNPs) have been comprehensively investigated in forensic research due to their usefulness in certain circumstances in complementing short tandem repeats (STRs) analysis, or even for use on their own when analysis of STRs fails. However, as with STRs, in order to properly use SNP markers in forensic casuistic we need to understand the population and forensic parameters in question. As a result of Portugal's colonial history during the time of empire, and the subsequent process of decolonization, some African individuals migrated to Portugal, giving rise to large African and African-descendent communities. One of these groups is the community originating from Guinea-Bissau, that in 2014, was enumerated to consist of more than 17,700 individuals with official residency status, more than the third major city of Guinea-Bissau.

In order to study the population and forensic parameters mentioned above for the two populations important to our casuistic, a total of 142 unrelated individuals from the South of Portugal and 90 immigrants from Guinea-Bissau (equally non related and all residing in Portugal) were typed with SNaPshot™ assay for all 52 loci included in the SNP*for*ID 52plex.

1. Population

Samples from 232 unrelated individuals involved in paternity testing casework were studied: 142 from the South of Portugal and 90 from Guinea-Bissauan immigrants residing in Portugal. The collection took place after informed consent.

2. DNA extraction and quantification

DNA was extracted from buccal swabs and/or blood samples. It was then collected and air dried prior to DNA extraction using the Chelex® 100 resin method [1]. DNA quantification was achieved using a sampling approach, in random selected samples, where DNA concentration was determined by Real-Time PCR using Quantifiler™ Human DNA Quantification Kit (Applied Biosystems® – AB) and AB 7500 Sequence Detection System (AB). This procedure was conducted in accordance with the manufacturer's instructions.

3. PCR amplification

PCR conditions to simultaneously amplify all 52 loci were arranged as originally described by Sanchez et al. [2]. PCR reactions were performed in 25 μL reaction volumes containing 1 × PCR buffer, 8 mM $MgCl_2$, 700 mM of each dNTP, 0.01–0.17 mM of each primer, 2 U AmpliTaq Gold DNA polymerase (Applied Biosystems) and 1 μL DNA in order to achieve a final concentration of around 1–10 ng in final reactions. Thermal cycling was performed in Applied Biosystems GeneAmp 9700 or Verity thermal cyclers with the following cycle programme: denaturation at 94 °C for 5 min followed by 35 cycles of 95 °C for 30 s, 60 °C for 30 s and 65 °C for 30 s, followed by 7 min at 65 °C. Excess primers and dNTPs

* Corresponding author at: INMLCF – National Institute of Legal Medicine and Forensic Sciences, Largo da Sé Nova, s/n, 3000-213 Coimbra, Portugal.
E-mail address: paulo.a.dario@inmlcf.mj.pt (P. Dario).

http://dx.doi.org/10.1016/j.legalmed.2016.11.004
1344-6223/

were removed by the addition of 1 μL USB® ExoSAP-IT® to 2.5 μL PCR product and incubation at 37 °C for 15 min, followed by 15 min at 80 °C.

4. Single base extension (SBE) reaction

SBE reactions were also conducted as described by Sanchez et al. [2] but with some slight modifications. The reactions were performed in total volumes of 8 μL containing 1 μL of purified PCR product, 4 μL of SNaPshot reaction mix (AB), 1 μL of SBE primer mix (0.01–0.27 mM) and 2 μL of Milli-Q water. The SBE primer mix was diluted in 160 mM of ammonium sulfate (Sigma-Aldrich) as pointed out by the authors. The SBE reaction was performed in Applied Biosystems GeneAmp 9700 or Verity thermal cyclers with 30 cycles of 96 °C for 10 s, 50 °C for 5 s and 60 °C for 30 s. Excess nucleotides were removed by the addition of 1 μL (1 U/μL) shrimp alkaline phosphatase to the SBE mix and incubation at 37 °C for 60 min, followed by 75 °C for 15 min.

5. SNP typing

SBE products were analyzed by capillary electrophoresis using AB 3130 Genetic Analyzer with 36 cm capillary array and POP-4 polymer. For analysis and allele calling, GeneMapper® ID 3.2.1 was used with peak thresholds set to a minimum of 100 RFUs, except for orange where 80 RFUs was used. Samples were analyzed together with reference sample 9947A and 007 results compared to the genotype established by Børsting et al. [3]. All samples were analyzed twice, and the results from the two experiments were compared. The data obtained is presented in Supplementary Data.

6. Analysis of data

Allele frequencies, observed (Ho) and expected (He) heterozygosity, matching probability (MP), power of discrimination (PD), polymorphism information content (PIC), Power of Exclusion (PE) and Typical Paternity Index (TPI) were assessed using PowerStats v1.2 [4]. Specific tests for Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) tests, F_{ST} values and locus by locus AMOVA were estimated using Arlequin v. 3.5 [5]. Population comparisons were later performed between the studied populations and the others referred to in SNP*for*ID browser [6], namely: Angola, Mozambique, Senegal (Mandenka), Somalia, Uganda, China (population resident in Germany), Denmark, France (French), Germany, Greenland, NW Spain, Slovenia, Sweden and Turkey, but also for the North of Portugal, later studied by Pontes and Pinheiro [7]. Genetic distances were graphically represented in a multidimensional scaling (MDS) using software SPSS v.22 for Windows [8] and evolutionary history was inferred in MEGA6 [9] using the Neighbor-Joining method [10].

7. Quality control

Analysis of the DNA polymorphisms was conducted following ISFG recommendations [11]. This Laboratory is ISO/IEC 17025:2005 accredited and participates regularly in the collaborative quality control and proficiency testing exercises of the Spanish and Portuguese Working Group from the International Society for Forensic Genetics (GHEP-ISFG) having participated in the 2009 and 2012 inter-laboratory collaborative exercises for autosomic SNPs promoted by the GHEP-ISFG group.

8. Results

Allele frequencies for the 52 SNPs from the Southern Portuguese and Guinea-Bissauan populations are presented in Table 1. All other studied population and forensic parameters for these two populations (Ho, He, MP, PD, PIC, PE, TPI and HWE), are presented in Supplementary Table I. The autosomal SNP A15 (rs2016276) presented the lowest PIC value for the Southern Portuguese population (0.2647), while autosomal SNP A17 (rs740910) was the lowest for the Guinea-Bissauan population resident in Portugal. The mean value for the observed heterozygosity for the 52 SNPs in the 52plex was 0.44631 ± 0.06844 for the Southern Portuguese population and 0.36267 ± 0.16503 for the Guinea-Bissauan (data not shown). After Bonferroni corrections, SNP A17 (rs740910) did not match the Hardy–Weinberg equilibrium (HWE) expectations for the Southern Portuguese population while the same occurred for A05 (rs717302), A07 (rs917118), A13 (rs1886510), A17 (rs740910) and A46 (rs1360288) for the Guinea-Bissauan population. The combined mean match probability using the studied SNPs was 3.4340×10^{-21} for the Southern Portuguese population and 4.9703×10^{-17} for the Guinea-Bissauan population, where mean exclusion probability was 99.980% and 99.941%, respectively.

In Arlequin analysis 3 groups were considered: European (consisting of Denmark, France (French), Germany, Greenland, NW Spain, Slovenia, Sweden, Turkey, Northern Portugal and Southern Portugal), African (consisting of Angola, Mozambique, Senegal (Mandenka), Somalia, Uganda and the Guinea-Bissauan immigrant population resident in Portugal) and Asian (consisting only of the Chinese population resident in Germany – the population that was later used to root the Phylogram tree). Obtained genetic distances were graphically represented in an MDS plot presented in Supplementary Fig. 1, while the inferred Phylogram using the Neighbor-Joining method is shown in Supplementary Fig. 2 (sum of branch length = 0.26371594).

9. Other remarks

SNP markers can be considered as important tools for analyzing highly degraded forensic samples, and for increasing the power of both kinship analyses and the identification of human remains where family reconstructions for missing and unidentified persons are required. They can also be particularly useful for providing additional investigative information in cases where there are no suspect(s) and/or no STR matches in the DNA databases [12].

Forensic genetics applies principles of population genetics to validate its methodologies. Samples have to be considered within the context of the population from which the sample's donor(s) belong to. In this way, it is necessary for forensic laboratories to have access to data regarding the population in question, or to study those populations themselves in order to better evaluate the forensic samples from any given case. In this study, population and forensic parameters of populations from the South of Portugal and from a Guinea-Bissauan immigrant population resident in Portugal were evaluated.

Until the middle of the 20th century some modern African nations were still Portuguese territories, namely; Cape Verde, Angola, Mozambique and Guinea-Bissau. During the decolonization process and the decades that followed it, many immigrants from these Portuguese-speaking African countries established themselves in Portugal, resulting in around 250,000 immigrants of African origin now residing in the country [13].

The results obtained (Table 1 and Supplementary Table 1) were compatible with the ones found in similar research for other populations previously studied with SNP*for*ID, namely; the Northwest region of Spain and Northern Portugal when compared to the

Table 1
Allele frequencies for a Southern Portuguese population sample (n = 142) and for a population of Guinea-Bissauan immigrants living in Portugal (n = 90).

Auto1							Auto2						
AutSNP	Allele	Frequencies S Portugal	Frequencies G Bissau	AutSNP	Allele	Frequencies S Portugal	Frequencies G Bissau	AutSNP	Allele	Frequencies S Portugal	Frequencies G Bissau	AutSNP	Allele	Frequencies S Portugal	Frequencies G Bissau
A1	C	0.60211	0.55000	A13	A	0.46071	0.46111	A24	C	0.39437	0.36667	A40	A	0.73239	0.82778
	T	0.39789	0.45000		G	0.53929	0.53889		T	0.60563	0.63333		G	0.26761	0.17222
A2	C	0.77817	0.84444	A14	A	0.60915	0.71667	A25	A	0.65845	0.72778	A41	C	0.56690	0.42222
	T	0.22183	0.15556		T	0.39085	0.28333		G	0.34155	0.27222		T	0.43310	0.57778
A3	A	0.31338	0.22778	A15	C	0.19504	0.05556	A26	G	0.35211	0.53333	A42	A	0.72535	0.82222
	T	0.68662	0.77222		T	0.80496	0.94444		T	0.64789	0.46667		G	0.27465	0.17778
A4	A	0.32394	0.58333	A16	G	0.60915	0.79444	A27	A	0.44718	0.67222	A43	A	0.55986	0.77778
	T	0.67606	0.41667		T	0.39085	0.20556		G	0.55282	0.32778		G	0.44014	0.22222
A5	A	0.40845	0.30556	A17	A	0.68214	0.96111	A28	A	0.28169	0.10000	A44	C	0.57801	0.21667
	G	0.59155	0.69444		G	0.31786	0.03889		G	0.71831	0.90000		T	0.42199	0.78333
A6	A	0.45357	0.59444	A18	G	0.35915	0.42222	A29	A	0.53521	0.38889	A45	C	0.63028	0.90556
	T	0.54643	0.40556		T	0.64085	0.57778		G	0.46479	0.61111		T	0.36972	0.09444
A7	A	0.34420	0.55147	A19	C	0.32624	0.17778	A30	A	0.56690	0.33889	A46	A	0.40141	0.04545
	G	0.65580	0.44853		T	0.67376	0.82222		C	0.43310	0.66111		G	0.59859	0.95455
A8	C	0.48944	0.55556	A20	G	0.65141	0.45000	A32	C	0.67958	0.63889	A48	C	0.44014	0.27778
	T	0.51056	0.44444		T	0.34859	0.55000		T	0.32042	0.36111		T	0.55986	0.72222
A9	C	0.76761	0.37778	A21	A	0.72535	0.21111	A33	C	0.20423	0.10556	A49	A	0.58099	0.47778
	G	0.23239	0.62222		G	0.27465	0.78889		T	0.79577	0.89444		G	0.41901	0.52222
A10	C	0.53546	0.71111	A22	A	0.26056	0.33333	A34	C	0.31338	0.19444	A50	A	0.45070	0.45000
	T	0.46454	0.28889		G	0.73944	0.66667		G	0.68662	0.80556		G	0.54930	0.55000
A11	C	0.28521	0.45000	A23	A	0.43662	0.11111	A35	A	0.61268	0.77222	A51	A	0.45070	0.55556
	T	0.71479	0.55000		G	0.56338	0.88889		G	0.38732	0.22778		G	0.54930	0.44444
A12	A	0.64789	0.71667					A36	A	0.46127	0.19444	A52	A	0.69366	0.14444
	G	0.35211	0.28333						T	0.53873	0.80556		T	0.30634	0.85556
								A37	C	0.39789	0.54444	A53	A	0.65493	0.13333
									T	0.60211	0.45556		G	0.34507	0.86667
								A38	C	0.61972	0.30556	A54	A	0.41197	0.05556
									G	0.38028	0.69444		G	0.58803	0.94444
								A39	A	0.44718	0.63889				
									T	0.55282	0.36111				

South of Portugal, and Senegal (Mandenka) when compared to the studied Guinea-Bissauan population. This fact is evidenced in Supplementary Fig. 1 and in Supplementary Fig. 2 where we can see the MDS plot and a phylogram based on F_{ST} distances between the Southern Portuguese and Guinea-Bissauan populations, and 15 others studied with SNP*for*ID.

In relation to the HWE deviations after Bonferroni correction for A17 in the Southern Portuguese sample and for A05, A07, A13, A17 and A46 in the Guinea-Bissauan, these may occur because we are not dealing with populations that perfectly comply with all the Hardy-Weinberg principles. With respect to Guinea-Bissau this is even more applicable because we are dealing with an immigrant subpopulation that may not have been totally random at the time of migration and therefore not reached HWE yet. This may be of particular relevance if we consider that certain loci such as A17 and A46 may have ancestral association and are unlikely to reach that equilibrium, probably evolving into the fixation of the most frequent allele.

In conclusion, the 52plex SNP assay is both satisfactorily robust and sufficiently sensitive enough for use in the forensic analysis of the Southern Portuguese and Guinea-Bissauan populations.

10. Access of data

Available upon request to paulo.a.dario@inmlcf.mj.pt.

Acknowledgements

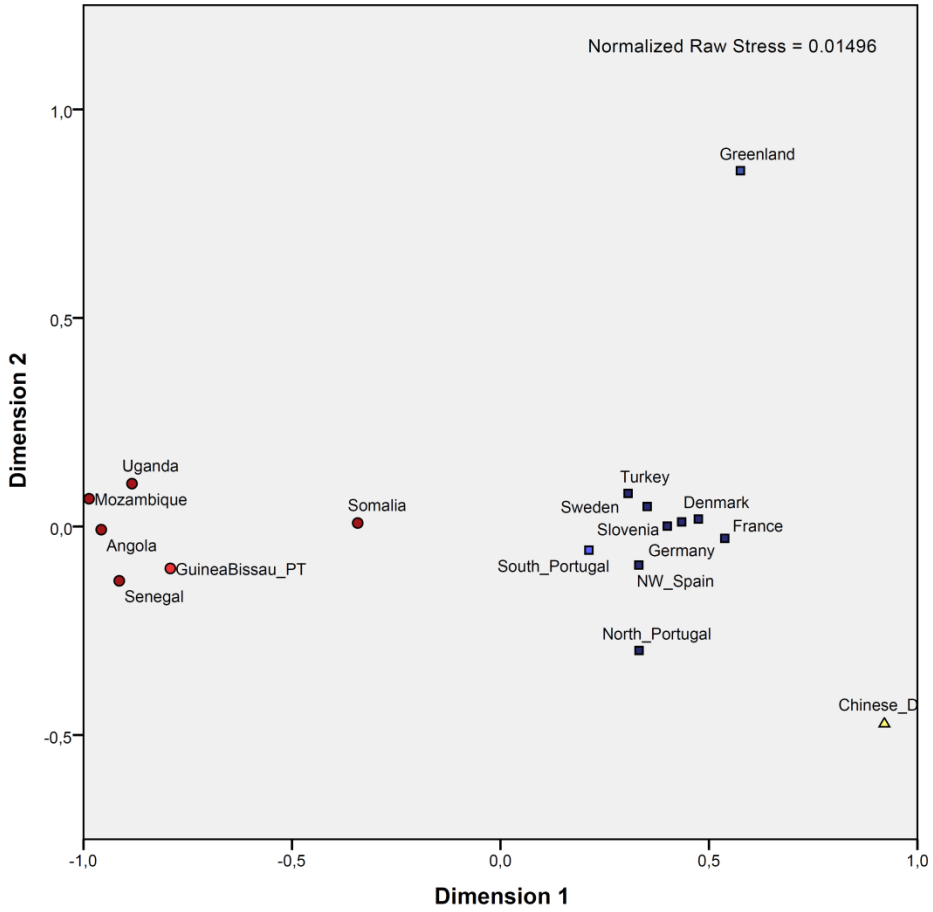
This work was supported by the National Institute of Legal Medicine and Forensic Sciences, Portugal.

Appendix A. Supplementary data

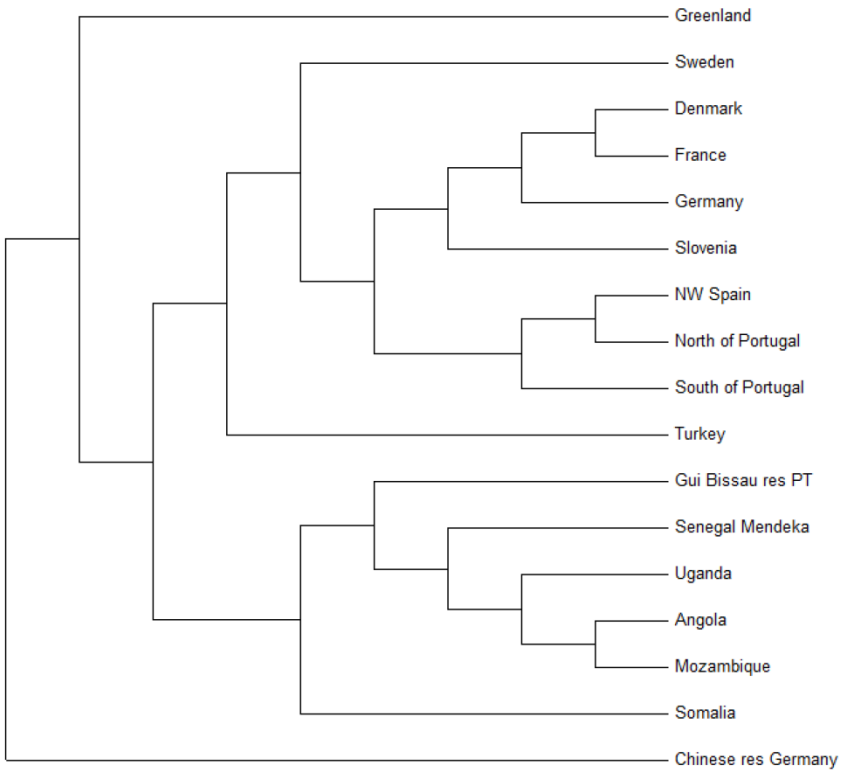
Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.legalmed.2016.11.004.

References

[1] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, Biotechniques 10 (1991) 506–513. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1867860.
[2] J.J. Sanchez, C. Phillips, C. Borsting, K. Balogh, M. Bogus, M. Fondevila, et al., A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724, http://dx.doi.org/10.1002/elps.200500671.
[3] C. Børsting, C. Tomas, N. Morling, C. Borsting, SNP typing of the reference materials SRM 2391b 1–10, K562, XY1, XX74, and 007 with the SNP*for*ID multiplex, Forensic Sci. Int. Genet. 5 (2011) e81–e82, http://dx.doi.org/10.1016/j.fsigen.2010.07.004.
[4] A. Tereba. Tools for analysis of population statistics. Profiles in DNA 2 14–16. pdf, (n.d.), 1999.
[5] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567, http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x.
[6] J. Amigo, C. Phillips, M. Lareu, A. Carracedo, The SNP*for*ID browser: an online tool for query and display of frequency data from the SNP*for*ID project, Int. J. Legal Med. 122 (2008) 435–440, http://dx.doi.org/10.1007/s00414-008-0233-7.
[7] M. Lurdes Pontes, M.F. Pinheiro, Autosomal SNPs study of a population sample from north of Portugal and a sample of immigrants from the Eastern Europe living in Portugal, Leg. Med. (Tokyo) 16 (2014) 118–120, http://dx.doi.org/10.1016/j.legalmed.2013.12.002.
[8] IBM Corp., IBM SPSS Statistics for Windows, Version 22.0, (n.d.).
[9] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729, http://dx.doi.org/10.1093/molbev/mst197.
[10] M. Nei, N. Saitou, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425. citeulike-article-id:93683.
[11] P.M. Schneider, Scientific standards for studies in forensic genetics, Forensic Sci. Int. 165 (2007) 238–243, http://dx.doi.org/10.1016/j.forsciint.2006.06.067.
[12] B. Budowle, A. van Daal, Forensically relevant SNP classes, Biotechniques 44 (2008). 603-608,610, doi:000112806 [pii] 10.2144/000112806.
[13] FFMS, População estrangeira com estatuto legal de residente: total e por algumas nacionalidades. PORDATA – Estatísticas, gráficos e indicadores de Municípios, Portugal e Europa. (2014). Retrieved May 18, 2016.



