X-chromosome markers in Forensic Genetics:

evaluation and development of a new 17 X-STR multiplex reaction

Endika Prieto Fernández

PhD Thesis

2017

eman ta zabal zazu

Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

eman ta zabal zazu

Universidad del País Vasco Euskal Herriko Unibertsitatea

X-CHROMOSOME MARKERS IN FORENSIC GENETICS:

Evaluation and development of a new 17 X-STR multiplex reaction.

ENDIKA PRIETO FERNÁNDEZ

PhD Thesis

Directed by:

Professor Doctor María de los Ángeles Martínez de Pancorbo Gómez

BIOMICs Research Group
Department of Zoology and Animal Cell Biology
University of the Basque Country UPV/EHU
Vitoria-Gasteiz, 2017

De todas formas se que es necesario,
andar contra corriente en esta tierra,
y que en el fondo merece la pena
estar loco, estar loco.

Acknowledgements

From the very outset of this report, I would like to extend my sincere and heartfelt gratitude to all the persons who have helped me in this endeavor. Without their active guidance, help, cooperation and encouragement, this project would not have been possible. Any omission in these brief acknowledgements does not mean lack of gratitude.

I am ineffably indebted to my mentor Prof. Dr. Marian M. de Pancorbo for conscientious guidance and encouragement to accomplish this assignment. In addition, I am extremely thankful to my lab colleagues, i.e. Andres, David, Maite, Lourdes, Patri, Melania, Leire, and Tamy, among others. Especially to Carol and Miriam, of course. *Sou les millors companyes que algú pot tenir. M'ho heu ensenyat gairebé tot. Gràcies per la vostra col·laboració desinteressada, els vostres consells, la vostra ajuda i sobretot el vostre somriure a l'altre costat de la taula.*

I also acknowledge with a deep sense of reverence my gratitude to my parents and family members, who have always supported me morally, as well as economically. *Ama, Aita, eskerrik asko horrela izateagatik!*

Last but not least, my gratitude goes to all of my friends who both directly and indirectly have supported me during these three years. Especially in the hardest days.

Thank you very much!

Endika Prieto

Motivation of the Present Study

The revolution in forensic profiling methodologies, as well as the easy accessibility to DNA testing, has triggered a growing popularity of genetic identification and kinship investigation, not only among the scientific community but also the present-day society. In Spain, the kinship testing has an additional utility for the identification of missing persons. Our group (BIOMICs Research Group UPV/EHU) is actively working on the identification of missing persons from the Spanish Civil War (July 17, 1936 – April 1, 1939) and posterior dictatorship. In this context, the relationships being investigated do not often correspond to close relatives but second or third generation family members. In these situations, the current sets of autosomal STR markers may not be enough to assure the relationship. In view of this, new markers located on the X-chromosome may help to solve certain complex kinship cases. However, the current multiplexes that are being applied in Forensic Genetics include a limited number of X-STRs. Therefore, new molecular tools able to obtain more reliable results from highly degraded DNA and low copy number are needed. With that in mind, the present work aimed to provide a new efficient set of microsatellite markers located on the X-chromosome (X-STRs) and the in silico evaluation of other kind of markers located on this chromosome, such as tri- and tetrallelic X-SNPs.

Abstract

Currently, the demand for paternity and kinship testing has become a hot topic in the present-day society. Likewise, the identification of missing persons due to both massive accidents and natural or man-made disasters is nowadays a burning issue. In Spain, a great number of skeletal remnants from executed and murdered individuals during the Spanish Civil War (July 17, 1936 – April 1, 1939) and posterior dictatorship remains yet unidentified. Therefore, this kind of genetic analysis has an additional utility for the identification of missing persons. In this context, the relationships being investigated do not often correspond to close relatives but second or third generation family members, such as grandparent-grandchildren, paternal half-sisters, maternal uncle-nephew, and etc.

The current sets of autosomal markers may not be enough to assure the relationship of some complex kinship cases. Therefore, new markers located on the sexual chromosomes, as well as the mitochondrial DNA have to be investigated and evaluated to their application in forensic casework. On this issue, the development of molecular tools able to obtain more reliable results is an ongoing process that requires a serious effort. In the last few years the study of the genetic markers located along the X-chromosome has aroused great interest, mainly due to their particular inheritance pattern. Females inherit one of their two X-chromosomes from their mother and the other from their father, whilst males receive their only X-chromosome from their mother. Thus, the study of the microsatellite markers located on the X-chromosome (X-STRs) can trace back long pedigrees unless marker transmission breaks down at father-son relationships.

During the last years, the most used multiplexes in the forensic field have been the decaplex of the GHEP-ISFG and the Investigator® Argus X-12 Kit that analyze 10 and 12 X-STRs, respectively. In terms of power of discrimination, the higher the number of polymorphic markers studied, the higher the resolution power of the case. With that in mind, a new 17 X-STR multiplex reaction has been developed and validated following the recommendations of the SWGDAM (Scientific Working Group on DNA Analysis Methods). This new panel has been designed to generate short amplification products, which facilitates the amplification of highly degraded DNA and low copy number.

The calculation in Forensic Genetics requires the establishment of allele and haplotype frequency databases. In the present work, the first allele frequency database for the

Iberian Peninsula was performed initially with the decaplex of the GHEP-ISFG and then with the newly designed panel of 17 X-STRs. In addition, several populations located on the Atlantic coast of Europe and North-West Africa have been analyzed and the corresponding genetic databases have been performed. On the other hand, the correct application of the X-STRs requires a precise knowledge of the linkage and linkage disequilibrium of these markers. With the objective of sheding light on this issue, the corresponding tests were performed and the cluster DXS7132-DXS10075-DXS10079 has been described.

Finally, the SNPs located on the X-chromosome (X-SNPs) have also demonstrated their utility in forensic casework, especially when dealing with highly degraded samples. Up to now, the forensic efficiency of the biallelic X-SNPs has been studied. However, the tri- and tetrallelic X-SNPs have not been forensically evaluated yet. Therefore, their efficiency has been *in silico* evaluated and some of these X-SNPs were postulated as candidate markers for being included in new X-SNP multiplexes or in MPS kits with forensic purposes.

Resumen

Actualmente, la demanda por los análisis de parentesco y la identificación genética se han convertido en un tema de interés social. Del mismo modo, la identificación de personas desaparecidas en desastres de masas, tanto naturales como producidos por el hombre, es un tema candente. En España, el número de personas no identificadas que fueron asesinadas y ejecutadas durante la Guerra Civil Española (17 de Julio de 1936 – 1 de Abril de 1939) y la posterior dictadura sigue siendo muy elevado. Es por ello que los análisis genéticos tienen un valor añadido en esta población. En este contexto, el tipo de relaciones estudiadas generalmente no se corresponden con familiares cercanos sino de segundo o tercer grado, tales como abuelo-nieto, medias hermanas por parte de padre, tía-sobrino, etc.

Los conjuntos de marcadores genéticos localizados en los cromosomas autosómicos actualmente empleados para la identificación genética, pueden resultar insuficientes a la hora de estudiar algunas relaciones de parentesco complejas. Por lo tanto, el estudio de los marcadores localizados tanto en los cromosomas sexuales como en el ADN mitocondrial suscita gran interés entre los investigadores que trabajan en Genética Forense. Durante los últimos años, los marcadores localizados en el cromosoma X han sido objeto de un número creciente de investigaciones. Su interés reside en el particular modelo de herencia de este cromosoma mediante el cual, las mujeres reciben una copia del cromosomas X de la madre y otra copia del padre mientras que los hombres únicamente recibirán una copia de este cromosoma por vía materna. Por esta razón, el estudio de los microsatélites localizados en el cromosoma X (X-STRs) puede ser de gran interés al investigar relaciones familiares de segundo o tercer grado en casos complejos de parentesco.

En los últimos años, las reacciones de análisis simultáneo de marcadores X-STR más comúnmente utilizadas han sido las desarrolladas por el GHEP-ISFG y el kit comercial Investigator® Argus X-12 Kit que analizan 10 y 12 marcadores X-STR en una única reacción de PCR, respectivamente. En términos de poder de discriminación, cuanto mayor sea el número de marcadores altamente informativos analizados, mayor será el poder de resolución. Teniendo esto en cuenta, en el presente trabajo de tesis doctoral se ha desarrollado una nueva reacción de análisis simultáneo de 17 X-STRs siguiendo las

recomendaciones establecidas por la SWGDAM (*Scientific Working Group on DNA Analysis Methods*). El nuevo panel de análisis desarrollado ha sido diseñado para generar productos de amplificación cortos que faciliten el análisis de ADN altamente degradado y/o escaso.

Por otro lado, el cálculo estadístico en el ámbito forense requiere del establecimiento de bases de datos de frecuencias alélicas y haplotípicas. En el presente trabajo, se desarrolló en primer lugar la primera base de datos para la península Ibérica con el set de marcadores del GHEP-ISFG y posteriormente se amplió y actualizó con el nuevo panel desarrollado de 17 X-STRs. Adicionalmente, se estudiaron varias poblaciones de la vertiente atlántica de Europa y el noroeste de África y se generaron las correspondientes bases de datos. Además, el correcto uso de los X-STRs requiere del conocimiento del estado de ligamiento y desequilibrio de ligamiento de cada uno de los marcadores genéticos. Con este objetivo, se llevaron a cabo los correspondientes análisis estadísticos y el grupo de marcadores DXS7132-DXS10075-DXS10079 fue descrito como un bloque de ligamiento.

Finalmente, los polimorfismos de base única o SNPs localizados en el cromosoma X (X-SNPs) también han demostrado su utilidad en el ámbito forense, especialmente cuando se trata de muestras altamente degradadas. Hasta ahora, únicamente había sido evaluada la eficiencia forense de los X-SNPs bialélicos. Sin embargo, los X-SNPs tri- y tetra- alélicos aun no habían sido estudiados. A la vista de ello, se realizó la evaluación *in silico* de la eficiencia forense de estos marcadores y algunos de ellos se postularon como X-SNPs candidatos para su inclusión en nuevos sets de análisis simultáneo de X-SNPs o en kits de secuenciación masiva (MPS).

Table of Contents

1. Introduction

Human Genetic Variability

The human genome is an elegant and well-organized library where the underlying code for human biology lies. Actually, more than 60 years have passed since the definitive structure of the salt deoxyribose nucleic acid was proposed [1]. Nonetheless, the understanding of all the secrets hidden along our genome is far from complete.

In 2001, the International Human Genome Sequencing Consortium (IHGSC) [2] and Celera Genomics [3] reported draft sequences providing a first overall view of the human genome but it was in april 2003 when this goal was completely achieved [4]. After that, the project known as Encyclopedia of DNA Elements (ENCODE) aimed to provide a more biologically informative representation of the human genome. Interestingly, of the approximately 3,000 billion base pairs encoded in the human genome, the 2.94% corresponds to exons located on protein-coding regions while only the 1.2% presents a defined function in protein coding [5–7].

Regarding the variability between individuals, only the 0.3% of the genome differs between human beings [8]. However, and despite the low percentage, this characteristic makes all of us unique, except for identical twins. Consequently, we are able to differentiate, as well as to establish biological relationships among individuals using the information contained in their DNA sequences through human identification and kinship testing analyses.

Variations in DNA sequences are known as genetic polymorphisms or markers. The efforts of the scientific community throughout history for understanding the human genome, has allowed the discovery of a great bunch of polymorphisms located along our genome. In this sense, the development of more powerful statistical methods, as well as the emergence of new experimental technologies, i.e. Massive Parallel Sequencing (MPS), have permitted to uncover more and more markers in the last years.

DNA typing consists in determining the genotype of a set of markers located along the genome of an individual in order to obtain its genetic profile. Genetic profiles are the basis of Forensic Genetics, which always requires the comparison between a profile obtained from a questioned sample and the genotypes obtained from known samples or databases containing DNA profiles from previous analyses [9].

Forensic Genetics: Evolution through the History

The ABO blood group system, which was described by the austrian scientist Karl Landsteiner in 1900, is considered the first genetic polymorphism with forensic capabilities in human identification [10–12].

After that, in 1955 Smithies opened a new era of Haemogenetics with the development of electrophoresis in starch gels and the detection of genetic polymorphisms in the haptoglobin. After that, other kind of polymorphisms have been analyzed e.g., serum proteins, erythrocyte enzyme polymorphisms, and the Human Leucocyte Antigen (HLA) system [13].

The year 1980 was the starting point in the new era of Forensic Genetics when the repetitive DNA was discovered in the regions outside the nuclear DNA genes. However, it was not until middle eighties when the real revolution broke out with the first publication about DNA fingerprinting [14]. The most variable loci discovered in the human genome consist of tandemly repeated minisatellites, also known as Variable Number of Tandem Repeats (VNTRs) [15], and provide the basis for most currently used DNA typing systems. DNA probes comprised of tandem repeats of the core sequence may detect multiple variable human DNA fragments by Southern blot hybridization, to produce an individual-specific DNA “fingerprint” [16]. Alec Jeffreys et al. have developed two multilocus probes, i.e. 33.6 and 33.15, which were extensively used in parentage testing [17,18] worldwide. In addition, other multilocus probes have been reported, e.g. [19,20]. The first use of DNA testing in a forensic case came in 1986 and was based on Jeffrey’s multi-locus probes. By this method was proved that the man who claimed to have sexually assaulted and then brutally murdered Lynda Mann and Dawn Ashworth in Leicestershire (England) was innocent. After a year, the police caught the real murderer and he confessed to committing the crimes. When the genetic analysis was carried out, his DNA profile matched the DNA from the crime scenes [21].

After that, Allele-Specific Oligonucleotide (ASO) probes were presented, which were short pieces of synthetic DNA that could detect single base substitutions in the human genomic DNA. The first locus examined extensively for identification was the HLA-DQα. Its analysis was made commercially available as a kit in a reverse-dot-format by Perkin-Elmer. This kit was followed by a second reverse dot blot format kit known as Polymarker (PM), which allowed the simultaneous amplification and analysis of five additional loci (LDLR, GYPA, HBGG, D7S8, and GC) [22]. To improve sensitivity, specificity and simplicity of this approach, the Polymerase Chain Reaction (PCR) was used to enzymatically amplify a specific segment of the ß-globin or HLA-DQα gene in human genomic DNA before hybridization with ASOs [23].

The development of the PCR was a key factor for the analysis of the microsatellite or Short Tandem Repeat (STR) regions [24]. Nowadays, the PCR systems based on STRs are widely used by the majority of the laboratories of Forensic Genetics for both human genetic identification and kinship testing. The main reasons for their success have been their high power of discrimination and their rapid analysis speed [8].

Since the beginning of the 21st century, there has been a gradual evolution in technologies that has allowed the discovery of new and more discriminative markers. In the last years, the trend has been to develop molecular tools capable of typing more and more markers in a single PCR reaction. In this sense, the higher the number of polymorphic markers studied, the higher the resolution power of the case. Apart from STRs, other kind of markers has also been studied, such as mitochondrial DNA, insertion-deletion polymorphisms (indels), and Single Nucleotide Polymorphisms or SNPs [13].

In the recent years, the use of Massive Parallel Sequencing or MPS has been debated and now, the first applications are beginning to emerge in forensics, especially in human identification and determination of phenotypical traits [25].

The most common markers in Forensic Genetics

Nowadays, the most common markers used by the forensic laboratories may be categorized into two groups:

First Group: sequence variation polymorphisms.

1. Hipervariable regions of the mitochondrial DNA: HVI, HVII, and HVIII:

The mitochondrial DNA (mtDNA) is a circular molecule of DNA that has a length of 16,569 base pairs (bp) and it is located inside the mitochondria [26]. Hypervariable segments HVI and HVII control regions have been used in forensic casework even though the power of discrimination is clearly below the values obtained from autosomal STRs [27]. This could be due to the fact that reliable results can be obtained from low copy number [8]. Moreover, the HVIII region has also showed its potential effectiveness in identifying individuals that present the same HVI-HVII haplotypes [28,29]. Nowadays, the mtDNA has been established as the referent maternal lineage marker for the study of both human phylogeny and evolutionary history of human populations [30–32].

2. Single Nucleotide Polymorphism (SNP):

A SNP is a single bp difference that occurs at a specific position in the genome. SNP markers will serve an important role in analyzing challenging forensic samples as they may increase the power of discrimination in kinship analyses, family reconstructions and/or human identification. There are four classes of SNPs that apply to forensic analyses: 1) identity-testing SNPs, 2) lineage informative SNPs, 3) ancestry informative SNPs, and 4) phenotypic SNPs. Despite this and the large number of SNP variants along the genome, they are not expected to replace the battery of STR loci in the foreseeable future [33].

Second group*:* length polymorphisms.

3. Short Tandem Repeats (STR):

A microsatellite or STR is a tract of repetitive DNA in which certain DNA motifs (ranging in length from 2-5 base pairs) are repeated, typically 5-50 times [34]. The appreciable success of STR typing in Forensic Genetics may be due to the fact that it is more sensitive than single-locus DNA probes, less prone to allelic dropout than DNA fingerprinting, and more discriminative than other PCR-based typing methods [21]. In the last years, the autosomal STRs (AS-STRs) have been the most used markers in kinship testing and human identification. Additionally, the STRs located on the gonosomes, i.e. X- and Y-chromosome STRs (X-STRs and Y-STRs, respectively), has gained popularity over the last years [35].

Microsatellite DNA in Forensic Genetics

Eukaryotic genomes are full of repeat DNA sequences, such as STRs [2]. In fact, it is estimated that around the 3% of the genome is composed by microsatellite DNA [36–38].

These markers may present different number of repetitions or alleles at a specific locus, e.g. 8 or 10. The two alleles of a diploid individual at a certain locus determine its genotype, e.g. 8/10 (Figure 1). Finally, all the genotypes of a set of markers make up the genetic profile of an individual.

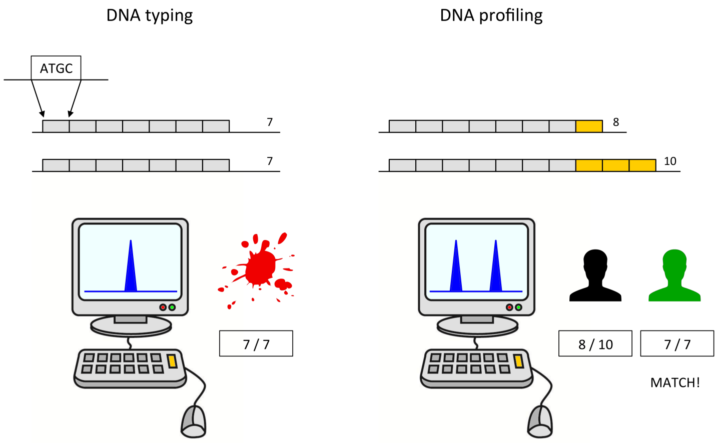


Figure 1. DNA typing (left) and profiling (right): from marker's variation to genetic identification.

Mutation of STR markers: origin of diversity

It is widely assumed that microsatellites are found scattered throughout the genome of eukaryotes and tend to vary among individuals but, how was their birth and expansion?

The evolution of each microsatellite may have started with a mutation in a nonreiterated sequence, probably due to a duplication event of a single short sequence motif. On this subject, Slipped Strand Mispairing (SSM) have played a major role in the generation of short proto-microsatellite regions by chance [39,40]. However, the answer to the question of when a DNA repeat sequence can be considered as a microsatellite has generated different points of view [40,41]. Some studies have suggested that a minimum number of repeats are required before SSM can extend the proto-microsatellites, i.e. threshold model [42]. Interestingly, other studies carried out in *Saccharomyces cerevisiae* suggest that no minimum number of repeats is necessary for their establishment [43].

During the replication of a microsatellite sequence, DNA polymerase pauses and temporarily dissociates from the DNA. Consequently, the terminal end of the newly synthesized DNA separates from the template sequence. After that, if the nascent strand realigns out of register, a small DNA 'loop' will be generated and, as consequence, a variation in the number of repeat units with regard to the template strand may occur. If the 'loop' is generated on the nascent strand, there will be an increase in length, while if it happens in the template strand a decrease will occur [36,39] (Figure 2). This mechanism may also take place *in vitro* during the PCR where minor products (<15% of the total height of the real peak), which differ in length from the main product, may be generated, i.e. 'stutter peaks'.

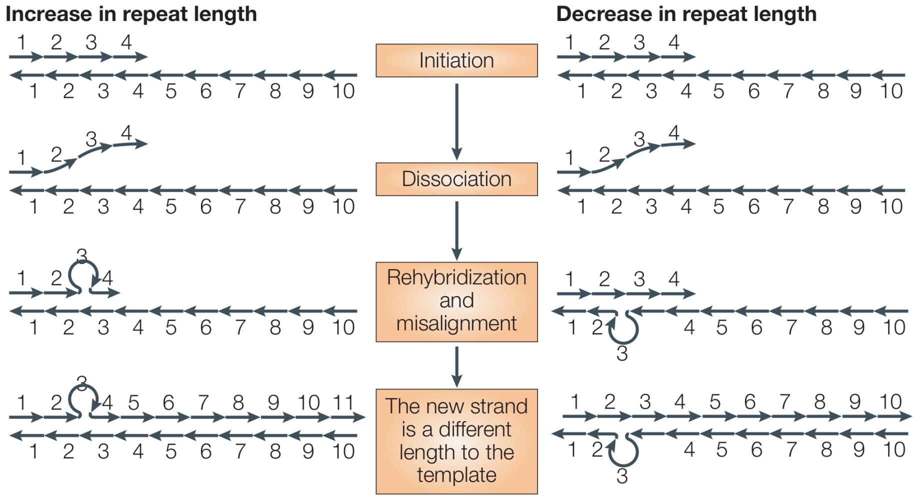

Increase in repeat length
Decrease in repeat length
Initiation
Dissociation
Rehybridization and misalignment
The new strand is a different length to the template
1 2 3 4
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10 11

Figure 2. Slipped Strand Mispairing (SSM): the origin of length variation in microsatellite DNA [36].

A great part of the above-mentioned replication errors are corrected *in vivo* by our organism through the Mismatch Repair System (MRS) [44]. However, if this enzymatic complex has not success, a mutation process will generate length changes in a microsatellite region of an individual. It may occur on the somatic level, affecting only individuals, or in the sex cells, affecting future generations by enriching genetic diversity of a certain population [45]. In this sense, there seems to be directionality in the evolution of these markers towards an increase in length, [46].

The mutation rate of a microsatellite locus is defined as the number of mutations that can occur in a single generation. Mutation rates of the autosomal microsatellites oscillate between 10^{-2} and 10^{-6} per locus and generation in human beings [47,48]. These high mutation rates may be due to both SSM and recombination between DNA strands [48]. Furthermore, sex bias may constitute another factor affecting mutation rates since it is often assumed that males mutate more often than females. This may be due to the fact that men have more germ-line cell divisions than women. Additionally, other factors affecting the mutation rate and the stability of the markers have been proposed, such as the number of repeat sequences, as well as the allele size [41, 49–53].

Type of STR markers

Microsatellite markers not only vary in the number of repetitions but also in the composition of the repeat motif that is not restricted to a single structure since mono, di, tri, and tetranucleotide structures are vastly overrepresented along our genome [36]. Moreover, attending to the rigor and degree of perfection of the repeat unit, STRs may be classified in 'simple' (units of identical length and sequence), 'compound' (two or more adjacent simple repeats), and 'complex' (several repeat blocks of variable unit length, as well as variable intervening sequences) [54] (Table 1).

Table 1. Variation in Short Tandem Repeat markers.

Repeat unit	*Motif structure*	*Example*
Simple		
Dinucleotide	$(AG)_n$	AGAGAGAG
Trinucleotide	$(ACG)_n$	ACGACGACGACG
Tetranucleotide	$(TCTA)_n$	TCTATCTATCTATCTA
Compound	$(GATA)_m$ – $(TAGA)_n$	GATAGATATAGATAGA
Complex	$(TATC)_k$ –$(ATCG)_m$ – $(CCAA)_n$	TATCATCGATCGCCAACCAA

However, not all alleles for an STR locus contain complete repeat units. In this way, simple repeats can contain non-consensus alleles that fall in between alleles with full repeat units. These allele variants are known as 'microvariants' [55]. A good example of this may be the complex repeat sequence composition of the marker DXS6803, which is located on the X-chromosome, $(TCTA)_n$-$(TCA)_{0\text{-}1}$-TCTA. In this structure, the presence of the trinucleotide motif $(TCA)_{0\text{-}1}$ may give rise to the alleles 10.3-14.3 [56].

Nomenclature of STRs and allele designation

A standardized designation of the structure of STR markers is crucial to both correctly assign the alleles of each marker and reach consensus among forensic laboratories worldwide. With that in mind, the DNA commission of the International Society of Forensic Haemogenetics (ISFH) recommended a nomenclature for STRs based on the number of repeat sequences present in each allele [57–59]. These recommendations seemed to be valid for markers with simple repeat sequences. However, and according to the European DNA Profiling group (EDNAP), the proposed method cannot accommodate complex repeat sequences. For that reason, the following issues for allele designation of STRs with complex repeat sequences were proposed [60]:

1. Reliable standard allelic ladders that cover the most common alleles for each locus should be developed.
2. The components of an allelic ladder should be sequenced and standardized.
3. Designation of alleles based on complex repeat sequences should be carried out by size comparison against the alleles represented in the reference ladders.

The most common STRs

National and international STR databases enable forensic laboratories to exchange and rapidly compare DNA profiles between individuals worldwide. These databases are widely recognized as some of the most effective and efficient crime fighting tools available to law enforcement. Hence, its use facilitates the resolution of forensic cases, as well as human identification in natural or man-made disasters [61–63].

Because of this, over the last years, the forensic community has been focused on developing simultaneous analysis methods considering the main STRs included in the principal forensic databases, i.e. Combined DNA Index System (CODIS) (https://www.fbi.gov/) and European Standard Set (ESS) database. To date, several multiplexes, which include the most common STRs, have been developed and commercialized. The most usual kits are Identifiler® Plus, MiniFiler™, NGM SElect™ (from Thermofisher Scientific, Waltham, MA, USA), Power Plex® 16, Power Plex® ESI-17, Power Plex® ESX-17 (from Promega Corporation, Madison, WI, USA), Investigator ESSplex SE QS Kit (from Qiagen, Valencia, CA, USA), BioPlex-11 [64], I-DNA1 [65], I-DNA2 [66], I-DNASE21 system [67], and Q8 [68], among others. It is noteworthy that some of these multiplexes have been specifically designed to be used in combination with others in order to type as many markers as possible in few PCR reactions [69] (Table 2).

Recently, the number of markers included in the above-mentioned databases, i.e. CODIS and ESS, has been enlarged with the following objectives: 1) to obtain more reliable results in forensic calculations and 2) to increase traceability between existing databases. First, the ESS was extended with five additional STR loci (D1S1656, D2S441, D10S1248, D12S391, and D22S1045) under the name ESS-Extended [70]. Then, and following the same strategy, the CODIS database included the ESS-Extended database's loci plus D19S433 and D22S1045 markers. In view of this scenario, multiplexes able to amplify the core STRs shared by both ESS and CODIS databases in a single PCR reaction are of great utility [67,71] (Table 2).

Table 2. Main STR loci included in the most common multiplex reactions in forensic casework. Abbreviations: BIO= BioPlex-11, ESS= Investigator ESSplex SE QS Kit, IDEN= Identifiler® Plus, MINI= MiniFiler™, NGM= NGM SElectTM, PP= Power Plex® 16, PP ESI= Power Plex® ESI-17, and PPESX= Power Plex® ESX-17. * indicates recently added markers in both ESS and CODIS databases. Modified from [69].

	Locus	*BIO [64]*	*ESS*	*IDEN*	*I-DNA1 [65]*	*I-DNA2 [66]*	*MINI*	*NGM*	*PP*	*PP ESI*	*PP ESX*	*Q8 [68]*	*IDNASE 21 [67]*	*24-Plex [71]*
CODIS + ESS	D16S539		X	X	X		X	X	X	X	X		X	X
	D18S51		X	X	X	X	X	X	X	X	X	X	X	X
	D21S11	X	X	X	X	X	X	X	X	X	X	X	X	X
	D3S1358	X	X	X	X			X	X	X	X	X	X	X
	D8S1179	X	X	X	X			X	X	X	X	X	X	X
	FGA	X	X	X	X	X	X	X	X	X	X	X	X	X
	TH01	X	X	X	X	X		X	X	X	X	X	X	X
	vWA	X	X	X	X	X		X	X	X	X	X	X	X
	D1S1656*		X					X		X	X		X	X
	D2S441*		X					X		X	X		X	X
	D2S1338*	X	X	X		X	X	X		X	X		X	X
	D10S1248		X					X		X	X		X	X
	D12S391*	X	X					X		X	X		X	X
	D19S433*		X	X	X			X		X	X		X	X
	D22S1045		X					X		X	X		X	X
CODIS	CSF1PO			X	X	X	X		X				X	X
	D13S317			X	X	X	X		X				X	X
	D5S818	X		X	X	X			X				X	X
	D7S820			X	X	X	X		X				X	X
	TPOX	X		X	X	X			X				X	X
Others	Penta-D								X					X
	Penta-E								X					X
	SE33		X					X		X	X	X	X	
	D6S1043													X

Selection criteria

Forensic specimens are often challenging to the analysis of STRs as the probability of successfully obtaining genetic results depends on several factors: 1) the amount of preserved DNA, 2) degradation of DNA, and 3) the absence / presence of PCR inhibitors, such as humic acid or hematin, among others [72–74]. This particularly applies to skeletal remains, such as those recovered from mass graves [75]. For all these reasons, the design of multiplex reactions requires a good selection of markers that can be properly amplified even under the above-mentioned conditions. Currently, the most extended markers in forensics are the tetranucleotide STR loci. Their success over other type of markers, such as di- or trinucleotide STRs may be due to [8]:

1. A narrow allele size range that permits multiplexing and reduces allelic dropout from preferential amplification of smaller alleles.

2. The capability of generating small PCR product sizes that benefit recovery of information from degraded DNA specimens.
3. Reduced stutter product formation compared to dinucleotide repeats that benefit the interpretation of sample mixtures.

A good selection of STR markers for multiplexing may be based on the following conditions [8,76,77]:

1. High heterozygosity (>70%) and power of discrimination (>0.9).
2. Simple STRs with reduced allelic range. Simple STR systems are more amenable to adopt as standard loci than complex systems [78,79].
3. Predicted length of amplification products <500 bp. It is undeniable that mini- and midiSTRs are more sensitive and robust in the analysis of highly degraded DNA and low copy number [80,81].
4. Low mutation rates. This characteristic is not as important for stain analysis as for kinship testing.
5. Distributed on different chromosomes. The markers used in Forensic Genetics are typically chosen from separate chromosomes to avoid linkage and linkage disequilibrium (LD), except for those located on the gonosomes, i.e. X and Y chromosomes.
6. Generation of a low number of stutters during PCR amplification [82,83].
7. Low percentage of null alleles.
8. Robustness in presence of amplification inhibitors [72–74].
9. Species specificity.
10. Concordance of results. This allows the creation of national and international databases.

Forensic databases and (inter)national co-working

In 1995, a comprehensive legislation enacted in the United Kingdom enabled to set up the first national DNA database with forensic purposes. Posteriorly, other countries started to perform their own databases [61]. In this context, the Technologies Group of the National Institute of Standards and Technology (NIST) created the 'STRBase' (http://www.cstl.nist.gov/biotech/strbase) in 1997 [84]. Its creation allowed the forensic community to widely share information about STR typing worldwide. Currently, more than 100 countries have incorporated their own databases for policing purposes [85].

The currently developed multiplexes share several STRs that aimed to standardize, as far as possible, the current forensic databases. However, the exchange of genetic profiles between countries remains a problem because an international legislation is not established yet [61]. On the other hand, realistic allele and haplotype frequencies of each region are necessary to carry out calculation in forensic cases. With that in mind, several populations worldwide are being typed every year with the most common multiplexes. This allows the constant development of allele and haplotype frequency databases that are of great utility in kinship testing and human identification.

Scientific Working Groups (SWG) consist of scientific subject-matter experts who collaborate to both determine best practices and develop consensus standards. Similarly, several national and international societies aim to advance the field of Forensic Genetics through dissemination of scientific results and opinions, communication amongst scientists and education or orientation. In the last years, several groups and commissions have been created in Europe and America, as it is shown in Table 3.

Table 3. Major scientific associations and working groups in the context of Forensic Genetics over the last few years in the United States of America (EEUU), Europe (EUR), and Latin America (LA).

	Acronym	*Name*	*Webpage*
EEUU	ASCLD/LAB	American Society of Crime Laboratory Directors – Laboratory Accreditation Board	http://www.ascld.org/
	DAB	DNA Advisory Board	*Not found*
	SWGDAM [A]	Scientific Working Group on DNA Analysis Methods	http://www.swgdam.org/
	AABB	American Association of Blood Banks	http://www.aabb.org/
	CAP	College of American Pathologists	http://www.cap.org/
	NIST	National Institute of Standard and Technology	http://www.cstl.nist.gov/
EUR	ISFG	International Society for Forensic Genetics	https://www.isfg.org/
	ENFSI	European Network of Forensic Science Institute	http://enfsi.eu/
	EDNAP	European DNA Profiling Group	https://www.isfg.org/EDNAP
	STADNAP	Standardization of DNA Profiling Techniques in the EU	http://www.stadnap.uni-mainz.de/
	IEWPDP	Interpol European Working Party on DNA Profiling	*Not found*
	GHEP-ISFG	Grupo de Habla Española y Portuguesa de la ISFG	http://www.gep-isfg.org/
LA	GITAD	Grupo Iberoamericano de Trabajo en Análisis de DNA	http://gitad.ugr.es/

[A] Originally TWGDAM (Technical Working Group on DNA Analysis Methods)

In Table 4 is shown a summary of the primary activities carried out by the above-mentioned groups and commissions according to John M. Butler [86]. In addition to this, other reviews and perspectives at the time on the past, present and future of Forensic Genetics have been published [21,87].

Table 4. Forensic DNA analysis phases and activities by decade [86].

Phase	*Time Frame*	*Description of activities*
Exploration	1985-1995	• Beginnings and first publications. • Different methods tried including multi- and single-locus VNTRs and early PCR assays, such as DQα and single-locus STR markers. • Need for standardization and quality control results in formation of EDNAP and SWGDAM.
Stabilization and standardization	1995-2005	• National databases launched for UK (1995), USA (1998) and many European countries. • Standardization of multiplex STR systems and capillary electrophoresis. • Initial autosomal STR and Y-STR kits released. • Selection of core loci for US and Europe. • Implementation of FBI Quality Assurance Standards in the USA. • ENFSI begins role in Europe to aid standardization and quality assurance.
Growth	2005-2015	• Rapid growth of DNA databases. • Expanded core loci in Europe and USA lead to new STR kits. • Y-STR use on the rise. • Extended applications being pursued, e.g. rapid DNA instruments, familial searching, NGS research into STR allele variability.
Sophistication	2015-2025	• Expanded set of tools with capabilities for rapid DNA testing outside of laboratories, greater depth of information from allele sequencing, higher sensitive methodologies applied to casework, and probabilistic software approaches to complex evidence. • The need to confront privacy concerns increases as knowledge of genomic information improves.

Likelihood ratio

In both kinship testing and human identification, statistics attempts to provide meaning to the DNA match or parentage between two individuals [88]. For this, the likelihood ratio (LR), which involves a comparison of the probabilities of the evidence under two alternative propositions, is calculated based on the genetic data obtained from the individuals under evaluation. In mathematical notation the LR is calculated by Eq. 1, where Pr(E/H_1) and Pr(E/H_2) correspond to the probability of the evidence (E) given the assumptions 1 and 2, respectively.

(Eq. 1) $$LR = \frac{Pr(E/H_1)}{Pr(E/H_2)}$$

A good example to understand the hypothesis approach is a paternity case where:

H_1 = ‘the alleged father is the biological father of the child’.
H_2 = ‘they are unrelated’.

Once the LR is calculated, the final value will indicate how much more likely is the hypothesis H_1 (i.e. true biological father) versus H_2 (i.e. unrelated).

Statistical Analyses

The calculation of the LR requires the establishment of population-specific forensic allele and haplotype frequency databases. In this context, the generation and validation of a population DNA database that can be used to estimate the frequency of an observed DNA profile in the population requires several steps [8]:

1. Deciding on the number of samples and ethnic/racial grouping.
2. Gathering samples.
3. Analyzing samples at desired genetic loci.
4. Summarizing DNA types.
5. Determining allele and haplotype frequencies.
6. Performing statistical tests on data:
    a. Hardy-Weinberg equilibrium (HWE) for allele independence.
    b. Determination of the linkage and LD state of the selected markers.
    c. Evaluating the parameters of forensic interest, i.e. PD_F, PD_M, MEC_T, and MEC_D.

All the above-mentioned parameters are based on both allele and haplotype frequencies and are described below.

Population genetic parameters

1. Heterozygosity and gene diversity:

Heterozygosity (H) is the proportion of heterozygous individuals in a certain population. A high heterozygosity means that more allele diversity exists, and therefore, there is less chance of random sample matching [89]. On the other hand, gene diversity (GD) is equal to the expected heterozygosity (H'). In other words, is the probability that two randomly chosen alleles from the population are different [89,90].

2. Allele and haplotype frequencies:

Relative allele frequencies are calculated by dividing the number of copies of an allele in a tested population by the total number of all alleles observed. These differ among population samples and therefore, site-specific frequency databases are needed for forensic calculation (Figure 3). Similarly, relative genotype frequencies refers to the number of individuals with a particular genotype divided by the total number of individuals in a certain population [88].

Relative haplotype frequencies cannot be calculated from allele frequencies but have to be estimated directly from appropriate population samples [91]. In autosomal markers, gametic phases are unknown but they could be estimated through a pseudo Bayesian approach known as ELB algorithm [92,93]. When dealing with sexual chromosomes, gametic phases are always known in men but not in women. However, the two X-chromosome gametic phases in women may be deduced through family studies composed by grandfather-daughter-grandson.

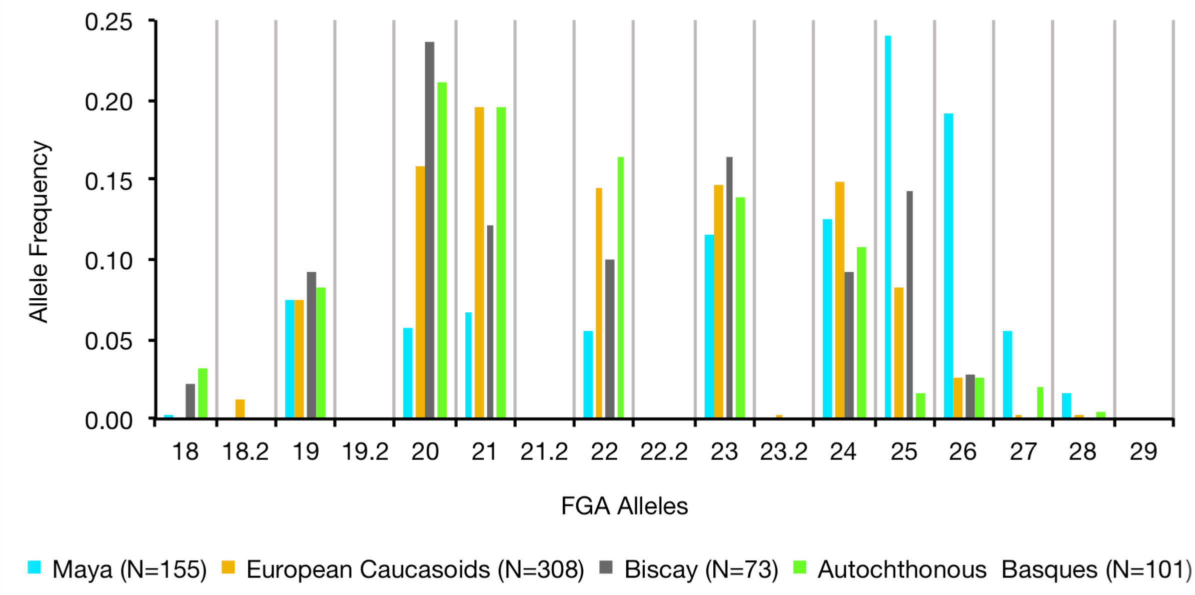


Figure 3. Comparison of FGA allele frequencies observed in four populations obtained from [67,94–96].

3. Hardy-Weinberg Equilibrium:

Hardy-Weinberg equilibrium (HWE) model relates both allele and genotype frequencies and indicates if these remain constant from generation-to-generation [88]. The ideal population is one that follows HWE assumptions and therefore, any allele combination is possible. Traditionally, geneticists have relied on test statistics with asymptotic χ^2-

distributions to test for goodness-of-fit with respect to HWE proportions. However, as pointed out by several author these asymptotic tests quickly become unreliable when samples are small or when rare alleles are involved [97]. Therefore, the exact test of HWE is usually used [98].

4. Determination of the linkage and LD state of the selected markers:

Linkage can be defined as the co-segregation of closely located loci within a family or pedigree [99]. On the other hand, Linkage Disequilibrium (LD) is simply a non-random association between two or more alleles. It means that these alleles appear together at rates that differ from what would be expected under independence [45,100]. These two concepts will be addressed more accurately in a following section.

5. Population differentiation:

Regarding population differentiation, Wright's F-statistics, and F_{ST} in particular, have been widely used as descriptive statistics to investigate processes that influence the distribution of genetic variation within and among populations [101]. F_{ST} is described as the correlation between two alleles chosen at random within subpopulations relative to alleles sampled at random from the total population [102,103]. In this sense, population geneticists have proposed several statistical measures that are related to F_{ST}, such as G_{ST}, R_{ST}, Φ_{ST}, and Q_{ST} [101]. Slatkin's R_{ST} is an analogue of F_{ST} assuming a Stepwise Mutation Model (SMM) [104]. However, there is no clear consensus over their relative accuracy when dealing with STR data. Their performance will depend on how well the STRs under study fit a SMM as it is only under a strict SMM that R_{ST} is independent of mutation [105].

Parameters of forensic interest

From 1995 to 2013, 1,404 articles on STR population data were published in the six main forensic journals. However, most of the authors publishing such data do not usually describe the parameters of forensic interest calculated to evaluate the forensic usefulness of each STR marker that has been typed [88]. In Table 5 the principal values that are currently being calculated from the allele frequencies of each STR marker are summarized.

Table 5. Summary of the main parameters of forensic interest that are usually estimated in forensic publications.

Parameter	*Description*
Power of discrimination (PD) [106]	• Is the potential power to differentiate between any two individuals taken randomly from the population.
Probability of Identity (P_I) [107,108]	• Is the probability that two individuals selected at random will have identical genotypes at a tested locus.
Power of Exclusion of paternity a priori (PE) [106,109]	• Is the probability that, given the mother and child genotypes, a non-father would be excluded from paternity.
Paternity Index (PI) [109]	• Is the probability that, given the genotype of the child, the man tested is the true biological father.
Polymorphisms Information Content (PIC) [110]	• Is the probability that a given offspring of a parent carrying a rare allele at a certain locus will allow deduction of the parental genotype at the locus.
Mean Exclusion Chance (MEC) [111]	• Is the probability of suspected parents falsely imputed whose paternity is excluded on the basis of a certain locus.

X-chromosome Short Tandem Repeats in Forensic Genetics

In recent years, the study of STRs located on the X-chromosome (X-STRs) has aroused a great interest in Forensic Genetics [35]. Due to its particular inheritance pattern, the application of X-chromosome markers may increase the chance of solving certain challenging cases that other forensic markers cannot [99]. In this sense, the X-STRs are of great utility in the identification of war victims, such as the Spanish Civil War (July 17, 1936 – April 1, 1939) and posterior dictatorship [75], where the majority of the profiles to compare correspond to second or third degree relatives, e.g. grandparents-grandchildren, maternal uncle-nephews, and etc. [112]. In addition, their usefulness is also noticeable in natural or man-made disasters where it is necessary to establish the relationship between the corpses and their relatives, which can be any degree family members. Currently, more than 45 X-STR markers have been studied and characterized [35]. Additionally, several populations worldwide have been typed with the most common markers included in the main multiplexes, i.e. the Investigator® Argus X-8 and Argus X-12 Kits, as well as the decaplex of the GHEP-ISFG [113,114], generating population frequency data essential for the application of this markers to complex kinship cases [35,115].

Characteristics of the X chromosome

The X-chromosome is one of the two sex-determining chromosomes in many mammal species, including humans. Regarding the position of the centromere, this human chromosome presents a submetacentric structure. It spans 155 million bp that represent approximately 5% of the human genome. Until now, 1,098 genes have been described along its sequence, which have shown to be essential for human beings as a large amount of genetic disorders has been related to this chromosome. However, a great part of its sequence corresponds to repeat DNA regions where the markers of forensic interest are located [116,117].

Each individual presents one pair of sexual chromosomes or allosomes, i.e. X- and Y-chromosomes. In terms of the X-chromosome, males have only one copy while females possess two copies of this chromosome. In females, one of these two copies remains silenced or inactivated during cell life except for the meiotic process. However, on a standard karyotype they look entirely normal and therefore they behave as a pair of autosomal chromosomes. Additionally, when genetic disorders occur, irregular karyotypes can be found that give rise to genetic aberrations, such as the Klinefelter's syndrome (XXY) [117].

Allosomes began their evolutionary process as a pair of homologous chromosomes that were identical except for one sex-determining locus that only one of them presented. However, sexually antagonistic selection, mutation, and random genetic drift led to the divergence of the X and Y chromosomes [118]. This evolutionary process over time has almost eradicated most traces of the ancestral relationship between the human X- and Y-chromosomes. However, there are 44 homologous genes that both chromosomes still share. These are distributed as follow: 24 and 5 genes in the pseudo-autosomal regions PAR1 and PAR2, respectively; 15 genes in the X-added region (XAR); 7 in the X-conserved region (XCR); and 3 in the X-transposed region (XTR) (Figure 4) [116].

Inheritance pattern

The forensic potential of the X-chromosome is mainly due to: 1) the ability to recombine only in women and 2) its particular inheritance pattern. As mentioned in the previous section, most of the X- and Y-chromosome regions are hemizygous and therefore, they cannot recombine in males. On the contrary, recombination events may occur in females as two homologous copies of the X-chromosome are involved in the meiotic process.

Regarding inheritance pattern, a female receives one non-recombinant copy from her father and a recombinant copy from her mother. On the other hand, a male receives only one recombinant copy from his mother that will transmit only to his female offspring without changes (Figure 5) [35].

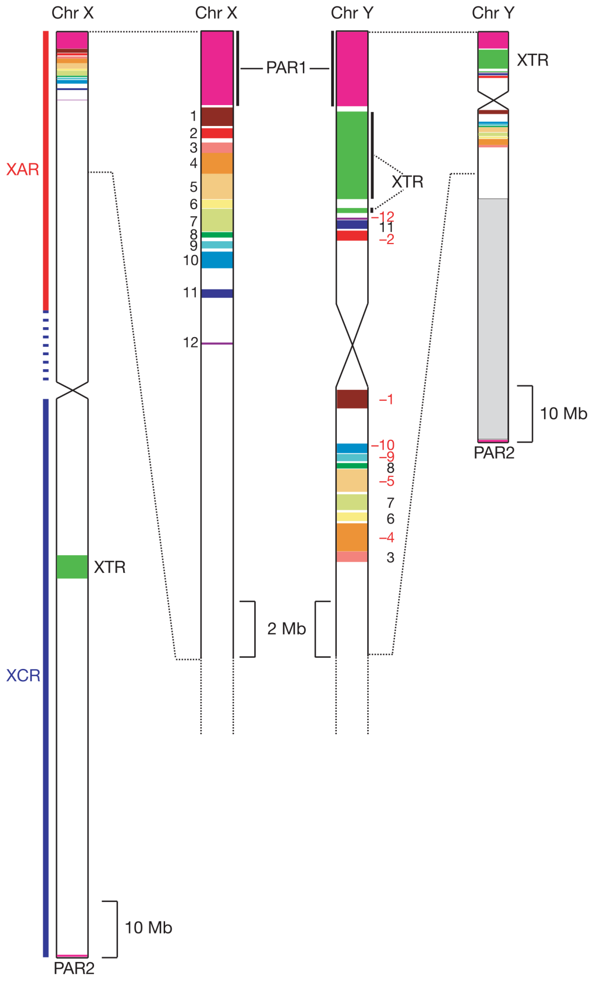


Figure 4. Schematic representation of major homologies between the human sex chromosomes [116].

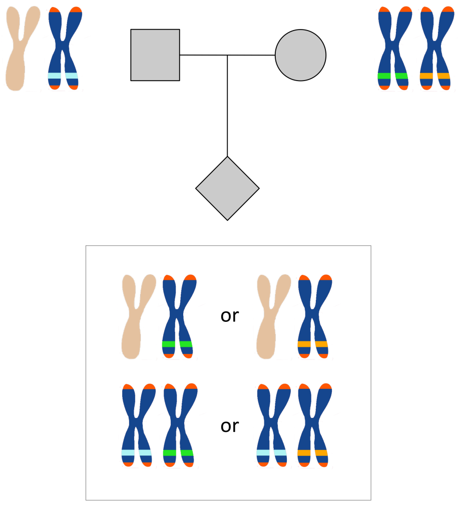


Figure 5. Inheritance pattern of the X-chromosome for both males (XY) and females (XX).