Supplementary Fig. 1 - MDS plot based on FST distances between a Southern Portuguese population, an immigrant Guinea-Bissauan population resident in Portugal and 15 other populations previously studied with SNPforID 52-plex by other authors (Angola, Mozambique, Senegal (Mandenka), Somalia, Uganda, China (population resident in Germany), Denmark, France, Germany, Greenland, NW Spain, Slovenia, Sweden, Turkey and the North of Portugal). European populations are represented in blue, African in red and Asian in yellow.



Supplementary Fig. 2 - Phylogram inferred using the Neighbor-Joining method to estimate the FST distances between the Southern Portuguese population, an immigrant Guinea-Bissauan population resident in Portugal and 15 other populations previously studied with SNPforID 52-plex by other authors (Angola, Mozambique, Senegal (Mandenka), Somalia, Uganda, China (population resident in Germany), Denmark, France, Germany, Greenland, NW Spain, Slovenia, Sweden, Turkey and the North of Portugal).

Supplementary Table 1 - Population and forensic data: Southern Portuguese population sample (n = 142).

Locus	**N**	**Ho**	**He**	**HWE**	**MP**	**PD**	**PIC**	**PE**	**TPI**
A01	142	0.47183	0.48084	0.86191	0.38296	0.61704	0.36436	0.16402	0.94667
A02	142	0.34507	0.34646	1.00000	0.48830	0.51170	0.28565	0.08382	0.76344
A03	142	0.37324	0.43187	0.11853	0.40538	0.59462	0.33775	0.09846	0.79775
A04	142	0.45070	0.43956	0.84940	0.41599	0.58401	0.34208	0.14789	0.91026
A05	142	0.40845	0.48495	0.08159	0.35856	0.64144	0.36648	0.11914	0.84524
A06	140	0.46429	0.49747	0.49691	0.36337	0.63663	0.37284	0.15812	0.93333
A07	138	0.35507	0.45310	0.01432	0.38259	0.61741	0.34955	0.08884	0.77528
A08	142	0.48592	0.50154	0.73876	0.36848	0.63152	0.37489	0.17547	0.97260
A09	142	0.30986	0.35804	0.15513	0.47739	0.52261	0.29313	0.06767	0.72449
A10	141	0.44681	0.49926	0.24089	0.35516	0.64484	0.37374	0.14504	0.90385
A11	142	0.44366	0.40917	0.40885	0.44386	0.55614	0.32461	0.14278	0.89873
A12	142	0.45070	0.45787	0.85451	0.39774	0.60226	0.35217	0.14789	0.91026
A13	140	0.46429	0.49869	0.49825	0.36214	0.63786	0.37345	0.15812	0.93333
A14	142	0.52817	0.47785	0.22104	0.41410	0.58590	0.36280	0.21336	1.05970
A15	141	0.31915	0.31511	1.00000	0.51964	0.48036	0.26470	0.07172	0.73438
A16	142	0.48592	0.47785	0.86230	0.39208	0.60792	0.36280	0.17547	0.97260
A17	140	0.30714	0.43520	0.00078	0.40071	0.59929	0.33962	0.06652	0.72165
A18	142	0.50704	0.46195	0.27638	0.41827	0.58173	0.35438	0.19374	1.01429
A19	141	0.41135	0.44118	0.44699	0.40285	0.59715	0.34298	0.12097	0.84940
A20	142	0.47183	0.45576	0.71400	0.40795	0.59205	0.35102	0.16402	0.94667
A21	142	0.50704	0.39984	0.00128	0.48016	0.51984	0.31906	0.19374	1.01429
A22	142	0.33803	0.38670	0.18683	0.44803	0.55197	0.31110	0.08041	0.75532
A23	142	0.54930	0.49370	0.23508	0.41133	0.58867	0.37095	0.23439	1.10938
A24	142	0.49296	0.47937	0.86029	0.39387	0.60613	0.36359	0.18141	0.98611
A25	142	0.45775	0.45138	1.00000	0.40676	0.59324	0.34863	0.15313	0.92208
A26	142	0.42254	0.45787	0.36173	0.38901	0.61099	0.35217	0.12822	0.86585
A27	142	0.50000	0.49617	1.00000	0.38058	0.61942	0.37219	0.18750	1.00000
A28	142	0.42254	0.40611	0.68266	0.44059	0.55941	0.32280	0.12822	0.86585
A29	142	0.47887	0.49928	0.73782	0.36759	0.63241	0.37376	0.16967	0.95946
A30	142	0.44366	0.49278	0.30467	0.36054	0.63946	0.37048	0.14278	0.89873
A32	142	0.47183	0.43704	0.43738	0.42660	0.57340	0.34067	0.16402	0.94667
A33	142	0.28169	0.32618	0.12156	0.51230	0.48770	0.27221	0.05628	0.69608
A34	142	0.42958	0.43187	1.00000	0.41688	0.58312	0.33775	0.13295	0.87654
A35	142	0.39437	0.47629	0.04820	0.36431	0.63569	0.36198	0.11053	0.82558
A36	142	0.45775	0.49876	0.39854	0.35955	0.64045	0.37350	0.15313	0.92208
A37	142	0.47183	0.48084	0.86222	0.38296	0.61704	0.36436	0.16402	0.94667
A38	142	0.42254	0.47300	0.21942	0.37393	0.62607	0.36026	0.12822	0.86585
A39	142	0.48592	0.49617	0.86620	0.37383	0.62617	0.37219	0.17547	0.97260
A40	142	0.35211	0.39337	0.28275	0.44188	0.55812	0.31516	0.08733	0.77174
A41	142	0.52817	0.49278	0.39808	0.39923	0.60077	0.37048	0.21336	1.05970
A42	142	0.38028	0.39984	0.67293	0.43821	0.56179	0.31906	0.10237	0.80682
A43	142	0.54225	0.49458	0.30663	0.40597	0.59403	0.37139	0.22722	1.09231
A44	141	0.48936	0.48956	1.00000	0.38202	0.61798	0.36884	0.17836	0.97917
A45	142	0.45775	0.46770	0.85715	0.39050	0.60950	0.35745	0.15313	0.92208
A46	142	0.49296	0.48226	0.86118	0.39099	0.60901	0.36509	0.18141	0.98611
A48	142	0.51408	0.49458	0.73261	0.38951	0.61049	0.37139	0.20012	1.02899
A49	142	0.57042	0.48860	0.05567	0.43077	0.56923	0.36836	0.25688	1.16393
A50	142	0.52113	0.49689	0.61307	0.39109	0.60891	0.37256	0.20666	1.04412
A51	142	0.56338	0.49689	0.12914	0.41758	0.58242	0.37256	0.24922	1.14516
A52	142	0.40141	0.42649	0.55523	0.41529	0.58471	0.33468	0.11478	0.83529
A53	142	0.45070	0.45359	1.00000	0.40200	0.59800	0.34984	0.14789	0.91026
A54	142	0.41549	0.48621	0.08593	0.35896	0.64104	0.36713	0.12362	0.85542

N - number of alleles, Ho - observed heterozigosity, He - expected heterozigosity, HWE - Hardy–Weinberg equilibrium (*p*-value), MP - matching probability, PD - power of discrimination, PIC - polymorphism information content, PE - power of exclusion, TPI - typical paternity index.

Supplementary Table 1 (cont.) - Population and forensic data: Guinea-Bissauan immigrants living in Portugal (n = 90).

Locus	**N**	**Ho**	**He**	**HWE**	**MP**	**PD**	**PIC**	**PE**	**TPI**
A01	90	0.47778	0.49777	0.83096	0.36963	0.63037	0.37249	0.16879	0.95745
A02	90	0.28889	0.26418	0.68538	0.57358	0.42642	0.22821	0.05907	0.70313
A03	90	0.38889	0.35376	0.54581	0.48617	0.51383	0.28991	0.10731	0.81818
A04	90	0.43333	0.48883	0.28721	0.36222	0.63778	0.36796	0.13552	0.88235
A05	90	0.21111	0.42675	0.00000	0.43136	0.56864	0.33433	0.03286	0.63380
A06	90	0.45556	0.48485	0.66181	0.37358	0.62642	0.36592	0.15148	0.91837
A07	68	0.10294	0.49837	0.00000	0.41825	0.58175	0.37234	0.00884	0.55738
A08	90	0.51111	0.49659	0.83226	0.38691	0.61309	0.37189	0.19741	1.02273
A09	90	0.62222	0.47275	0.00340	0.48840	0.51160	0.35962	0.31840	1.32353
A10	90	0.35556	0.41316	0.20244	0.42321	0.57679	0.32646	0.08908	0.77586
A11	90	0.41111	0.49777	0.13286	0.34741	0.65259	0.37249	0.12082	0.84906
A12	90	0.43333	0.40838	0.61312	0.44222	0.55778	0.32365	0.13552	0.88235
A13	90	0.87778	0.49975	0.00000	0.78099	0.21901	0.37348	0.75029	4.09091
A14	90	0.41111	0.40838	1.00000	0.43630	0.56370	0.32365	0.12082	0.84906
A15	90	0.11111	0.10552	1.00000	0.80247	0.19753	0.09943	0.01018	0.56250
A16	90	0.32222	0.32843	1.00000	0.50691	0.49309	0.27327	0.07309	0.73770
A17	90	0.01111	0.07517	0.00001	0.91432	0.08568	0.07196	0.00012	0.50562
A18	90	0.51111	0.49063	0.82832	0.39284	0.60716	0.36888	0.19741	1.02273
A19	90	0.26667	0.29398	0.46606	0.54765	0.45235	0.24961	0.05072	0.68182
A20	90	0.47778	0.49777	0.83157	0.36963	0.63037	0.37249	0.16879	0.95745
A21	90	0.33333	0.33495	1.00000	0.50025	0.49975	0.27761	0.07819	0.75000
A22	90	0.40000	0.44693	0.34952	0.39556	0.60444	0.34568	0.11392	0.83333
A23	90	0.22222	0.19863	0.59090	0.65432	0.34568	0.17802	0.03611	0.64286
A24	90	0.42222	0.46704	0.37279	0.38074	0.61926	0.35659	0.12802	0.86538
A25	90	0.38889	0.39845	0.79664	0.44173	0.55827	0.31773	0.10731	0.81818
A26	90	0.44444	0.50056	0.29707	0.35407	0.64593	0.37389	0.14334	0.90000
A27	90	0.47778	0.44314	0.48299	0.42395	0.57605	0.34358	0.16879	0.95745
A28	90	0.20000	0.18101	0.59202	0.68000	0.32000	0.16380	0.02976	0.62500
A29	90	0.42222	0.47796	0.27451	0.36988	0.63012	0.36235	0.12802	0.86538
A30	90	0.45556	0.45059	1.00000	0.40765	0.59235	0.34770	0.15148	0.91837
A32	90	0.50000	0.46400	0.49988	0.41358	0.58642	0.35497	0.18750	1.00000
A33	90	0.14444	0.18988	0.05174	0.69802	0.30198	0.17100	0.01645	0.58442
A34	90	0.34444	0.31502	0.50695	0.52025	0.47975	0.26420	0.08352	0.76271
A35	90	0.30000	0.35376	0.22596	0.48321	0.51679	0.28991	0.06354	0.71429
A36	90	0.30000	0.31502	0.73499	0.52173	0.47827	0.26420	0.06354	0.71429
A37	90	0.55556	0.49882	0.29685	0.41136	0.58864	0.37302	0.24090	1.12500
A38	90	0.38889	0.42675	0.45837	0.41358	0.58642	0.33433	0.10731	0.81818
A39	90	0.50000	0.46400	0.49989	0.41358	0.58642	0.35497	0.18750	1.00000
A40	90	0.23333	0.28672	0.12546	0.56321	0.43679	0.24448	0.03951	0.65217
A41	90	0.60000	0.49063	0.05170	0.45210	0.54790	0.36888	0.29088	1.25000
A42	90	0.22222	0.29398	0.02941	0.55951	0.44049	0.24961	0.03611	0.64286
A43	90	0.33333	0.34761	0.76102	0.48765	0.51235	0.28593	0.07819	0.75000
A44	90	0.30000	0.34134	0.34463	0.49556	0.50444	0.28183	0.06354	0.71429
A45	90	0.16667	0.17200	0.56719	0.70395	0.29605	0.15642	0.02135	0.60000
A46	88	0.02273	0.08727	0.00009	0.89127	0.10873	0.08301	0.00049	0.51163
A48	90	0.42222	0.40348	0.79409	0.44395	0.55605	0.32074	0.12802	0.86538
A49	90	0.51111	0.50180	1.00000	0.38173	0.61827	0.37451	0.19741	1.02273
A50	90	0.52222	0.49777	0.67429	0.39185	0.60815	0.37249	0.20770	1.04651
A51	90	0.53333	0.49659	0.52401	0.39951	0.60049	0.37189	0.21837	1.07143
A52	90	0.24444	0.24854	1.00000	0.59802	0.40198	0.21662	0.04308	0.66176
A53	90	0.15556	0.23240	0.00694	0.64963	0.35037	0.20440	0.01883	0.59211
A54	90	0.11111	0.10552	1.00000	0.80247	0.19753	0.09943	0.01018	0.56250

N - number of alleles, Ho - observed heterozigosity, He - expected heterozigosity, HWE - Hardy–Weinberg equilibrium (*p*-value), MP - matching probability, PD - power of discrimination, PIC - polymorphism information content, PE - power of exclusion, TPI - typical paternity index.

Forensic Science International: Genetics Supplement Series 5 (2015) e118–e120

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

SNP*for*ID 52-plex in casework samples: "Cracking" bones and other difficult samples

P. Dario[a,*], A.R. Oliveira[b], T. Ribeiro[a], M.J. Porto[a], J. Costa Santos[c], D. Dias[b], F. Corte Real[d]

[a] *INMLCF—National Institute of Legal Medicine and Forensic Sciences, Portugal*
[b] *University of Lisbon, Faculty of Science, Lisbon, Portugal*
[c] *University of Lisbon, Faculty of Medicine, Lisbon, Portugal*
[d] *University of Coimbra, Faculty of Medicine, Coimbra, Portugal*

ARTICLE INFO

Article history:
Received 27 August 2015
Accepted 14 September 2015
Available online 16 September 2015

Keywords:
SNP*for*ID 52-plex
Degraded samples
STR incomplete profiles

ABSTRACT

Casework samples can present difficulties to forensic scientists in criminal and identification investigations. Some challenging samples like bones, teeth and crime scene samples, often contain low DNA quantity which can even be degraded. In these cases, obtained STR profiles are many times incomplete or even null. This is partly due to relative bigger size of commonly used STRs in forensic analysis. In order to bypass this problem, other strategies of analysis have been developed in the past based on mini-STRs and biallelic markers, such as Indels and SNPs. Although each marker type has its advantages and disadvantages, SNPs benefit from the fact of having smaller amplification products and its analysis can be realized analyzing simultaneously a great number of loci using large multiplexes. One of such multiplex is SNP*for*ID 52-plex which analyzes 52 loci, providing good results, as reported by some authors. Taking this in consideration, we compared the amplification success of 53 real casework samples from our casuistic consisting of bones, teeth and other samples using the 52-plex and the Identifier® Plus kit. Mean amplification success rate by loci was of 73% and 43% respectively and 16 out of 36 samples in which STR profiles were not obtained or in which these were poor, generated complete or almost complete SNP profiles. We conclude that the 52-plex can be a valuable tool in the analysis of different types of challenging forensic samples when STRs fail to provide necessary genetic information for identification.

1. Introduction

In routine casework, forensic scientists have to deal with different kinds of samples where type, availability, and state of its preservation can make difficult to obtain a genetic profile [1]. Samples such as bones, teeth and crime scene samples often present challenging situations due to their DNA quantities and their degradation state, usually when it is no longer possible to obtain more sample for further studies. In these cases, obtained STR profiles can be incomplete or even null due to amplicon sizes of commonly used STRs in forensic routine. In order to overcome this problem, other genetic markers have successfully started to being used: mini-STRs and biallelic markers, such as Indels and SNPs [2]. Advantages of SNP typing in forensic genetics are well known and includes the use of smaller PCR products and the possibility of multiplexing many loci in a single PCR reaction. Sanchez et al. developed the SNP*for*ID 52-plex assay [3], that allows simultaneously amplification of 52 autosomal SNPs and which was validated for forensic use [4]. This multiplex was reported as having an higher performance when compared to STRs and even to Indels [2], resulting in full profiles from DNA extracts that yielded no or few STR loci during challenging criminal samples analysis [5].

The aim of this work was to study suitability of the SNP*for*ID 52-plex to complement standard STRs methodologies in bone and teeth analysis but also in other challenging samples related to criminal and individual identification. Taking this in consideration we compared amplification success of 53 real casework samples from our casuistic consisting of bones, teeth and other challenging samples using the 52-plex and the Identifier® Plus® kit.

2. Material and methods

A total of 53 samples were investigated comprising 24 bones, 16 teeth and 13 diverse crime scene and identification samples. Different extraction methods were used: Phenol-Chloroform,

* Corresponding author at: INMLCF, I.P. (Del. Sul). Rua Manuel Bento de Sousa, 3. 1150-219—Lisboa, Portugal. Fax: +351 218 864 493.
E-mail address: pdario@dlinml.mj.pt (P. Dario).

http://dx.doi.org/10.1016/j.fsigss.2015.09.048
1875-1768/

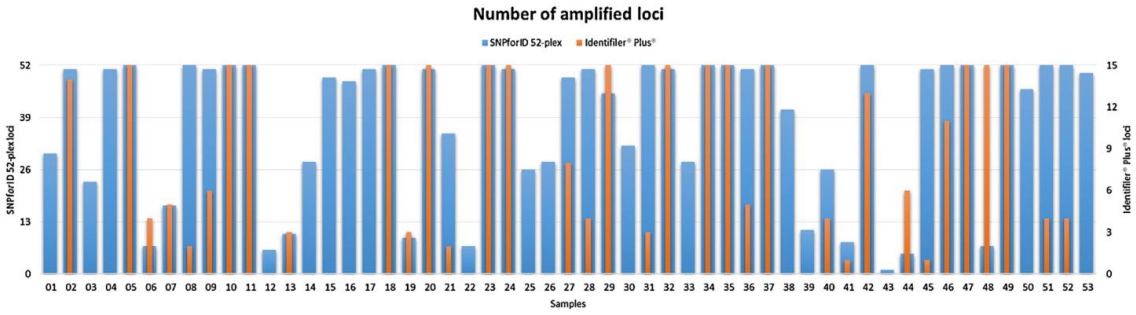


Fig. 1. Number of amplified loci for each one of tested samples, with 52-plex and Identifiler® Plus®.

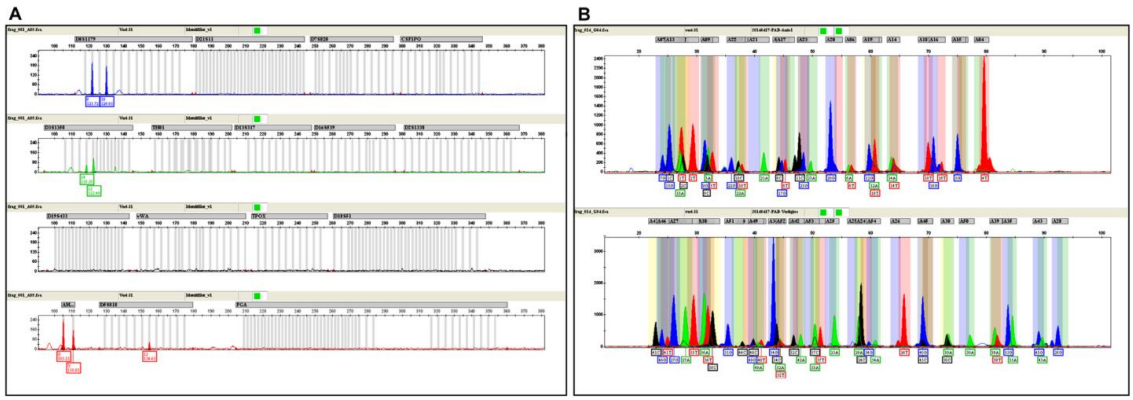


Fig. 2. Bone sample analyzed with Identifiler® Plus® (A) and with 52-plex (B) [sample 31, concentration of 0.02 ng/μl].

Machery Nagel Nucleospin, Qiagen Mini, Micro and Investigator kits depending on type, availability and quality of samples presented for analysis. Same extract was used both for STR and SNP amplification. DNA quantification was performed using Quantifiler Duo DNA Quantification kit in an Applied Biosystems® (AB) 7500 Real Time PCR System. STR typing was done using Identifier® Plus® (AB) and SNPs were typed with SNaPshot® Multiplex kit (AB) using SNP*for*ID 52-plex [3]. All amplifications were performed in AB GeneAmp® PCR System 9700 (gold block) and capillary electrophoresis was accomplished in AB 3130/3130xl Genetic Analyzers. Obtained data was analyzed with GeneMapper® ID v3.2.1 (AB).

3. Results and discussion

Mean amplification success rate of the 52-plex and the Identifiler® Plus® kit by loci was of 73% and 43% respectively, and 16 out of 36 samples in which STR profiles were not obtained or in which these were poor, generated complete or almost complete SNP profiles, as can be visualized in Fig. 1 (e.g. samples 04, 08 and 09).

An example of these samples is presented in Fig. 2, where a poor profile was obtained with Identifiler® Plus® but a full profile was obtained with 52-plex.