The most common X-STRs

Over the last years, the number of X-STRs studied in Forensic Genetics has undergone a rapid growth. Currently, more than 45 markers have been studied in several populations worldwide [35]. The main markers, as well as the most common multiplex panels for the analysis of ≥ 8 X-STRs in a single PCR reaction are summarized in Table 6.

Table 6. Main X-STR loci currently used in Forensic Genetics according to Diegoli et al. (2015) [35]. The markers included in the following multiplexes are remarked in grey: Mentype® Argus X-8 PCR Amplification Kit (Argus X-8), Investigator® Argus X-12 (Argus X-12), and the decaplex of the GHEP-ISFG (GHEP). n/a = not available.

Marker	*Cytogenetic localization* ^{B}	*Position (cM)* ^{B}	*Argus X-12*	*Argus X-8*	*GHEP [113,114]*	*[119]*	*[120]*	*[121]*	*[122]*	*[123]*	*[124]*	*[125]*	*[126]*
DXS6807	p22.3	14.76				X		X		X	X_2		
DXS9895	p22.3	17.09							X		X_1		
DXS10148	P22.3	19.84	X										X
DXS8378	p22.3	20.21	X	X	X	X	X	X		X			X
DXS10135	p22.3	20.03	X	X									X
DXS9902	p22.2	32.32			X	X	X			X			
DXS6795	p22.1	44.24											
DXS6810	p11.3	75.12											
DXS10078	p11.2	85.07											
DXS10159	Centromere	90.01											X
DXS7132	Centromere	90.75	X	X	X		X	X	X	X	X_2	X	X
HumARA ^{A}	q12	90.81											
DXS10079	q12	90.82	X									X	X
DXS10074	q12	90.83	X	X								X	X
DXS10075	q12	90.83										X	X
DXS981	q13.1	92.81									X_1	X	
DXS6800	q13.3	97.49							X		X_2	X	
DXS6803	q21.2	99.40									X_2		
DXS9898	q21.3	101.29			X				X	X		X	
DXS6801	q21.3	106.08										X	
DXS6789	q21.3	108.47			X		X	X	X	X	X_2	X	X
DXS6809	q21.3	110.71			X	X		X		X	X_2	X	X
DXS6799	q21.3	110.71					X						
DXS7424	q22.1	115.25					X	X		X	X_1	X	X
DXS101	q22.1	*116.15*					X	X		X	X_2	X	X
DXS6797	q22.3	117.74											
DXS7133	q22.3	118.18			X			X	X		X_2	X	
DXS6804	q23	122.32											
GATA172D05	q23	124.36			X	X	X	X		X	X_1		
DXS7130	q24	130.28							X				
GATA165B12	q25	136.18										X	
DXS6854	q25	n/a				X							
HPRTB	q26.2	*149.66*	X	X		X	X	X	X	X	X_1		X
DXS10101	q26.3	*149.75*	X	X									X

DXS10103	q26.2	*149.37*	X										X
GATA31E08	q27.1	160.54			X		X	X	X		X_1	X	
DXS8377	q28	183.66				X				X	X_1		
DXS10134	q28	183.96											X
DXS10147	q28	184.01											
DXS7423	q28	184.19	X	X	X	X	X	X	X	X	X_1		X
DXS10134	q28	183.96	X	X									
DXS10146	q28	183.72	X										
DXS10011	q28	188.70							X				
DXS6793	n/a	n/a											
DXS6808	n/a	n/a											
DXS10162	n/a	n/a											
DXS10164	n/a	n/a											

[A] The HumARA marker is linked to spinal and bulbar muscular dystrophy and some further disease risks and should no longer be used as a DNA marker for forensic purposes [127,128].

[B] Cytogenetic localization and genetic localization (cM Kosambi) according to Rutgers Map v.2 obtained from http://www.chrx-str.org/.

[124] 16 X-STRs are analyzed in two octaplex PCR reactions (X_1 and X_2).

Highly degraded DNA and low copy number, may be more successfully typed using short amplification products [128]. In this context, the approach of mini- (<200 bp) and midiSTRs (<300 bp) located on the X-chromosome may be of great utility in forensics. With that in mind, new multiplexes that generate low size amplicons are being developed and (re)designed [129].

X-STRs in kinship testing

Due to its particularities, X-STRs have recently gained recognition as a powerful tool to complement the information provided by autosomal STRs, particularly in kinship cases where the majority of the profiles to compare correspond to second or third degree relatives, such as grandparents-grandchildren, paternal half-sisters, paternal aunt/uncle niece, and maternal uncle-nephew [112,130]. Even when incestuous situations occur, i.e. parent plus grandparent-child, parent plus half-sib-child, and parent plus uncle-aunt-child, the use of X-STRs may help distinguish these relationships [112].

Kinship testing is phrased in terms of probabilities that sets of genes have descended from a single ancestral gene. That is, the probability they are identical-by-descent or IBD [131]. In this sense, Pinto et al. (2011) derived new formulae that accommodate the particular mode of transmission of the X-chromosome [112]. However, linkage and LD along the X-chromosome are two hot topics in forensic casework as a precise knowledge of these issues is required for the standardization of the X-STRs in daily routine. In this context, Kling et al. (2015) have developed an algorithm for likelihood calculations that accommodates the particular transmission mode of the X-chromosome, as well as linkage, LD, and mutation [45,132,133]. The algorithm is represented in Eq. 2 where L(H) is the likelihood for the hypothesis H and for its calculation all the possible inheritance patterns V= $V_1 \ldots V_N$ and the possible founder allele sets F=$F_1 \ldots F_N$ are taken into account [133].

(Eq. 2)

$$L(H) = \sum_{v_1} \ldots \sum_{v_n} \sum_{F_1} \ldots \sum_{F_n} \Pr(V_1)\Pr(F_1) \prod_{i=2}^{N} \Pr\,(v_i|v_{i-1}) \prod_{i=2}^{N} \Pr\,(F_i|F_{i-1}, \ldots, F_{i-L}) \prod_{i=1}^{N} \Pr\,(D_i|V_i, F_i)$$

Linkage and LD along the X-chromosome

The most common multiplexes that analyze autosomal markers include STRs located in different chromosomes and consequently, they are segregated independently. However, X-STRs are located in the same chromosome and therefore, the alleles of two or more markers may be transmitted together. In other words, they may be linked. This condition will make them to be inherited together from a single parent. In this context, linkage can be defined as the co-segregation of closely located loci within a family or pedigree [99]. Additionally, linkage disequilibrium (LD) is simply a non-random association between two or more alleles. It means that these alleles appear together at rates that differ from what would be expected under independence [45,132]. Therefore, two linked markers tend to show significant LD [134]. However, LD do not ensure either linkage or lack of equilibrium [100]. Hence, the use of X-STRs requires a precise knowledge not only of allele and haplotype frequencies but also of the genetic linkage and LD along the X-chromosome since the product rule for the calculation of the LR of a genetic profile, commonly used for autosomal STRs, is invalid when dealing with linked markers.

Linkage groups

For practical reasons the X-chromosome has been divided into four linkage groups located on the following positions Xp22.2, Xq12, Xq26, and Xq28. This division has been based on their physical location along the X-chromosome [128] as closely located markers are likely to segregate as a stable haplotype [135]. Up to now, different viewpoints regarding the minimum genetic distance to consider two or more markers as a stable haplotype have been discussed , i.e. from <3 cM [135–138] to 6 cM [134].

Nonetheless, even if the physical distance among two markers is small, recombination events may occur in case of existence of hotspots of recombination [134]. These hotspots are regions where the frequency and rate of recombination are most favorable [139]. Single Nucleotide Polymorphisms (SNPs) present low mutation rates and therefore, tend to show more LD than STRs [128]. In view of this, linkage and LD of certain regions along the genome can be estimated by HapMap LD plots based on SNP data. This allows determining LD blocks along the chromosomes. SNPs located within a LD block tend to show strong LD with each other. (Figure 6). In other words, if two loci are located within the same LD block, their alleles are most likely being transmitted together. On the contrary, if two loci are in different but closely located LD blocks, recombination and crossing over may occur during meiosis [134]. A measure of the linkage between two LD blocks is given by the recombination rate (θ).

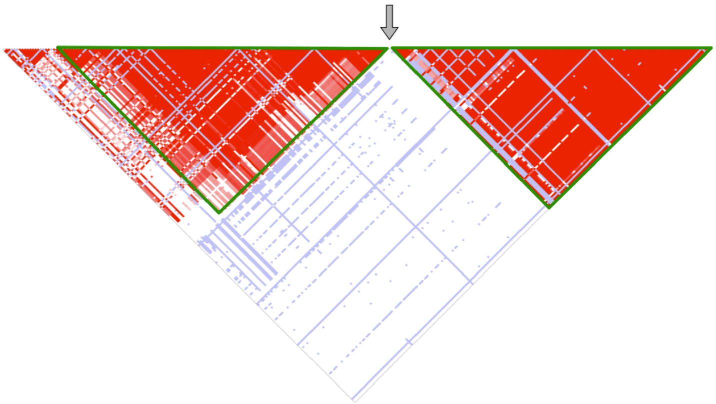


Figure 6. HapMap LD plots in a certain region along the genome. The two LD blocks and a hotspot of recombination are indicated by triangles and an arrow (↓), respectively.

Recombination rates can be estimated through extrapolation of genetic distances from haplotype maps, such as HapMap [140]. However, this method, which is based on mathematical estimation, only provides a provisionary substitute of recombination rates. In this sense, realistic rates can be measured collecting data from family studies composed by grandfather-daughter-grandson trios [138,141]. But, why investigate this relationship and not others? In this pedigree, the female inherits a non-recombinant copy of the X-chromosome from her father what will allow inferring the other phase of her genotype, i.e. what she has inherited from her mother. Then, as the haplotype of her son is known, by mere comparison can be checked if recombination events have occurred in the female as this phenomenon is limited to female meiosis [128] (Figure 7).

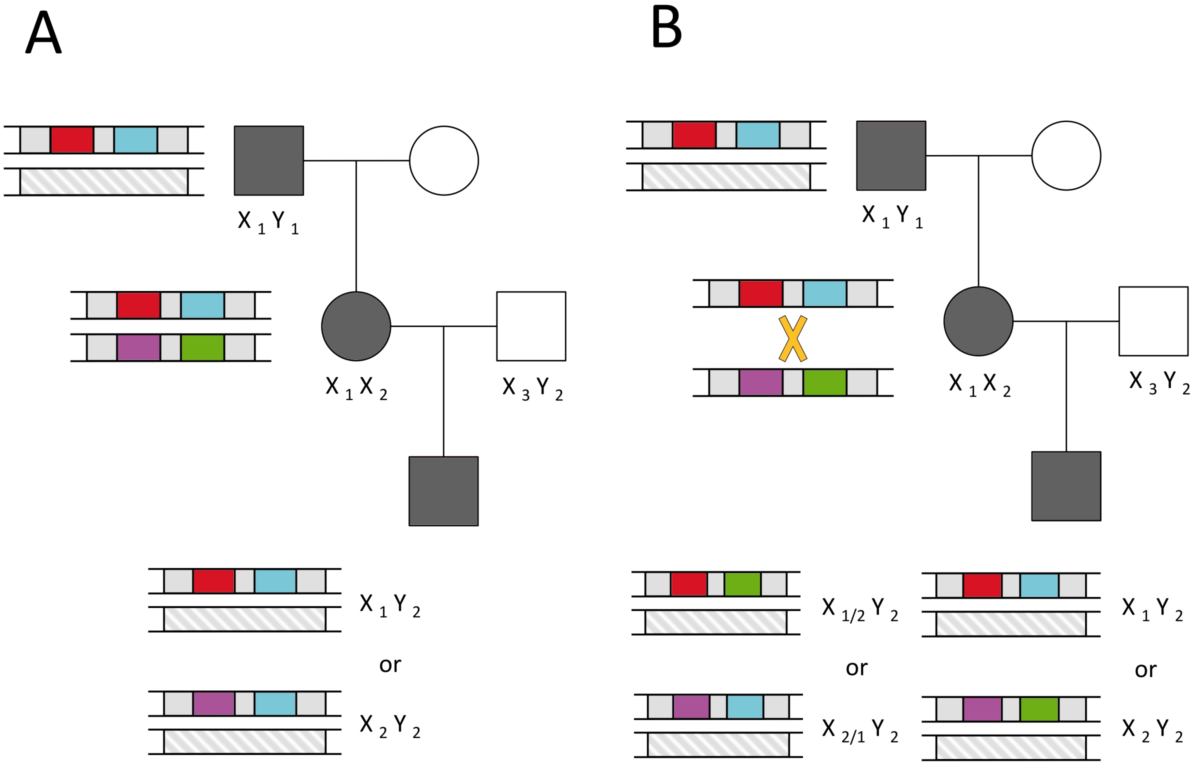


Figure 7. Grandfather-daughter-grandson trios may be valid to measure recombination rates directly through population sampling. X indicates a hotspot of recombination where recombination may occur.

Gene mapping

The mathematical relationship between recombination rates (θ) and genetic distances (ω) is described by mapping functions. One of the most widely used for human mapping is Kosambi's function [142] (Eq. 3) that considers interference during meiosis. In other words, it takes into account the fact that the presence of one genetic exchange between two homologous non-sister chromatids interferes with the coincident occurrence of a second genetic exchange [143].

(Eq. 3) $$\omega = \frac{1}{4} ln\left[\frac{(1+2\theta)}{(1-2\theta)}\right] \quad or \quad \theta = \frac{1}{2}\frac{[\exp^{4\omega}-1]}{[\exp^{4\omega}+1]}$$

On the other hand, Haldane's mapping function (Eq. 4) assumes random recombination [117].

(Eq. 4) $$\omega = -\frac{1}{2}\ln(1-2\theta) \quad \text{or} \quad \theta = \frac{1}{2}[1-\exp^{-2\omega}]$$

It is widely assumed that two loci that show a θ of 0.01 (1%) are separated by a genetic distance of 1 cM. In this context, Phillips et al. assumed that this relationship is valid until a genetic distance of ~18 cM [140].

Linkage and LD in forensic calculation

As previously mentioned, an algorithm that accommodates the inheritance mode of the X-chromosome, as well as linkage, LD, and mutation has been developed [45,132,133] (Eq. 2). In this way, the forensic calculation with X-STRs requires allele and haplotype frequencies, as well as empirical recombination rates between the markers that are being investigated.

When using loci that are in LD, haplotype frequencies cannot be estimated from allele frequencies but have to be measured directly from the population data. However, this is not always possible since it would be necessary to study a huge number of individuals in order to observe all the possible haplotypes that can be generated for each cluster [91]. In view of this, Kling et al. (2015) described the following formula that can estimate haplotype frequencies under conditions of LD (Eq. 5), where, H_k is the estimated probability for haplotype k, given c_k number of observations; α_k is the prior probability of the haplotype in linkage equilibrium (LE) that is calculated from the allele frequencies; $\sum c_k$ is the total number of haplotype observations in the database, and λ is a parameter giving weight to the prior haplotype probabilities [132].

(Eq. 5) $$H_k = \frac{(c_k+\lambda\alpha_k)}{(\sum c_k+\lambda)}$$

Additionally, in kinship testing, consideration of haplotypes may be more informative than considering individual markers as sharing rare X-STR haplotypes is strongly indicative of relatedness [144].

X-chromosome Single Nucleotide Polymorphisms in Forensic Genetics

The analysis of Single Nucleotide Polymorphisms (SNPs) may complement the results obtained with other kind of markers in forensics, such as STRs. The power of discrimination (PD) of SNPs is much lower than that of STRs. Therefore, a great number of well-balanced SNPs is required to obtain similar PD values to those obtained with the current STR multiplexes, i.e. ≥ 60 SNP markers. Despite this, SNPs are being successfully used in Forensic Genetics for solving certain deficient cases as they present some advantages that other kind of markers lacked, such as: 1) low mutation rates and 2) ability to obtain results from low copy number and highly degraded DNA samples [145–147].

SNaPshot™ minisequencing technology

Up to now, several SNP typing methodologies have been described, such as: 1) allele specific hybridization; 2) primer extension methods, e.g. SNaPshot™ minisequencing, MALDI-TOF minisequencing, and pyrosequencing; 3) allele specific oligonucleotide ligation; and 4) invasive cleveage. A good review of all these methodologies was published by Sobrino et al. [147].

SNaPshot™ minisequencing (Thermofisher Scientific, Waltham, MA, USA) is currently one of the most extended methodologies. Its success may be due to the fact that the detection is performed on an automatic capillary electrophoresis instrument which may be considered as the flagship of the laboratories specialized in kinship testing [147]. This technology is based on the dideoxy (ddNTP) single base extension of an unlabeled oligonucleotide at the 3'-end of the base immediately upstream to the SNP of interest [148]. Each ddNTP is labeled with one fluorescent dye and additionally, different length tails of non-human DNA can be added to the 5'-end of the primers. This allows multiplexing several SNPs in a single reaction as it is shown in Figure 8 [147,149–151].

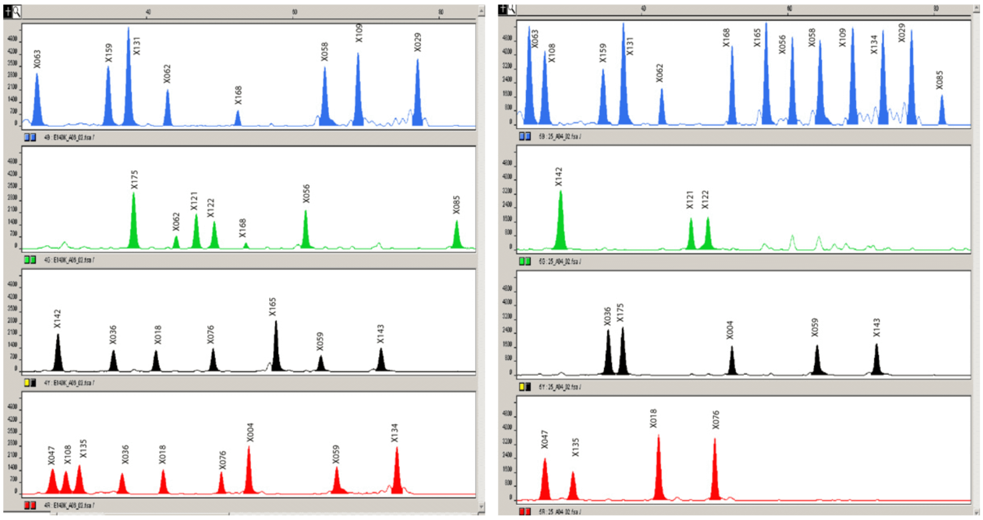


Figure 8. Genetic profile of a female (left) and a male (right) analyzed using 25 X-chromosome SNPs [150].

Forensic usefulness of X-SNPs

In previous sections the utility of the X-STRs in Forensic Genetics has been mentioned. In the same track, the SNPs located on the X-chromosome (X-SNPs) have demonstrated their utility in forensic casework [145,146], especially when dealing with highly degraded samples as only 60-80 bp fragments in length are necessary for PCR amplification [33]. In this sense, the combination of STR and SNP markers may help to solve certain complex kinship cases that autosomal markers cannot, e.g. sisters in motherless paternity cases [145] (Figure 9). In other words, they may play an important role in forensic casework for increasing the probability of parentage and PD in complex kinship cases, family reconstructions and/or human identification [33]. In this context, new MPS kits that will combine SNP and STRs located on the X-chromosome will be of great interest in forensic casework.

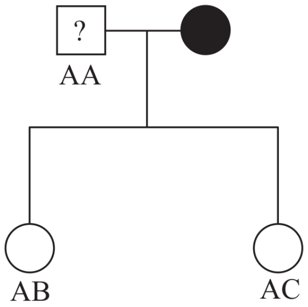


Figure 9. A motherless paternity case where AA claimed to be the father of AB and AC. ◆ indicates individuals who are not available for typing [145].

Currently, MPS allows distinguishing between Identical By State (IBS) alleles from those that are Identical By Descent (IBD) since the sequence of the motif structure of the X-STRs of interest is provided. However the conventional technology is not capable of making this differentiation. In this context, SNPs linked to STRs constitute a potential approach to help determine whether that variation is due to a mutation caused by polymerase or to non-paternity. Because of the low rate of mutation per generation of the SNP markers, i.e. 10^{-8}, this method will support for the determination of biological paternity and other kinship analyses in which mutation needs to be ruled out as grounds for exclusion [152].

During the last two decades, the genetic polymorphisms located on the autosomal chromosomes have been widely studied, such as AS-STRs and AS-SNPs. Furthermore, in recent years the study of STR and SNP markers located on the X-chromosome has become a burning issue in Forensic Genetics since their particular inheritance pattern may help to solve certain cases that other markers cannot. For that reason, the purpose of this doctoral thesis work is to deepen the forensic applicability and efficiency of the X-STR and X-SNP markers.

2. Hypothesis and objectives

Hypothesis

The analysis of Short Tandem Repeat (STR) regions or microsatellites has become the referent technique used by the great majority of the forensic laboratories for both human identification and biological kinship determination. Additionally, and due to its particular inheritance pattern, the analysis of the STRs located on the X-chromosome (X-STRs) has undergone a rapid evolution in the last few years, as they can be decisive for the resolution of some complex kinship cases.

Up to now, the decaplex of the GHEP-ISFG, as well as the Investigator® Argus X-8 and Argus X-12 Kits (Qiagen, Valencia, CA, USA) have been the most applied multiplexes in practice by the forensic community. Therefore, several allele and haplotype frequency databases have been performed with at least, one of these three sets of markers.

Nonetheless, in certain cases the resolution power obtained with these multiplexes may be insufficient. This may be in part due to the fact that few markers are included in each PCR reaction, which generally results in low power of discrimination values. In this sense, the higher the number of markers included in each reaction, the higher the chance of getting more accurate kinship determinations.

Another limitation of the above-mentioned kits is the generation of large amplification products that could not always allow obtaining optimal results, especially when dealing with highly degraded DNA samples and/or low copy number. This might be attributed to the fact that DNA degradation does not enable the obtainment of amplicons bigger in size that the available DNA targets. In this way, the mini- (<200 bp) and midiSTRs (<300 bp) located on the X-chromosome may be of great utility in Forensic Genetics.

In addition, there are certain cases where even autosomal markers are not able to distinguish between two close relatives. In these situations, the markers located on the X-chromosome may also shed some light to their resolution. Up to now, X-STRs have been much more studied than the SNP markers located on the X-chromosome (X-SNPs). However, X-SNPs have demonstrated their utility in forensic casework. In this sense, only biallelic X-SNPs have been studied and evaluated with forensic purposes. Therefore, it may be interesting to carry out the *in silico* evaluation of the forensic efficiency of the tri- and tetrallelic X-SNPs for their application in forensic casework.

In view of these, there is a demand for more efficient multiplexes that will allow the analysis of a higher number of X-chromosome markers than the currently available, as well as the obtainment of reliable results from highly degraded DNA and low copy number.

Objectives

Main objective

The main objective of this doctoral thesis work is to develop new molecular tools, based on polymorphic regions of the X-chromosome that will allow increasing the power of resolution in complex kinship cases. Moreover, the design of these tools will be focused on the obtainment of smaller-sized PCR amplification products to improve their application in samples with highly degraded DNA and/or low copy number.

Specific objectives

1. Estimating the diversity of ten X-STRs in order to determine the absence of heterogeneity among Spanish regional subpopulations, which will allow the formation of the first global Spanish allele frequency database with application in forensic casework.

2. Designing, developing, optimizing, and validating a new multiplex reaction that allows the simultaneous analysis of 17 STRs located on the X-chromosome, with applicability in both complex kinship cases and highly degraded remains.

3. Performing allele and haplotype frequency databases from Spain and other populations located on the Atlantic coast of Europe and North-West Africa by using the previously developed 17 X-STR panel that will allow its application in forensic casework.

4. Studying the efficiency of the tri- and tetrallelic SNPs located on the X-chromosome with the aim of evaluating their utility as a complement for the developed X-STR panel.

3. Materials and methods

Development of Multiplex Systems for the Analysis of X-STRs

Marker selection and primer design

Until now, many forensic databases have been performed with the markers included in the decaplex of the GHEP [113,114] and/or the Investigator® Argus X-12 Kit. Therefore, eleven markers included in these two multiplexes were selected for the new X-STR panel (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902, DXS6789, and DXS10079). In addition, six markers were included to increase the resolution power of the current multiplexes (DXS6801, DXS6799, DXS6800, DXS10075, DXS6807, and DXS6803). The main criterion for selection of new markers was a heterozygosity value of ≥ 0.6 [56, 141,153–155].

The amplification primers for the markers included in the decaplex of the GHEP-ISFG were the same as described in Gusmao et al. (2009) [114], except for DXS7423, DXS7132, and DXS9902 that were modified by the addition of a nonspecific nucleotide tail [156]. In that way, the primer sequences for the markers DXS6801, DXS10079 and DXS10075 were obtained from Castañeda et al. (2013) [129]. The remaining primers were newly designed for this study.

DNA sequences of the flanking regions of the X-STRs of interest were obtained from the GenBank® genetic sequence database (https://www.ncbi.nlm.nih.gov/genbank/) [157]. Allele range of each marker was determined according to the literature [56, 114, 141, 153–155,158–162]. The number of alleles of each locus was taken into account to avoid overlapping between neighboring markers. Primer sequences were designed using PerlPrimer v1.1.21 software [163]. Specificity for the X-chromosome, as well as the lack of interaction between primers were evaluated by using BLASTN alignment tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Autodimer v.1.0 software [164], respectively. All forward primers were modified by the addition of a fluorescent dye at their 5' end with the exception of the marker DXS10075, labelled at the 5' end of the reverse primer due to a punctual mutation in the hybridization region. The fluorescent dyes used to modify the primers were: 5-FAM, YAKIMA YELLOW, ATTO 550, and ATTO 565 (Eurofins Genomics, Ebersberg, Germany). Definitive primer sequences are shown in Table 7.

Table 7. Primer sequences and fluorescent dyes for each locus included in the newly designed multiplex. ^{1} and ^{2} indicate the markers included in the decaplex of the GHEP-ISFG and the Investigator® Argus X-12 Kit, respectively. The non-specific nucleotide tails added to the primer sequences described in Gusmao et al. (2009) [114] are underlined.

Locus	*Dye*	*Forward sequence*	*Reverse sequence*
DXS8378 $^{1, 2}$	5'-FAM	TTAGGCAACCCGGTGGTCC	ACAAGAACGAAACTCCAACTC
DXS9898 ^{1}	5'-FAM	CGAGCACACCTACAAAAGCTG	TAGGCTCACCTCACTGAGCA
DXS7133 ^{1}	5'-FAM	CACTTCCAAAAGGGGAAAAA	ACTTGTACTTGGTGGGAGGAA
GATA31E08 ^{1}	5'-FAM	GCAAGGGGAGAAGGCTAGAA	TCAGCTGACAGAGCACAGAGA
GATA172D05 ^{1}	5'-YAKYE	TAGTGGTGATGGTTGCACAG	ATAATTGAAAGCCCGGATTC
DXS6801	5'-YAKYE	CATAATCACATGAGTCATTTCCT	ATCTGTATTAGTTATGAGTTTCCAG
DXS7423 $^{1, 2}$	5'-YAKYE	GTCTTCCTGTCATCTCCCAAC	ACGTCGTGAAAGTCTGACAATAGCTTAGCGCCTGGCACAT
DXS6809 ^{1}	5'-YAKYE	TCCATCTTTCTCTGAACCTTCC	TGCTTTAGGCTGATGTGAGG
DXS6799	5'-ATTO 550	ACTAGCAAACTGAATTTAGTAATGTG	GCACATGGGATGGATGGATA
DXS7132 $^{1, 2}$	5'-ATTO 550	TCCCCTCTCATCTATCTGACTG	CACGTCGTGAAAGTCTGACAAAACCACTCCTGGTGCCAAACTCTATT
DXS9902 ^{1}	5'-ATTO 550	CTGGGTGAAGAGAAGCAGGA	AAGTCTGACAAGGCAATACACATTCATATCAGGA
DXS6800	5'-ATTO 550	TTCAGAGGGCCTATTGTGG	TCAGACTGGCTGACACTTAGG
DXS6789^{1}	5'-ATTO 550	CTTCATTATGTGCTGGGGTAAA	ACCTCGTGATCATGTAAGTTGG
DXS10075	5'-ATTO 565	AGTTATTGCAGAGAAGAATCATATC AGATATTGCAGAGAAGAATCATATC	GACTACCTCTGCTCCCTT
DXS10079 ^{2}	5'-ATTO 565	GTGACCAAGTGAGACCAA	TTGTTGAGAACTTTTGCATCA
DXS6807	5'-ATTO 565	TTTCACTTGAGTTTAGTAGTGTTTG	ATCATAAGTAAACATGTATAGGAAAAAGC
DXS6803	5'-ATTO 565	CTAGAAATGTGCTTTGACAGGA	GAGTAAGACTGTTAAACAGGCA

Human DNA control samples

Control DNAs 2800M and K562 (Promega® Corporation, Madison, WI, USA), as well as 9947A from AmpFLSTR™ PCR Amplification Kit (ThermoFisher Scientific, Waltham, MA, USA) were used to: 1) set up the PCR amplification conditions and 2) carry out sensitivity and stability studies.

Primer optimization

With the aim of covering the highest allele diversity worldwide, the following populations were analyzed with the original panel: Asians from Thailand, European Caucasoids from Spain, Africans from Malawi and Equatorial Guinea, and Hispanics from Colombia (Table 8).

After analyzing the above-mentioned population samples, some of the originally designed primers were modified by adding non-specific nucleotide tails [156] to avoid allele overlapping between neighboring markers (Table 7).

Population samples

Several populations have been studied (Table 8). All the samples were obtained from volunteer donors following the ethical standards of Helsinki Declaration.

Table 8. Summary of the population samples analyzed in the present doctoral thesis work. N= number of individuals and BNADN= *Banco Nacional de ADN Carlos III* - Spanish national DNA bank (BNADN Ref. 12/0031).

Study	*Population*	*Sample Size*	*Provided by*
Study number 1	Alicante	(N= 141, XX= 51, XY= 90)	Miguel Hernández University – UMH – Pathology and Surgery Dept.
(N= 1136)	Andalusia	(N= 198, XX= 100, XY= 98)	BNADN
	Asturias	(N= 131, XX= 63, XY= 68)	BNADN
	Barcelona	(N= 199, XX= 98, XY= 101)	BNADN
	Basque Country	(N= 147, XY= 147)	Sampling [A]
	Galicia	(N= 121, XX= 51, XY= 70)	BNADN
	Madrid	(N= 199, XX= 100, XY= 99)	BNADN
Study number 2	Asians from Thailand	(N=138, XY= 138)	Colorado College – Molecular Biology Dept.
(N= 488)	Europeans from Spain	(N= 101, XX= 36, XY= 65)	BNADN
	Africans from Malawi and Equatorial Guinea	(N= 135, XX= 73, XY= 62)	University of Santiago de Compostela - USC
	Hispanics from Colombia	(N= 114, XX= 57, XY= 57)	University of Antioquia
Study number 3	Alicante	(N= 56, XY= 56)	Miguel Hernández University – UMH – Pathology and Surgery Dept.
(N= 593)	Andalusia	(N= 123, XX= 59, XY= 64)	BNADN
	Aragon	(N= 50, XY= 50)	University of Zaragoza – Forensic Medicine Dept.
	Barcelona	(N= 112, XX= 58, XY= 54)	BNADN
	Basque Country	(N= 75, XX= 9, XY= 66)	Sampling [A]
	Galicia	(N= 74, XX= 41, XY= 33)	BNADN
	Madrid	(N= 103, XX= 35, XY= 68)	BNADN
Study number 4	Brittany	(N= 179, XY= 179)	University of Bretagne Occidentale – UBO
(N= 513)	Ireland	(N=100, XY= 100)	Trinity Biomedical Sciences Institute – Academic Unit of Neurology
	Northern Portugal	(N=79, XY= 79)	National Institute of Legal Medicine and Forensic Sciences of Porto
	Casablanca	(N= 155, XX= 13, XY= 142)	Université Mohammed VI des Sciences de la Santé

[A] The sampling was carried out by BIOMICs Research Group once favorable ethical reports were obtained (Faculty of Pharmacy UPV/EHU, September 2008 (CEISH/119/2012).

DNA extraction

DNA from bloodstains of the African population was isolated by organic extraction (study number 2). Additionally, DNA of a sample set of Alicante population was extracted from buccal swabs with Chelex-100® chelating resin suspension (Sigma-Aldrich Corporation, St. Louis, MO, USA) (study number 3). The remaining populations' DNA samples had already been previously extracted.

The protocols applied for each of the above-mentioned methods are described below (Figure 10):

1. Organic extraction:

- Cut 1 cm^2 of the bloodstain and put it into a centrifuge tube.
- Add 470 µl of lysis buffer (Tris 10 mM, EDTA 10 mM, NaCl 0.2 M, SDS 2%), pH 8.
- Add 10 µl of proteinase K (20 mg/ml).
- Incubate overnight at 37 ºC with shaking (850 rpm).
- Mix for 5-10 seconds using a vortex mixer and centrifuge.
- Add 200 µl of phenol and shake by hand thoroughly for 5-7 minutes.
- Centrifuge for 10 minutes at 13,000 rpm.
- Carefully remove the upper aqueous phase, and transfer the layer to a new tube.
- Add 200 µl of phenol: chloroform: isoamyl alcohol (25:24:1) and shake by hand thoroughly for 5-7 minutes.
- Centrifuge for 10 minutes at 13,000 rpm.
- Carefully remove the upper aqueous phase, and transfer the layer to a new tube.
- Add one volume of chloroform: isoamyl alcohol (24:1) and shake by hand thoroughly for 5-7 minutes.
- Centrifuge for 10 minutes at 13,000 rpm.
- Carefully remove the upper aqueous phase, and transfer the layer to a new tube.
- Add the following reagents to the aqueous phase (in the listed order) and turn over the tubes 20-30 times:
    - 1/10 volumes of AcNa 2 M.
    - 1 µl of glycogen (20 mg/ml).
    - 2 volumes of absolute ethanol at -20 ºC.
- Place the tube at -20 ºC for 60 minutes to precipitate the DNA from the sample.

- Centrifuge the sample at -10 ºC for 20 minutes (13,000 rpm) to pellet the DNA.
- Carefully remove the supernatant without disturbing the DNA pellet.
- Add 1 ml of ethanol at 80% and centrifuge at -10 ºC for 2 minutes (13,000 rpm).
- Discard the supernatant.
- Repeat the previous two steps with 1 ml of ethanol at 70%.
- Remove the remaining ethanol in the DNA concentrator (about 20 minutes) at room temperature.
- Resuspend the pellet in 40 µl of sterile Milli-Q water. This volume now contains the DNA.

2. Chelex-100® chelating resin suspension:

- Incubate the buccal swab in 1 ml of sterilized milli-Q water at room temperature for about 30 minutes, with shaking every 5 minutes.
- Centrifuge for 1 minute at 13,000 rpm.
- Remove the supernatant without disturbing the pellet, except for 50 µl.
- Resuspend the pellet in the remaining volume.
- Add 150 µl of Chelex-100® resin at 5%.
- Incubate for 15 minutes at 56 ºC.
- Mix for 5-10 seconds using a vortex mixer.
- Incubate for 8 minutes at 100 ºC.
- Centrifuge for 3 minutes at 13,000 rpm.
- Transfer the supernatant containing the DNA to a new tube.

DNA quantification

DNA was quantified by using the Scientific NanoDrop™ 1000 Spectrophotometer (Thermofisher Scientific, Wilmington, DE, USA) and then diluted in Milli-Q water to a 1-3 ng/µl concentration, except for DNA extracted with Chelex-100®.

PCR amplification, capillary electrophoresis and data analysis

PCR reaction consisted of 5 µl of QIAGEN Multiplex PCR Kit (Qiagen, Valencia, CA, USA), 0.5 µl of primer mix, and Milli-Q water in order to reach a final reaction volume of 10 µl. The final concentration of each primer in the primer mix was 0.2 µM, except for the markers DXS6799 (0.4 µM), DXS7132 (0.4 µM), and DXS6809 (0.6 µM). 1-3 ng of genomic DNA were used in standard reactions to avoid allelic dropouts in female samples.

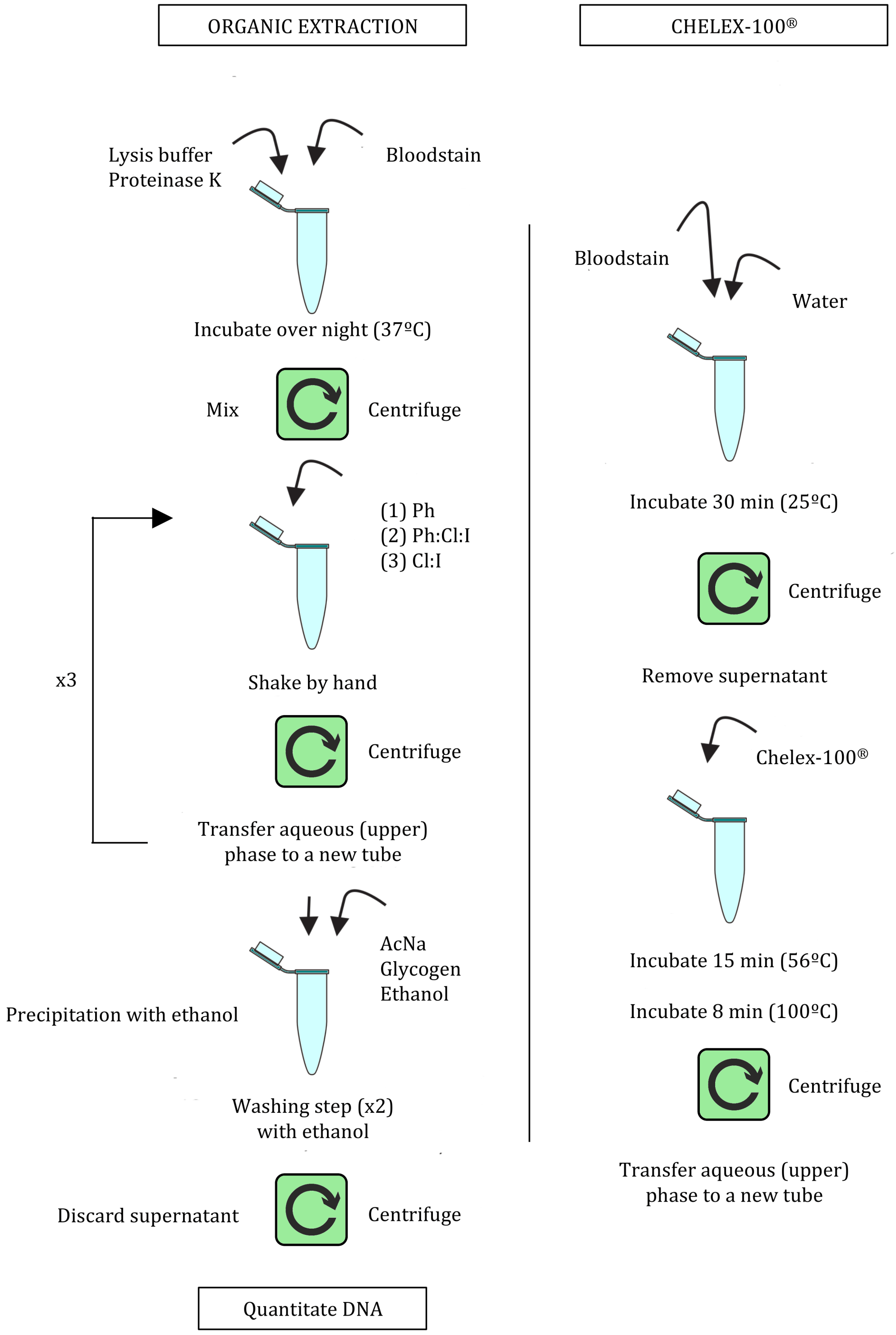


Figure 10. A schematic representation of the two extraction methods conducted. On the left, organic extraction and on the right, Chelex-100® chelating resin suspension. Abbreviations: Ph = phenol, Ph:Cl:I = phenol: chloroform: isoamyl alcohol (25:24:1), and Cl:I = chloroform: isoamyl alcohol (24:1).

The amplification was carried out on the GeneAmp® 9800 PCR System (Thermofisher Scientific, Waltham, MA, USA) under the same thermocycling conditions used for the decaplex of the GHEP-ISFG [114], which consisted in: pre-incubation for 15 min at 95 ºC, followed by ten cycles of 30 s at 94 ºC, 90 s at 60 ºC, 60 s at 72 ºC, and 20 cycles of 30 s at 94 ºC, 90 s at 58 ºC, and 60 s at 72 ºC with a final incubation for 60 minutes at 72 ºC. If adenilation is detected, the final incubation at 72º C can be extended to 90 minutes.

Amplification of samples and performance of the PCR reaction were verified by conventional agarose gel electrophoresis. The amplification products were analyzed by mixing 9 µl of Hi-Di™ Formamide (ThermoFisher Scientific, Waltham, MA, USA), 1 µl of each PCR amplification product, and 0.25 µl of GeneScan™ LIZ-500™ Size Standard (ThermoFisher Scientific, Waltham, MA, USA). After denaturation, capillary electrophoresis of PCR products was conducted on an ABI PRISM 3130 Genetic Analyzer with the POP7 polymer (Thermofisher Scientific, Waltham, MA, USA). Electrophoresis data were analyzed with GeneMapper ID software version 4.0 (Thermofisher Scientific, Waltham, MA, USA).

Allelic ladder development

To assure allele identification an allelic ladder was developed. Each allele was independently amplified by PCR. The reaction consisted of 20 µl of QIAGEN Multiplex PCR Kit (Qiagen, Valencia, CA, USA), 0.25 µl of both forward and reverse primers at 10 µM concentration, and 2 ng of genomic DNA. To measure the efficiency of the PCR reaction, a migration on an agarose gel was carried out. The amplicon concentration of each allele was estimated by comparison to the electrophoretic ladder. According to the intensity of the bands, the amplification products were proportionally pooled. Then, in order to assure the equal height of the alleles into each marker, the mixture was analyzed on an ABI PRISM 3130 Genetic Analyzer (Thermofisher Scientific, Waltham, MA, USA). Finally, when all the alleles of each X-STR were balanced, all the markers were pooled together according to their relative fluorescence units (RFUs).Artifact removal was performed using the MinElute® PCR Purification Kit (Qiagen, Valencia, CA, USA).

Validation of Multiplex Systems for the Analysis of X-STRs

The new panel has been developed and validated taking into account the SWGDAM Validation Guidelines for DNA Analysis Methods (approved in December 2012) [77].

Precision and accuracy

Precision characterizes the degree of mutual agreement among a series of individual measurements, values and/or results. Precision depends only on the distribution of random errors and does not relate to the true value or specified value. The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results. On the other hand, accuracy is the degree of conformity of a measured quantity to its actual (true) value. Accuracy of a measuring instrument is the ability of a measuring instrument to give responses close to a true value [77].

Repeatability of results was evaluated by five independent electrophoresis migrations of the allelic ladder on an ABI PRISM 3130 Genetic Analyzer (Thermofisher Scientific, Waltham, MA, USA). The mean of bp size and standard deviation for each allele was calculated.

Concordance of results was evaluated for the markers included in the decaplex of the GHEP-ISFG (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902, and DXS6789) by comparisons with the results obtained in study number 1 for the Iberian Peninsula population. Additionally, the control DNAs 2800M and K562 (Promega® Corporation, Madison, WI, USA), as well as the 9947A (ThermoFisher Scientific, Waltham, MA, USA) were used to test the concordance of the rest of the markers included in the newly designed panel (DXS6799, DXS6800, DXS6801, DXS6803, DXS6807, DXS10075, and DXS10079).

Sensitivity and stability

Sensitivity indicates the ability to obtain reliable results from a range of DNA quantities, to include the upper and lower limits of the assay [77]. To evaluate the minimum quantity of DNA required to obtain reliable results, 2800M and 9947A control DNAs were analyzed in triplicate using the following amounts of DNA: 30, 20, 10, 1.6, 1, 0.4, 0.2, 0.1, 0.05, and 0.025 ng.

Stability indicates the ability to obtain results from DNA recovered from biological samples deposited on various substrates and subjected to various environmental and chemical insults [77]. Stability of the new panel was evaluated by amplifying 1 ng of 9947A control DNA in presence of two common inhibitors in forensic casework: haematin and humic acid (Sigma-Aldrich Corporation, St. Louis, MO, USA). This study was performed in triplicate using the following concentrations of inhibitors: 5,000, 3,000, 1,500, 1,000, 750, 500, 300, 150, and 100 µM of haematin, as well as 3,000, 2,000, 1,000, 500, 300, 250, 200, 100, 50, and 25 ng/µl of humic acid.

Determination of stutter percentage and heterozygous peak height ratio

The percentage of observed stutter peaks at each locus was examined by dividing the stutter peak height by the corresponding allele peak height [8]. Additionally, the standard deviation for each marker was calculated. On the other hand, Peak Height Ratio (PHR), i.e. the proportion of the less intense allele relative to the more intense allele at a given heterozygous genotype, was evaluated by dividing the peak height of the allele with the lower RFU value by the peak height of the allele with the higher value [165]. Additionally, the standard deviation for each marker was calculated.

Off-ladder alleles

Off-ladder alleles were checked by size comparison with other samples that presented neighboring alleles. With this purpose, genomic DNA of the compared samples was blended, amplified together in a single PCR reaction and checked by capillary electrophoresis. However, some off-ladder variants did not present neighboring alleles to compare by size. In such cases, sequencing was carried out. Sample preparation steps for sequencing were carried out as follows:

- Amplification of DNA considering the previously described PCR conditions.
- Mix 10 µl of a post-PCR reaction with 0.5 µl of Exonuclease I and 2.5 µl of SAP enzymes (Takara Clontech, Kyoto, Japan).
- Incubate for 45 minutes at 37 ºC to degrade remaining primers and nucleotides.
- Incubate for 15 minutes at 80 ºC to inactivate the enzymes.
- Mix 3.5 µl of the cleaned PCR product with 6.5 µl of BrightDye® Terminator Cycle Sequencing Kit v3.1 Mix (Nimagen, Nijmegen, The Nertherlands) composed by 1 µl of BDT reaction Mix, 0.25 µl of primers 10 at µM concentration, and 5.25 µl of ddH_2O.
- Incubate for 1 minute at 96 ºC followed by 24 cycles of 10 s at 96 ºC, 5 s at 50 ºC, and 75 s at 60 ºC.

- Mix 10 µl of BDT reaction product with the reagents of the BigDye XTerminator® Purification Kit: 22.5 µl of SAM solution and 5.5 µl of resin solution (Thermofisher Scientific, Waltham, MA, USA).
- Mix for 35 minutes using a vortex mixer (2,250 rpm).
- Centrifuge for 270 s at 2,000 rpm.
- Mix 5 µl of XTerminator® reaction with 5 µl of Hi-Di™ Formamide (Thermofisher Scientific, Waltham, MA, USA). DNA is ready for sequencing.

Additionally, to confirm a previously undescribed allele variant observed in a heterozygous genotype that did not present neighboring alleles to compare, sequencing was performed from the DNA bands of each allele separated by electrophoresis in an agarose gel. The DNA bands were carefully cut with a scalpel from the gel, extracted from the agarose by using the QIAEX II® Gel Extraction Kit (Qiagen, Valencia, CA, USA), and purified with the MinElute PCR Purification Kit (Qiagen, Valencia, CA, USA). The DNA was extracted from the agarose gel as follows:

- Excise the DNA band from the agarose gel with a scalpel and introduce it in a pre-weighed centrifuge tube.
- Weigh the gel slice and add 6 volumes of Buffer QX1.
- Add 30 µl of QIAEX II® to the sample.
- Incubate for 10 minutes at 50 ºC with shaking every 2 minutes.
- Centrifuge the sample for 30 s and carefully remove supernatant.
- Wash the pellet with 500 µl of Buffer QX1. Resuspend, centrifuge and remove all traces of supernatant.
- Wash the pellet twice with 500 µl of Buffer PE. Resuspend, centrifuge and remove all traces of supernatant.
- Air-dry the pellet for 30 minutes.
- Add 20 µl of water and resuspend the pellet by vortexing.
- Centrifuge for 30 s, and carefully pipet the supernatant into a clean tube. The supernatant now contains the purified DNA.

Statistical Analyses

In all the formulae described along this section, $p_{i/j}$ is the frequency of i^{th} / j^{th} allele in a population of n samples and x_i is the frequency of i^{th} genotype.

Population genetic parameters

Allele frequencies, as well as expected heterozygosity (Eq. 6) were calculated from both male and female samples. On the other hand, male samples were used to calculate haplotype frequencies. All aforementioned parameters were calculated by using Arlequin software v.3.5.1.2 [93].