This superior amplification success of the 52-plex relatively to Identifiler® Plus® kit was observed in most bone samples (e.g. 08 and 09), teeth samples (e.g. 04 and 15) and also in crime scene samples (52 and 53). There were two termination of pregnancy samples tested which presented opposite results – 48 and 50 – what makes believe these must be the result from preservation methods used. Nevertheless, better results were obtained with 52-plex where full profiles were obtained from DNA extracts that generated no or few STR loci.

4. Conclusion

This study demonstrates utility of SNP analysis as a complement or as an alternative to STR typing in challenging samples analysis such as bones, teeth and other problematic samples where STR amplification fails.

Conflict of interest

None.

Acknowledgment

The authors thank INMLCF for financing this work.

References

[1] S.R. Hughes-Stamm, K.J. Ashton, A. Van Daal, Assessment of DNA degradation and the genotyping success of highly degraded samples, Int. J. Legal Med. 125 (2011) 341–348, doi:http://dx.doi.org/10.1007/s00414-010-0455-3.

[2] C. Romanini, M.L. Catelli, A. Borosky, R. Pereira, M. Romero, M. Salado Puerto, et al., Typing short amplicon binary polymorphisms: supplementary SNP and Indel genetic information in the analysis of highly degraded skeletal remains, Forensic Sci. Int. Genet. 6 (2012) 469–476, doi:http://dx.doi.org/10.1016/j.fsigen.2011.10.006.

[3] J.J.J. Sanchez, C. Phillips, C. Børsting, M. Bogus, A. Carracedo, D. Syndercombe-Court, et al., Development of a multiplex PCR assay detecting 52 autosomal SNPs, Int. Congr. Ser. 1288 (2006) 67–69, doi:http://dx.doi.org/10.1016/j.ics.2005.10.039.

[4] E. Musgrave-Brown, D. Ballard, K. Balogh, K. Bender, B. Berger, M. Bogus, et al., Forensic validation of the SNPforID 52-plex assay, Forensic Sci. Int. Genet. 1 (2007) 186–190 doi: S1872-4973(07)00041-5 [pii] 10.1016/j.fsigen.2007.01.004..

[5] S. Alimat, S. Hadi, W. Goodwin, SNP genotyping of forensic casework samples using the 52 SNPforID markers, Forensic Sci. Int. Genet. Suppl. Ser. 4 (2013) e178–e179, doi:http://dx.doi.org/10.1016/j.fsigss.2013.10.092.

Chapter IV

Lineage Informative SNP multiplex in casuistic

Paper 7.

Dario, P., Bom, J., Ribeiro, T. and Geada, H. (2009) MtSNP typing before mtDNA sequencing: Why do it? ***Forensic Science International: Genetics Supplement Series*****, 2(1), pp. 187–188. doi:10.1016/j.fsigss.2009.08.137.**

Forensic Science International: Genetics Supplement Series 2 (2009) 187–188

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

Research article

MtSNP typing before mtDNA sequencing: Why do it?

Paulo Dario [a,b], Joana Bom [b], Teresa Ribeiro [a], Helena Geada [a,b,*]

[a] Genetics and Forensic Biology Department, National Institute of Legal Medicine, South Branch, Lisbon, Portugal
[b] Faculty of Medicine, University of Lisbon, Portugal

ARTICLE INFO

Article history:
Received 15 August 2009
Accepted 25 August 2009

Keywords:
MtSNPs
SNaPshot
Forensic casework

ABSTRACT

Analysis of control mitochondrial DNA (mtDNA) hypervariable regions is sometimes the only available method to study hair evidence in forensic casework although being a laborious technique. Nowadays there is a huge interest in new genetic markers such as single nucleotide polymorphisms (SNPs) to type degraded forensic samples. For that purpose, a 10-Plex mitochondrial SNP for haplogroup typing, chosen from several SNP studies and useful to study the most common populations in our laboratory was applied in forensic casework. Hair shafts from three forensic cases with different ethnic backgrounds were studied with mtDNA sequencing and compared with mitochondrial SNPs (mtSNPs) study. Coding mtSNP typing prior to sequencing can allow for a rapid screening in forensic casework, which is emphasized in the first two cases. Moreover, in cases in which mtDNA sequencing fails, mtSNPs can still be detected. This 10 SNP *loci* multiplex provides a less expensive and simpler method for mitochondrial typing compared to control region mtDNA sequencing, especially when used as a fast screening method.

1. Introduction

In recent years, interest in autosomal, Y chromosome and mitochondrial single nucleotide polymorphisms (SNPs) had increased in forensic area. Mitochondrial DNA (mtDNA) is useful for identity testing and, above all, for analysis of samples with few or degraded DNA, such as skeletal remains or hair shafts. Analysis of mtDNA HVI and HVII hypervariable regions is sometimes the only available method in forensic casework but provide limited power of discrimination besides being a laborious technique. In many forensic cases, the only evidence found in crime-scenes consists on hairs, usually in telogenic phase. Due to difficulties in obtaining autosomic DNA in these samples, forensic laboratories still exploit mtDNA analysis to give some information to police forces.

Nowadays, there is a huge interest in the study of mitochondrial SNPs (mtSNPs) for forensic purposes. We developed a 10 SNP-plex assay for mitochondrial DNA haplogroup typing of the two most common populations in our casework—European and African ancestry populations. This set of coding region mtSNPs were chosen from different SNP studies [1–3]. The aim of this study is to demonstrate the advantage of typing coding region mtSNPs before mtDNA sequencing.

2. Materials and methods

DNA was extracted using QIAmp DNA Mini Kit (QIAGEN) in reference samples and QIAmp DNA Micro Kit (QIAGEN) in evidence ones. MtDNA sequences from reference samples were studied and compared with sequenced hair shafts in three forensic cases with different ethnic backgrounds. HVI region was amplified using 15997L and 017H primers, and all sequences have been sequenced in more than 500 bp, which have increased the discrimination power of HVI mtDNA methodology [4,5]. BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) was used for DNA sequencing.

Ten mtSNP *loci*—G1719A, C3594T, T4216C, G4580A, C7028T, G8251A, A10398G, C10400T, C12705T and A12308G, chosen from different SNP works [1–3] were studied with single base extension using SNaPshot® methodology (Applied Biosystems). Selected SNPs were combined in one multiplex reaction defining the most common European and African haplogroups—H, U, J, T, I, K, X, W, M, N, V, HV, L1/L2 and L3.

Samples from both methodologies were run in a 3130 Genetic Analyzer, using POP7™ polymer for MtDNA sequencing and POP4™ polymer for mtSNPs. Results were analysed, respectively, with SeqScape® Software v2.5 and GeneMapper® ID v3.5 software (Applied Biosystems).

3. Results and discussion

There is an increasing tendency to use mtDNA analysis in criminal investigations, particularly when the only evidence found

* Corresponding author. Tel.: +351 218 811 800; fax: +351 218 864 493.
E-mail address: hgeada@dlinml.mj.pt (H. Geada).

1875-1768/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigss.2009.08.137

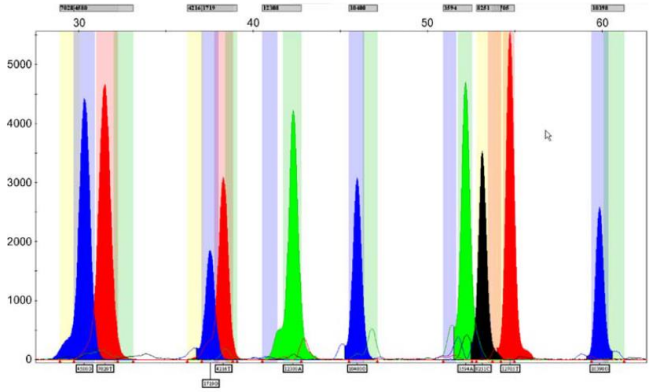


Fig. 1. Electropherogram of 10 mtSNP coding region *loci* showing L1/L2 haplogroup characterized by 4580G, 7028T, 1719G, 4216T, 12380A, 10400G, 3594A, 8251C, 12705T and 10398G.

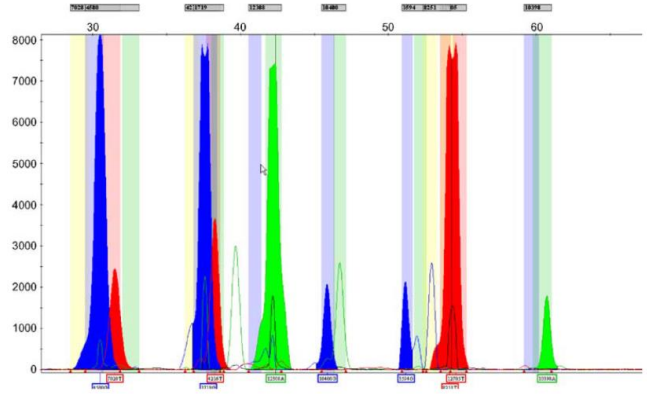


Fig. 2. MtSNPs obtained from a degraded hair sample in one case where mtDNA sequencing failed—haplogroup W was characterized by 4580G, 7028T, 1719G, 4216T, 12380A, 10400G, 3594G, 8251T, 12705T, 10398A. Some extra-peaks were also detected in this sample.

consists on hair shafts or telogenic hairs, many often subjected to harsh environments, being very difficult to recover DNA, except mtDNA.

In the first case of three criminal investigation cases, the victiḿs sample was typed as L1/L2 haplogroup (Fig. 1) and one hair found in the body as H* haplogroup. In thirty-eight reference samples, more than twenty different mitochondrial haplotypes were found, although the forensic sample did not match any of these haplotypes. According to the different ethnic background of the suspects, a rapid screening with mtSNPs subdivided reference samples into 11 haplogroups, being H*, U, T, X, J, L3 and N, haplogroups with higher frequencies. H*, the evidence sample haplogroup, was detected in 42% of all samples, showing that 58% reference samples (22 samples) were unnecessarily sequenced. More, sub-typing haplogroup H*, with a second multiplex (data not shown), the number of samples to perform DNA sequencing were drastically reduced.

In case 2, mtSNPs from two hairs have been studied. Haplogroup U characterized by 7028T, 12308G and haplogroup H* characterized by 7028C were obtained. These samples were easily excluded from belonging to the victim (haplogroup L1/L2–3594T, 10308G, 7028T). Case 3 emphasizes SNPs usefulness when mtDNA sequencing fails as in degraded hair samples (Fig. 2).

Prior to mtDNA sequencing, haplogroup characterization with mtSNPs is a very useful technique for fast screening of samples, reducing costs and time used in casework analysis. Furthermore, this is a simple and less expensive methodology using genetic markers that are the near future for most Forensic Laboratories.

Conflict of interest

None.

Funding source

This work was financially supported by Instituto Nacional de Medicina Legal, I.P. and Faculty of Medicine, University of Lisbon, Portugal.

References

[1] A. Brandstatter, T.J. Parson, W. Parson, Rapid screening of mtDNA coding region SNPs for the identification of West European Caucasian haplogroups, Int. J. Legal Med. 117 (2003) 291–298.
[2] B. Quintáns, V. Álvarez-Iglesias, A. Salas, C. Phillips, M.V. Lareu, A. Carracedo, Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SnaPshot minisequencing, Forensic Sci. Int. 140 (2004) 251–257.
[3] T.M. Nelson, R.S. Just, O. Loreille, M.S. Schanfield, D. Podini, Development of a multiplex single base extension assay for mitochondrial DNA haplogroup typing, Croat. Med. J. 48 (2007) 460–472.
[4] W. Parson, H.J. Bandelt, Extended guidelines for mtDNA typing of population data in forensic science, Forensic Sci. Int. Genet. 1 (2007) 13–19.
[5] V. Álvarez-Iglesias, J.C. Jaime, A. Carracedo, A. Salas, Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups, Forensic Sci. Int. Genet. 1 (2007) 44–55.

Chapter V

Development and application of a Phenotype Informative SNPs multiplex in Portuguese population

Paper 8.

Dario, P., Mouriño, H., Oliveira, A. R., Lucas, I., Ribeiro, T., Porto, M. J., Costa Santos, J., Dias, D. and Corte Real, F. (2015). Assessment of IrisPlex-based multiplex for eye and skin color prediction with application to a Portuguese population. *International Journal of Legal Medicine*, 129(6), pp. 1191–1200. doi:10.1007/s00414-015-1248-5

Paper 9.

Dario, P., Oliveira, A. R., Marques, M., Ribeiro, T., Porto, M. J., Dias, D., Costa Santos, J. and Corte Real, F. (in preparation). Forensic phenotyping of a mummified body – genetics contribution to a glimpse from the past. *To be submitted.*

Int J Legal Med (2015) 129:1191–1200
DOI 10.1007/s00414-015-1248-5

ORIGINAL ARTICLE

Assessment of IrisPlex-based multiplex for eye and skin color prediction with application to a Portuguese population

Paulo Dario[1,2,3,4] · **Helena Mouriño**[2] · **Ana Rita Oliveira**[2,4] · **Isabel Lucas**[1] · **Teresa Ribeiro**[1,3] · **Maria João Porto**[1,3] · **Jorge Costa Santos**[1,3,5] · **Deodália Dias**[2,4] · **Francisco Corte Real**[1,3,6]

Received: 18 March 2015 / Accepted: 12 August 2015 / Published online: 20 August 2015

Abstract DNA phenotyping research is one of the most emergent areas of forensic genetics. Predictions of externally visible characteristics are possible through analysis of single nucleotide polymorphisms. These tools can provide police with "intelligence" in cases where there are no obvious suspects and unknown biological samples found at the crime scene do not result in any criminal DNA database hits. IrisPlex, an eye color prediction assay, revealed high prediction rates for blue and brown eye color in European populations. However, this is less predictive in some non-European populations, probably due to admixing. When compared to other European countries, Portugal has a relatively admixed population, resulting from a genetic influx derived from its proximity to and historical relations with numerous African territories. The aim of this work was to evaluate the utility of IrisPlex in the Portuguese population. Furthermore, the possibility of supplementing this multiplex with additional markers to also achieve skin color prediction within this population was evaluated. For that, IrisPlex was augmented with additional SNP loci. Eye and skin color prediction was estimated using the multinomial logistic regression and binomial logistic regression models, respectively. The results demonstrated eye color prediction accuracies of the IrisPlex system of 90 and 60 % for brown and blue eye color, respectively, and 77 % for intermediate eye color, after allele frequency adjustment. With regard to skin color, it was possible to achieve a prediction accuracy of 93 %. In the future, phenotypic determination multiplexes must include additional loci to permit skin color prediction as presented in this study as this can be an advantageous tool for forensic investigation.

Keywords Eye color prediction · Skin color prediction · EVC · Forensic DNA phenotyping · IrisPlex · Multinomial logistic regression

Electronic supplementary material The online version of this article (doi:10.1007/s00414-015-1248-5) contains supplementary material, which is available to authorized users.

✉ Paulo Dario
pdario@dlinml.mj.pt

1 INMLCF - National Institute of Legal Medicine and Forensic Sciences, Largo da Sé Nova, 3000-213 Coimbra, Portugal

2 Faculty of Sciences, University of Lisbon, Campo Grande, 1749-016 Lisboa, Portugal

3 CENCIFOR - Forensic Sciences Centre, Largo da Sé Nova, 3000-213 Coimbra, Portugal

4 CESAM - Centre for Environmental and Marine Studies, Edifício C2, Campo Grande, 1749-016 Lisboa, Portugal

5 Faculty of Medicine, University of Lisbon, Av. Professor Egas Moniz, 1649-028 Lisboa, Portugal

6 Faculty of Medicine, University of Coimbra, Rua Larga, 3004-504 Coimbra, Portugal

Introduction

Forensic DNA phenotyping, also referred to as "DNA intelligence," is still not widely used but can provide important information in criminal casework [1–3]. However, externally visible characteristics (EVCs) are difficult to understand due to their high complexity derived from their multifactorial nature, involving various genes, gene–environment interactions, and epistasis phenomena [4, 5]. EVCs are also not widely used at present because some phenotypic loci may provide ancestral or clinical information that may raise ethical concerns in the forensic field. However, these markers could be used in crime solving in a similar way to conventional eyewitness testimony, helping police in the determination of a

suspect's profile and probably with a lower risk of exacerbating social pressure [6].

One multiplex designed to determine visible traits is IrisPlex assay [7] developed by Walsh et al. to predict eye color through the genotyping of 6 SNPs: rs12913832 (HERC2), rs1800407 (OCA2), rs12896399 (SLC24A4), rs16891982 (SLC45A2 (MATP), rs1393350 (TYR), and rs12203592 (IRF4). This multiplex based on the work of Liu et al. [8], later validated for forensic use [9], was reported as having precisions ranging from 91 to 98 % on the prediction of blue and brown eye color, depending on the European population under consideration [10].

As for implementation and internal validation of any forensic multiplex, studies need to be performed in the population of interest in order to satisfactorily evaluate its applicability to that population [11]. Since IrisPlex's publication, several studies were made to address its application on different populations, namely a Slovenian population [12], a variety of Eurasian populations [13], and in a US population [14]. Even though IrisPlex reliably predicted blue and brown eye color from DNA samples in Dutch, Slovenian, and East Asian populations [10, 12, 13], this assay turned out to be only moderately predictive for Asian and US populations [13, 14]. The authors of the works carried out on the Eurasian and US populations explained this phenomenon as an effect of IrisPlex testing being conducted using mainly European samples. Also, the populations they studied were highly admixed when compared to European ones [13, 14]. These authors point out the importance of studying additional populations, particularly from areas outside of Europe, for the evaluation and optimization of predictions.

Due to the close proximity of the Mediterranean region with that of North Africa, and the genetic flux resultant from multiple interactions between these two regions throughout the centuries [15], Mediterranean countries, especially the southern regions, have a stronger background of African ancestry when compared to Central and Eastern European countries [16]. The presence of African gene flow into southern Europe dates back to, at least, 1600 years ago and historically results from the trade relations and slave trading across the Western Sahara during the Roman occupation (200–600 AD) [15]. In Portugal, and also in Spain, this process took on even greater significance due to the existence of further possible sources of African admixture, namely the invasion of Iberia by the Moorish armies after 711 AD that lasted until the sixteenth century [17] and both Portugal and Spain's imperial colonial history and subsequent decolonization processes [18].

Taking into consideration that IrisPlex loci are not only related to eye color, but are also associated with hair color and skin pigmentation [19–23], adding further SNP loci to IrisPlex should provide genetic information able to predict these other human traits. Walsh et al. accomplished this with the development of HirisPlex, which includes the original 6 IrisPlex loci together with 18 additional SNPs that allow simultaneous prediction of hair and eye color [3]. Similarly, it is conceivable that IrisPlex, if analyzed together with additional loci, would allow the prediction of skin color in a similar way to the multiplex reported by Spichenok [24]. The association of these loci to different human traits occurs because the corresponding genes are involved in the same physiologic processes, namely melanosome biogenesis and melanin biosynthetic pathways. These are the main processes responsible for human pigmentation variation that affect eye and hair color as well as skin pigmentation [25].

The original aim of this work was to evaluate IrisPlex within a Portuguese population, but later, due to the refereed above, it was decided to test if it was possible to supplement the multiplex with additional loci in order to achieve simultaneous eye and skin color predictions and apply this to the study population. In order to do so, three loci were selected: rs1129038 (HERC2), rs2424984 (ASIP), and rs1426654 (SLC24A5). Sturm et al. described SNP rs1129038 as having a high degree of association with eye color [26]. Leite et al. also reported it as being involved in skin color pigmentation [27]. Valenzuela et al. categorized ASIP rs2424984 as the third most significant genetic contributor to variance in phenotypic skin reflectance [28]. SLC24A5 is a gene already known to be involved in skin color determination for some time [29]. Moreover, rs1426654 locus explains per se the difference in skin melanin index between 25 and 38 % amongst people of European vs. West African ancestry, respectively [30–32]. Valenzuela et al. identified rs1426654 and rs2424984, together with IrisPlex rs16891982, as being responsible for 45.7 % of the reflectance variance and an indicator of skin reflectance in human phenotypes across different populations [28].

In this work, eye color prediction was based on the multinomial logistic regression model. The predictors were the usual IrisPlex. In order to predict skin color, a binomial logistic regression model was developed. In this case, the predictors were IrisPlex and the three additional loci referred to above.

Material and methods

Sample collection and phenotypic characteristics record and classification

Buccal swabs and blood samples were collected from 192 unrelated volunteers residing in Portugal and under informed consent. This sample consisted of 95 males and 97 females with ages ranging from 16 to 72. For each individual, information was collected regarding age and sex (in Supplementary Table 1) and also their country of birth and ethnic origin, together with those of his/her parents.

Subsequently, photographic images of phenotypic characteristics were documented. Pictures were taken of the eyes and the skin of the inner upside part of the arm. Photographs were taken using a DSLR Nikon D90 camera equipped with a Nikon AF-S Nikkor 35 mm lens and a Speedlight SB-800 flash mounted on a tripod. For normalization, similar distance and light conditions were used for each photo.

Iris characteristics of all the individuals were classified independently by three different observers into three categories: 1: "blue eyes", 2: "neither blue nor brown" (which included blue/green, green, and green/brown eyes), and 3: "brown eyes." The majority classification was used to individually classify the phenotype of each volunteer's eyes.

Using a similar approach, individual skin color classification was performed following Fitzpatrick classification I to VI [33], but for simplification purposes, only three categories were applied: 1: "white and beige skin color" (equivalent to I and II in Fitzpatrick classification), 2: "light brown to medium brown" (equivalent to III and IV in Fitzpatrick classification), and 3: "dark brown to black skin" (equivalent to V and VI in Fitzpatrick classification). The majority classification was also used to individually classify each volunteer's skin tone (Supplementary Table 1). It is worth emphasizing here that there was not enough data to develop a logistic regression model where the response variable had six categories. As can be seen in the following, when describing the skin color prediction model, even with just three categories for skin color classification, it was necessary to combine the skin categories 1 and 2 just cited. This factor was due to the sparseness of the respective contingency table and gave rise to highly unstable parameter estimates.

DNA extraction and quantification

DNA from both buccal swabs and blood samples was extracted from each individual using Chelex® method [34]. Whenever necessary, samples were quantified with Quantifiler Duo Quantification kit in a 7500 Real-Time System using HID Real-Time PCR Software v1.1 and following the manufacturer's instructions (Applied Biosystems).

SNP loci selection and multiplex design

As a basis for multiplex design, IrisPlex development validation conditions were used [9]. This multiplex analyzes six SNP loci: rs12913832, rs1800407, rs12896399, rs16891982, rs1393350, and rs12203592 from the HERC2, OCA2, SLC24A4, SLC45A2 (MATP), TYR, and IRF4 genes previously identified as having a high degree of association with eye color [19, 20, 26]. Additionally, three other loci were added to complement the eye color analysis and to predict information about the sample donor skin tone: rs1129038, rs2424984, and rs1426654 from the HERC2, ASIP, and SLC24A5, respectively. The selection of these three markers was based on revision of the literature, especially in GWAS regarding SNP association with human phenotypic visible color characteristics [19, 20, 28, 35]. As would be expected, all these GWAS studies of SNP association with skin and hair color indicate IrisPlex loci as having a high degree of association with visible color characteristics, and not only eye color, since all these components are interrelated.

The three PCR primer pairs for SNP loci not present in IrisPlex were empirically designed using the rules of Dieffenbach et al. [36], and AutoDimer software was used to test the resulting multiplex design and to ensure little interaction between all primer pairs [37]. Single base extension (SBE) primer design followed a similar procedure whilst ensuring that primer melting temperatures were approximately 55 °C so that the SBE reaction would occur under the same conditions as those of IrisPlex [9]. The PCR and SBE primer sequences can be found in Supplementary Table 2.

SNP loci amplification and detection

The conditions used for amplification and minisequencing were the same as those indicated for IrisPlex PCR [7]. A total of 1 μl genomic DNA extract from each sample was amplified in a 12-μl PCR reaction containing 1× PCR buffer, 2.7 mM $MgCl_2$, 200 mM of each dNTP, 0.416 mM concentration of each primer (Supplementary Table 2), and 0.5 U of AmpliTaq Gold DNA polymerase (Applied Biosystems). PCR thermal cycling was performed on a gold-plated 96-well GeneAmp® PCR system 9700 (Applied Biosystems) using the following conditions: (1) 95 °C for 10 min, (2) 33 cycles of 95 °C for 30 s and 60 °C for 30 s, and (3) 5 min at 60 °C. After PCR amplification, excess primers and unincorporated deoxynucleotides were removed by adding 1 μL of USB® ExoSAP-IT® PCR Product Clean-Up (Affymetrix) to 2.5 μL of PCR product and incubating for 30 min at 37 °C, followed by 15 min at 80 °C for enzyme inactivation. SNP loci minisequencing was achieved by SBE in 5 μL reaction volumes containing 1 μL of SNaPshot® reaction mix (Applied Biosystems), 1 μl of purified PCR product, and 1 μL of SBE primer mix (Supplementary Table 2). Thermal cycling for SBE was conducted on a gold-plated GeneAmp® PCR system 9700 (Applied Biosystems) using the following thermocycling program: (1) 96 °C for 2 min and (2) 25 cycles of 96 °C for 10 s, 50 °C for 5 s, and 60 °C for 30 s. Excess fluorescently labeled ddNTPs were inactivated by incubation of SBE reaction with 1 μL of USB shrimp alkaline phosphatase (Affymetrix) at 37 °C for 45 min followed by enzyme inactivation at 80 °C for 15 min. Multiplex extension products were run on an Applied Biosystems 3130 Genetic Analyser using 1 μl of cleaned SBE product with 0.5 μl 120 LIZ size standard with 8.5 μl Hi-Di formamide (Applied Biosystems).

Allele calling was achieved with GeneMapper v. 3.7 software (Applied Biosystems).

Eye color prediction model

Eye color predictions were made using the Multinomial Logistic Regression Model [3, 8, 38]. Eye (iris) color was chosen as the response variable and categorized into three groups: blue, brown, and intermediate (that is, green to hazelnut eyes). The covariates corresponded to the classification of subjects based on population minor allele frequencies. Nine SNP loci were used in this study: rs12913832, rs1800407, rs12896399, rs16891982, rs1393350, rs12203592, rs1129038, rs2424984, and rs1426654. Details of the mathematical model are provided in the supplementary material.

Models were fitted by maximum likelihood. More precisely, the parameters of the models were estimated by the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, which belongs to the class of the Quasi-Newton Methods. To achieve this goal, the package "mlogit" [39], from the R software, version 3.0.3 [40] was used.

The significance of each variable to the fitted model was assessed by means of the Wald test [38]. The relevance of each predictor in the DNA phenotyping context was also taken into account in selecting the optimal subset of predictors. The likelihood ratio test was computed to evaluate the overall significance of the adjusted model.

The estimated outcome for the ith subject was given by:

$$\hat{Y}_i = \max\left\{\hat{\pi}_0(\boldsymbol{x}_i),\ \hat{\pi}_1(\boldsymbol{x}_i),\ \hat{\pi}_2(\boldsymbol{x}_i)\right\},\quad i = 1,\ \ldots,\ n.$$

In this context, it was crucial to evaluate the accuracy of the model in discriminating between the different outcomes. It is worth stressing that in cases where the three probabilities are relatively similar to each other, and the probability for the selected category is only marginally higher, there is an increased risk of an incorrect prediction being made [10]. The probability thresholds of 0.5 and 0.7 were applied to avoid potentially misleading classifications. Therefore, only the estimates that satisfied the restriction $\hat{Y}_i \geq c$, i=1, …, n, — where c represents the cutoff probability—was considered in the remaining analysis.

In regard to the goodness of fit, the focus of interest was the classification table. The estimated model was evaluated in relation to its ability to discriminate between the different categories of the outcome. From the 3×3 confusion matrix, the usual measures were computed for each category: sensitivity (SENS), specificity (SPEC), positive predictive value (PPV), and negative predictive value (NPV). In this context, SENS and PPV play a special role because these measures rely only on the comparison between the estimated values and the raw data for each category of the outcome. More precisely, for each eye color category j=0, 1, 2, SENS corresponds to the probability of an individual being classified correctly in category j; PPV represents the probability of an individual who has been classified in category j actually belonging to that category.

For each eye color category, the receiver operating characteristic (ROC) curve was computed. The respective area under the curve (AUC)—that is, the integral of the ROC curve—was measured to assess the overall performance of the estimated model in terms of the sensitivity and specificity [38, 41]. Bearing in mind that there were three classes under consideration, an extension of the AUC was also calculated for more than two classes. In brief, the AUC for each class against the rest was computed and then the results averaged [42, 43].

As stated previously, the fitted model will provide eye color predictions for future subjects. Therefore, it was crucial to evaluate the performance of the estimated model in the context of cross-validation. This procedure will also decrease the possibility of over-fitting [44]. Due to the reduced sample size, it was not possible to split the data into two subsets: the training set for modeling purposes and the validation set for predictions (usually 20 % of the sample size). To overcome the small sample size limitation, the leave-one-out cross-validation procedure was utilized. Hence, for each run, the model was estimated from the whole data set except one subject, and the prediction was made for the individual that had been left out. The procedure was repeated n times, where n is the sample size. The ROC curve and the respective AUC were computed from the probabilities generated by the cross-validation. The results were compared to those obtained from the whole data set to assess the goodness-of-fit of the model.

Skin color prediction model

Initially, the multinomial logistic regression model was applied to model skin color. The outcome variable was divided into the three categories referred to earlier in this section: 1: very white, 2: white beige, 3: dark brown. People with very white skin are extremely rare in the Portuguese population. The sampled data confirmed this to be true, and thus, there were very few subjects classified in this category. This led to very inaccurate estimation results. Therefore, the two categories "very white" and "white beige" were combined into a single category called "white."

The binomial logistic regression model was used to make skin color predictions [38, 44]. The response variable was the skin color divided into two categories: white and dark brown skin. The predictors were the same as those used in modeling eye color. Details of the mathematical model are provided in the supplementary material.

The models were fitted by maximum likelihood. More precisely, the parameters of the models were estimated using the iteratively reweighted least squares. For details, see McCullagh et al. [45]. The R software, version 3.0.3 [40],

was used to accomplish this. The significance of each variable in the fitted model was assessed by means of the Wald test [38]. The relevance of each predictor in the context of DNA phenotyping was also taken into account when selecting the optimal subset of predictors. The likelihood ratio test was computed to evaluate the overall significance of the adjusted model.

The estimated outcome for the *i*th subject was given by $\hat{Y}_i = \hat{\pi}(\boldsymbol{x}_i)$, $i = 1, \ldots, n$. Nevertheless, our interest relied on predicting the outcome (in a binary way). Following the methodology reported previously for modeling eye color, the probability threshold of 0.5 was defined to distinguish an individual between dark and white skin color. Therefore, only the estimates that satisfied the restriction $\hat{Y}_i \neq 0.5$, $i=1, \ldots, n$, were considered when assessing the fit of the model.

The goodness-of-fit measures were similar to those developed for modeling eye color, with appropriate adaptations, namely SENS, SPEC, PPV, and NPV. Additionally, the ROC curve was computed, and the respective AUC determined. This measure quantifies the ability of the model to distinguish between dark brown and white skin colors. As the fitted model will provide skin color predictions for future subjects, the leave-one-out cross-validation procedure was also applied in this context. The ROC curve and the respective AUC were computed from the probabilities generated by the cross-validation so as to compare the results with those obtained from the whole data set.

Results

Multiplex genotype assay design

The addition of three SNP loci to IrisPlex assay was conducted whilst simultaneously ensuring that all characteristics of the original multiplex were maintained [9]. The results obtained are demonstrated in an electropherogram of a brown-eyed individual, as shown in Fig. 1. Results for all the remaining individuals in the study can be found in Supplementary Table 1.

Eye color

Using a sample of the Portuguese population (sample size, $n=192$), a multinomial logistic regression model was developed using IrisPlex loci. The estimated multinomial logistic regression model to describe eye color is shown in Table 1. It is well established in the literature that the six IrisPlex SNP genotypes are very important to predict eye color [7, 8, 10]. Although the SNPs rs1393350 and rs12203592 are not statistically significant, they remain in the final estimated model due to their relevance in the phenotyping context. The likelihood ratio test was calculated to assess the overall significance of the model. The results showed that the covariates under consideration are particularly relevant when describing eye color (chi-squared test statistic=60.41, p value=0.00).

The probability thresholds discussed by Walsh et al. [17] and Dembinski and Picard [21] were applied here so as to evaluate the accuracy of the estimated model. Accordingly, two threshold values were analyzed: 0.5 and 0.7. The former of the two values was chosen as a threshold of 0.7 gave rise to a large number of inconclusive results. Moreover, when using a threshold of 0.5, 138 out of 192 subjects (72 %) were correctly predicted and only 22 individuals (11 %) incorrectly. There was a loss of 32 people (17 %) from the sample as their predictions were inconclusive.

Some goodness-of-fit measures are shown in Table 2, namely sensitivity, specificity, PPV, NPV, and AUC. Computations are based on the comparison between each of the classes.

The percentage of blue, intermediate, and brown eyes correctly classified is 60, 57, and 96 %, respectively. The positive predictive value for blue eyes is not very high (60 %), but the model behaves much better when predicting intermediate and brown eyes: 77 and 90 %, respectively.

The ROC curves and the AUCs are displayed in Fig. 2. The areas under the ROC curves are very high (0.94, 0.84, and 0.87 for blue, intermediate, and brown eyes, respectively). In the literature, values such as these for the AUCs are considered to result in excellent discrimination for each eye color category under consideration [42]. Furthermore, the overall AUC computed from the three classes under consideration is 0.84, which means that the predicted model is highly accurate in discriminating between blue, intermediate, and brown eye color.

The predicted probabilities generated by the leave-one-out cross-validation procedure have led to slightly lower values for the AUC, especially for blue eyes (0.84, 0.81, and 0.85 for blue, intermediate, and brown eyes, respectively). Nevertheless, they are still considered as excellent discrimination [42]. The overall AUC, based on the three classes under consideration, has a value of 0.82, which is similar to that obtained from the complete data set. In summation, the goodness-of-fit measures accurately support the behavior of the predicted model.

Skin color

The binomial logistic regression model used to describe skin color was estimated from the same sample of the Portuguese population (sample size, $n=192$). The modeling procedure began with all nine of the SNPs referred to in "Material and Methods." However, there was the need to exclude the SNP-ID rs12913832 due to its strong linkage with SNP-ID rs1129038.

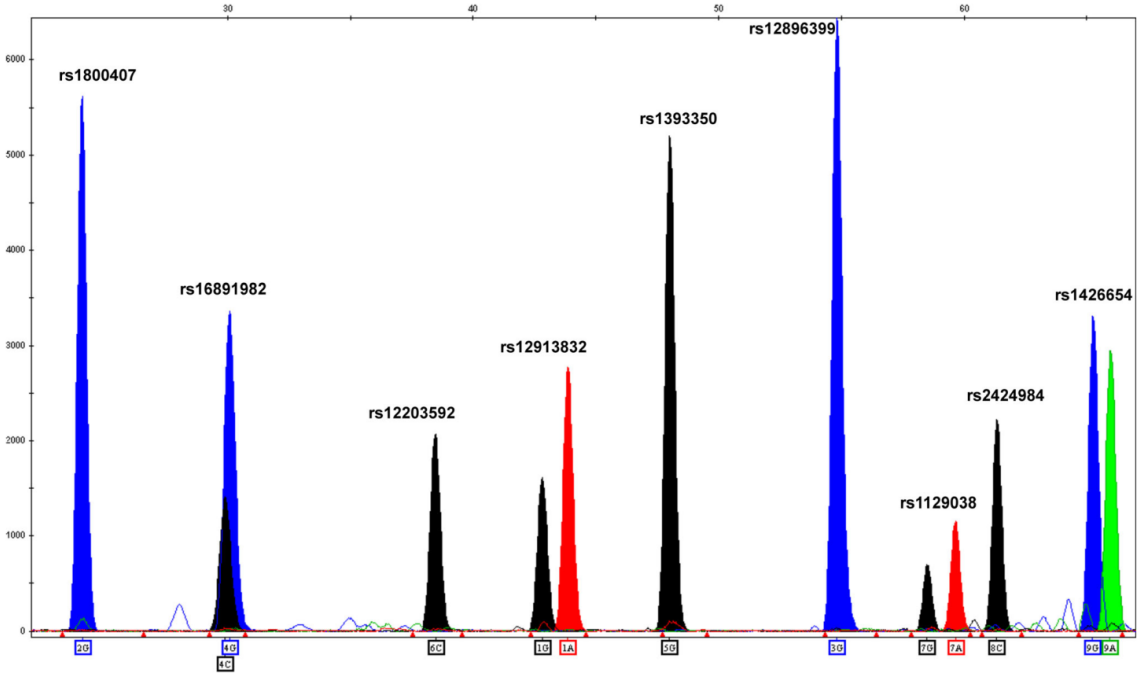


Fig. 1 Electropherogram of an individual genotyped for eye and skin color prediction. The three new loci added to IrisPlex are located at the right on the electropherogram

The estimated binomial logistic regression model is shown in Table 3. The results have revealed that the SNPs rs1800407, rs12896399, rs1393350, and rs1129038 are not statistically significant. Nevertheless, they are retained in the final estimated model because of their importance in the phenotyping context. In order to evaluate the overall significance of the model, the likelihood ratio test was computed. The results show that the covariates are relevant to describe skin color (chi-squared test statistic= 158.00, p value=0.00).

The accuracy of the estimated model in predicting the outcome in a dichotomous way was evaluated. Following the approach for modeling eye color, individuals were differentiated between having dark brown and white skin using the cutoff probabilities of 0.5 and 0.7. Due to the large number of inconclusive results obtained from the higher cutoff, the former probability threshold was chosen. Using this threshold of 0.5, the estimated model correctly predicted 183 out of 192 subjects (95 %) and incorrectly only nine individuals (5 %).

Table 1 Estimated coefficients, standard errors, and p values for the fitted model to the eye color

SNP - ID	Coefficient	Standard error	p value
logit: g_1			
rs12913832	3.126	0.626	0.000
rs1800407	1.005	1.124	0.372
rs12896399	-2.577	0.640	0.000
rs16891982	-5.084	1.101	0.000
rs1393350	-0.455	0.672	0.499
rs12203592	-0.872	0.999	0.383
logit: g_2			
rs12913832	1.941	0.425	0.000
rs1800407	1.717	0.571	0.003
rs12896399	-1.024	0.315	0.001
rs16891982	-2.165	0.353	0.000
rs1393350	0.021	0.447	0.962
rs12203592	0.319	0.548	0.560

logit: g_1 and g_2, comparison of blue and intermediate eye color to the reference level, respectively

Table 2 Goodness-of-fit measures for the eye color: sensitivity, specificity, PPV, NPV, and AUC

	Eye color		
Measure	Blue	Intermediate	Brown
Sensitivity (%)	60	57	96
Specificity (%)	97	96	68
Positive predicted value (%)	60	77	90
Negative predicted value (%)	97	91	84
Area under the curve	0.94	0.84	0.87

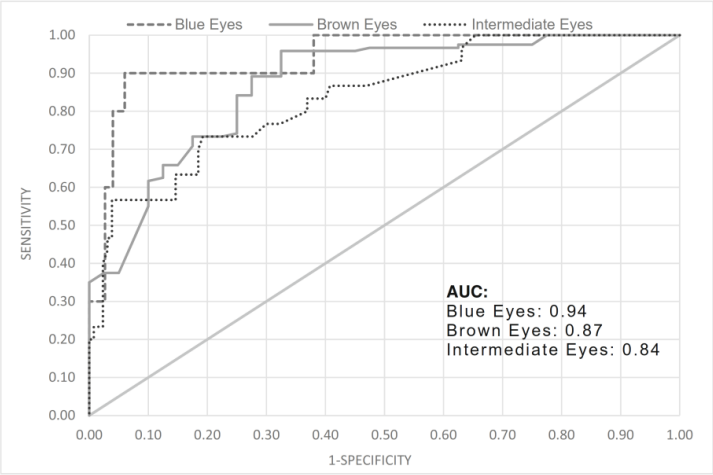


Fig. 2 Goodness-of-fit for the eye color: ROC curve and AUC, computed for each class against the remaining classes

There were no inconclusive predictions, and therefore, no individuals were excluded from the analysis.

Table 4 shows some goodness-of-fit measures, namely sensitivity, specificity, PPV, NPV, and AUC. The percentage of dark brown color correctly classified is 86 %. The positive predictive value is 93 %. Furthermore, both specificity and negative predicted values are very high (98 and 96 %, respectively). These results are very promising, showing that the estimated model performs very well in distinguishing between persons with white skin color and those with dark brown.

The ROC curve, as well as the AUC, is displayed in Fig. 3. The area under the ROC curve gives a very high value (AUC=0.99). The predicted probabilities generated by the leave-one-out cross-validation procedure have led to a similar AUC (AUC=0.96). According to the literature [42], these results clearly show an outstanding performance in discriminating between white and dark brown skin colors.

Table 3 Estimated coefficients, standard errors, and *p* values for the fitted model to the skin color

SNP - ID	Coefficient	Standard error	*p* value
Constant	-7.327	1.736	0.000
rs1800407	-0.982	1.203	0.414
rs12896399	-0.762	0.641	0.235
rs16891982	3.076	0.933	0.001
rs1393350	0.754	0.929	0.417
rs12203592	1.993	1.379	0.148
rs1129038	-0.824	0.683	0.227
rs2424984	1.778	0.726	0.014
rs1426654	4.658	1.089	0.000

Discussion

The 192 samples used are representative of the Portuguese population. Even though it would be advantageous, from a statistical point of view, to have a larger sample size, this number is in line with others used in similar works [12–14]. The eye and skin color classifications were carried out by visual qualitative determination, as stated by Dembinski et al. [21]. This practical manner of classification forms the basis of eyewitness testimonies, and therefore, it is essential that the obtained results correlate with the visual determinations [14]. In order to reduce any ambiguity in color classification, the grouped independent observer classification was used.