(Eq. 6) $$H \frac{n}{(n-1)}\left[1-\sum_{j=1}^{k}({}^{n_j}/_{n})^2\right]=\frac{n}{(n-1)}\left[1-\sum_{j=1}^{k}(p_j)^2\right]$$

The following steps were taken to calculate both allele frequencies and expected heterozygosity:

- Open project.
- Choose the input file.
- Go to “Settings” and select the following options:
    - Molecular diversity indices.
        - Standard diversity indices.
            - Output sample allele frequencies for all loci.
        - Molecular diversity indices.
- Press “Start” button.

On the other hand, haplotype frequencies were calculated as follow:

- Open project.
- Choose the input file.
- Go to “Settings” and select the following options:
    - Haplotype inference.
        - Estimate haplotype frequencies by mere counting.
- Press “Start” button.

The input files for the calculation of both allele (1) and haplotype (2) frequencies, are shown below:

```
#==================================
#Project file created by Arlequin
#==================================

[Profile]

    NbSamples= N   # Number of populations.
    DataType=MICROSAT
    GenotypicData=0
    GameticPhase=1
    LocusSeparator=TAB
    RecessiveData=0
```

```
     MissingData='?'

#    Some advanced settings the experienced user can uncomment.
#    Frequency= ABS
#    FrequencyThreshold= 1.0e-5
#    EpsilonValue= 1.0e-7

[Data]
```

(1) Allele frequencies and expected heterozygosity:

```
[[Samples]]

SampleName= "P1 name"     # Name of the first population.
SampleSize= N             # Number of haplotypes considered, i.e. no. of men + no. of women (x2).
SampleData= {             # In the case of samples of women, genotypes were divided into A and B.

Ind. 1      1      10      8.3     10      13      6       11      15      31      11
  12        11     19      20      13      19      11      12.3
Ind. 2A     1      10      8.3     14      12      6       12      14      30      13
  14        11     16      23      18      19      14      13
Ind. 2B     1      10      11      12      12      9       12      14      35      10
  13        12     16      20      18      20      11      11
}
```

(2) Haplotype frequencies:

```
[[Samples]]

SampleName= "P1 name"     # Name of the first population.
SampleSize= N             # Number of haplotypes considered, i.e. no. of men.
SampleData= {

Ind. 1      1      13      14      17
Ind. 2      1      14      15      19
Ind. 3      1      12      17      19
}
```

Hardy-Weinberg Equilibrium

In the present work, the exact test of HWE has been used [98]. This test is analogous to Fisher's exact test on a two-by-two contingency table but extended to a contingency table of arbitrary size. First a contingency table is built where the k x k entries of the table are the observed allele frequencies and k is the number of alleles. The probability to observe the table under the null-hypothesis of no association is given by Eq. 7, where H is the number of heterozygote individuals. The HWE test was performed by using Arlequin software v.3.5.1.2 [93].

(Eq. 7)

$$L_0 = \frac{n!\prod_{i=1}^{k} n_{i*}!}{(2n)!\prod_{i=1}^{k}\prod_{j=1}^{i} n_{ij}!}2^H$$

The following steps were taken to perform the HWE test:

- Open project.
- Choose the input file.
- Go to “Settings” and select the following options:
  - Hardy-Weinberg.
  - Perform exact test of Hardy-Weinberg equilibrium.
    - Set no. of steps in Markov chain to 1000000.
    - Set no. of dememorization steps to 100000.
  - HWE test type.
    - Locus by locus.
- Press “Start” button.

The input file for the HWE test is shown below:

```
#=================================
#Project file created by Arlequin
#=================================

[Profile]

    NbSamples= N   # Number of populations.
    DataType=MICROSAT
    GenotypicData=1
    GameticPhase=1
    LocusSeparator=TAB
    RecessiveData=0
    MissingData='?'

[Data]

[[Samples]]

SampleName= "P1 name"      # Name of the first population.
SampleSize= N              # Number of haplotypes considered, i.e. no. of women.
SampleData= {

Ind. 001    1       10      8.3     10      13      6       11      15      31      11
  12        11      19      20      13      19      11      12.3    10      12      10
  14        10      12      15      33      12      17      12      19      22      17
  20        11      13
Ind. 002    1       11      11      9       9       6       11      14      32      11
  13        11      19      19      17      18      11      11      12      13      9
  12        10      11      16      33      12      15      12      20      21      18
  19        14      12
}
```

Determination of the Linkage and LD of the Selected Markers

Linkage and LD analyses were based on both (1) SNP data provided by the HapMap Project and (2) X-STR population data derived from our studies. The first provides information about linkage and LD within a certain region along the genome while the second provides information about the LD between two different loci.

1. Linkage and LD analyses based on SNP data:

Physical location of markers

To situate the markers along the X-chromosome, the hybridization regions of both forward and reverse primers of each locus were identified through Primer-BLAST alignment tool (http://blast.ncbi.nlm.nih.gov/tools/primer-blast). This search indicates the position of each X-STR according to the reference sequence GRCh38 provided by the Genome Reference Consortium (GRC). However, the genome positions in the HapMap Project are given according to the reference sequence NCBI36 (hg18). Hence, a conversion between the two sequences is necessary to be able to localize the X-STRs of interest in the HapMap Project. For that, the rs number of the SNPs located upstream and downstream of the X-STRs of interest was used as reference since it does not change between reference sequences.

HapMap SNP genotype data

We used the physical location of all the loci included in the new panel to download the HapMap SNP genotype data of all the regions among consecutive markers (HapMap Data Rel 28 PhaseII+III, August 10, on NCBI B36 assembly dbSNP b126) for the Utah residents with Northern and Western European ancestry from CEPH collection (CEU).

Hotspots of recombination

The data provided by HapMap Project were used to determine the LD blocks between consecutive markers by using Haploview software v4.2 [166] in order to obtain the number of hotspots of recombination between two neighboring loci. The following steps were taken:

- Open Haploview software v4.2.
- Click on “File” button and select “Open new data”.
- Choose “HapMap Format” option and browse the file of interest (format: Unix Executable File).
- Set the following conditions:

    - Ignore pairwise comparisons of markers > 500 kb apart.
    - Exclude individuals with > 50% missing genotypes.
- Click on “Ok” button to perform the analysis.
- To display the LD blocks click on the “LD Plot” tab. More information about the Haplotype’s composition is provided by the “Haplotypes” tab.

2. Pairwise linkage disequilibrium test:

Linkage disequilibrium (LD) can be defined as the non-random association of alleles at two or more loci in a population [99]. This occurs when two alleles appear together at rates that differ from what would be expected under independence [45]. LD is specific for a certain population and may reflect its evolutionary past [100]. The calculation of LD depends on whether the phase composition of the sample is known or not. In the first case, the test is an extension of Fisher exact probability test [167]. First, a contingency table of k_1 x k_2 is generated where k_1 and k_2 are the number of alleles at locus 1 and 2, respectively. The test consists in obtaining the probability of finding a table with the same marginal totals and which has a probability equal or less than the observed table. Under the null hypothesis of no association between the two tested loci, the probability of the observed table is given by Eq. 8, where n_{ij} represents the counts of the haplotypes that have the i^{th} allele at the first locus and the j^{th} allele at the second locus, n_{i*} is the overall frequency of the i^{th} allele at the first locus (i= 1,… k_1), and n_{*1} is the count of the i^{th} alleles at the second locus (i= 1,… k_2). In order to study a large amount of all possible contingency tables, the Marcov’s chain is usually used [92].

(Eq. 8) $$L_0 = \frac{n!}{\prod_{i,j} n_{ij}!} \prod_i (n_{i^*}/n)^{n_{i^*}} \prod_j (n_{*j}/n)^{n_{*j}}$$

Alternatively, when the gametic phase is unknown, the likelihood ratio (LR) test of LD between a pair of loci is tested from genotypic data by Eq. 9, where the likelihood of the data assuming linkage equilibrium (L_{H*}) is computed based on the haplotype frequencies that may be estimated directly from the allele frequencies. On the other hand, the likelihood of the data not assuming linkage equilibrium (L_H) is obtained by applying the EM algorithm to estimate haplotype frequencies [168–170].

(Eq. 9) $$S = -2\log\left(\frac{L_{H^*}}{L_H}\right)$$

In the present study, the pairwise LD test was performed applying the Eq. 8 by using Arlequin software v.3.5.1.2 [93] from the X-STR genotypic data. The following instructions were followed:

- Open project.
- Choose the input file.
- Go to "Settings" and select the following options:
  - Pairwise linkage.
    - Linkage disequilibrium between all pairs of loci.
  - LD settings.
    - Set no. of steps in Markov Chain to 10000.
    - Set no. of dememorization steps to 10000.
  - LD coefficients between pairs of alleles at different loci.
    - Compute D, D', and r2 coefficients.
      - Generate histograms and table in file "LD_DIS.XL".
        - Set the significance level considering the Bonferroni correction.
- Press "Start" button.

The input file for the analysis is shown below:

```
#================================
#Project file created by Arlequin
#================================

[Profile]

    NbSamples= N   # Number of populations.
    DataType=MICROSAT
    GenotypicData=1
    GameticPhase=1
    LocusSeparator=TAB
    RecessiveData=0
    MissingData='?'

SampleName= "P1 name"     # Name of the first population.
SampleSize= N             # Number of haplotypes considered, i.e. no. of men.
SampleData= {

Ind. 001   1    10    11    10    13    6     11    14    33    11
  12       11   19    17    13    19    11    10.3
Ind. 002   1    12    11    12    13    7     11    15    33    11
  12       13   19    17    15    19    11    11.3
}
```

Population differentiation

The genetic distances based on F_{ST} were calculated with Arlequin software v.3.5.1.2 [93] from the male and female samples. The following steps were taken:

- Open project.
- Choose the file of interest.
- Go to “Settings” and select the following options:
  - Population comparisons.
    - Compute pairwise F_{ST}.
      - Genetic distance settings.
        - Slatkin’s distance.
        - Reynold’s distance.
        - Compute pairwise differences (pi).
          - Estimate relative population sizes.
            - Set no. of permutations to 10000.
            - Set the significance level considering Bonferroni correction.
  - Population differentiation
    - Exact test of population differentiation.
      - Exact test settings.
        - No. of steps in Markov Chain: 10000.
        - No. of dememorization steps: 10000.
          - Set the significance level considering the Bonferroni correction.
- Press “Start” button.

The input file for the analysis is shown below:

```
#==================================
#Project file created by Arlequin
#==================================

 [Profile]

    NbSamples= N   # Number of populations.
    DataType=STANDARD
    GenotypicData=0
    GameticPhase=0
    LocusSeparator=TAB
    RecessiveData=0
    MissingData='?'

SampleName="P1 name"      # Name of the first population.
SampleSize= N             # Number of haplotypes considered, i.e. no. of men + no. of women x2.
SampleData= {             # In the case of samples of women, genotypes were divided into A and B .

Ind. 1      1     10    8.3   10    13    6     11    15    31    11
  12        11    19    20    13    19    11    12.3
Ind. 2A     1     10    8.3   14    12    6     12    14    30    13
  14        11    16    23    18    19    14    13
Ind. 2B     1     10    11    12    12    9     12    14    35    10
  13        12    16    20    18    20    11    11
}
```

In order to obtain a representation of the genetic distances, 2D- and 3D-nonmetric multidimensional scaling (NMDS) analysis were carried out using PAST software v.3.04

[171] and the x-y-z coordinates were represented using the rgl package [172] for R software [173]. The R script used in the study no. 4 is shown below:

```
> setwd(“~/Desktop/R”) # First insert the path of the file.
> install.packages(“rgl”) # Install and open the “rgl” package.
> library(“rgl”)
# Introduce below the coordinates obtained through the 3D-MDS analysis for each of the studied population.
> POP1 <- c(x1, y1, z1)
> POP2 <- c(x2, y2, z2)
> POP3 <- c(x3, y3, z3)
> POPn <- c(xn, yn, zn)
> data <- matrix(c(POP1 , POP2 , POP3 , ⋯ , POPn), ncol= n, nrow= 3)
> cnames < c("POP1","POP2","POP3", ... "POPn”)
> rnames <- c("Axis 1", "Axis 2", "Axis 3")
> colnames(data) <- cnames
> rownames(data) <- rnames
> data
> data2 <- t(data)
# The following parameters may be modified to obtain different visual representations.
> plot3d(data2, xlab= "Axis 1", ylab= "Axis 2", z= "Axis 3", type= "h", col = rainbow(n), lwd = 2.5)
> spheres3d(data2, radius = 0.007, col= rainbow (n))
> grid3d(side= "z", at=list(z=0), col= "gray", lwd = 1.5, lty= 1.5, n=6)
> text3d(data2, text=rownames(data2), adj=1.4, font=7, cex= 0.75)
> snapshot3d(filename="filename.png", fmt='png')
```

Parameters of Forensic Interest

Mean exclusion index in father/daughter duos (MEC$_D$) (Eq. 10) and in trios involving daughters (MEC$_T$) (Eq. 11) [174,175], as well as power of discrimination in males (PD$_M$) (Eq. 12) and females (PD$_F$) (Eq. 13) [106] were calculated for the analyzed populations by using the online tool of the Forensic ChrX Research database by considering the allele frequencies for each locus. The formulae applied in each case are described below.

(Eq. 10) $$1 - 2\sum_i f_i^2 + \sum_i f_i^3$$

(Eq. 11) $$1 - \sum_i f_i^2 + \sum_i f_i^4 - \left(\sum_{i<j} \cdot\ f_i^2\right)^2$$

(Eq. 12) $$1 - \sum_i f_i^2$$

(Eq. 13) $$1 - 2\left(\sum_i \cdot\ f_i^2\right)^2 + \sum_i f_i^4$$

The following steps were followed to calculate the above-mentioned parameters of forensic interest:

- Go to the following URL: http://www.chrx-str.org/

- Click on “Evaluation & Calculate” button.
    - Calculate.
    - Set the “Allele count” to the number of allele variants in each case.
    - Insert the allele designation and their corresponding frequencies.
- Click on “Calculate” button.

Search for Tri- and Tetrallelic SNPs along the X-chromosome

The search for polymorphic bi-, tri- and tetrallelic X-SNPs was performed through Ensembl genome browser (http://www.ensembl.org/biomart) based on the human genome assembly GRCh38.p5 of the Genome Reference Consortium (GRC). Two criteria were laid down in order to ensure that the selected X-SNPs displayed heterozygosity in all the described alleles. In the first stage we chose those X-SNP markers that displayed a 1000 Genomes Global MAF (Minor Allele Frequency) ≥0.01. Currently, the Single Nucleotide Polymorphism Database (dbSNP) (https://www.ncbi.nlm.nih.gov/SNP) reports the MAF as the frequency value of the second most common allele, in order to distinguish common polymorphism from rare variants. Furthermore, to select the final candidate X-SNPs a second criterion based on allele frequencies ≥0.01 for the third or fourth allele in at least one population was established. This second criterion was applied to ensure that the selected polymorphic X-SNPs displayed three or four variants. Only those markers that met the above-mentioned criteria were considered as candidate X-SNPs in this study.

Allele frequencies of each selected marker were compiled through the Ensembl genome browser for both the overall population and the following five major populations: African, American, East Asian, European, and South Asian. Only those tri- and tetrallelic markers that met the second criterion in each major population were considered. Parameters of forensic interest, i.e. the power of discrimination in females (PD_F) and males (PD_M), as well as the mean exclusion chance in trios (MEC_T) and duos (MEC_D) were calculated by using the Forensic ChrX Research database (http://www.chrx-str.org). Moreover, a total of five sets of markers were built and an *in silico* evaluation was performed to assess their forensic efficiency in each major population. For that end the combined parameters of forensic interest ($cMEC_D$, $cMEC_T$, cPD_M, and cPD_F) were calculated.

4. Results

Study number 1

‘Iberian allele frequency database for 10 X-STRs’

The present study corresponds to the attainment of the objective 1: *Estimating the diversity of ten X-STRs in order to determine the absence of heterogeneity among Spanish regional subpopulations, which will allow the formation of the first global Spanish allele frequency database with application in forensic casework*.

As mentioned in this publication, along the last years, the most used multiplexes in the forensic field have been the decaplex of the GHEP-ISFG and the Investigator® Argus X-12 Kit that analyze 10 and 12 markers, respectively. The Iberian Peninsula comprises a wide territory of cultural, historical, socioeconomic, and genetic singularity. Since the appearance of the first multiplexes for the simultaneous study of STRs located on the X-chromosome, several populations of this territory have been typed. However, until the release of this article, there was not any global database available that encompass a great representation of the most populated regions of the Iberian Peninsula.

In this study, 1,136 individuals from Alicante, Andalusia, Asturias, Barcelona, Basque Country, Galicia, and Madrid were analyzed with the decaplex of the GHEP-ISFG. Once checked that all the populations were in Hardy Weinberg Equilibrium, pairwise F_{ST} values between the studied populations were calculated. Our results showed that no differentiation was observed among the populations of the present study, dismissing a possible genetic substructure. Therefore, the presented global Iberian allele frequency database can be safely applied for matching or kinship probabilities since the populations analyzed herein did not present heterogeneity among them.

However, the up-to-date efforts of the forensic community regarding the X-chromosome variation with forensic purposes are focused on the development of new molecular tools and techniques that allow the simultaneous analysis of a high number of polymorphic markers. In this sense, the necessity of both to expand the current global databases and to develop new methodological approaches for the analysis of X-STRs can be taken as the background that motivates the development of the current doctoral thesis work.

This study has resulted in an international publication in the journal *Forensic Science International: Genetics* under the heading '*Iberian allele frequency database for 10 X-STRs'* in June 2015. Q1, IP: 4.988. The publication is shown below.

Forensic Science International: Genetics 19 (2015) 76–78

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Letter to the Editor

Iberian allele frequency database for 10 X-STRs

Miriam Baeta [a], María José Illescas [a], Luis García [a], Carolina Núñez [a], Endika Prieto-Fernández [a], Susana Jiménez-Moreno [b], Marian M. de Pancorbo [a,*]

[a] BIOMICs Research Group, Centro de Investigación “Lascaray” Ikergunea, Universidad del País Vasco UPV/EHU, Vitoria-Gasteiz, Spain.
[b] División Medicina Legal y Forense, Departamento Patología y Cirugía, Universidad Miguel Hernández. Elche, Alicante, Spain.
* Corresponding author

Received 13 May 2015, Revised 23 June 2015, Accepted 24 June 2015, Available online 2 July 2015.

Dear Editor,

The study of X-chromosomal short tandem repeat markers (X-STRs) constitutes a valuable tool for forensic casework, in particular in complex kinship deficiency cases [1]. Due to its particular characteristics [2], the X-chromosome efficiently complements the analysis of autosomal and Y-chromosome STRs, as well as mitochondrial DNA. Currently, numerous worldwide populations have been typed by different X-STR multiplexes [3-5], being the decaplex developed by the GHEP-ISFG, one of the most extensively used panels [6]. Further studies are still necessary to determine the allele frequency distribution in heterogeneous countries in order to establish more accurate reference databases for forensic purposes.

Iberian Peninsula comprises a wide territory of cultural, historical, socioeconomic and genetic singularity. From the genetic viewpoint, X-STR allele frequencies have been studied in different populations of this territory [3, 7-12]. However, a comprehensive database for statistical evaluation of forensic casework is not yet available. Aiming to fulfill this gap, here we present an extensive genetic characterization of 10 X-STRs in Iberian populations in order to deepen in their genetic structure, and evaluate the suitability of a unique allele database. The present work reports the frequency data and forensic efficiency parameters for 10 X-STRs in a sample set of 1,136 unrelated individuals (465 men and 671 women), from seven different regional populations of the Iberian Peninsula. Individuals from Alicante (N= 141, XX= 51, XY= 90), Andalusia (N= 198, XX= 100, XY= 98), Asturias (N= 131, XX= 63, XY= 68), Barcelona (N= 199, XX= 98, XY= 101), Basque Country (N= 147, XY= 147), Galicia (N= 121, XX= 51, XY= 70), and Madrid (N= 199, XY= 99, XX= 100) were selected under informed consent, following the ethical standards of the Helsinki Declaration.

DNA extraction from individuals of the Basque Country and Alicante was performed from saliva swabs, using a standard organic procedure [13]. DNA was quantified with the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, USA) in a DTX880 Multimode Detector (Beckman Coulter, Fullerton, USA). DNA from the rest of populations was obtained through the Spanish National DNA Bank. PCR was performed using the decaplex X-STR system (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902, and DXS6789) described in Gusmao et al. [6]. The electrophoresis of PCR products was conducted on an ABI Prism 3130 Genetic Analyzer (AB/LT/TFS: Applied Biosystems™, Life Technologies, ThermoFisher Scientific, Waltham, MA, USA) and allele designation was performed using GeneMapper® Software v.4.0 (AB/LT/TFS). For statistical analysis, software Arlequin v.3.5 [14] was used to calculate allelic frequencies, genetic diversity, Hardy-Weinberg equilibrium (females), linkage disequilibrium (males), exact test of differentiation, and pairwise F_{ST} genetic distances with other Iberian populations [7-11]. Phylogenetic comparisons were visualized with a multidimensional scaling (MDS) analysis, based on F_{ST} distances, using SPSS Statistics v.22.0.0 (http://www.spss.com.hk/statistics). Significant differences between allele frequencies among the seven Iberian populations were computed with a chi square test using PAST software v.3.04 [15]. Statistics for forensic efficiency evaluation, namely power of discrimination in females (PD_F) and males (PD_M), mean exclusion chance in trios involving daughters (MEC_T) and father/daughter duos (MEC_D) was calculated following Desmarais et al. [16].

Population genetic profiles for all the individuals from Alicante, Andalusia, Barcelona, Basque Country, Galicia, and Madrid are presented in Supplementary Table S1. All profiles were unique in the different populations. Male and female samples were pooled in each population to calculate allelic frequencies after the exact test of differentiation revealed no significant differences. Allele frequencies obtained for the 10 X-STR loci studied in the seven Iberian populations are shown in Supplementary Table S2. Furthermore, non-significant differences were found, based on a chi square test, in allele frequencies of the 10 X-STRs among the seven Iberian populations. Consequently, a global allele database could be safely used for their analysis (Supplementary Table S3).

No deviations from Hardy-Weinberg equilibrium were observed for any of the analyzed loci in female samples (Supplementary Table S4), for a significance level of 0.005 (after Bonferroni correction). Moreover, no significant association was determined for linkage disequilibrium between loci after Bonferroni correction ($p< 0.0011$, for 45 comparisons inside each population), except in the pair DXS7133-DXS9902 ($p= 0.0008$) in Madrid population (Supplementary Table S5). This disequilibrium might be spurious or consequence of sampling effects, since these two markers are physically separated by > 90Mb [7]. However, other factors such as, random genetic drift, founder effects, recent interethnic admixture or population stratification cannot be ruled out [17].

Average genetic diversity values for all loci were above 0.7418, being DXS6809 and DXS7133 the markers displaying the highest and lowest average diversity values (0.8099 and 0.6492), respectively (Supplementary Table S3). The studied populations displayed similar gene diversity for the whole set of 10 X-STRs, the highest and lowest diversity values were respectively observed in the populations from Andalusia (0.7452 ± 0.3863) and the Basque Country (0.7363 ± 0.3832) (Table 1). Likewise, there were not marked differences in forensic efficiency statistical parameters among the seven Iberian populations (Table 1). This set of markers proved to be highly discriminative for all the studied populations, with PD values ranging from 0.9999999997–0.99999999985 for females and 0.999998–0.9999990 for males. Furthermore, it can be also particularly useful in kinship testing, since MEC values of 0.999993–0.999995 for trios and 0.9997–0.9998 for duos were observed. The forensic efficiency statistics for each X-STR loci are shown in Supplementary Table S6.

Pairwise F_{ST} values between the studied populations and other nine populations from the Iberian Peninsula for the 10 X-STRs were calculated (Supplementary Table S7). No differentiation was observed among the seven populations of the present study ($p> 0.005$, after Bonferroni correction), dismissing a possible genetic substructure. However, in comparison with other Iberian populations, differences were observed with respect to groups, such as the autochthonous groups from Basque Country, Navarre or Pas Valley (Cantabria).

Table 1. Statistical parameters of forensic interest of the 10 X-STRs panel in the seven Iberian populations under study. Gene diversity (GD), power of discrimination in females (PD$_F$) and males (PD$_M$), and mean paternity exclusion probability for trios (MEC$_T$) and duos (MEC$_D$) for the decaplex system.