Walsh et al. provide prediction accuracies for eye color determination at the 0.7 threshold of 91.6 and 56 % for blue and brown eye colors, respectively, and 91.6 and 87.5 % at the 0.5 threshold [7]. In the Portuguese population, using adjusted parameters, prediction accuracies for eye color determination of 60 and 90 % were obtained at the 0.5 threshold for blue and brown eye colors, respectively. Due to the large number of inconclusive results generated, it was not possible to rely on the estimation at the 0.7 threshold. However, the AUCs performed similarly: 0.94, 0.84, and 0.87 (0.5 threshold) compared with 0.97, 0.84, and 0.95 (0.7 threshold) for blue,

Table 4 Goodness-of-fit measures for the skin color: sensitivity, specificity, PPV, NPV, and AUC

Measure	Skin color
Sensitivity (%)	86
Specificity (%)	98
Positive predicted value (%)	93
Negative predicted value (%)	96
Area under the curve	0.99

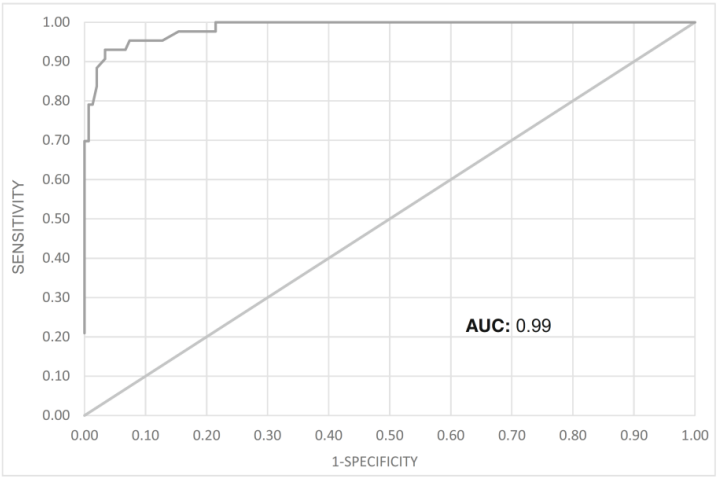


Fig. 3 Goodness-of-fit for the skin color: ROC curve and AUC

intermediate, and brown. These results suggest that the ability of the estimated model to make accurate predictions is only slightly reduced when using the 0.5 threshold. Another important factor is that for the data sets under consideration, TYR rs1393350 and IRF4 rs12203592 do not seem to play such an important role in the differentiation of brown and intermediate eye color as suggested by Walsh et al. [7] because they are not statistically significant. This conclusion may reflect the existence of some admixing just as Dembinski et al. reported for a US population [14]. On the other hand, this could also be a consequence of using a smaller sample size than Walsh et al. [7]. Nevertheless, the AUCs clearly show that the estimated model is well fitted to the Portuguese population. This conclusion is verified by a correct eye color prediction rate of 72 %, with 17 % inconclusive, and only 11 % incorrect. Additionally, the results from cross-validation have also showed that the model developed in this paper is very accurate in predicting eye color for future observations. More precisely, the AUCs based on the cross-validation methodology were similar to those obtained from the whole data set.

Concerning skin color prediction, the modeling procedure began with multinomial logistic regression so as to conduct a similar analysis to eye color. The estimated model did not give accurate results. The reason for this is that Portugal is the most southwestern region of Europe, and as a result, it is difficult to encounter individuals of the types I and II Fitzpatrick phototype that are much more commonplace in northern Europe. For the sample under analysis, the number of individuals who belonged to those two categories was significantly reduced. Consequently, in order to satisfactorily carry out the statistical analysis, these individuals were amalgamated into the same category as types III and IV Fitzpatrick phototypes. This category mainly corresponds to Mediterranean individuals. Unfortunately, this procedure prevented predictions between phototypes of classes I/II, III/IV, and V/VI as initially desired. However, it was possible to differentiate between class V/VI (dark brown skin) and the two others (white skins). If a larger sample was used, the multinomial logistic regression model would lead to more accurate results. In this case, it would be possible to develop a model to predict all skin types.

Thus, skin color prediction was based on the binomial logistic regression model. The results were very promising. In fact, the statistical analysis has led to an estimated model with a prediction accuracy of 93 %. The AUC value is 0.99, which gives an excellent discrimination between skin colors. This model has proved its value with a 95 % correct prediction, which can be considered a very good estimation. Moreover, the model is highly adequate for predicting skin color for future individuals. The veracity of this statement relies on the fact that the AUC by cross-validation is similar to that attained by analyzing the whole data set.

To summarize, in this paper, the multinomial and binomial logistic regression models were used to predict eye and skin color, respectively. One of the major advantages of these methodologies is that they rely on a strong theoretical framework. As a consequence, the contribution of each SNP to the prediction of the outcome (either eye or skin color), as well as its precision, can be obtained (see Tables 1 and 3). Moreover, the models developed in the paper have revealed excellent prediction accuracies. These methodologies are largely applied in forensic biology, as can be seen, for instance, in [11, 14–17, 19, 21]. In future research, classification methods, such as neural networks for instance, may be applied to these data sets so as to predict eye and skin colors. Liu et al. [15] used multinomial logistic regression to predict eye color. However, they also used classification methods (Fuzzy c-means clustering, neural networks, and classification tree) to

attain the same objective. In their Supplemental Data [15], they demonstrated that multinomial logistic regression behaves better than the classification methods in respect to the model's ability to correctly discriminate between three different categories of eye color.

Conclusions

EVCs such as eye and skin color can be difficult to classify due to the small difference between adjacent categories, especially when classification is done by visual determination due to differences in the observers color perception. Nevertheless, visual determination is still the basis for eyewitness testimonies used in the field of criminal investigation. These characteristics are even more difficult to investigate due to their multigenic nature that is frequently subjected to the influence of environment.

Despite the limitations cited previously, good results have been obtained for the prediction of eye color, a characteristic that is the result of a relatively small number of genes and with little influence of environment. Still, eye color is very informative about the appearance of an individual. A good example of eye color investigation is the development of IrisPlex, a multiplex developed for eye color determination [7]. However, genes involved in eye color determination are also involved in other "color" characteristics, such as hair color [3]. In this work, the IrisPlex has been supplemented with three additional SPNs in order to obtain a simultaneous test for eye and skin color determination.

In addition to eye color analysis, different loci from other candidate genes can be used to complement IrisPlex and enhance prediction models, as discussed by Dembinski and Picard [21]. Thus, in the future development of this kind of multiplex, it is intended to add different loci so as to infer other phenotypic characteristics. This procedure will certainly lead to the establishment of more informative "intelligence" tools. Together, these will not only enhance the prediction accuracy independently of the studied populations and the admixing present, but will also provide combined information about eye, hair and skin color, or even about other characteristics too.

Acknowledgments The authors would like to express their sincere gratitude to all the volunteers who contributed to this work by providing samples. The authors would also like to thank two anonymous reviewers who gave valuable comments and advice on how to improve the content of this article.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards All experiments were approved by the Governing board of the National Institute of Legal Medicine and Forensic Sciences, Portugal.

References

1. Poetsch M, Blohm R, Harder M et al (2013) Prediction of people's origin from degraded DNA—presentation of SNP assays and calculation of probability. Int J Leg Med 127:347–357. doi:10.1007/s00414-012-0728-0
2. Keating B, Bansal AT, Walsh S et al (2013) First all-in-one diagnostic tool for DNA intelligence: genome-wide inference of biogeographic ancestry, appearance, relatedness, and sex with the Identitas v1 Forensic Chip. Int J Leg Med 127:559–572. doi:10.1007/s00414-012-0788-1
3. Walsh S, Liu F, Wollstein A et al (2013) The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int Genet 7:98–115. doi:10.1016/j.fsigen.2012.07.005
4. Pospiech E, Draus-Barini J, Kupiec T et al (2011) Gene-gene interactions contribute to eye colour variation in humans. J Hum Genet 56:447–455. doi:10.1038/jhg.2011.38
5. Pośpiech E, Wojas-Pelc A, Walsh S et al (2014) The common occurrence of epistasis in the determination of human pigmentation and its impact on DNA-based pigmentation phenotype prediction. Forensic Sci Int Genet 11C:64–72. doi:10.1016/j.fsigen.2014.01.012
6. Kayser M, Schneider PM (2009) DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. Forensic Sci Int Genet 3:154–61. doi:10.1016/j.fsigen.2009.01.012
7. Walsh S, Liu F, Ballantyne KN et al (2011) IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet 5:170–80. doi:10.1016/j.fsigen.2010.02.004
8. Liu F, van Duijn K, Vingerling JR et al (2009) Eye color and the prediction of complex phenotypes from genotypes. Curr Biol 19:R192–3. doi:10.1016/j.cub.2009.01.027
9. Walsh S, Lindenbergh A, Zuniga SB et al (2011) Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence. Forensic Sci Int Genet 5:464–71. doi:10.1016/j.fsigen.2010.09.008
10. Walsh S, Wollstein A, Liu F et al (2012) DNA-based eye colour prediction across Europe with the IrisPlex system. Forensic Sci Int Genet 6:330–40. doi:10.1016/j.fsigen.2011.07.009
11. (2004) Scientific Working Group on DNA analysis methods, revised validation guidelines. Forensic Sci. Commun. 6
12. Kastelic V, Pospiech E, Draus-Barini J et al (2013) Prediction of eye color in the Slovenian population using the IrisPlex SNPs. Croat Med J 54:381–386
13. Yun L, Gu Y, Rajeevan H, Kidd KK (2014) Application of six IrisPlex SNPs and comparison of two eye color prediction systems in diverse Eurasia populations. Int J Leg Med. doi:10.1007/s00414-013-0953-1
14. Dembinski GM, Picard CJ (2014) Evaluation of the IrisPlex DNA-based eye color prediction assay in a United States population. Forensic Sci Int Genet 9:111–117. doi:10.1016/j.fsigen.2013.12.003
15. Moorjani P, Patterson N, Hirschhorn JN et al (2011) The history of African gene flow into Southern Europeans, Levantines, and Jews. PLoS Genet 7, e1001373. doi:10.1371/journal.pgen.1001373
16. Plaza S, Calafell F, Helal A et al (2003) Joining the Pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. Ann Hum Genet 67:312–328. doi:10.1046/j.1469-1809.2003.00039.x
17. Kennedy H (1996) Muslim Spain and Portugal. A Political History of al-Andalus. Longman
18. Pinto AC, Lloyd-Jones S (2003) The last empire: thirty years of Portuguese decolonization. ix, 156 p

19. Han J, Kraft P, Nan H et al (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet 4, e1000074. doi:10.1371/journal.pgen.1000074
20. Sulem P, Gudbjartsson DF, Stacey SN et al (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet 39:1443–1452. doi:10.1038/ng.2007.13, ng.2007.13 [pii]
21. Sulem P, Gudbjartsson DF, Stacey SN et al (2008) Two newly identified genetic determinants of pigmentation in Europeans. Nat Genet 40:835–7. doi:10.1038/ng.160
22. Branicki W, Brudnik U, Wojas-Pelc A (2009) Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype. Ann Hum Genet 73:160–170. doi:10.1111/j.1469-1809.2009.00504.x
23. Nan H, Kraft P, Qureshi A a et al (2009) Genome-wide association study of tanning phenotype in a population of European ancestry. J Invest Dermatol 129:2250–7. doi:10.1038/jid.2009.62
24. Spichenok O, Budimlija ZM, Mitchell AA et al (2011) Prediction of eye and skin color in diverse populations using seven SNPs. Forensic Sci Int Genet 5:472–478. doi:10.1016/j.fsigen.2010.10.005
25. Sturm RA, Frudakis TN (2004) Eye colour: portals into pigmentation genes and ancestry. Trends Genet 20:327–32. doi:10.1016/j.tig.2004.06.010
26. Sturm RA, Duffy DL, Zhao ZZ et al (2008) 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 82:424–31. doi:10.1016/j.ajhg.2007.11.005
27. Leite TK, Fonseca RM, de Franca NM et al (2011) Genomic ancestry, self-reported "color" and quantitative measures of skin pigmentation in Brazilian admixed siblings. PLoS One 6, e27162. doi:10.1371/journal.pone.0027162
28. Valenzuela RK, Henderson MS, Walsh MH et al (2010) Predicting phenotype from genotype: normal pigmentation. J Forensic Sci 55:315–322. doi:10.1111/j.1556-4029.2009.01317.x, JFO1317 [pii]
29. Lamason RL, Mohideen MA, Mest JR (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science (80-) 310:1782–1786. doi:10.1126/science.1116238
30. Stokowski RP, Pant PV, Dadd T et al (2007) A genomewide association study of skin pigmentation in a South Asian population. Am J Hum Genet 81:1119–1132. doi:10.1086/522235
31. Dimisianos G, Stefanaki I, Nicolaou V et al (2009) A study of a single variant allele (rs1426654) of the pigmentation-related gene SLC24A5 in Greek subjects. Exp Dermatol 18:175–177. doi:10.1111/j.1600-0625.2008.00758.x
32. Soejima M, Koda Y (2007) Population differences of two coding SNPs in pigmentation-related genes SLC24A5 and SLC45A2. Int J Leg Med 121:36–39. doi:10.1007/s00414-006-0112-z
33. Fitzpatrick TB (1988) The validity and practicality of sun-reactive skin types I through VI. Arch Dermatol 124:869. doi:10.1001/archderm.1988.01670060015008
34. Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10:506–513
35. Sturm R (2009) Molecular genetics of human pigmentation diversity. Hum Mol Genet 18:R9–17. doi:10.1093/hmg/ddp003
36. Dieffenbach CW, Lowe TM, Dveksler GS (1993) General concepts for PCR primer design. PCR Methods Appl 3:S30–7
37. Vallone PM, Butler JM (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques 37:226–31
38. Hosmer DW Jr, Lemeshow S, Sturdivant RX (2013) Applied logistic regression, third edit. Wiley, New Jersey
39. Croissant Y (2013) mlogit: multinomial logit model. R package
40. R Core Team (2014) R: A language and environment for statistical computing
41. Lee K, Ahn H, Moon H et al (2013) Multinomial logistic regression ensembles. J Biopharm Stat 23:681–94. doi:10.1080/10543406.2012.756500
42. Ferri C, Hernández-Orallo J, Salido M (2003) Volume under the ROC Surface for Multi-class Problems. In: Lavrač N, Gamberger D, Blockeel H, Todorovski L (eds) Mach. Learn. ECML 2003 SE - 12. Springer Berlin Heidelberg, pp 108–120
43. Hand DJ, Till RJ (2001) A simple generalisation of the area under the ROC curve for multiple class classification problems. Mach Learn 45:171–186. doi:10.1023/A:1010920819831
44. Hilbe JM (2009) Logistic regression models. Chapman & Hall/CRC Press, Boca Raton
45. McCullagh P, Nelder JA (1989) Generalized linear models, second Edi. Chapman & Hall, London

Supplementary Material for

"Assessment of IrisPlex based multiplex for eye and skin color prediction with application to a Portuguese population"

Paulo Dario [a,b,c,d,*], Helena Mouriño [b], Ana Rita Oliveira [b,d], Isabel Lucas [a], Teresa Ribeiro [a,c], Maria João Porto [a,c], Jorge Costa Santos [a,c,e], Deodália Dias [a,d], Francisco Corte Real [a,c,f].

[a] INMLCF - National Institute of Legal Medicine and Forensic Sciences, Portugal
Largo da Sé Nova, s/n. 3000-213 Coimbra, Portugal.

[b] University of Lisbon, Faculty of Science, Portugal
Campo Grande. 1749-016 Lisboa, Portugal.

[c] CENCIFOR - Forensic Science Centre, Portugal
Largo da Sé Nova, s/n. 3000-213 Coimbra, Portugal.

[d] CESAM - Centre for Environmental and Marine Studies
Pólo de Lisboa. Edifício C2. Campo Grande. 1749-016 Lisboa, Portugal.

[e] University of Lisbon, Faculty of Medicine, Portugal
Av. Professor Egas Moniz. 1649-028 Lisboa, Portugal

[f] University of Coimbra, Faculty of Medicine, Portugal
Rua Larga. 3004-504 Coimbra, Portugal

* Corresponding author e-mail address: pdario@dlinml.mj.pt (Paulo Dario).
Tel: +351 218 811 866; Fax: +351 218 864 493.

Eye color prediction model

Let Y_i represent the response for subject i, $i = 1, \ldots, n$, with three levels, where n is the sample size. It is assumed that the categories of this variable are coded 0 (reference value, which is the brown eyes), 1 (blue eyes) or 2 (intermediate color). Thus, the outcome variable follows a Multinomial distribution. Let the conditional probability of each outcome category, given the covariate vector for the i-th subject, $\boldsymbol{x}_i$, be denoted by $\pi_j(\boldsymbol{x}_i) = P(Y_i = j \mid \boldsymbol{x}_i)$, $j = 0, 1, 2$. The multinomial logistic regression model relates a transformation of the outcome (eye color) to a linear combination of the predictors. The transformation is referred to as the logit link function. In this context, two logit functions are supplied because there is the need to compare each category to the baseline, that is,

$$g_j(\boldsymbol{x}_i) = \ln\left(\frac{\pi_j(\boldsymbol{x}_i)}{\pi_0(\boldsymbol{x}_i)}\right) = \beta_{j1}x_{1i} + \ldots + \beta_{j9}x_{9i}, \quad j = 1, 2, \tag{1}$$

where the quotient $\frac{\pi_j(\boldsymbol{x}_i)}{\pi_0(\boldsymbol{x}_i)}$ represents the odds of having blue eyes, or intermediate eye color respectively for $j = 1, 2$, relative to brown eyes; the predictor x_{ki} is the number of minor alleles of the k-th SNP, for the i-th subject, with the respective regression coefficient denoted by $\beta_k, k = 1, \ldots, 9, i = 1, \ldots, n$.

Solving Eq. 1 for the probabilities $\pi_j(\boldsymbol{x}_i)$, $j = 1, 2$, the conditional probability of each outcome category takes the form,

$$\pi_0(\boldsymbol{x}_i) = P(Y_i = 0 \mid \boldsymbol{x}_i) = \frac{1}{1 + \exp\left(\sum_{k=1}^{9}\beta_{1k}x_{ki}\right) + \exp\left(\sum_{k=1}^{9}\beta_{2k}x_{ki}\right)}$$

$$\pi_1(\boldsymbol{x}_i) = P(Y_i = 1 \mid \boldsymbol{x}_i) = \frac{\exp\left(\sum_{k=1}^{9}\beta_{1k}x_{ki}\right)}{1 + \exp\left(\sum_{k=1}^{9}\beta_{1k}x_{ki}\right) + \exp\left(\sum_{k=1}^{9}\beta_{2k}x_{ki}\right)}$$

$$\pi_2(\boldsymbol{x}_i) = P\left(Y_i = 2 \mid \boldsymbol{x}_i\right) = \frac{\exp\left(\sum_{k=1}^{9} \beta_{2k} x_{ki}\right)}{1 + \exp\left(\sum_{k=1}^{9} \beta_{1k} x_{ki}\right) + \exp\left(\sum_{k=1}^{9} \beta_{2k} x_{ki}\right)}.$$

Skin color prediction model

Let Y_i represent the response for subject $i,\ i = 1, \ldots, n,$ with two levels. The categories of this variable are coded 0 (reference value: white skin) or 1 (dark brown skin). Thus, the outcome variable follows a Bernoulli distribution. Let $\pi(\boldsymbol{x}_i) = P\left(Y_i = 1 \mid \boldsymbol{x}_i\right), i = 1, \ldots, n,$ represent the conditional probability that the outcome takes the value one, given the covariate vector for the i-th subject, $\boldsymbol{x}_i$. In the same way as modeling eye color, the logit link function is used to describe the linear relationship between the outcome (skin color) and the predictors:

$$g(\boldsymbol{x}_i) = \ln\left(\frac{P\left(Y_i = 1 \mid \boldsymbol{x}_i\right)}{P\left(Y_i = 0 \mid \boldsymbol{x}_i\right)}\right) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_9 x_{9i}, \quad i = 1, \ldots, n, \qquad (2)$$

where β_0 is the constant term. The remaining parameters have similar meanings to those described for the eye color prediction model. The same is true for the explanatory variables (see Eq. 1).

Solving equation (2) for the probability $\pi(\boldsymbol{x}_i),$ the conditional probability of dark brown skin is then given by

$$\pi(\boldsymbol{x}_i) = P\left(Y_i = 1 \mid \boldsymbol{x}_i\right) = \frac{\exp\left(\sum_{k=0}^{9} \beta_k x_{ki}\right)}{1 + \exp\left(\sum_{k=0}^{9} \beta_k x_{ki}\right)}, \quad \text{with } x_{0i} = 1,\ i = 1, \ldots, n.$$

Table S1 - Information of sample donors, eye and skin color classification and genotyping results

Sample #	Sex	Age	Eye Color Classification	Skin Color Classification	F1 1 rs12913832	F1 2	F2 1 rs1800407	F2 2	F3 1 rs12896399	F3 2	F4 1 rs16891982	F4 2	F5 1 rs1393350	F5 2	F6 1 rs12203592	F6 2	F7 1 rs1129038	F7 2	F8 1 rs2424984	F8 2	F9 1 rs1426654	F9 2
1	M	33	2	2	1G	1G	2G	2G	3G	3T	4G	4C	5G	5A	6C	6C	7G	7A	8T	8T	9A	9A
2	M	35	3	2	1A	1A	2G	2G	3T	3T	4G	4G	5G	5G	6C	6C	7G	7G	8C	8T	9A	9A
3	F	21	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
4	M	37	1	2	1G	1G	2G	2A	3G	3T	4G	4G	5G	5G	6C	6C	7A	7A	8T	8T	9A	9A
5	M	38	1	2	1G	1G	2G	2G	3G	3T	4G	4G	5G	5A	6C	6C	7A	7A	8C	8T	9A	9A
6	M	37	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9G	9A
7	F	38	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
8	M	29	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8C	9G	9A
9	F	25	3	3	1G	1A	2G	2G	3G	3T	4G	4C	5G	5A	6C	6C	7G	7A	8C	8T	9G	9A
10	M	23	3	3	1A	1A	2G	2G	3G	3T	4C	4C	5G	5G	6C	6C	7G	7G	8T	8T	9G	9A
11	M	23	3	3	1G	1A	2G	2G	3G	3T	4C	4C	5G	5A	6C	6T	7G	7A	8C	8T	9G	9A
12	F	22	2	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5A	6T	6T	7G	7A	8T	8T	9A	9A
13	M	41	3	2	1G	1A	2G	2G	3G	3T	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
14	F	30	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5A	6C	6C	7G	7G	8C	8T	9A	9A
15	F	30	3	2	1A	1A	2G	2G	3T	3T	4G	4G	5G	5G	6C	6T	7G	7G	8T	8T	9A	9A
16	M	47	3	3	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8C	8T	9G	9G
17	M	35	3	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
18	F	24	3	2	1A	1A	2G	2G	3G	3T	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
19	M	22	3	2	1G	1A	2G	2G	3T	3T	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9G	9A
20	F	28	3	2	1A	1A	2A	2A	3G	3G	4C	4C	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
21	M	29	1	2	1G	1G	2G	2A	3G	3T	4G	4G	5G	5G	6C	6T	7A	7A	8C	8T	9A	9A
22	M	25	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
23	F	21	3	3	1G	1A	2G	2G	3T	3T	4C	4C	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A

Sample #	Sex	Age	Eye Color Classification	Skin Color Classification	F1 1 rs12913832	F1 2	F2 1 rs1800407	F2 2	F3 1 rs12896399	F3 2	F4 1 rs16891982	F4 2	F5 1 rs1393350	F5 2	F6 1 rs12203592	F6 2	F7 1 rs1129038	F7 2	F8 1 rs2424984	F8 2	F9 1 rs1426654	F9 2
24	M	51	3	2	1G	1A	2G	2G	3T	3T	4G	4G	5G	5G	6C	6C	7G	7A	8C	8T	9A	9A
25	M	36	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8C	9G	9G
26	F	25	3	3	1G	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7A	8C	8C	9G	9A
27	M	28	2	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8C	8T	9A	9A
28	F	20	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8C	8T	9A	9A
29	M	54	2	2	1G	1A	2G	2A	3G	3G	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
30	F	29	2	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5A	6T	6T	7G	7G	8T	8T	9A	9A
31	F	58	3	3	1A	1A	2G	2A	3G	3T	4C	4C	5G	5G	6C	6C	7G	7G	8C	8T	9G	9A
32	M	58	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8T	9G	9A
33	M	23	2	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6T	7G	7A	8T	8T	9A	9A
34	F	21	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6T	7G	7G	8T	8T	9A	9A
35	M	25	3	2	1G	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7G	7A	8C	8T	9A	9A
36	F	32	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
37	M	49	3	2	1A	1A	2G	2G	3T	3T	4G	4C	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
38	M	23	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
39	M	44	3	3	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8C	8C	9G	9A
40	F	44	3	3	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8C	8C	9G	9A
41	M	41	3	2	1A	1A	2G	2A	3G	3T	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
42	F	24	2	2	1G	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
43	M	28	2	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7A	8C	8T	9A	9A
44	F	18	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
45	F	24	3	3	1G	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7A	8C	8C	9G	9G
47	M	44	3	3	1A	1A	2G	2G	3G	3T	4G	4C	5G	5G	6C	6C	7G	7G	8C	8T	9G	9A
48	F	43	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8C	9G	9G
49	F	34	3	2	1G	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
50	F	24	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8C	9G	9A
51	M	29	1	1	1G	1G	2G	2G	3T	3T	4G	4G	5G	5A	6C	6C	7A	7A	8T	8T	9A	9A
52	M	25	2	2	1G	1A	2G	2A	3G	3T	4G	4G	5G	5G	6C	6T	7G	7A	8C	8T	9A	9A
53	F	26	3	2	1A	1A	2G	2G	3G	3T	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
54	M	40	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7G	7G	8C	8T	9A	9A
55	M	60	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8C	8T	9A	9A
56	F	38	2	3	1G	1G	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7A	7A	8C	8C	9A	9A
57	M	24	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
58	F	24	3	2	1A	1A	2G	2A	3G	3T	4C	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
59	M	21	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5A	6C	6T	7G	7G	8T	8T	9A	9A
60	F	21	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6T	7G	7G	8C	8T	9A	9A
61	M	24	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5A	6C	6C	7G	7G	8C	8T	9A	9A
62	F	23	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
63	M	42	3	2	1G	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
64	F	36	2	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
65	M	26	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5A	6C	6C	7G	7G	8C	8T	9A	9A
66	M	32	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
67	F	34	3	2	1G	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7G	7A	8C	8C	9A	9A
68	M	33	3	3	1A	1A	2G	2G	3T	3T	4C	4C	5G	5G	6C	6C	7G	7G	8C	8T	9G	9G
69	F	24	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8T	9G	9G
70	F	24	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8T	9G	9A
71	M	43	3	2	1G	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7A	8T	8T	9G	9A

72	F	36	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
73	F	26	2	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
74	F	29	3	3	1G	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9G	9A
75	M	72	1	1	1G	1G	2G	2G	3T	3T	4G	4C	5G	5A	6C	6C	7A	7A	8T	8T	9A	9A
76	F	46	3	2	1G	1A	2G	2G	3T	3T	4C	4C	5G	5A	6C	6C	7G	7A	8C	8T	9A	9A
77	M	25	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
78	F	38	3	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
79	M	48	3	3	1G	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7A	8C	8T	9A	9A
80	M	46	3	2	1A	1A	2G	2A	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
81	F	32	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
82	F	36	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8T	9G	9G
83	F	38	3	2	1G	1A	2G	2G	3G	3T	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
84	M	22	3	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
85	F	19	2	2	1G	1A	2G	2A	3G	3T	4G	4G	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
86	F	30	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
87	M	40	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
88	F	37	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
89	M	69	2	3	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8C	8T	9A	9A
90	F	30	3	3	1A	1A	2G	2G	3G	3G	4G	4C	5G	5A	6C	6C	7G	7G	8C	8T	9G	9A
91	M	42	2	2	1G	1A	2G	2G	3G	3T	4G	4C	5G	5A	6C	6T	7G	7A	8T	8T	9A	9A
92	F	30	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
93	M	34	3	2	1G	1A	2G	2G	3G	3T	4G	4C	5G	5A	6C	6C	7G	7A	8T	8T	9A	9A
94	F	33	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
95	F	23	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7G	8C	8T	9A	9A
96	M	21	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5A	6C	6C	7G	7G	8C	8T	9A	9A
97	F	18	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8C	9G	9G
98	M	45	2	2	1G	1G	2G	2G	3G	3T	4G	4C	5G	5A	6C	6C	7A	7A	8T	8T	9A	9A
99	F	38	3	2	1G	1A	2G	2G	3T	3T	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
100	M	26	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
101	F	18	3	2	1G	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9G	9A
102	M	20	1	2	1G	1G	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7A	7A	8C	8C	9A	9A
103	F	20	2	2	1G	1A	2G	2G	3T	3T	4G	4G	5G	5A	6C	6C	7G	7A	8T	8T	9A	9A
104	F	45	3	2	1A	1A	2G	2G	3G	3T	4G	4C	5A	5A	6C	6C	7G	7G	8T	8T	9A	9A
105	M	18	2	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
106	F	16	3	2	1A	1A	2G	2G	3T	3T	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
107	F	17	3	3	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9G	9A
108	M	33	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
109	M	27	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8C	8T	9A	9A
110	F	27	1	2	1G	1G	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7A	7A	8C	8T	9A	9A
111	M	42	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8T	9G	9G
112	F	20	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5A	6C	6C	7G	7G	8C	8T	9G	9A
113	M	35	3	2	1A	1A	2G	2A	3G	3T	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
114	F	40	3	2	1G	1A	2G	2G	3G	3T	4C	4C	5G	5G	6C	6C	7G	7A	8C	8T	9A	9A
115	M	27	3	3	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8C	8C	9G	9G
116	F	25	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8C	9G	9G
117	M	37	1	1	1G	1G	2G	2G	3G	3T	4G	4G	5G	5A	6C	6C	7A	7A	8T	8T	9A	9A
118	F	28	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8C	9G	9A

Sample #	Sex	Age	Eye Color Classification	Skin Color Classification	F1 1 rs12913832	F1 2	F2 1 rs1800407	F2 2	F3 1 rs12896399	F3 2	F4 1 rs16891982	F4 2	F5 1 rs1393350	F5 2	F6 1 rs12203592	F6 2	F7 1 rs1129038	F7 2	F8 1 rs2424984	F8 2	F9 1 rs1426654	F9 2
119	M	23	3	3	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8C	8T	9G	9G
120	M	28	3	2	1G	1A	2G	2G	3G	3T	4G	4C	5G	5G	6C	6C	7G	7A	8C	8T	9A	9A
121	F	26	3	2	1A	1A	2G	2G	3G	3T	4G	4C	5G	5A	6C	6T	7G	7G	8T	8T	9A	9A
122	M	33	3	2	1G	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
123	M	42	3	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
124	F	38	3	3	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9G	9A
125	M	66	2	2	1G	1A	2G	2G	3G	3T	4G	4C	5G	5A	6C	6T	7G	7A	8T	8T	9A	9A
126	F	49	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5A	6C	6C	7G	7G	8C	8T	9A	9A
127	M	34	2	2	1G	1A	2G	2G	3T	3T	4G	4C	5G	5A	6C	6C	7G	7A	8T	8T	9A	9A
128	F	30	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
129	M	19	2	2	1G	1G	2G	2G	3G	3T	4C	4C	5G	5A	6C	6C	7A	7A	8C	8T	9A	9A
130	F	29	3	2	1G	1A	2G	2G	3G	3G	4G	4G	5G	5A	6C	6T	7G	7A	8C	8T	9A	9A
131	M	28	2	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
132	F	27	3	2	1G	1A	2G	2G	3G	3G	4G	4C	5G	5A	6C	6C	7G	7A	8T	8T	9A	9A
133	M	21	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8T	9G	9G
134	F	16	3	3	1A	1A	2G	2G	3G	3T	4C	4C	5G	5G	6C	6C	7G	7G	8C	8C	9G	9A
135	M	44	2	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9G	9A
136	F	48	3	2	1G	1A	2G	2G	3G	3T	4G	4C	5G	5G	6C	6C	7G	7A	8C	8T	9A	9A
137	M	34	3	2	1A	1A	2G	2G	3G	3T	4C	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
138	F	29	3	2	1G	1A	2G	2G	3G	3T	4G	4C	5G	5A	6C	6C	7G	7A	8T	8T	9A	9A
139	M	35	2	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5A	6C	6C	7G	7A	8T	8T	9A	9A
140	F	31	3	2	1A	1A	2G	2G	3T	3T	4G	4G	5G	5A	6C	6T	7G	7G	8T	8T	9A	9A
141	M	28	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
142	F	28	2	2	1G	1A	2G	2A	3G	3G	4G	4G	5G	5G	6C	6T	7G	7A	8T	8T	9G	9A
143	M	34	2	2	1G	1A	2G	2G	3G	3T	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
144	M	26	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
145	M	35	3	3	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8C	8C	9G	9G
146	F	22	1	1	1G	1G	2G	2G	3G	3T	4G	4G	5G	5A	6C	6T	7A	7A	8C	8T	9A	9A
147	M	23	2	2	1G	1A	2G	2A	3G	3T	4G	4G	5G	5G	6C	6T	7G	7A	8T	8T	9A	9A
148	F	23	3	2	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8C	8T	9A	9A
149	M	21	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
150	F	22	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7G	7G	8C	8T	9A	9A
151	M	18	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6T	7G	7G	8T	8T	9A	9A
152	F	21	3	2	1A	1A	2G	2A	3T	3T	4G	4C	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
153	M	25	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7A	8C	8T	9A	9A
154	F	33	3	2	1G	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9G	9A
155	M	26	2	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
156	F	22	3	3	1A	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7G	8T	8T	9G	9A
157	M	18	3	2	1G	1A	2G	2G	3G	3T	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
158	F	17	2	2	1A	1A	2A	2A	3G	3T	4G	4C	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
159	M	36	2	2	1G	1A	2G	2A	3G	3T	4G	4G	5G	5A	6C	6C	7G	7A	8T	8T	9A	9A
160	F	29	3	2	1G	1A	2G	2G	3T	3T	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
161	M	32	2	2	1G	1A	2G	2G	3T	3T	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
162	F	23	3	2	1A	1A	2G	2A	3T	3T	4G	4C	5G	5A	6C	6C	7G	7G	8C	8T	9A	9A
163	M	18	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6T	7G	7G	8C	8T	9A	9A
164	F	18	3	2	1A	1A	2G	2A	3G	3T	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
165	M	41	2	2	1G	1A	2G	2A	3G	3T	4G	4G	5G	5G	6C	6T	7G	7A	8T	8T	9A	9A

166	F	42	2	2	1A	1A	2G	2A	3G	3T	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
168	M	26	2	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5A	6C	6T	7G	7G	8T	8T	9A	9A
169	F	32	3	3	1A	1A	2G	2G	3G	3G	4G	4C	5G	5A	6C	6T	7G	7G	8T	8T	9A	9A
170	M	29	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
171	F	28	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
172	F	33	3	2	1G	1A	2G	2G	3G	3G	4G	4C	5G	5A	6C	6C	7G	7A	8C	8T	9A	9A
173	M	37	3	3	1G	1A	2G	2G	3G	3G	4C	4C	5G	5G	6C	6C	7G	7A	8C	8T	9G	9A
174	M	40	2	3	1G	1A	2G	2A	3G	3T	4G	4C	5G	5A	6C	6C	7G	7A	8C	8T	9G	9A
175	F	37	3	3	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8C	8T	9G	9A
182	F	43	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5A	5A	6C	6C	7G	7G	8C	8T	9A	9A
183	F	33	2	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5A	6C	6C	7G	7A	8T	8T	9A	9A
184	F	28	2	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
185	F	25	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7G	7G	8C	8T	9A	9A
186	F	23	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5A	6C	6C	7G	7G	8T	8T	9A	9A
187	F	27	3	2	1A	1A	2G	2G	3G	3T	4G	4G	5G	5A	6C	6T	7G	7G	8C	8T	9A	9A
188	F	25	3	2	1A	1A	2G	2G	3T	3T	4C	4C	5A	5A	6C	6C	7G	7G	8C	8T	9A	9A
189	M	45	1	1	1G	1G	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7A	7A	8C	8T	9A	9A
190	M	35	2	2	1G	1A	2G	2G	3T	3T	4G	4G	5G	5G	6C	6C	7G	7A	8C	8T	9A	9A
191	F	25	3	2	1G	1A	2G	2G	3G	3T	4G	4C	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
192	F	50	2	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7A	7A	8T	8T	9A	9A
193	F	53	2	2	1G	1A	2G	2A	3G	3G	4C	4C	5G	5A	6C	6C	7G	7A	8T	8T	9A	9A
194	F	32	3	2	1A	1A	2G	2G	3G	3T	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
195	M	38	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
196	F	35	3	2	1A	1A	2G	2G	3G	3G	4G	4G	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A
197	F	38	3	2	1A	1A	2G	2G	3G	3G	4G	4C	5A	5A	6C	6C	7G	7G	8T	8T	9A	9A
198	F	39	2	2	1G	1A	2G	2G	3G	3T	4G	4G	5G	5G	6C	6C	7G	7A	8T	8T	9A	9A
199	F	23	3	2	1A	1A	2G	2G	3G	3T	4G	4C	5G	5A	6C	6C	7G	7G	8C	8C	9A	9A
200	F	38	3	2	1A	1A	2G	2A	3G	3G	4G	4C	5G	5G	6C	6C	7G	7G	8T	8T	9A	9A

Table S2 - SNP markers used in the IrisPlex plus skin color prediction assay, including multiplex genotyping details

SNP-ID	Chr position	Gene	PCR product (bp)	Forward PCR primer (5'-3') and Reverse PCR primer (5'-3')	Each primer concentration (mM)	SBE primers	Primer length no tail (bp)	Total primer length (bp)	Alleles detected	Primer direction	Concentration (mM)
rs12913832	15-26039213	HERC2	87	TGGCTCTCTGTGTCTGATCC GGCCCCTGATGATGATAGC	0.416	ttttttttttttttttttttttGCGTGCAGAAC-TTGACA	17	41	T/C	Reverse	0.2
rs1800407	15-25903913	OCA2	127	TGAAAGGCTGCCTCTGTTCT CGATGAGACAGAGCATGATGA	0.416	ttttttGCATACCGGCTCTCCC	16	23	G/A	Forward	0.2
rs12896399	14-91843416	SLC24A4	104	CTGGCGATCCAATTCTTTGT CTTAGCCCTGGGTCTTGATG	0.416	ttttttttttttttttttttttttttttttaTCTTTAGG-TCAGTATATTTTGGG	23	53	G/T	Forward	0.3
rs16891982	5-33987450	SLC45A2 (MATP)	128	TCCAAGTTGTGCTAGACCAGA CGAAAGAGGAGTCGAGGTTG	0.416	ttttttttttAAACACGGAGTTGATGCA	18	29	C/G	Forward	0.6
rs1393350	11-88650694	TYR	80	TTCCTCAGTCCCTTCTCTGC GGGAAGGTGAATGATAACACG	0.416	ttttttttttttttttttttaTTTGTAAAAGA-CCACACAGATTT	23	47	T/C	Reverse	0.1
rs12203592	6-341321	IRF4	115	ACAGGGCAGCTGATCTCTTC GCTAAACCTGGCACCAAAAG	0.416	ttttttttttttttaTTTGGTGGGTAAAAGAAGG	19	35	C/T	Forward	0.6
rs1129038	15-26030454	HERC2	156	GACGATTCAGATAACGAGGATGTC ACACCAGGCAGCCTACAGTCTA	0.416	ttttttttttttttttttttttttttttttttttttACAG-TCTACACAGCAGCGAG	20	58	C/T	Reverse	0.4
rs2424984	20-32314036	ASIP	112	CACTTAGGTCGAGGGACCAG AGCCAGGCACTGATAGCAAC	0.416	ttttttttttttttttttttttttttttttttttttttGGG-ACCAGGCCCCACAAGAG	20	60	C/T	Forward	0.2
rs1426654	15-46213776	SLC24A5	53	TTCAGCCCTTGGATTGTCTC CCGCTGCCATGAAAGTTG	0.416	ttttttttttttttttttttttttttttttttttttttttaT-GTCTCAGGATGTTGCAGGC	20	64	G/A	Forward	0.1

In preparation to be submitted

Lessons from the Museum

Forensic phenotyping of a mummified body – genetics contribution to a glimpse from the past

Paulo Dario[1,2,3,4], Ana Rita Oliveira[2,4], Manuela Marques[1], Teresa Ribeiro[1,3], Maria João Porto[1,3], Deodália Dias[1,4], Jorge Costa Santos[1,3,5], Francisco Corte Real[1,3,6]

(1) Instituto Nacional de Medicina Legal e Ciências Forenses, I.P., Portugal

(2) Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal

(3) CENCIFOR – Centro de Ciências Forenses, Coimbra, Portugal

(4) CESAM – Centro de Estudos do Ambiente e do Mar, Lisboa, Portugal

(5) Faculdade de Medicina da Universidade de Lisboa, Lisboa, Portugal

(6) Faculdade de Medicina da Universidade de Coimbra, Coimbra, Portugal

Paulo Dario

Email: paulo.a.dario@inmlcf.mj.pt

In preparation to be submitted

Lessons from the Museum

Forensic phenotyping of a mummified body – genetics contribution to a glimpse from the past

Abstract

The museum of the South Branch of the Portuguese National Institute of Legal Medicine and Forensic Sciences holds a mummified corpse of a female individual in its collection, of which very little is known. In order to collect more information about the individual to whom this remains belong, a DNA investigation was initiated. A muscle fragment from the knee joint region was collected and DNA was extracted using QIAamp DNA Mini Kit and quantified with Quantifiler® Duo. We analyzed phenotypic IrisPlex SNP *loci* for determination of iris color rs12913832 (HERC2), rs1800407 (OCA2), rs12896399 (SLC24A4), rs16891982 (SLC45A2), rs1393350 (TYR) and rs12203592 (IRF4) along with three other *loci* rs1129038 (HERC2), rs2424984 (ASIP) and rs1426654 (SLC24A5) also involved in skin color determination. Based on obtained results, it was found that the deceased remains belonged to an individual that probably had brown eyes and white skin. In this work it was found that it is possible to obtain information about the phenotype of a given sample donor, even if the sample is relatively old and eventually degraded. Although the presented results have more historical and scientific interest than a forensic one, this study aims to illustrate that phenotypic markers may serve not only as a tool in criminal investigation but also as an instrument in the anthropology field. Concluding, phenotypic markers may provide additional information when is necessary to study cadaveric remains of forensic or historical interest.