	ALI	*AND*	*AST*	*BAS*	*BCN*	*GAL*	*MAD*
GD	0.7441 ± 0.3864	0.7452 ± 0.3863	0.7411 ± 0.3849	0.7363 ± 0.3832	0.7403 ± 0.3839	0.7379 ± 0.3837	0.7439 ± 0.3856
PD$_F$	0.9999999998	0.9999999999	0.9999999998	0.9999999997	0.9999999998	0.9999999997	0.9999999998
PD$_M$	0.9999989	0.999999	0.9999987	0.9999985	0.9999988	0.999998	0.9999989
MEC$_T$	0.999995	0.999995	0.999994	0.999993	0.999994	0.999993	0.999995
MEC$_D$	0.9998	0.9998	0.9997	0.9997	0.9998	0.9997	0.9998

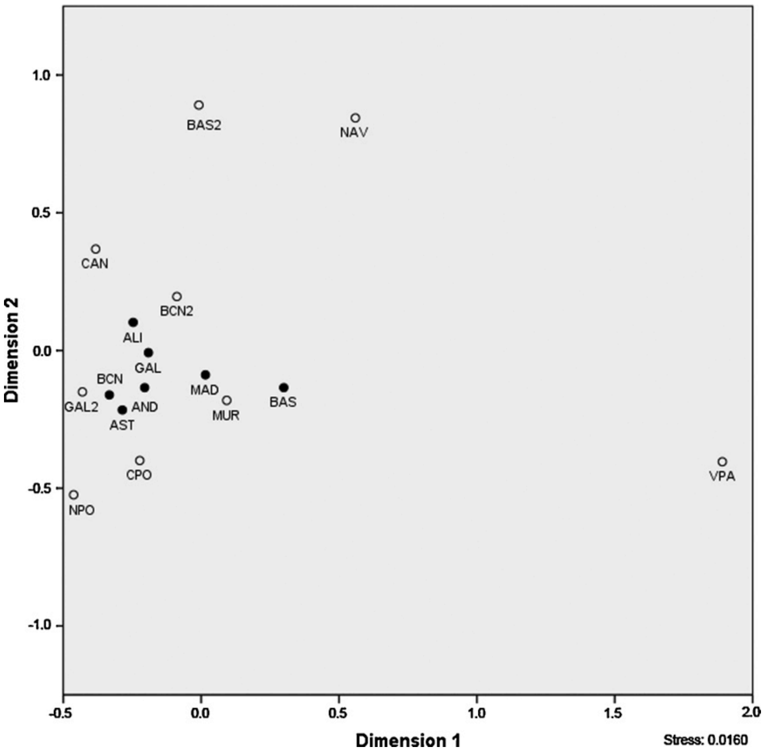


Fig. 1. MDS based on pairwise F_{ST} genetic distances calculated between the seven populations from the present study (bold dots) and other Iberian populations. Studied populations: Alicante (ALI), Andalusia (AND), Asturias (AST), Barcelona (BCN), Basque Country (BAS), Galicia (GAL), and Madrid (MAD). Other Iberian populations included for comparisons: North Portugal (NPO), Central Portugal (CPO), Galicia (GAL2) [7], Murcia (MUR) [8], Barcelona (BCN2) [9], Cantabria (CAN), autochthonous Pas Valley (VPA), autochthonous Basque Country (BAS2) [10], and autochthonous Navarre (NAV) [11]. Stress: 0.0160.

This genetic scenario is also shown in the two-dimensional MDS plot (Fig. 1), where the Iberian populations tended to group altogether with the exception of the autochthonous groups from the Basque Country and Navarre, as well as Pas Valley (Cantabria). The microdifferentation observed for these three populations has previously been reported for X-STRs [10,11] and other markers [18-20], and attributed to their genetic singularity mainly derived from a historical isolation. As discussed by Zarrabeitia et al. [10], in forensic or kinship testing, specific databases or F_{ST} -based corrections should be applied for these particular populations. However, the seven Iberian populations analyzed herein did not present heterogeneity among them and neither with the rest of Iberian populations compared, with the exception of the above mentioned isolated groups. Consequently, a global Iberian database can be safely applied for matching or kinship probabilities in most of the cases, since evidences of homogeneity have been found for these markers.

In conclusion, the 10 X-STR panel offers a highly discriminative tool for forensic identification and kinship testing. The lack of significant differences among the studied Iberian populations supports the use of the allele database presented herein for statistical evaluation of the results, with the exception of some isolated groups which due to their genetic uniqueness, may require specific databases or F_{ST}-based corrections.

Acknowledgements

Funds were provided by the Basque Government (Grupo Consolidado IT833-13). The authors are grateful to PhD Maite Álvarez for her technical and human support provided by the DNA Bank Service (SGIker) of the University of the Basque Country (UPV/EHU), and the Spanish National DNA Bank for the sample collection provided.

Conflict of interest

Authors declare no competing interest in the content of this manuscript.

Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.fsigen.2015.06.009.

References

[1] R. Szibor, M. Krawczak, S. Hering, J. Edelmann, E. Kuhlisch, D. Krause, Use of X-linked markers for forensic purposes, Int. J. Legal Med. 117 (2003) 67–74.

[2] R. Szibor, X-chromosomal markers: past, present and future, Forensic Sci. Int. Genet. 1 (2007) 93–99.

[3] J.F. Ferragut, K. Bentayebi, J.A. Castro, C. Ramon, A. Picornell, Genetic analysis of 12 X-chromosome STRs in Western Mediterranean populations, Int. J. Legal Med. 129 (2015) 253–255.

[4] Q.L. Liu, J.Z. Wang, L. Quan, H. Zhao, Y.D. Wu, X.L. Huang, D.J. Lu, Allele and haplotype diversity of 26 X-STR loci in four nationality populations from China, PLoS One 8 (6) (2013) e65570.

[5] A. Bekada, S. Benhamamouch, A. Boudjema, M. Fodil, S. Menegon, C. Torre, C. Robino, Analysis of 21 X-chromosomal STRs in an Algerian population sample, Int. J. Legal Med. 124 (4) (2010) 287–294.

[6] L. Gusmao, C. Alves, P. Sanchez-Diz, Results of the GEP-ISFG collaborative study on a X-STR decaplex, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 677–679.

[7] L. Gusmao, P. Sanchez-Diz, C. Alves, I. Gomes, M.T. Zarrabeitia, M. Abovich, et al., A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations, Inter. J. Legal Med. 123 (2009) 227–234.

[8] M.J. Illescas, J.M. Aznar, S. Cardoso, A. López-Oceja, D. Gamarra, J.F. Sánchez- Romera, et al., Genetic diversity of 10 X-STR markers in a sample population from the region of Murcia in Spain, Forensic Sci. Int.: Genet. Suppl. Ser. 3 (2011) e437–e438.

[9] B. García, M. Crespillo, M. Paredes, J.L. Valverde, Population data for 10 X-chromosome STRs from north-east of Spain, Forensic Sci. Int.: Genet. 6 (2012) e13–e15.

[10] M.T.Zarrabeitia, F. Pinheiro, M. M. de Pancorbo, L. Caine,S. Cardoso, L. Gusmao, et al., Analysis of 10 X-linked tetranucleotide markers in mixed and isolated populations, Forensic Sci. Int. Genet. 3 (2009) 63–66.

[11] M.J. Illescas, A. Perez, J.M. Aznar, L. Valverde, S. Cardoso, J. Algorta, et al., Population genetic data for 10 X-STR loci in autochthonous Basques from Navarre (Spain), Forensic Sci. Int. Genet. 6 (2012) e146–8.

[12] J.C. Pinto, V. Pereira, S.L. Marques, A. Amorim, L. Alvarez, M.J. Prata, Mirandese language and genetic differentiation in Iberia: a study using X chromosome markers, Ann. Hum. Biol. 42 (2015) 20–25.

[13] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning A Laboratory Manual, 2nd ed. 1989.

[14] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis, Evol. Bioinform. Online 1 (2005) 47–50.

[15] Ø. Hammer, D.A.T. Harper, P. Ryan, Paleontological Statistics software package for education and data analysis, Paleontol Electron. 4 (9) (2001) .

[16] D.Desmarais, Y. Zhong, R. Chakraborty, C. Perreault, L. Busque, Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), J. Forensic Sci. 43 (1998) 1046–1049.

[17] J.A. Martins, J.C. Costa, G.G. Paneto, R.F. Figueiredo, L. Gusmao, P. Sanchez-Diz, et al., Genetic profile characterization of 10 X-STRs in four populations of the southeastern region of Brazil, Int. J. Legal Med. 124 (2010) 427–432.

[18] M.M. de Pancorbo, M. Lopez-Martinez, C. Martinez–Bouzas, A. Castro, I. Fernandez-Fernandez, G.A. de Mayolo, et al., The Basques according to polymorphic Alu insertions, Hum. Genet. 109 (2001) 224–233.

[19] S. Cardoso, M.J. Villanueva-Millan, L. Valverde, A. Odriozola, J.M. Aznar, S. Pineiro-Hermida, et al., Mitochondrial DNA control region variation in an autochthonous Basque population sample from the Basque Country, Forensic Sci. Int. Genet. 6 (2012) e106–e108.

[20] S. Cardoso, M. T. Zarrabeitia, L. Valverde, A. Odriozola, M. A. Alfonso-Sanchez, M. M. de Pancorbo, Variability of the entire mitochondrial DNA control region in a human isolate from the Pas Valley (northern Spain), J. Forensic Sci. 55 (2010) 1196–1201.

Study number 2

‘Development of a new highly efficient 17 X-STR multiplex for forensic purposes’

The study number 2 corresponds to the attainment of the objective 2 of the present doctoral thesis work: *Designing, developing, optimizing, and validating a new multiplex reaction that allows the simultaneous analysis of 17 STRs located on the X-chromosome, with applicability in both complex kinship cases and highly degraded remains*.

In this work a new 17 X-STR multiplex has been developed and validated following the recommendations of the SWGDAM (Scientific Working Group on DNA Analysis Methods) for an internal validation of STR systems.

This panel has been designed taking into account both the decaplex of the GHEP-ISFG and the Investigator® Argus X-12 Kit multiplexes, since a great part of the allele and haplotype frequency databases with application in forensics have been performed by using, at least, one of these two sets of markers. In this sense, the newly developed 17 X-STR panel has been raised as a complement for the decaplex. For this reason, all the markers included in the decaplex of the GHEP-ISFG (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902, and DXS6789) were selected. On the other hand, a great part of the forensic laboratories utilize the Investigator® Argus X-12 Kit for solving cases in their daily routine. With that in mind, and in order to keep the traceability of the samples previously analyzed with this kit, the markers DXS8378, DXS7423, DXS7132, and DXS10079 were also taken into account when designing the panel. Additionally, six extra markers (DXS6801, DXS6799, DXS6800, DXS10075, DXS6807, DXS6803) were included based on a heterozygosity ≥ 0.6 to increase the resolution power of the current panels.

Of the 17 X-STRs included in this reaction (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS6801, DXS7423, DXS6809, DXS6799, DXS7132, DXS9902, DXS6800, DXS6789, DXS10075, DXS10079, DXS6807, and DXS6803), six of them present amplification products of less than 200 bp in length, while the rest generate amplicons between 200 and 300 bp in length. The small size of the generated PCR products allows considering them as mini- (<200 bp) and midi-STR markers (<300 bp), respectively.

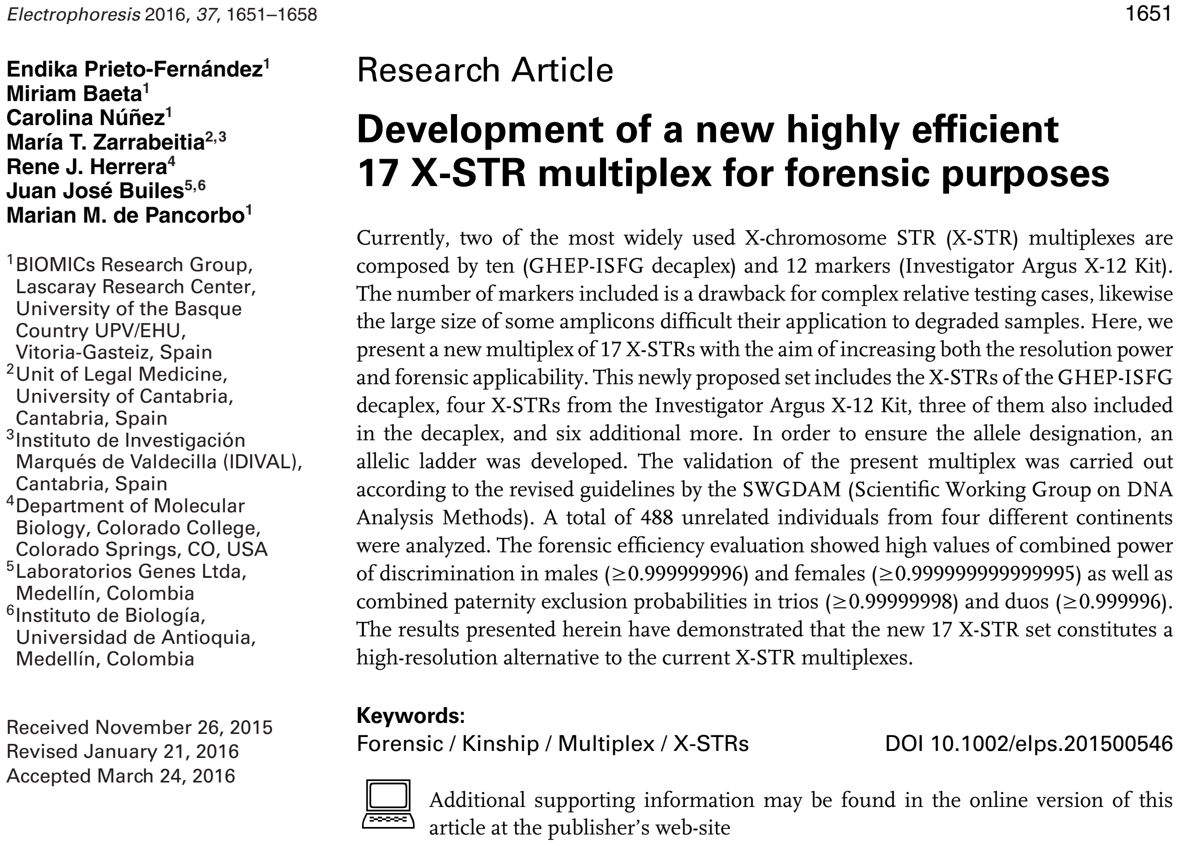
As mentioned, the validation process was carried out according to the revised guideline by the SWGDAM. This validation procedure includes precision, concordance, sensitivity, and sensibility studies, as well as determination of stutter percentage and heterozygous peak height ratio. After evaluation, the developed multiplex has proven to be a high-resolution alternative to the current X-STR multiplexes.

In short, the developed panel is a set of 17 mini- and midi-X-STR markers that can be amplified in a single PCR reaction. Besides, its combined use with the Investigator® Argus X-12 Kit allows the amplification of 25 of the most common X-STRs in the forensic community for the resolution of both kinship and human identification cases.

This study has resulted in:

1. An international publication in the journal *Electrophoresis* in March 2016 under the heading '*Development of a new highly efficient 17 X-STR multiplex for forensic purposes*'. Q2, IP: 2.482.
2. A publication reporting the first stages of the development in the journal *Forensic Science International: Genetics Supplement Series* in September 2015 under the heading '*A new 17 X-STR multiplex for forensic purposes*'.

The above-mentioned publications are shown below.



Electrophoresis 2016, *37*, 1651–1658 1651

Endika Prieto-Fernández[1]
Miriam Baeta[1]
Carolina Núñez[1]
María T. Zarrabeitia[2,3]
Rene J. Herrera[4]
Juan José Builes[5,6]
Marian M. de Pancorbo[1]

[1] BIOMICs Research Group, Lascaray Research Center, University of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain
[2] Unit of Legal Medicine, University of Cantabria, Cantabria, Spain
[3] Instituto de Investigación Marqués de Valdecilla (IDIVAL), Cantabria, Spain
[4] Department of Molecular Biology, Colorado College, Colorado Springs, CO, USA
[5] Laboratorios Genes Ltda, Medellín, Colombia
[6] Instituto de Biología, Universidad de Antioquia, Medellín, Colombia

Received November 26, 2015
Revised January 21, 2016
Accepted March 24, 2016

Research Article

Development of a new highly efficient 17 X-STR multiplex for forensic purposes

Currently, two of the most widely used X-chromosome STR (X-STR) multiplexes are composed by ten (GHEP-ISFG decaplex) and 12 markers (Investigator Argus X-12 Kit). The number of markers included is a drawback for complex relative testing cases, likewise the large size of some amplicons difficult their application to degraded samples. Here, we present a new multiplex of 17 X-STRs with the aim of increasing both the resolution power and forensic applicability. This newly proposed set includes the X-STRs of the GHEP-ISFG decaplex, four X-STRs from the Investigator Argus X-12 Kit, three of them also included in the decaplex, and six additional more. In order to ensure the allele designation, an allelic ladder was developed. The validation of the present multiplex was carried out according to the revised guidelines by the SWGDAM (Scientific Working Group on DNA Analysis Methods). A total of 488 unrelated individuals from four different continents were analyzed. The forensic efficiency evaluation showed high values of combined power of discrimination in males (≥0.999999996) and females (≥0.999999999999995) as well as combined paternity exclusion probabilities in trios (≥0.99999998) and duos (≥0.999996). The results presented herein have demonstrated that the new 17 X-STR set constitutes a high-resolution alternative to the current X-STR multiplexes.

Keywords:
Forensic / Kinship / Multiplex / X-STRs DOI 10.1002/elps.201500546

Additional supporting information may be found in the online version of this article at the publisher's web-site

Abstract

Currently, two of the most widely used X-chromosome STR (X-STR) multiplexes are composed by ten (GHEP-ISFG decaplex) and 12 markers (Investigator Argus X-12 Kit). The number of markers included is a drawback for complex relative testing cases, likewise the large size of some amplicons difficult their application to degraded samples. Here, we present a new multiplex of 17 X-STRs with the aim of increasing both the resolution power and forensic applicability. This newly proposed set includes the X-STRs of the GHEP-ISFG decaplex, four X-STRs from the Investigator Argus X-12 Kit, three of them also included in the decaplex, and six additional more. In order to ensure the allele designation, an allelic ladder was developed. The validation of the present multiplex was carried out according to the revised guidelines by the SWGDAM (Scientific Working Group on DNA Analysis Methods). A total of 488 unrelated individuals from four different continents were analyzed. The forensic efficiency evaluation showed high values of combined power of discrimination in males (≥0.999999996) and females (≥0.999999999999995) as well as combined paternity exclusion probabilities in trios

(≥0.99999998) and duos (≥0.999996). The results presented herein have demonstrated that the new 17 X-STR set constitutes a high-resolution alternative to the current X-STR multiplexes.

Keywords: forensic, kinship, multiplex, X-STRs

Introduction

The technology for the analysis of autosomal microsatellite markers or Short Tandem Repeats (STRs) has experienced a rapid development in the last decade [1], becoming the referent analysis technique used by the great majority of the forensic laboratories for both to carry out human identification studies and to determine human kinship. However, in some cases, the use of autosomal STRs may be inconclusive. As a result, the analysis of STRs located in the sexual chromosomes is becoming more and more usual [2].

The use of STRs located in the X chromosome (X-STRs) is very efficient for determining kinship between fathers and daughters, since it increases the power of discrimination as well as the paternity exclusion probability obtained by analyzing only autosomal STRs [3]. Furthermore, the analysis of X-STRs can strongly benefit the investigation of some complex kinship cases e.g. paternal half-sisters, paternal aunt/uncle-niece as well as maternal uncle-nephew. Even when incestuous situations occur [4], the use of X-STRs can help to distinguish these relationships.

Two of the most used kits for the analysis of X-STRs until now are the decaplex of the GHEP-ISFG [2,5], which includes ten markers, and the Investigator Argus X-12 Kit (Qiagen, Valencia, CA) that combines three of the markers from the decaplex and nine additional X-STRs. The disadvantages of both kits are the limited number of markers and the large size of some of the amplification products that make them inappropriate for the application in highly degraded DNA.

Consequently, the aim of this study was to design a new multiplex of 17 X-STRs combining all the markers included in the decaplex of the GHEP-ISFG (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902, and DXS6789), four X-STRs from the Investigator Argus X-12 Kit (DXS8378, DXS7423, DXS7132, and DXS10079), three of them also included in the decaplex, and six additional more (DXS6801, DXS6799, DXS6800, DXS10075, DXS6807, and

DXS6803) to increase both the power of discrimination and the paternity exclusion probabilities. Amplicon sizes were also taken into account for application in highly degraded DNA.

Here we presents the development and validation of the 17 X-STR panel according to the revised guidelines issued by the Scientific Working Group on DNA Analysis Methods (SWGDAM) [6]. Assays for PCR and capillary electrophoresis optimization, evaluation of the fragment size precision, concordance studies, sensitivity and stability to inhibitors and determination of stutter percentage as well as heterozygous peak-height ratio were performed. In order to ensure the allele designation, an allelic ladder was developed.

Materials and methods

Human DNA control samples

In order to set up the PCR amplification conditions and run the appropriate sensitivity and stability studies, three DNA control samples were used: 2800M and K562 (Promega® Corporation, USA) as well as the 9947A from AmpFLSTR™ PCR Amplification Kit (ThermoFisher Scientific, Waltham, MA, USA).

Population samples

Population samples for the validation of the 17 X-STRs were obtained from 488 unrelated healthy individuals (322 men and 166 women) including: Asians from Thailand (N= 138; XY= 138), European Caucasoids from Spain (N= 101; XX= 36, XY= 65), Africans from Malawi and Equatorial Guinea (N= 135; XX= 73, XY= 62), and Hispanics from Colombia (N= 114; XX= 57, XY= 57). Male and female individuals were sampled for each population whenever possible. DNA from bloodstains of the African individuals was isolated by a phenol-chloroform-isoamyl alcohol method. DNA from the buccal swabs samples of the Asians and the peripheral blood of the Hispanics were obtained using the Gentra Buccal Cell Kit (Puragene, Gentra Systems, Inc., MN, USA) and the Qiamp DNA Micro Kit (Qiagen, Valencia, CA), respectively.

DNA from the European Caucasoids was obtained through the Spanish National DNA Bank. All the samples were obtained from volunteer donors following the ethical standards of the Helsinki Declaration. DNA was quantified by using the Scientific

NanoDrop$^{\text{TM}}$ 1000 Spectrophotometer (ThermoFisher Scientific Inc., Wilmington, DE) and then diluted to a 1-3 ng/µl concentration.

Selection of markers and primer design

17 X-STRs were selected for the new multiplex described in the present work (Table 1). This panel combines eleven markers commonly used (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902, DXS6789, and DXS10079) and six additional markers (DXS6801, DXS6799, DXS6800, DXS10075, DXS6807, and DXS6803) that were chosen based on a heterozygosity ≥ 0.6 [3, 9, 11, 13, 15].

Table 1. Motif structure, bibliography reference, and alleles of the control DNAs 9947A, 2800M, and K562 for each of the X-STR loci included in the new 17 X-STR multiplex. Markers included in decaplex are represented with [1] while the ones included in Argus X-12 Kit with [2].

Marker	*Motif structure*	*Ref.*	*9947A*	*2800M*	*K562*
DXS8378 [1,2]	$(CTAT)_n$	[7]	10/11	12	10
DXS9898 [1]	$(TATC)_2$-(ATC)-$(TATC)_{0\text{-}1}(ATC)_{0\text{-}1}$-$(TATC)_n$	[8]	12/15	8.3	12
DXS7133 [1]	$(ATAG)_n$	[7]	9/10	11	10
GATA31E08 [1]	$(AGGG)_{2\text{–}3}(AGAT)_n$	[5]	13	14	13
GATA172D05 [1]	$(TAGA)_n$	[7]	10	11	12
DXS6801	$(ATCT)_n$-N_7-$(ATCT)_2$	[9]	11	11	11
DXS7423 [1,2]	$(TCCA)_3$- N_8 -$(TCCA)_n$	[7]	14/15	15	17
DXS6809 [1]	$(CTAT)_n$-$(ATCT)_3$-N_9-$(TATC)_n$-$(ATCT)_n$ - N_{10}-$(ATCT)_n$	[10]	31/34	31	34
DXS6799	$(TATC)_n$	[3]	11/12	12	11
DXS7132 [1,2]	$(TCTA)_n$	[7]	12	13	13
DXS9902 [1]	$(TAGA)_n$	[5]	12	12	12/13
DXS6800	$(TAGA)_n$–CA–GATA–GAT–$(GATA)_4$-GG-$(TAGA)_3$-TC-$(GATA)_3$	[11]	18/19	19	21
DXS6789 [1]	$(TATC)_{0\text{-}1}$-$(TATG)_n$-$(TATC)_n$	[12]	21/22	21	21
DXS10075	$(TAGA)_n$–N_3-$(TAGA)_n$	[13]	17/18	18	18
DXS10079 [2]	$(AGAA)_n$-AGAG-$(AGAA)_3$	[13]	20/23	19	17
DXS6807	$(GATA)_1$-N_7-$(GATA)_2$-GAC-$(GATA)_n$	[14]	12/14	14	11
DXS6803	$(TCTA)_n$-$(TCA)_{0\text{-}1}$-TCTA	[15]	11.3/12	13	10

The primers used to amplify the markers included in this panel are shown in Supporting Information Table S1. For newly designed primers (for markers DXS6800, DXS6803, and DXS6807) the software PerlPrimer v.1.1.21 [16] was used.