Keywords: Phenotypic investigation; Ancient DNA; Archeo-anthropology.

Introduction

The South Branch of the Portuguese National Institute of Legal Medicine and Forensic Sciences (INMLCF) possesses a museological collection under study, from which a part is patent in a small museum space open to medical students, researchers, academics and to the public in general upon request. This collection possesses, for example, one of the more impressive collection of preserved tattoos from last century 10's to 40's or a collection of busts of individuals who died from asphyxiation, most of them by hanging. However some of the specimens of the collection are not cataloged and their provenience is not well known and, in the recent years, there has been an attempt to restore information about these pieces.

One of the most interesting pieces is a complete female mummified body that very few is known about it except that would be from last century 30's or 40's. Because of this, many stories have come out during the years… some connected to the intense spy activity that occurred during the WWII or the years that preceded it. One of the most interesting stories was that this was the body of a German spy that deceased in a hotel room without no one noticing it… or not wanting to notice it, during a long period, long enough to provide the conditions for the mummification of the cadaver. However this was only a forensic "urban myth" and the truth remained undiscovered.

SNP phenotypic markers are being studied for *loci* such as the ones responsible for human eye, hair and skin color leading to the development of phenotype informative SNP multiplexes [1–3], which some have been tested in different populations such as the Slovenian [4] or the U.S. population [5] but also in the Portuguese one [6]. These may provide information for identification of phenotypic traits from samples of unknown donors found in criminal casework where there are no suspects [2,7,8] or in the identification of missing persons or human skeletal remains [9,10]. This kind of information may be even more relevant when the sample donor possesses phenotypic characteristics which distinguish him from the population in which he is inserted.

In this work, a DNA sample from a mummified corpse with historical and scientific interest was studied in order to discover more information about the person which this corps belonged to. In this investigation an IrisPlex [1] based multiplex designed for eye and skin color prediction, which was previously applied to study Portuguese population [10] was

used in order to infer about the eye and skin color of the mummified corpse. This study demonstrates how these methodologies of phenotypic investigation may be useful, not only in forensic investigation but also in similar areas of science such as in archeo-anthropology.

Specimen description

At the observation of the mummy it was observed that the specimen consisted of a complete human individual body, presumably a female. It was in a dry, odorless state and presented a pale yellow-brownish-grey color (Fig. 1). The tissue and skin were very dehydrated and hard and the left lower part of the leg was separated from the rest of the body at the knee joint articulation due to this rigidity but in the overall the specimen was in a well-preserved condition and presented no apparent injuries.

Fig. 1 - Photographic overview of the specimen. A) Lateral view of the mummified body. B) Aerial ventral view of the mummified body. The left lower part of the leg separation at the knee joint articulation can be better visible at this angle.

At specimen observation a small folded paper was discovered between the 3rd and 4th fingers from the left hand. The paper was also very dehydrated and was presumably put in the location right after individuals' obit. This paper appeared to be an identification card from Miguel Bombarda Mental Hospital, an old Mental Hospital Located in Lisbon, and to be dated from the 30's (Fig. 2).

Fig. 2 - Miguel Bombarda Mental Hospital card discovered between the 3rd and 4th fingers from the left hand. Presumably this was put in place after individuals' obit.

Genetic and phenotypic analysis

For genetic analysis, taking part of the previous leg separation and with the objective of maintaining body preservation, a muscle fragment from the inner part of the knee joint region was collected from the mummified corpse (Figs. 1B and 3).

The DNA from this tissue was extracted using QIAamp DNA Mini kit (Qiagen) and quantified with Quantifiler® Duo in Applied Biosystems 7500 Real-Time PCR System. SNPs were genotyped with a multiplex that combines IrisPlex *loci* [1] (HERC2 rs12913832, OCA2 rs1800407, SLC24A4 rs12896399, SLC45A2 rs16891982, TYR rs1393350, and IRF4

rs12203592) along with three other *loci* for determination of skin color (HERC2 rs1129038, ASIP rs2424984 and SLC24A5 rs1426654) and that was previously studied in Portuguese population [6]. PCR amplification was done in Applied Biosystems GeneAmp® PCR System 9700 thermocycler and subsequent capillary electrophoresis and analysis performed in Applied Biosystems 3130 system. Eye color estimation was based on the model of Liu et al. [12].

Fig. 3 – Collection of soft tissue sample from the inner part of the knee joint for genetic analysis. This procedure was made taking part of the previous leg separation to maintain piece integrity.

Results and Discussion

Based on information found on the corpse, the Miguel Bombarda Mental Hospital card discovered between the 3rd and 4th fingers from the left hand, it is believed that this individual was deceased in the years 30-40 of the last century. This is consistent with the corpse being from a pre or even beginning of World War II period in that, according to documents of that time, the Institute of Legal Medicine of Lisbon (now the South Branch of the National Institute) was taking procedures for the acquisition of mortuary cold

chambers and ending some parts of the building that only finished in the end of the 30's. This can explain that in such period, a troubled and difficult period in Europe and in Portugal, a corpse coming from a mental hospital could be left in such conditions that could lead to its mummification without no one reclaiming the body.

Based on the phenotyping profile obtained from the mummy's sample (Electrophoretogram (EPG) presented in Fig. 4 - 1G/A, 2C, 3G/T, 4G, 5G, 6C, 7C, 8C/T, 9A), and based on Liu et al. model for IrisPlex system (*loci* 1 to 6) [12], it was estimated that the individual would have a 71% probability of having brown eyes (Table 1). This model used for this estimate has an accuracy of 92.5% for the class of brown eyes.

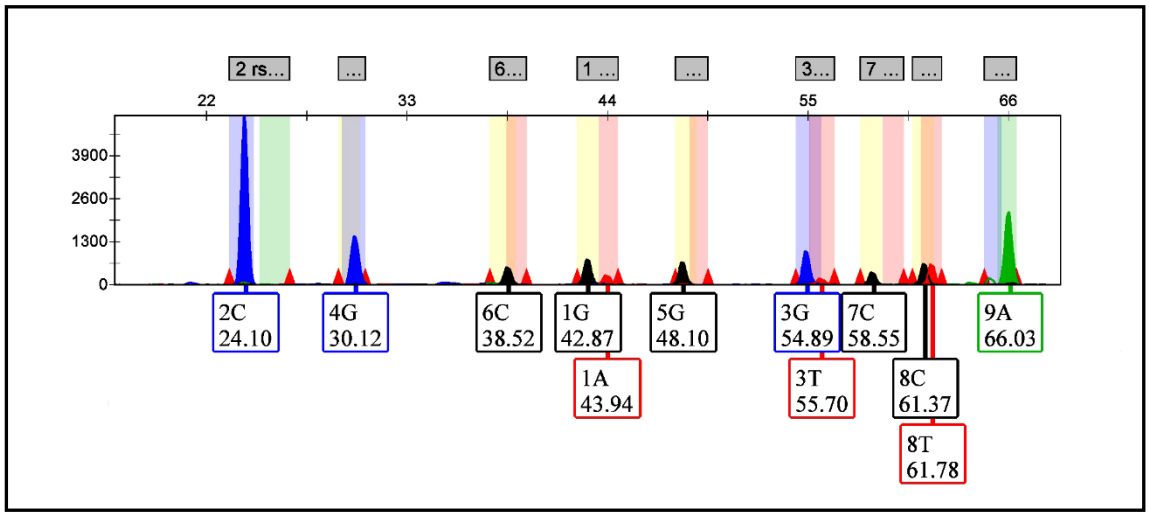


Fig. 4 – Phenotyping EPG of mummy's sample.

Using for comparison a female individual with the same IrisPlex profile as the one genotyped on the mummy´s sample (Fig. 5, with estimation probabilities indicated in Table 1), we perceive that the mummified individual may have had brown eyes similar to the ones found on that women (Fig. 6).

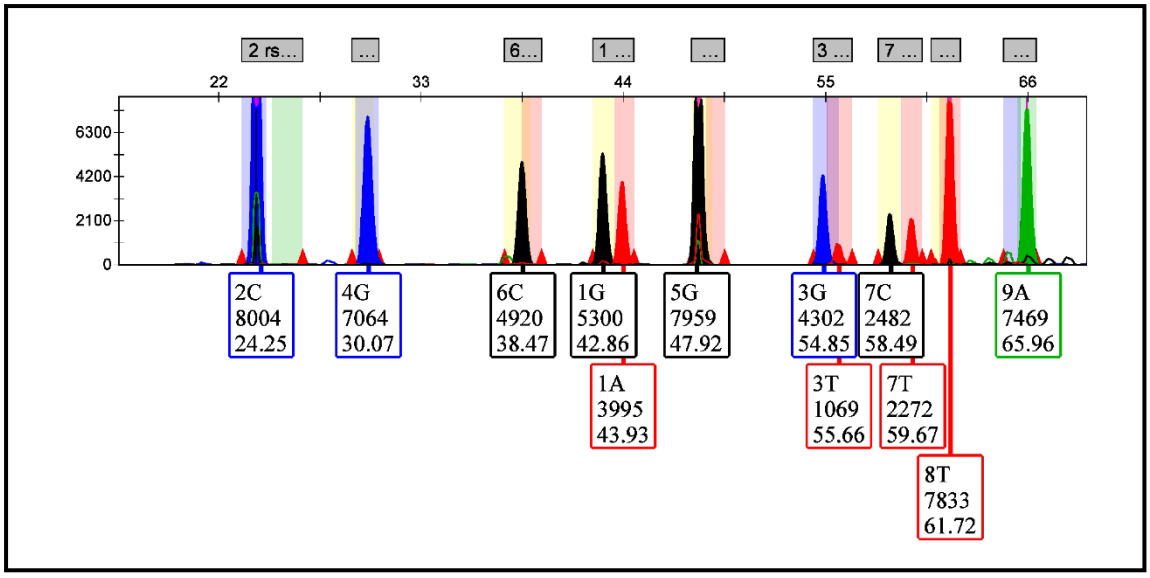


Fig. 5 – Phenotyping EPG of female individual presenting same IrisPlex phenotype than mummy's sample (first 6 *loci*).

Table 1 – Probabilities for eye color of the referred individuals. Mummified corpse and the female individual used for exemplification.

	P. blue	P. intermediate	P. brown
Mummified corpse (♀)	15 %	14 %	71 %
Female sample (♀)	15 %	14 %	71 %

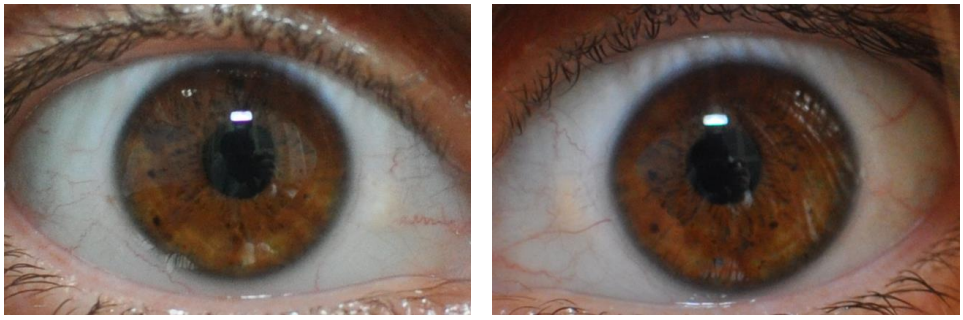


Fig. 6 – Eyes from female individual having same IrisPlex phenotype than mummy's sample.

This sample belongs to the female with the most similar phenotype for all *loci* found in a recent study for the Portuguese population, differing in rs1129038 for having C/T instead of homozygous C and in rs2424984 for having homozygous T instead of C/T. This female had white skin, as presented in Fig. 7.

Fig. 7 – Skin color from female individual having the more similar genotype to the one obtained from mummy's sample.

However, in this same study it was found an individual having a phenotype even more similar to the one found in mummy's sample, differing only in rs1129038 for having C/T instead of homozygous C (Fig. 8). This was a male individual whose eyes and skin color are presented in Fig. 9.

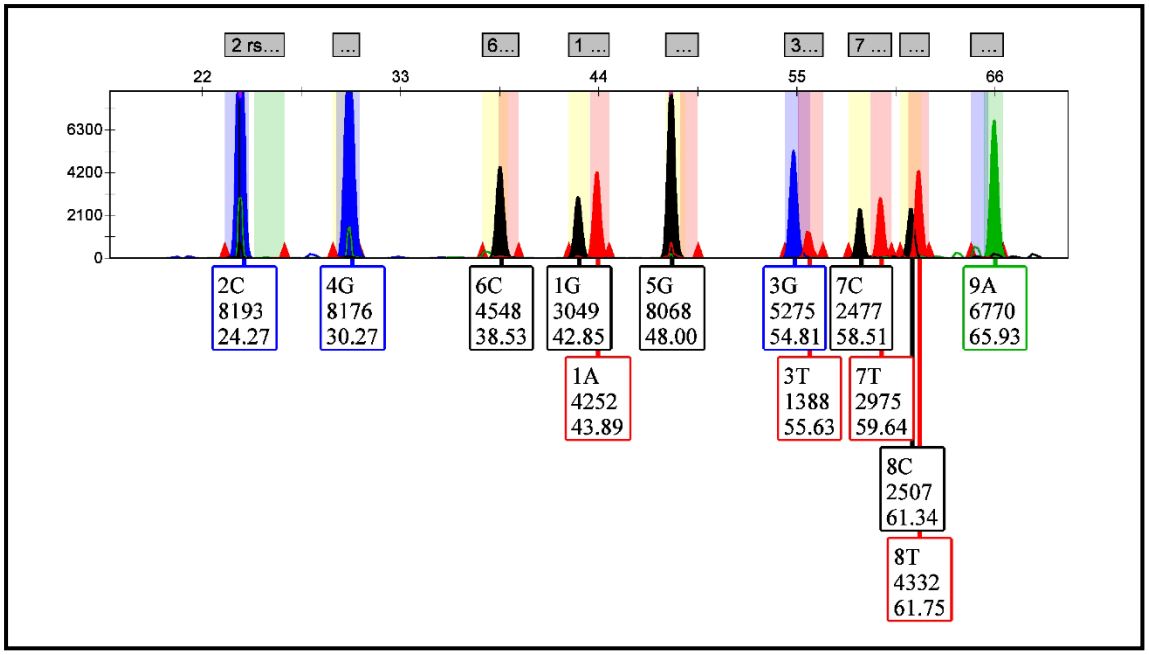


Fig. 8– Phenotyping EPG of the individual having the more similar phenotype to the one found in than mummy's sample (first 6 *loci*).

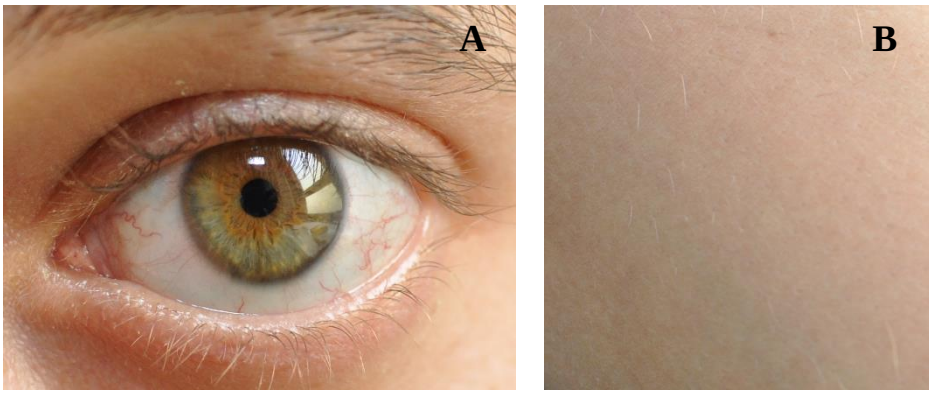


Fig. 9 – Eye (A) and skin (B) color from male individual having the more similar genotype to the one obtained from mummy's sample.

As can be seen in the image from the eye of the male individual having the same IrisPlex *loci*, this presents an intermediate phenotype of green/brown eye color that can be considered as belonging to the not blue and not brown eye color category. This difference in the color of the eyes of the male individual can be explained by the effect found by

Martinez-Cadenas et al. for IrisPlex, of the bias of males tending to have lighter eye colors than females [13] what leads to consider that the eyes of the female individual as more prone to being representative of the mummy's eyes instead of the eyes of the male individual.

These results indicate that the female individual to which the mummified corpse belonged to should have had brown eyes and white skin, probably whiter as the individuals compared in this study… probably not a German spy or at least probably not German.

Conclusions

Genetic and phenotypic information about a mummified individual was obtained, presumably a corpse from years 1930-1940. From the sample, which was relatively old and that was in an altered state - mummification, was possible to infer about their eye and skin colors. Although the presented results have more historical and scientific interest than a forensic one, this study aims to illustrate that phenotypic markers may serve not only as a tool in criminal investigation but also as an instrument in the anthropology field. Concluding, phenotypic markers may provide additional information for the study of cadaveric remains of forensic or historical interest when necessary. This demonstrates the increasing importance of genetics in the study of anthropology or archeo-antropology.

References

1. Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M, et al. IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci. Int. Genet. 2011; 5:170–80.

2. Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A, et al. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci. Int. Genet. 2013; 7:98–115.

3. Hart KL, Kimura SL, Mushailov V, Budimlija ZM, Prinz M, Wurmbach E. Improved eye- and skin-color prediction based on 8 SNPs. Croat. Med. J. 2013; 54:248–56.

4. Kastelic V, Pospiech E, Draus-Barini J, Branicki W, Drobnic K. Prediction of eye color in the Slovenian population using the IrisPlex SNPs. Croat Med J. 2013; 54:381–6.

5. Dembinski GM, Picard CJ. Evaluation of the IrisPlex DNA-based eye color prediction assay in a United States population. Forensic Sci. Int. Genet. 2014; 9:111–7.

6. Dario P, Mouriño H, Oliveira AR, Lucas I, Ribeiro T, Porto MJ, et al. Assessment of IrisPlex-based multiplex for eye and skin color prediction with application to a Portuguese population. Int. J. Legal Med. 2015; 129:1191–200.

7. Poetsch M, Blohm R, Harder M, Inoue H, von Wurmb-Schwark N, Freitag-Wolf S. Prediction of people's origin from degraded DNA - presentation of SNP assays and calculation of probability. Int J Leg. Med. 2013; 127:347–57.

8. Keating B, Bansal AT, Walsh S, Millman J, Newman J, Kidd K, et al. First all-in-one diagnostic tool for DNA intelligence: genome-wide inference of biogeographic ancestry, appearance, relatedness, and sex with the Identitas v1 Forensic Chip. Int J Leg. Med. 2013; 127:559–72.

9. Bouakaze C, Keyser C, Crubezy E, Montagnon D, Ludes B, Crubézy E. Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis. Int J Leg. Med. 2009; 123:315–25.

10. Draus-Barini J, Walsh S, Pośpiech E, Kupiec T, Głąb H, Branicki W, et al. Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. Investig. Genet. 2013; 4:3.

11. Pereira L, Alshamali F, Andreassen R, Ballard R, Chantratita W, Cho N, et al. PopAffiliator: online calculator for individual affiliation to a major population group based on 17 autosomal short tandem repeat genotype profile. Int. J. Legal Med. 2011; 125:629–36.

12. Liu F, van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, Janssens a CJW, et al. Eye color and the prediction of complex phenotypes from genotypes. Curr. Biol. 2009; 19:R192-3.

13. Martinez-Cadenas C, Peña-Chilet M, Ibarrola-Villava M, Ribas G. Gender is a major factor explaining discrepancies in eye colour prediction based on HERC2/OCA2 genotype and the IrisPlex model. Forensic Sci. Int. Genet. 2013; 7:453–60.

Chapter VI

General Discussion

This chapter presents an integrative discussion of the results obtained in the papers that compose Chapters II to V of this thesis.

SNPs are important tools in several areas of biomedical sciences and are gaining a prominent role in forensic genetics. In forensics, these genetic markers besides providing genetic information that can be used for (i) individual identification or for the (ii) investigation of familial lineages similarly to the current STR *loci* used in routine investigation they can provide information about the (iii) ancestry and admixture of an individual or even about its (iv) physical traits such as eye color or skin color when nothing else is known except for the fact that the individual left behind a sample on a specific place that is under investigation. Because of these characteristics these markers can be classified as (i) IISNPs, (ii) LISNPs, (iii) AISNPs and (iv) PISNPs.

Nevertheless, despite the intense study on the use of SNPs for forensic identification, as exemplified by the creation of the SNP*for*ID consortium, STRs should continue to be the most used genetic markers in forensic genetics, given that they have high polymorphism, which makes them very useful for the analysis of mixtures, but mainly because the data obtained by analysis of SNPs cannot be compared with genetic profiles of STRs which are the basis of the existing forensic DNA databases.