The lack of interactions between primers and the specificity for the X chromosome was evaluated by using Autodimer v.1.0 [17] software and the BLASTN (http://blast.ncbi.nlm.nih.gov/Blast.cgi) alignment tool, respectively.

All the forward primers of each marker were modified by the addition of a fluorescent dye at their 5' end with the exception of the marker DXS10075, labelled at the 5' end of the reverse primer. The fluorescent dyes used to modify the primers were: *5-FAM* (Abs.= 495nm; Em.= 520nm), *YAKIMA YELLOW* (Abs.= 530nm; Em.= 549nm), *ATTO 550* (Abs.= 554nm; Em.= 576nm), and *ATTO 565* (Abs.= 563nm; Em.=592 nm) (Eurofins Genomics, Ebersberg, Germany) (Supporting Information Table S1).

PCR amplification, electrophoresis and data analysis

The amplification was carried out on the GeneAmp® 9800 PCR System (Thermofisher Scientific, Waltham, MA, USA) under the same thermocycling conditions used for the GHEP-ISFG decaplex [5], which consists in: pre-incubation for 15 min at 95 ºC, followed by ten cycles of 30 s at 94 ºC, 90 s at 60 ºC, 60 s at 72 ºC, and 20 cycles of 30 s at 94 ºC, 90 s at 58 ºC, and 60 s at 72 ºC with a final incubation for 60 min at 72 ºC. If adenilation is detected, especially in the largest fragments, we recommend lengthening the final incubation at 72 ºC for additional 30 min.

PCR reaction consisted of 5 µl of QIAGEN Multiplex PCR kit (Qiagen, Valencia, CA), 0.5 µl of primer mix, and Milli-Q water in order to reach a final reaction volume of 10 µl. 1-3 ng of genomic DNA were used in standard reactions for avoiding allelic dropouts in female samples. The final concentration of each primer in the primer mix was 0.2 µM except for the markers DXS6809 (0.6 µM), DXS6799 (0.4 µM), and DXS7132 (0.4 µM) (Supporting Information Table S1).

The amplification products were analyzed by mixing 9 µl of Hi-Di^{TM} Formamide (Thermofisher Scientific, Waltham, MA, USA), 1 µl of each PCR amplification product, and 0.25 µl of GeneScan^{TM} LIZ-500^{TM} Size Standard (Thermofisher Scientific, Waltham, MA, USA). After denaturation, PCR products were separated and detected by capillary electrophoresis with the polymer POP7 on an ABI PRISM 3130 Genetic Analyzer (Thermofisher Scientific, Waltham, MA, USA). Electrophoresis data were analyzed with GeneMapper ID software version 4.0 (Thermofisher Scientific, Waltham, MA, USA).

Allelic ladder

To assure allele designation an allelic ladder was constructed based on all the different detected alleles for each marker in this study. Each allele was separately amplified in a singleplex PCR reaction that consisted of 20 µl of QIAGEN Multiplex PCR kit (Qiagen, Valencia, CA), 0.25 µl of each primer at 10 µM, and 2 ng of genomic DNA. To confirm the efficiency of the PCR reaction, a migration on a 1.5% agarose gel was carried out. The amplicon intensity of each allele was estimated by comparison to the electrophoretic ladder. According to the intensity of the bands, the allele amplicons of each marker were proportionally pooled. Then, in order to assure the equal height of the alleles into each marker, the mixture was analyzed on an ABI PRISM 3130 Genetic Analyzer (Thermofisher Scientific, Waltham, MA, USA). Finally, when all the alleles of each marker were balanced, the 17 X-STRs were pooled together according to their relative fluorescence units (RFUs). Artifact removal was performed using MinElute® PCR Purification Kit (Qiagen, Valencia, CA).

Precision

The precision of the allelic ladder from 17 X-STRs was calculated from five independent electrophoresis migrations on an ABI PRISM 3130 Genetic Analyzer (ThermoFisher Scientific, Waltham, MA, USA). The mean of base pair size and standard deviation for each allele was calculated.

Concordance

The concordance of the X-STR loci DXS8378, DXS9902, DXS7132, DXS9898, DXS6809, DXS6789, DXS7133, GATA172D05, GATA31E08, and DXS7423 included in the 17 X-STR panel was established by comparison to the European Caucasoid population profiles (N= 101) obtained with the decaplex of the GHEP-ISFG in a previous study carried out in our laboratory [18]. Additionally, positive controls (2800M, 9947A, and K562) were used to test the concordance of the other seven markers (DXS6799, DXS6800, DXS6801, DXS6803, DXS6807, DXS10075, and DXS10079).

Sensitivity and stability studies

To evaluate the minimum quantity of DNA required to obtain reliable results, 2800M and 9947A control DNAs were analyzed in triplicate with the following DNA quantities: 30 ng, 20 ng, 10 ng, 1.6 ng, 1 ng, 400 pg, 200 pg, 100 pg, 50 pg, and 25 pg.

Stability studies were also carried out by adding to the PCR reaction 1 ng of 9947A and different concentrations of two common inhibitors in forensic science: haematin (Sigma-Aldrich Corporation, St. Louis, MO, USA) and humic acid (Sigma-Aldrich Corporation, St. Louis, MO, USA). This study was performed in triplicate using the following range of inhibitor concentrations: 100 µM, 150 µM, 300 µM, 500 µM, 750 µM, 1000 µM, 1500 µM, 3000 µM, and 5000 µM of haematin and 25 ng/µl, 50 ng/µl, 100 ng/µl, 200 ng/µl, 250 ng/µl, 300 ng/µl, 500 ng/µl, 1000 ng/µl, 2000 ng/µl, and 3000 ng/µl of humic acid.

Determination of stutter percentage and heterozygous peak height ratio

The percentage of observed stutter at STR locus was examined by calculating the ratio of the stutter peak height (in RFU) compared to the corresponding allele peak height [19]. On the other hand, the proportion of the less intense allele relative to the more intense allele at a given heterozygous genotype, i.e. peak height ratio (PHR), was calculated by dividing the peak height of the allele with the lower RFU value by the peak height of the allele with the higher value, and then multiplying this value by 100 to express the PHR as a percentage [20]. All the heterozygous genotypes observed in each marker of the analyzed women were taken into account (the total number of heterozygous peaks for each marker are shown in Table 2).

Forensic parameters and statistical analysis

Allelic frequencies and genetic diversity (GD) of the four populations analyzed herein were calculated. Haplotype frequencies for the three closely linked X-STR markers DXS7132-DXS10075-DXS10079 (http://www.chrx-str.org) were also determined. Hardy-Weinberg equilibrium (HWE) was tested in the female population samples and pairwise linkage disequilibrium (LD) in the male population samples. All aforementioned parameters were estimated by using the Arlequin software v.3.5.1.2 [21].

In addition, power of discrimination for males (PD_M) and females (PD_F), as well as the paternity exclusion index in *duos* (MEC_D) and *trios* (MEC_T) by *Desmarais et al.* [22] was estimated for each population by using the online tool of the Forensic ChrX Research database (http://www.chrx-str.org).

Results and discussion

A 17 X-STR multiplex has been developed and validated taking into account the recommendations of the SWGDAM for an internal validation of STR systems [6].

Primer set, PCR, and capillary electrophoresis optimization

A panel of 17 X-STRs has been developed which includes eleven markers of the GHEP-ISFG decaplex or Argus X-12 panels and six additional markers. The final primer set design generates amplicons of less than 294 bp (Fig. 1). Particularly, product sizes for 12 of the X-STR markers were less than 228 bp.

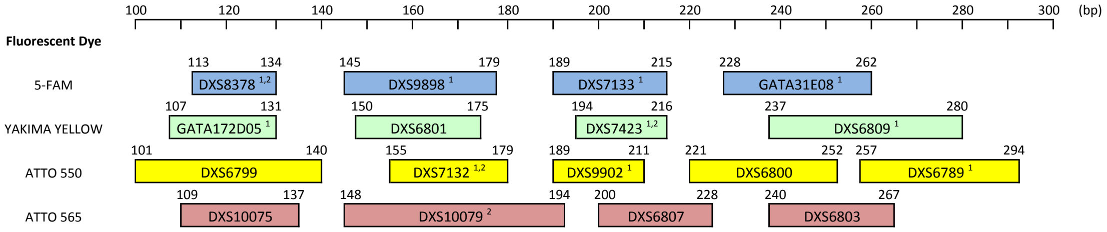


Fig. 1. Final design of the 17 X-STR multiplex. The boxes represent the expected fragment sizes for each locus. The scale at the top represents the size (bp). The fluorescent dye for the corresponding loci is shown on the left. Markers included in decaplex are represented with [1] while the ones included in Argus X-12 Kit with [2].

The analyzed PCR product sizes of each STR locus matched the expected reference sequence. Additionally, for the DXS6803 marker some samples were sequenced to confirm the presence of the frequently observed 11.3, 12.3, 13.3, and 14.3 microvariants and their corresponding sizes. For the creation of the panel on the GeneMapper software, control DNAs were considered (Table 1). An electropherogram of the standard DNA 2800M is shown in Fig. 2.

Allelic ladder

A total number of 152 different alleles were included in the 17 X-STR ladder (Fig. 3). Hemizygous and homozygous samples were used in order to balance the allele peaks height wherever possible. However, in some cases, heterozygous samples had to be used because some rare alleles were only found in these samples. It should be noted that allele 15 for the DXS6799 marker is not represented in the ladder because it was not detected in the analyzed populations. Previously described allele microvariants were found in this study for the markers DXS9898 (8.3 and 13.3), DXS9902 (12.1) [5], and DXS6803 (10.3, 12.3, 13.3, and 14.3) [15] (www.chrx-str.org database). New alleles were also found for the markers DXS10075 (16.1, 17.1, 17.2, and 18.2), DXS10079 (18.2), and DXS6803 (11.3) that were not previously described (Supporting Information Table S2). Dye blobs and residual primers were successfully removed during the development of the allelic ladder by using MinElute® PCR Purification Kit (Qiagen, Valencia, CA).

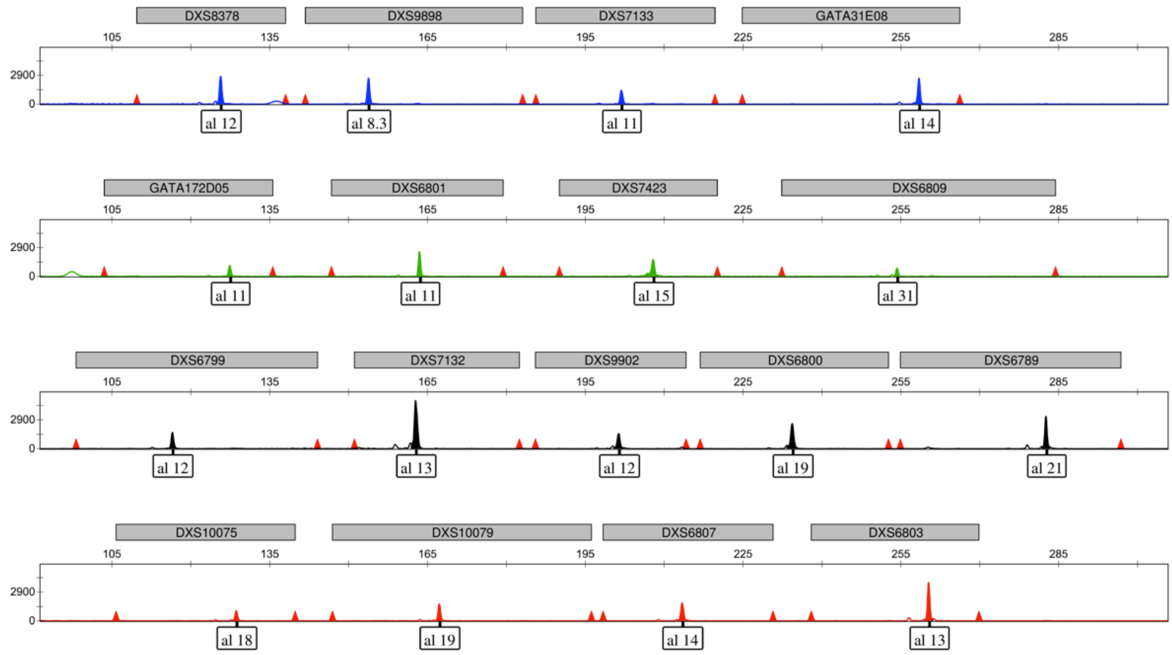


Fig. 2. Electropherogram of the positive control 2800M (Promega® Corporation, USA) for the 17 X-STR multiplex system.

Precision

To evaluate the precision of the allelic ladder, five independent electrophoresis migrations were carried out under identical conditions obtaining a standard deviation value of 0.12 bp (Supporting Information Table S2).

Concordance studies

A total of 101 European Caucasoid profiles were compared with the previously obtained genotypes by using the decaplex of the GHEP-ISFG [18]. A complete concordance was observed. The concordance of the additional seven markers, non-included in the decaplex, was confirmed through the typing of positive controls with known genotypes (2800M, 9947A, and K562). These results confirmed the reliability of the 17 X-STR multiplex as a genotyping system.

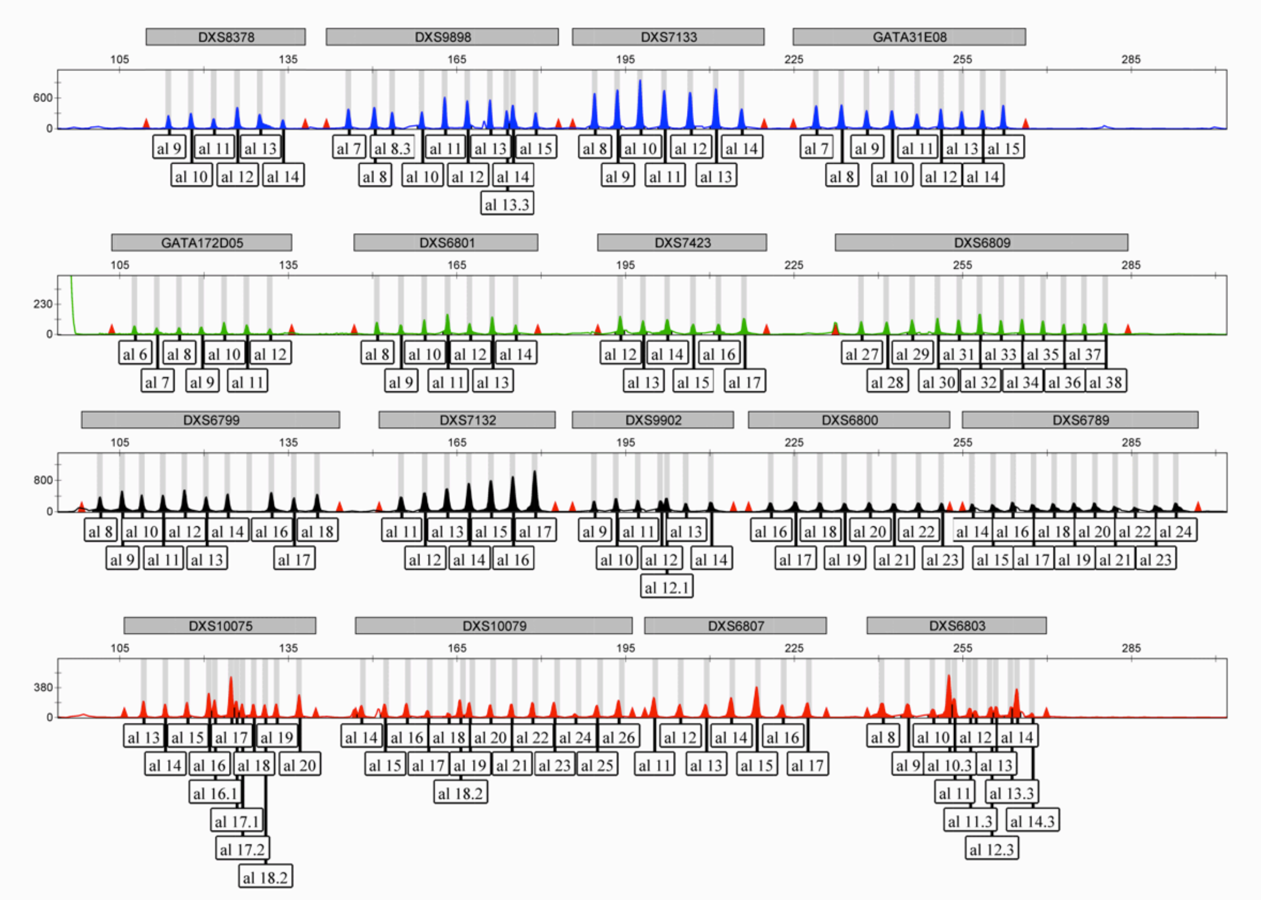


Fig. 3. Electropherogram of the allelic ladder for the 17 X-STR multiplex system.

Sensitivity and stability studies

The sensitivity was assayed by analyzing in triplicate the control DNAs 2800M and 9947A at different concentrations. It was determined that the minimum quantity of DNA to obtain complete genetic profiles was 100 pg whereas a lower amount of DNA showed dropouts in some markers (Supporting Information Table S3).

The stability of the 17 X-STR panel was evaluated by adding haematin or humic acid that are two common inhibitors in forensic casework. Complete genetic profiles from 1 ng of DNA were obtained with ≤ 250 ng/µl of humic acid or 150-300 µM of haematin for most of the replicas. Values over these concentrations resulted in allele dropouts. No amplification results were detected when concentrations were higher than 500 ng/µl of humic acid or 500 µM of haematin. A summary of the obtained results is shown in Supporting Information Table S3. To sum up, these results proved the sensitivity and robustness of the new 17 X-STR panel.

Determination of stutter percentage and heterozygous peak height ratio

The determination of the stutter percentage showed a mean value of 8.1%. The percentages for each marker are shown in Table 2. The DXS6800 was the locus with the lowest stutter average percentage (4.0%) while the higher average percentage was showed by DXS6809 (15.2%).

All markers showed a mean heterozygous PHR of 82.6% being the minimum and maximum values for the loci GATA172D05 (71.8%) and DXS7133 (87.9%), respectively (Table 2).

Table 2. Stutter percentage and heterozygous peak height ratio mean was calculated for each of the loci included in the new 17 X-STR multiplex. N_S=number of alleles with stutter peaks; N_A= number of observed alleles; H_s/H_a=mean ratio between height of the stutter (H_s) and allele (H_a); H_l / H_h = mean ratio between height of the lowest allele (H_l) and the highest allele (H_h); SD=standard deviation.

Marker	*Stutter percentage*				*Heterozygous peak height ratio*			
	N_S	H_s / H_a	*SD*	*%*	N_A	H_l / H_h	*SD*	*%*
DXS8378	301	0.0861	0.0604	8.6	230	0.8651	0.1258	86.5
DXS9898	247	0.0896	0.0546	9.0	234	0.8514	0.1304	85.1
DXS7133	273	0.0530	0.0218	5.3	204	0.8786	0.1290	87.9
GATA31E08	282	0.0773	0.0496	7.7	262	0.8683	0.1293	86.8
GATA172D05	138	0.0680	0.0549	6.8	244	0.7179	0.1798	71.8
DXS6801	226	0.0731	0.0519	7.3	204	0.8536	0.0938	85.4
DXS7423	270	0.0748	0.0282	7.5	216	0.8315	0.0972	83.2
DXS6809	238	0.1524	0.0482	15.2	244	0.8051	0.1129	80.5
DXS6799	181	0.0685	0.0314	6.9	186	0.7662	0.1280	76.6
DXS7132	190	0.1038	0.0605	10.4	192	0.8231	0.1337	82.3
DXS9902	219	0.0613	0.0172	6.1	206	0.8299	0.1318	83.0
DXS6800	180	0.0404	0.0243	4.0	196	0.8473	0.1287	84.7
DXS6789	267	0.1077	0.0347	10.8	272	0.7930	0.1353	79.3
DXS10075	279	0.0989	0.0393	9.9	250	0.8452	0.0904	84.5
DXS10079	288	0.0890	0.0428	8.9	274	0.8306	0.1186	83.1
DXS6807	241	0.0539	0.0341	5.4	216	0.7709	0.1111	77.1
DXS6803	258	0.0714	0.0148	7.1	260	0.8618	0.1009	86.2
Mean		0.0805	0.0261	8.1		0.8259	0.0430	82.6

Forensic parameters

Genotypes for each population are showed in the Supporting Information Table S4. The allele and haplotype frequencies obtained are detailed in Supporting Information Table S5 and S6, respectively. No significant deviation from HWE (Supporting Information Table S7) was found for any of the loci after Bonferroni correction (p= 0.0029). No detectable evidence of LD was found for the same pair of markers in all populations after Bonferroni correction (p= 0.0004) (Supporting Information Table S8). In fact, only LD was detected for the pairs DXS10075-DXS10079 (p< 0.0000) and DXS9898-DXS6803 (p< 0.0000) in the African population as well as DXS9898-DXS6807 (p< 0.0000) in the Hispanic population. LD in the pair DXS10075-DXS10079 was expected since they belong to the same linkage group along with DXS7132. Although evidence of LD was no detected in all populations, these three markers were treated as a haplotype. The pair DXS9898-DXS6807 cannot be considered as a haplotype since their location in the X-chromosome is very distant (87,682 Mb and 4,753 Mb, respectively) (http://www.chrx-str.org), therefore the obtained results in LD test can be attributed to spurious effects. Finally, the linkage between the pair of loci DXS9898-DXS6803 was investigated by using the HapMap recombination map data (http://hapmap.ncbi.nlm.nih.gov/). We identified the closest HapMap SNP site upstream and downstream of the STR positions and used these markers as the STRs proxies (rs5923861 and rs5984892 close to DXS9898 and DXS6803, respectively). This analysis showed that these markers belong to different linkage blocks, consequently they are separated by enough genetic distance for linkage not to interfere with the treatment as independent loci [23]. The evaluated forensic parameters GD, PD_M and PD_F as well as the MEC_D and MEC_T are summarized in Supporting Information Table S9. DXS10079 showed the highest GD in Hispanic (0.8515), European Caucasoid (0.8170), and Asian (0.8119) populations and, consequently, displayed the greatest values of forensic parameters in these populations, whereas for the African population, DXS6809 (0.8484) showed the highest GD. On the other hand, the loci with the lower GD were DXS7133 in the Hispanic (0.6063) and African (0.6189) populations, DXS6801 (0.5134) in the European Caucasoid population and DXS6800 (0.4651) in the Asian population.

Combined values of all the aforementioned forensic parameters were clearly higher for the 17 X-STR multiplex than for the GHEP-ISFG decaplex in all cases. These results were expected since more markers were included on the new X-STR panel. All the forensic parameters generally reached one order of magnitude higher in the African population than in the Hispanic and European Caucasoid ones, and up to three orders in the case of Asiatic populations (Table 3).

Table 3. Combined values of forensic parameters obtained for the GHEP-ISFG decaplex and the 17 X-STR panel for each population.

	Population	*GHEP-ISFG decaplex*	*17 X-STR panel*
MEC_D	Hispanic	0.9997	0.9999991
MEC_T		0.999992	0.999999997
PD_M		0.999998	0.9999999995
PD_F		0.9999999997	0.99999999999999987
MEC_D	African	0.99989	0.99999990
MEC_T		0.999998	0.9999999998
PD_M		0.9999995	0.99999999997
PD_F		0.99999999997	0.9999999999999999986
MEC_D	European Caucasoid	0.9998	0.999998
MEC_T		0.999991	0.999999994
PD_M		0.999998	0.9999999993
PD_F		0.9999999995	0.9999999999999995
MEC_D	Asian	0.9993	0.999996
MEC_T		0.99998	0.99999998
PD_M		0.999995	0.999999996
PD_F		0.999999998	0.999999999999995

Conclusion

In this work we describe a 17 X-STR multiplex system including the most commonly used X chromosome markers in a single PCR reaction that has revealed to be robust and highly discriminative. The concordance studies show the reliability of this tool for forensic genotyping purposes. Moreover, the results obtained in the sensitivity and stability studies, as well as the values obtained for power of discrimination and combined paternity exclusion probabilities in duos and trios, have proved the suitability of this multiplex for paternity cases. Additionally, the reduced size of amplicons of the new panel markers increases the potential applicability for typing highly degraded DNA samples.