However, even if the profile analysis of SNPs is unlikely to play the main role in forensics in the future it will certainly have an important part. Compared to STRs, SNPs are much more abundant in the human genome and although having a smaller number of alleles because they are normally biallelic systems, the combination of being abundant and biallelic allows for these to be investigated with a higher degree of automation and a more standardized and facilitated analysis. Thus, the reduced number of alleles is counterbalanced by examining a larger number of *loci* that, on the whole, ends up having a power of analysis equals to or higher than the ones obtained through the study of a relatively small number of STRs. This can be particularly important if we consider that SNPs can be analyzed through a great number of techniques such as through the use of big PCR multiplexes analyzed by capillary electrophoresis, by massive analysis techniques

such as MALDI-TOF or as DNA microarrays but especially by the recent methods of MPS systems. However, the more important is that SNPs can be used for niche applications for which the STRs are unsuitable like the analysis of heavily degraded samples, for phenotyping or even for the determination of ancestry and admixture.

Now it's time to discuss the results obtained during this PhD thesis and this is done analyzing the three types of SNPs studied in this project: IISNPs, LISNPs and PISNPs.

IISNPs

IISNP markers can be considered as important tools for analyzing highly degraded forensic samples, and for increasing the power of both kinship analyses and the identification of human remains where family reconstructions for missing and unidentified persons are required. They can also be particularly useful for providing additional investigative information in cases where there are no suspect(s) and/or no STR matches in the DNA databases. Forensic genetics applies principles of population genetics to validate its methodologies. Samples have to be considered within the context of the population from which the sample's donor(s) belong to. In this way, it is necessary for forensic laboratories to have access to data regarding the population in question, or to study those populations themselves in order to better evaluate the forensic samples from any given case.

Initially, one of the aims of this PhD project was to develop, optimize and implement a small IISNPs multiplex based on SNP*for*ID 52-plex for use on more complex cases but that could also be used in parentage testing in cases related to South Portuguese population and immigrant populations residing in this area. This way and using SNP*for*ID browser (Amigo *et al.*, 2008), two 10-plex were designed to analyze a total of 20 SNPs selected from the SNP*for*ID 52-plex (Sanchez *et al.*, 2006), choosing the most polymorphic *loci* in the nearby populations but also taking care to ensure that chosen *loci* were also polymorphic in African ones, due to its possible application to some of the most representative immigrant populations resident in Portugal due to historical reasons, namely Cape-Verde, Angola and Guinea-Bissau.

This multiplex was internally validated for forensic use (Dario *et al.*, 2012) following the SWGDAM guidelines (Scientific Working Group on DNA analysis methods, revised validation guidelines, 2004) and the results of this validation confirmed that the assay presented correct genotypes of known and non-probative evidence samples, with good reproducibility and match criteria, although sometimes with the quality being affected by room temperature, probably due to the small SNP amplicons. This small multiplex revealed to be human specific; with no cross contamination just as current procedures for STR analysis when strictly followed.

This application of this 20-plex was evaluated as a complement to standard STR methodologies in paternal testing (Dario *et al.*, 2009b), in kinship casework (Dario *et al.*, 2011b) and on familial complex casework analysis (Dario *et al.*, 2011a). Results obtained with SNP methodology were consistent with STR ones, except for one paternity case where the alleged father could not be excluded, probably due to the small number of SNP *loci* analyzed. Nevertheless, a similar situation like this has also been referred in the past for STRs which are much more polymorphic than SNPs (González-Andrade *et al.*, 2009). However, when these 20 SNPs were analyzed together with one standard commercial STR kit, likelihood ratios obtained were always equivalent or higher than those obtained with a combination of two overlapping commercial STR kits.

These studies demonstrated the effectiveness of these 20 IISNPs multiplex in familial testing, from paternal testing to complex casework analysis. Besides being a very fast technique, which enables a better and a quick response from the Forensic Laboratory, this 20 SNPs methodology is relatively simple and economical and can be easily used in the resolution of this kind of casework. This also demonstrated to be a very sensitive method and, although providing a relatively low PE, demonstrated to be very informative specially when analyzed together with routine standard STR *loci*.

Later on, and related to the necessity of studying a greater number of SNPs it became aware that would be important to expand the aim of this PhD thesis to the study of the complete SNP*for*ID multiplex. This way, the 52-plex population and forensic parameters

of the population of the South of Portugal and also from a Guinea-Bissauan immigrant population resident in this area were evaluated (Dario *et al.*, 2017). The results obtained were in line with others reported on similar studies, performed on different populations previously studied with this multiplex, namely: the Northwest region of Spain and Northern Portugal when compared to the South of Portugal, and Senegal when compared to the studied Guinea-Bissauan population.

Besides the population and forensic evaluation of the SNP*for*ID 52-plex in the South of Portugal population, this thesis had the objective of studying the suitability of the application of this multiplex to complement the standard STRs methodologies in the analysis of real case samples consisting of degraded bone, teeth and also other different challenging samples related to criminal and individual identification casework. Taking this in consideration, in 2015 (Dario *et al.*, 2015b) it was compared the amplification success of 53 casework challenging samples using the 52-plex and the Identifiler® Plus® kit and it was obtained a mean amplification success rate of 73% for the 52-plex and of 43% for the Identifiler®. More, in 16 out of 36 samples in which profiles were not obtained with STR analysis or in which these were poor it was possible to obtain a complete or almost complete SNPs profile.

These studies provided the population and forensic data necessary for forensic scientists to apply the SNP*for*ID 52-plex in criminal and identification cases occurring in the area of the South of Portugal and demonstrated its utility as a complement or as an alternative to STR typing in challenging samples analysis where STR amplification fails.

LISNPs

Often there is the necessity of using mtDNA analysis in forensic examinations, especially in criminal investigations. This is particularly true when the only evidence found consists on hair shafts or telogenic hairs, often subjected to harsh environments, because in these cases it is very difficult to recover DNA. However there is an exception, which is mtDNA. Still, mtDNA analysis is a time consuming and expensive technique that cannot be used to

identify a sample donor. This serves to merely verify if an individual presents the same haplotype found on a sample, making it possible for him or her to be the sample donor, or if it present a different haplotype what excludes him or her from being the one responsible for leaving that sample.

For mtDNA analysis, DNA is sequenced in the hypervariable regions in order to uncover variations relatively to a consensus mitochondrial sequence named rCRS or revised Cambridge Reference Sequence. After sequencing these regions, it is necessary to identify which base pairs may present variants relatively to the consensus region what can range from nil to more than a dozen alterations. However, due to human population evolution some of these variations were established in an earlier time and others were acquired more recently, but more important they were established during times of human migration establishing lineages characteristic of different populations. This way some of these variations that are more linked to the population from which the individual is related to form the haplogroups and the other variations that are more related to maternal lineage are defined as haplotypes, resembling the hours and minutes pointers from a clock of human evolution. Yet, being this a difficult technique and with the quantity of samples that can enter a forensic lab in just a single criminal case, becomes important to have a technique which can be used to firstly identify which samples can present a particular haplogroup and then evaluate the best strategy for sequencing and analyzing the evidence samples brought to the forensic laboratory.

With that purpose in mind and based on previously mtDNA SNP studies developed by other authors, a 10-Plex mitochondrial SNP for haplogroup typing was published by Dario and coworkers (Dario *et al.*, 2009a) that could be applied to the study of the most common populations in our laboratory circumscription and applied to our forensic casework. Hair shafts from three forensic cases with different ethnic backgrounds were studied by mtDNA sequencing and compared with mitochondrial SNPs analysis using those 10 LISNPs.

The analysis of these cases demonstrated that mtDNA LISNPs typing analysis prior to mtDNA sequencing can allow for a rapid screening in forensic casework. This technique

also has the additional advantage of providing results when mtDNA sequencing fails. This 10 SNP *loci* multiplex provides a less expensive and rather simpler method for mitochondrial typing compared to control region mtDNA sequencing, especially when used as a fast screening method.

PISNPs

Relatively to PISNPs its use may be very important in the near future for the discovery of the truth in forensic cases, through the estimation of EVCs. This strategy could serve as intelligence to the police forces in a manner similar to the visual determination, which is still the basis for eyewitness testimonies used in the field of criminal investigation. PISNPs based EVC prediction could serve to corroborate eyewitness testimonies and to inform investigators when eyewitnesses do not exist.

Nevertheless, EVCs such as eye and skin color are difficult to classify due to the small difference between adjacent categories and are difficult to investigate due to their multigenic nature that is frequently subjected to the influence of environment. Good examples of PISNPs based EVC prediction investigation were the development of IrisPlex and later of HirisPlex which can be used to infer about the characteristics of a sample donor: eye and eye / hair color, respectively. Nevertheless, these multiplexes are subjected to some variations when used on different populations besides the ones originally used for its development and need to be studied / validated in the pretended populations in order to be used on them (Kastelic *et al.*, 2013; Yun *et al.*, 2014; Dembinski and Picard, 2014).

During this PhD thesis work and taking in consideration that the genes involved in eye color determination are also involved in other “color” characteristics, such as hair color (Sulem *et al.*, 2007), a multiplex was developed for eye and skin color determination through the optimization of IrisPlex supplementing this assay with three additional SNPs. This strategy was used to obtain a simultaneous test for the determination of both eye and skin color EVCs and at the same time validate its use in Portuguese population (Dario *et al.*, 2015a).

In this study, solid results were obtained for the prediction of eye color in the Portuguese population. This trait is the result of a relatively small number of genes and with little influence of environment but very informative about the appearance of an individual. Using a multinomial logistic regression model, prediction accuracies for eye color determination of 60% and 90% were obtained at the 0.5 threshold for blue and brown eye colors, respectively which is similar to the results obtained by Walsh and coworkers (Walsh *et al.*, 2011), although these authors applied a 0.7 threshold. More, the AUCs also performed similarly: 0.94, 0.84, and 0.87 (0.5 threshold) compared with 0.97, 0.84, and 0.95 (0.7 threshold) for blue, intermediate, and brown, suggesting that the ability of the estimated model to make accurate predictions is only slightly reduced when using the 0.5 threshold. Another important factor is that for the data sets under consideration, TYR rs1393350 and IRF4 rs12203592 do not seemed to play such an important role in the differentiation of brown and intermediate eye color as suggested by Walsh and coworkers (Walsh *et al.*, 2011) because they were not statistically significant. This conclusion may reflect the existence of some admixing just as reported by Dembinski and Picard for a US population (Dembinski and Picard, 2014). On the other hand, this could also be a consequence of using a smaller sample size than Walsh and coworkers (Walsh *et al.*, 2011) but that is in line with similar studies presented by other authors. Nevertheless, the AUCs clearly confirm that the estimated model is well fitted to the Portuguese population, a conclusion that is verified by a correct eye color prediction rate of 72%, with 17% inconclusive and only 11% incorrect. Additionally, the results from cross-validation have also showed that the model developed in this thesis is very accurate in predicting eye color for future observations. More precisely, the AUCs based on the cross-validation methodology were similar.

Concerning skin color prediction, the modeling procedure began with multinomial logistic regression so as to conduct a similar analysis to eye color. The estimated model did not give accurate results. The reason for this is that Portugal is the most southwestern region of Europe, and as a result, it is difficult to encounter individuals of the types I and II Fitzpatrick phototype that are much more commonplace in northern Europe. For the

sample under analysis, the number of individuals who belonged to those two categories was significantly reduced. Consequently, in order to satisfactorily carry out the statistical analysis, these individuals were amalgamated into the same category as types III and IV Fitzpatrick phototypes. This category mainly corresponds to Mediterranean individuals. Unfortunately, this procedure prevented predictions between phototypes of classes I/II, III/IV, and V/VI as was initially desired in the design of the experimental work of this thesis. However, it was possible to differentiate between class V/VI (dark brown skin) and the two others (white skins). If a larger sample was used, the multinomial logistic regression model would lead to more accurate results. In this case, it would be possible to develop a model to predict all skin types. Despite that, the statistical analysis was done with a binomial logistic regression model and led to a prediction accuracy of 93% to differentiate between Fitzpatrick phototypes classes V-VI (dark brown skin) and the Fitzpatrick phototypes classes I-IV (white skins). The AUC value was 0.99, which gives an excellent discrimination between whiter and darker skin colors. This model proved its value with a 95% correct prediction, which can be considered a very good estimation proving to be highly adequate for predicting skin color for future individuals.

The genetic information about the phenotypic characteristics of a mummified individual present in the museum space of the INMLCF was also obtained during this investigation about the genetic information necessary for phenotyping the eye and skin color in individuals from the Portuguese population. This mummy is presumably a corpse from years 1930-1940, and from a sample tissue taken from this piece was possible to infer about the eyes and skin colors of that individual (Dario *et al.*, in preparation). Phenotypic markers may provide additional information for the study of cadaveric remains of forensic or historical interest as demonstrated in this study as also by other authors (Bouakaze *et al.*, 2009; Draus-Barini *et al.*, 2013). This study illustrated that these may serve as a tool in criminal investigation and in connected fields such as anthropology or archeo-antropology.

References

Amigo, J., Phillips, C., Lareu, M. and Carracedo, A. (2008) The SNPforID browser: an online tool for query and display of frequency data from the SNPforID project. *International Journal of Legal Medicine*, 122(5), pp. 435–40. doi: 10.1007/s00414-008-0233-7.

Bouakaze, C., Keyser, C., Crubezy, E., Montagnon, D., Ludes, B. and Crubézy, E. (2009) Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis. *International Journal of Legal Medicine*, 123(4), pp. 315–325. doi: 10.1007/s00414-009-0348-5.

Dario, P., Bom, J., Ribeiro, T. and Geada, H. (2009a) MtSNP typing before mtDNA sequencing: Why do it? *Forensic Science International: Genetics Supplement Series*, 2(1), pp. 187–188. doi: 10.1016/j.fsigss.2009.08.137.

Dario, P., Mouriño, H., Oliveira, A. R., Lucas, I., Ribeiro, T., Porto, M. J., Costa Santos, J., Dias, D. and Corte Real, F. (2015a) Assessment of IrisPlex-based multiplex for eye and skin color prediction with application to a Portuguese population. *International Journal of Legal Medicine*, 129(6), pp. 1191–1200. doi: 10.1007/s00414-015-1248-5.

Dario, P., Oliveira, A. R., Marques, M., Ribeiro, T., Porto, M. J., Dias, D., Costa Santos, J. and Corte Real, F. (in preparation) Forensic phenotyping of a mummified body – genetics contribution to a glimpse from the past. *To be submitted*.

Dario, P., Oliveira, A. R., Ribeiro, T., Porto, M. J., Dias, D. and Corte Real, F. (2017) *Autosomal SNPs study of a population sample from Southern Portugal and from a sample of immigrants from Guinea-Bissau residing in Portugal*, *Legal Medicine*. doi: 10.1016/j.legalmed.2016.11.004.

Dario, P., Oliveira, A. R., Ribeiro, T., Porto, M. J., Santos, J. C., Dias, D. and Corte Real, F. (2015b) SNP*for*ID 52-plex in casework samples: “Cracking” bones and other difficult samples. *Forensic Science International: Genetics Supplement Series*, 5, pp. 118–120. doi: 10.1016/j.fsigss.2015.09.048.

Dario, P., Ribeiro, T., Dias, D., Corte-Real, F. and Geada, H. (2011a) Complex casework using single nucleotide polymorphisms. *Forensic Science International: Genetics Supplement Series*, 3(1), pp. e379–e380. doi: 10.1016/j.fsigss.2011.09.051.

Dario, P., Ribeiro, T., Espinheira, R., Dias, D., Geada, H. and Corte-Real, F. (2011b) 20 SNPs as supplementary markers in kinship testing. *Forensic Science International: Genetics Supplement Series*, 3(1), pp. e508–e509. doi: 10.1016/j.fsigss.2011.09.111.

Dario, P., Ribeiro, T., Espinheira, R. and Geada, H. (2009b) SNPs in paternity investigation: The simple future. *Forensic Science International: Genetics Supplement Series*, 2(1), pp. 127–128. doi: 10.1016/j.fsigss.2009.08.136.

Dario, P., Ribeiro, T., Oliveira, A.R., Dias, D., Geada, H., Corte Real, F. and Costa Santos, J. (2012) Internal Validation of 20 SNP Multiplex for Forensic Genetics. International Journal of Legal Medicine, 126(S1), pp. 357. doi:10.1007/s00414-012-0711-9 (Only abstract published).

Dembinski, G. M. and Picard, C. J. (2014) Evaluation of the IrisPlex DNA-based eye color prediction assay in a United States population. *Forensic Science International: Genetics*, 9(0), pp. 111–117. doi: 10.1016/j.fsigen.2013.12.003.

Draus-Barini, J., Walsh, S., Pośpiech, E., Kupiec, T., Głąb, H., Branicki, W. and Kayser, M. (2013) Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. *Investigative genetics*, 4(1), p. 3. doi: 10.1186/2041-2223-4-3.

González-Andrade, F., Sánchez, D., Penacino, G. and Martínez Jarreta, B. (2009) Two fathers for the same child: a deficient paternity case of false inclusion with autosomic STRs. *Forensic Science International: Genetics*, 3(2), pp. 138–40. doi: 10.1016/j.fsigen.2008.09.008.

Kastelic, V., Pospiech, E., Draus-Barini, J., Branicki, W. and Drobnic, K. (2013) Prediction of eye color in the Slovenian population using the IrisPlex SNPs. *Croatian Medical Journal*, 54(4), pp. 381–386. doi: 10.3325/cmj.2013.54.381.

Sanchez, J. J., Phillips, C., Borsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P. M., Carracedo, A., Morling, N. and Børsting, C. (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis*, 27(9), pp. 1713–1724. doi: 10.1002/elps.200500671.

Scientific Working Group on DNA analysis methods, revised validation guidelines (2004) *Forensic Science Communications*, 6(3).

Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Magnusson, K. P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G., Jakobsdottir, M., Steinberg, S., Pálsson, S., Jonasson, F., Sigurgeirsson, B., Thorisdottir, K., Ragnarsson, R., Benediktsdottir, K. R., Aben, K. K., Kiemeney, L. a, Olafsson, J. H., Gulcher, J., Kong, A., Thorsteinsdottir, U., Stefansson, K. and Palsson, S. (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature Genetics*, 39(12), pp. 1443–1452. doi: 10.1038/ng.2007.13.

Walsh, S., Liu, F., Ballantyne, K. N., van Oven, M., Lao, O., Kayser, M. and Oven, M. Van (2011) IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics*, 5(3), pp. 170–80. doi: 10.1016/j.fsigen.2010.02.004.

Yun, L., Gu, Y., Rajeevan, H. and Kidd, K. K. (2014) Application of six IrisPlex SNPs and comparison of two eye color prediction systems in diverse Eurasia populations. *International Journal of Legal Medicine*, 128(3):447-53. doi: 10.1007/s00414-013-0953-1.

Conclusions

At the end of this PhD thesis it is possible to conclude that:

- A small IISNPs multiplex, like the 20-plex investigated in this thesis, can be used with success in familial testing, from paternal testing to complex casework analysis. Besides being a rapid technique this type of methodology is relatively simple but effective.

- In criminal and identification casework more IISNP *loci* need to be analyzed in order to provide insight on forensic casework where STR analysis fails. This PhD thesis work provided the SNP*for*ID 52-plex population and forensic data and demonstrated its utility in challenging samples in order to this methodology can be applied to South Portuguese population casuistic.

- In mtDNA analysis, LISNP typing before mtDNA control region sequencing provides a simple and economical method that can be used as a fast screening method.

- PISNPs can provide valuable information about phenotypic traits of a sample donor. During this PhD thesis work, a phenotypic multiplex was developed based on IrisPlex in order to predict not only the eye color but also the skin color of a sample donor. This multiplex proved to be useful to accurately predict eye and skin color of a Portuguese individual trough the analysis of a sample.

- The PISNP multiplex developed during this thesis demonstrated its sensitivity through the positive analysis of a much degraded sample of muscle of a mummified individual what evidenced that this type of analysis can also be used in areas such as archeo-anthropology.

Perspectives

A set of technologies that allows for the identification of individuals have emerged and will continue to emerge. These tools enable the forensic geneticist to inform the society, particularly the judicial institutions and police investigations, with a growing build of information that assists such bodies in the search for truth and justice.

It is possible to say that for sure SNPs analysis is a tool that will play a major role in forensic investigation, especially in casework related to criminal and identification analysis. This is true because, in addition to these markers allowing for the identification of an individual, particularly when severely degraded samples are involved and STR analysis fails, they can also be used for the discovery of information that STR analysis cannot uncover like the discovery of the physical characteristics of the sample donor. Thus, the analysis of these genetic markers may be of great importance for forensic investigation in the upcoming years.

Thus, considering both the improvement of the scientific and technological methods but also the advancement of society itself to embrace these, it is possible to anticipate that in the future, with the demystification of the idea of these technologies as being unethical and with the demonstration that these can help society in the resolution of more situations, more quickly and with greater certainty, the analysis of SNPs will be used in a greater number of cases. This is evidenced by the increasing number of studies conducted in recent years about the use of SNPs, also demonstrated by the complexity of those and the growing numbers of samples used in those studies.

Certainly, more studies will be needed before SNPs can achieve this major role in forensics discussed above but, with the development of new phenotypic methodologies with more *loci* giving more accurate information and evidence about other human traits, probably with application of the more recent methodologies, this will permit to establish more informative “intelligence” tools that will certainly be used in a routine basis by forensic laboratories around the world.