Acknowledgements

Funds were provided by the Basque Government (Grupo Consolidado IT833-13). The authors are grateful to Prof. Dr. José Luis Blázquez Caeiro at the University of Santiago de Compostela (Spain) for the African population samples, to the Spanish National DNA

Bank for the European Caucasoid sample collection, and to PhD Maite Álvarez for her technical and human support provided by the DNA Bank Service (SGIker) of the University of the Basque Country (UPV/EHU) and European funding (ERDF and ESF).

Conflict of interest

The authors have declared no conflict of interest.

Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1002/elps.201500546.

References

[1] Aznar, J.M., Celorrio, D., Odriozola, A., Köhnemann, S., Bravo, M.L., Builes, J.J., Pfeiffer, H., Herrera, R.J., de Pancorbo, M.M., Forensic Sci. Int. Genet. 2014, 8, 10–19.

[2] Gusmão, L., Alves, C., Sánchez-Diz, P., Zarrabeitia, M.T., Abovich, M.A., Aragón, I., Arce, B., Arrieta, G., Arroyo, E., Atmetlla, I., Baeza, C., Bobillo, M.C., Cainé, L., Campos, R., Caraballo, L., Carvalho, E., Carvalho, M., Cicarelli, R.M.B., Comas, D., Corach, D., Espinoza, M., Espinheira, M.R., Rendo, F., García, O., Gomes, I., González, A., Hernández, A., Hidalgo, M., Lozano, P., Malaghini, M., Manzanares, D., Martínez, B., Martins, J.A., Maxzud, K., Miguel, I., Modesti, N., Montesino, M., Ortiz, R., Pestano, J.J., Pinheiro, M.F., Prieto, L., Raimondi, E., Riancho, J.A., Rodríguez, M.B., Salgado, I., Salgueiro, N., Sánchez, J.J., Silva, S., Toscanini, U., Vidales, C., Silva, C.V., Villalobos, M.C., Vullo, C., Yurrebaso, I., Zubillaga, A.I., Carracedo, A., Amorim, A., Forensic Sci. Int. Genet. Suppl. Ser. 2008, 1, 677–679.

[3] Castañeda, M., Universidad de Cantabria 2013 (PhD Thesis).

[4] Pinto, N., Gusmão, L., Amorim, A., Forensic Sci. Int. Genet. 2011, 5, 27–32.

[5] Gusmão, L., Sánchez-Diz, P., Alves, C., Gomes, I., Zarrabeitia, M.T., Abovich, M., Atmetlla, I., Bobillo, C., Bravo, L., Builes, J., Cainé, L., Calvo, R., Carvalho, E., Carvalho, M., Cicarelli, R., Catelli, L., Corach, D., Espinoza, M., García, O., Malaghini, M., Martins, J., Pinheiro, F., João Porto, M.J., Raimondi, E., Riancho, J.A., Rodríguez, A., Rodríguez, A., Cardozo, B.R., Schneider, V., Silva, S.,

Tavares, C., Toscanini, U., Vullo, C., Whittle, M., Yurrebaso, I., Carracedo, A., Amorim, A., Int. J. Legal Med. 2009, 123, 227–234.

[6] SWGDAM Validation Guidelines for DNA Analysis Methods, http://swgdam.org/SWGDAM_Validation_Guidelines_APPROVED_Dec_2012.pdf (accessed 09.08.15).

[7] Edelmann, J., Deichsel, D., Hering, S., Plate, I., Szibor, R., Forensic Sci. Int. 2002, 129, 99–103.

[8] Hering, S., Szibor, R., J. Forensic Sci. 2000, 45, 929–931.

[9] Edelmann, J., Szibor, R., Forensic Sci. Int. 2005, 148, 219–220.

[10] Edelmann, J., Deichsel, D., Plate, I., Käser, M., Szibor, R., Int. J. Legal Med. 2003, 117, 241–244.

[11] Edelmann, J., Hering, S., Michael, M., Lessig, R., Deichsel, D., Meier-Sundhausen, G., Roewer, L., Plate, I., Szibor, R., Forensic Sci. Int. 2001, 124, 215–218.

[12] Hering, S., Kuhlisch, E., Szibor R., Forensic Sci. Int. 2001, 119, 42–46.

[13] Hering, S., Augustin, C., Edelmann, J., Heidel, M., Dressler, J., Rodig, H., Kuhlisch, E., Szibor, R., Int. J. Legal Med. 2006, 120, 337–345.

[14] Edelmann, J., Szibor, R., Electrophoresis 1999, 20, 2844–2846.

[15] Diegoli, T.M., Coble, M.D., Forensic Sci. Int. Genet. 2011, 5, 415–421.

[16] Marshall, O.J., Bioinformatics 2004, 20, 2471–2472.

[17] Vallone, P.M., Butler, J.M., Biotechniques 2004, 37, 226–231.

[18] Baeta, M., Illescas, M.J., García, L., Núñez, C., Prieto-Fernández, E., Jiménez-Moreno, S., de Pancorbo, M.M., Forensic Sci. Int. Genet. 2015, 19, 76–78.

[19] Butler, J.M. Forensic DNA Typing. 2nd edition, Elsevier, New York, 2005, pp. 400-401.

[20] SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories https://www.fbi.gov/about-us/lab/biometric-analysis/codis/swgdam.pdf (accessed 13.01.16).

[21] Excoffier, L., Laval, G., Schneider, S., Evol. Bioinform. Online 2005, 1, 47–50.

[22] Desmarais, D., Zhong, Y., Chakraborty, R., Perreault, C., Busque, L., J. Forensic Sci. 1998, 43, 1046–1049.

[23] Phillips, C., Applications of Autosomal SNPs and Indels in Forensic Analysis, in: Shewale, J.R. and Liu, H., (Eds.), Forensic DNA Analysis: Current Practices and Emerging Technologies, 1st edition CRC Press, Boca Raton, FL, 2014, pp. 280-307.

[24] Castañeda, M., Odriozola, A., Gómez, J., Zarrabeitia, M.T., Int. J. Legal Med. 2013, 127, 735–739.

Forensic Science International: Genetics Supplement Series 5 (2015) e283–e285

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

Supplement Series

A new 17 X-STR multiplex for forensic purposes

Endika Prieto-Fernández [a], Miriam Baeta [a], Carolina Núñez [a], Susana Jiménez-Moreno [b,*], Marian M. de Pancorbo [a]

[a] BIOMICs Research Group, Lascaray Research Center, University of the Basque Country UPV/EHU, Avda. Miguel de Unamuno, 3. 01006 Vitoria-Gasteiz, Spain.

[b] Área Medicina Legal y Forense, Dpto. Patología y Cirugía, Universidad Miguel Hernández de Elche, Elche, Alicante, Spain.

* Corresponding author

Received 29 August 2015, Accepted 18 September 2015, Available online 30 September 2015.

Abstract

The use of STRs located in the X chromosome (X-STRs) is very efficient for determining kinship between fathers and daughters as well as other complex kinship cases. Currently, X-STR markers are analyzed by using two different multiplexes, the decaplex of the GHEP-ISFG and the Investigator Argus X-12 Kit. Here, we present a new multiplex whose advantage is the analysis of 17 X-STRs in a single reaction. Population samples from four different continents were analyzed. A sensitivity test showed the efficacy of this new multiplex for samples with DNA quantities as low as 100 pg.

Keywords: X-STRs, multiplex, forensic, kinship

Introduction

The technology for the analysis of autosomal microsatellite markers or short tandem repeats (STRs) has become the reference technique used by the great majority of forensic laboratories for both carrying out human identification studies and for determining human kinship. However, in some complex cases, as well as in individual identification cases based on kinship relationships, the use of autosomal chromosome STRs may be inconclusive. As a result, the analysis of STRs located in the sexual chromosomes is becoming more and more usual [1].

The use of STRs located on the X chromosome (X-STRs) is highly efficient for determining kinship between fathers and daughters since it increases the power of discrimination and the paternity exclusion probability provided by the analysis of autosomal STRs alone [2]. Furthermore, the analysis of X-STRs can strongly benefit the investigation of some complex kinship cases, e.g. paternal half-sisters, paternal aunt/uncle-niece, as well as maternal uncle-nephew. Even when incestuous situations, such as parent plus grandparent-child, parent plus half-sib-child, and parent plus uncle/aunt-child occur [3], the use of X-STRs can help distinguish these relationships.

Two of the kits used most for the analysis of X-STRs are the decaplex of the GHEP-ISFG [1,4], which includes ten X-STRs, and the Investigator Argus X-12 Kit (Argus X-12, Qiagen, Valencia, CA, USA) that analyzes three of the markers included in the decaplex and nine additional X-STRs. The disadvantages of both kits are the limited number of markers and the large size of some of the amplification products that make them inappropriate for application in highly degraded DNA.

Consequently, the aim of this study was to design a new multiplex of 17 X-STRs combining markers included in the decaplex of the GHEP-ISFG or Argus X-12 and six more to increase both the power of discrimination and the paternity exclusion probabilities for its application in well preserved or highly degraded DNA.

Materials and methods

In order to set up the PCR amplification conditions and optimize the new 17 X-STR multiplex, three DNA control samples were used: 2800M and K562 from Promega (Promega® Corporation, USA) as well as the AmpFLSRT™ PCR Amplification 9947A (Thermo Fisher Scientific, Waltham, MA, USA).

For the validation of the 17 X-STRs, 200 unrelated healthy individuals were obtained from four population samples including Asians from Thailand (N=50), European Caucasoids from Spain (N=50), Africans from Malawi and Equatorial Guinea (N=50), and Hispanics from Colombia (N=50).

17 X-STRs were selected; eleven of them were included in the decaplex of the GHEP-ISFG or Argus X-12 panels (DXS8378, DXS9902, DXS7132, DXS9898, DXS6809, DXS6789, DXS7133, GATA172D05, GATA31E08, DXS10079, and DXS7423). In addition, six markers were also incorporated (DXS6800, DXS6801, DXS6803, DXS6807, DXS6799, and DXS10075).

For the amplification of some of the markers, new primers were designed by using PerlPrimer v1.1.21 software [5] while for the others, the primers were the same as described in the decaplex of the GHEP-ISFG. The lack of interactions between primers was checked with the Autodimer [6] and BLASTN (http://blast.ncbi.nlm.nih.gov/Blast.cgi) programs. All forward primers of each marker were modified by the addition of a fluorescent dye at their 5' end with the exception of the DXS10075 marker, labeled at the 5' end of the reverse primer. The fluorescent dyes used to label the primers were *5-FAM*, *YAKIMA YELLOW*, *ATTO 550*, and *ATTO 565* (Eurofins Genomics, Ebersberg, Germany).

The PCR reaction consisted of 5 µl of the QIAGEN Multiplex PCR Kit (Qiagen, Valencia, CA, USA), 3.5 µl of water, 1 µl of DNA previously diluted to 1-3 ng/µl, and 0.5 µl of primer mix.

The amplification was carried out on the GeneAmp® 9800 PCR System (Thermo Fisher Scientific, Waltham, MA, USA) under the same thermocycling conditions used for the GHEP-ISFG decaplex [4]. Finally, the PCR products were separated and detected on an ABI PRISM 3130 Genetic Analyzer. Electrophoresis data were analyzed with GeneMapper ID Software v4.0 (Thermo Fisher Scientific, Waltham, MA, USA).

To facilitate allele designation, an allelic ladder was constructed by pooling different samples that covered the known allelic range.

In order to check the sensitivity of the new multiplex, a range from 30 ng to 25 pg of two positive controls (2800M and 9947A) were analyzed in triplicate.

Finally, to verify the utility of this new multiplex, the following population parameters were analyzed by using Arlequin software v.3.5.1.2 [7]: allelic frequencies, genetic diversity (GD), Hardy-Weinberg equilibrium, and pairwise linkage disequilibrium. In addition, forensic parameters power of discrimination for males (PD_M) and females (PD_F), as well as the paternity exclusion index in duos (MEC_D) and trios (MEC_T) by Desmarais *et al.* [8], were estimated from the allelic frequencies for each population by using the online tool of the Forensic ChrX Research database (http://www.chrx-str.org).

Results and discussion

A panel of 17 X-STRs has been developed. The distribution of the markers in this panel is shown in Fig. 1, which corresponds to one of the samples included in the study for the 17 X-STR multiplex system.

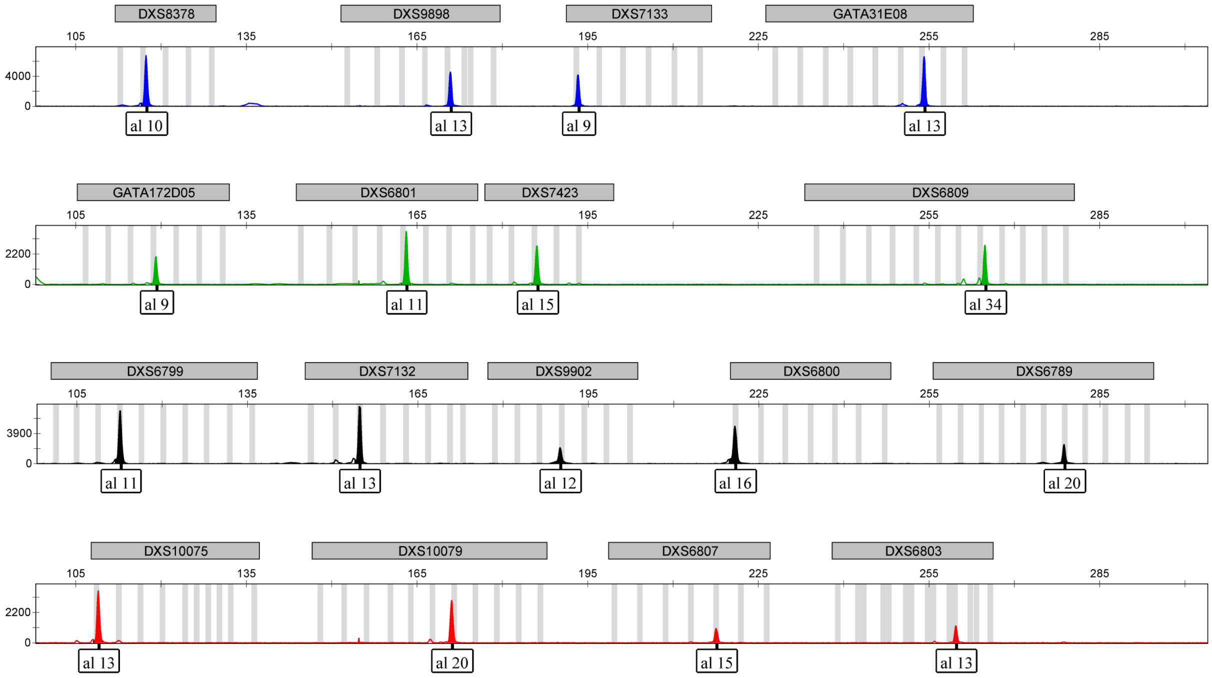


Fig. 1. Electropherogram of one of the samples included in this study for the 17 X-STR multiplex system.

An allelic ladder has been constructed considering all the known alleles for each locus in this study.

The sensitivity study has proven that the minimum quantity of DNA to obtain complete genetic profiles was 100 pg.

The combined values of forensic parameters obtained with the 17 X-STRs were compared with those obtained by taking into account the markers included in the decaplex of the GHEP-ISFG. The forensic efficiency evaluation for the new multiplex showed high values of combined power of discrimination in males ($cPD_M \geq 0.999999994$)

and females ($cPD_F \geq 0.99999999999998$), as well as combined paternity exclusion probabilities in trios ($cMEC_T \geq 0.99999994$) and duos ($cMEC_D \geq 0.999990$), proving the discriminatory strength of this panel. These values were at least two orders of magnitude higher for the combined paternity exclusion probabilities in trios ($cMEC_T$), 3 orders of magnitude higher for the combined paternity exclusion probabilities in duos ($cMEC_D$) and combined power of discrimination in males (cPD_M), and 5 orders of magnitude higher for the power of discrimination in females (cPD_F) than those obtained by considering only the ten markers included in the decaplex of the GHEP-ISFG.

Conclusion

In this work, we describe a highly discriminative multiplex system that permits the analysis of 17 of the most commonly used X-STRs in a single PCR reaction.

Acknowledgements

Funds were provided by the Basque Government (Grupo Consolidado IT833-13). The authors are grateful to Prof. Dr. José Luis Blázquez Caeiro at the University of Santiago de Compostela (Spain) for the African population samples, the Spanish National DNA Bank for the European Caucasoid sample collection, and to PhD Maite Álvarez, for her technical and human support provided by the DNA Bank Service (SGIker) of the University of the Basque Country (UPV/EHU).

Conflict of interest

The authors have declared no conflict of interest.

References

[1] L. Gusmão, C. Alves, P. Sánchez-Diz, et al., Results of the GEP-ISFG collaborative study on an X-STR Decaplex, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 677–679.

[2] M.C. Fernández, Estudio de los microsatélites y miniSTRs del cromosoma X de aplicación forense, (2013).

[3] N. Pinto, L. Gusmão, A. Amorim, X-chromosome markers in kinship testing: A generalisation of the IBD approach identifying situations where their contribution is crucial, Forensic Sci. Int. Genet. 5 (2011) 27–32.

[4] L. Gusmão, P. Sánchez-Diz, C. Alves, et al., A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: Data on 15 Iberian and Latin American populations, Int. J. Legal Med. 123 (2009) 227–234.

[5] O.J. Marshall, PerlPrimer: Cross-platform, graphical primer design for standard, bisulphite and real-time PCR, Bioinformatics. 20 (2004) 2471–2472.

[6] P.M. Vallone, J.M. Butler, AutoDimer: A screening tool for primer-dimer and hairpin structures, Biotechniques. 37 (2004) 226–231.

[7] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis, Evol. Bioinform. Online. 1 (2005) 47–50.

[8] D. Desmarais, Y. Zhong, R. Chakraborty, et al., Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), J. Forensic Sci. 43 (1998) 1046–1049

Study number 3

'Forensic Spanish allele and haplotype database for a 17 X-STR panel'

The study number 3 corresponds to the attainment of the first part of the objective 3: *Performing allele and haplotype frequency databases from Spain and other populations located on the Atlantic coast of Europe and North-West Africa by using the previously developed 17 X-STR panel that will allow its application in forensic casework*.

Forensic calculation based on STR data requires updated allele and haplotype frequency databases of each heterogeneous population. In this context, the forensic community is continuously performing and upgrading allele and haplotype frequency databases with the most used multiplexes. Regarding the X-chromosome, a great majority of the frequency databases correspond to those performed with the decaplex of the GHEP-ISFG and/or the Investigator® Argus X-12 Kit.

In terms of the Iberian Peninsula, many population samples have been typed, at least, with one of the above-mentioned multiplexes, e.g. Zamora, Galicia, Cantabria, autochthonous Basques from Navarre, north-east of Spain, Valencia, Murcia, and Balearic Islands, among others. Despite this, only a global Spanish allele database was available for the markers of the decaplex of the GHEP-ISFG. In view of this point, the main objective of the present study was to update the only allele frequency database available and to broaden the forensic applicability of the 17 X-STR panel to the Spanish population. With this purpose, 593 unrelated individuals from seven different regional populations of Spain, i.e. Alicante, Aragon, the Basque Country, Andalusia, Galicia, Madrid, and Barcelona, have been typed with the 17 X-STRs and frequency data, as well as efficiency parameters of forensic interest were calculated. The determination of genetic distances based on F_{ST}, revealed no significant differences among the analyzed populations. This lack of significance supports the use of the allele and haplotype frequency database presented herein as a global Spanish population sample for statistical evaluation with the 17 X-STR panel.

In addition, the use of X-STRs in the forensic field requires a precise knowledge of both physical linkage and linkage disequilibrium between markers in each population. Typically, the X-chromosome has been divided into linkage groups 1–4 located on the following regions: Xp22.2, Xq12, Xq26, and Xq28. However, and despite the physical distance between two loci, recombination and crossing over might occur in case of existence of hotspots of recombination. With the aim of studying the linkage state of the markers included in the 17 X-STR panel, analyses of LD plots derived from SNP data (based on HapMap Project) and pairwise LD tests were carried out. After evaluation, the cluster DXS7132-DXS10075-DXS10079 was considered and therefore, haplotype frequencies are presented.

This study has resulted in an international publication in the journal *Forensic Science International: Genetics* in June 2016 under the heading '*Forensic Spanish allele and haplotype database for a 17 X-STR panel'.* Q1, IP: 4.988. The publication is shown below.

Forensic Science International: Genetics 24 (2016) 120–123

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Short communication

Forensic Spanish allele and haplotype database for a 17 X-STR panel

Endika Prieto-Fernández [a], Carolina Núñez [a], Miriam Baeta [a], Susana Jiménez-Moreno [b], Begoña Martínez-Jarreta [c], Marian M. de Pancorbo [a,*]

[a] BIOMICs Research Group, Lascaray Research Center, University of the Basque Country UPV/EHU, Avda. Miguel de Unamuno, 3. 01006 Vitoria-Gasteiz, Spain.
[b] Área Medicina Legal y Forense, Dpto. Patología y Cirugía, Universidad Miguel Hernández de Elche, Elche, Alicante, Spain.
[c] Department of Forensic Medicine, University of Zaragoza, Zaragoza, Spain.
* Corresponding author

Received 22 April 2016, Revised 20 June 2016, Accepted 22 June 2016, Available online 22 June 2016.

Abstract

The currently developed 17 X-STR panel (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS6801, DXS7423, DXS6809, DXS6799, DXS7132, DXS9902, DXS6800, DXS6789, DXS10075, DXS10079, DXS6807, and DXS6803) offers a highly discriminative tool for forensic identification and kinship testing. With the aim of providing a global Spanish population X-STR database, we present haplotype and allele frequencies and parameters of forensic interest for the 17 X-STR panel obtained from 593 unrelated individuals from Alicante, Aragon, the Basque Country, Andalusia, Galicia, Madrid, and Barcelona that represent the most populated regions of the Spanish Peninsular territory. The seven populations were compared to test possible population genetic substructures. The lack of significant differences among the studied Spanish populations supports the use of the allele and haplotype frequency database presented herein as a global Spanish population sample useful for statistical evaluation in forensic casework. After conducting the LD plots derived from HapMap and pairwise linkage disequilibrium tests, DXS7132, DXS10075, and DXS10079 markers were included in a cluster and haplotype frequencies were calculated. The improvement in the forensic parameters for the Spanish population using 17 X-STRs in comparison to the previous 10 X-STR allele frequencies database is also shown.

Keywords: 17 X-STRs, forensic, database, linkage disequilibrium, haplotype

Introduction

The study of X-chromosomal short tandem repeats (X-STRs) has been established as a valuable tool for forensic routine practice in complex kinship cases as a result of the particular inheritance of the X chromosome [1]. Females inherit one of their two X chromosomes from their mother and the other from their father whilst males receive their only X chromosome from their mother [2]. Thus, the study of X-STRs can trace back long pedigrees unless marker transmission breaks down at father-son relationships [3]. In the same way, due to their higher mean exclusion chance (MEC) [4], the study of X-STRs provides an excellent complement to the analysis of autosomal (AS-STRs) and Y chromosome STRs (Y-STRs) markers, as well as mitochondrial DNA [1] in some family cases, e.g. testing father-daughter, grandparents-grandchildren or determining the exclusion of the paternity of two sisters or half-sisters even if DNA of the parents is not available. Moreover, due to the contribution of the X-STRs to the determination of complex relationships, the application of these markers is particularly interesting in the identification of war and mass disaster victims [4].

The forensic application of X-STRs requires the establishment of allele frequency databases from different populations. Currently, numerous worldwide populations have been typed by different X-STR multiplexes, being the decaplex developed by the GHEP-ISFG [5,6] and the Investigator^{TM} Argus X-12 Kit (Qiagen GmbH, Hilden, Germany) two of the most generally used panels. Up to now, X-STR allele frequencies have been studied in different populations of Spain with the decaplex of the GHEP-ISFG (Murcia [7,8], autochthonous Basques from Navarre [9], Galicia [6], Cantabria [6], and north-east of Spain [10]) and with the Investigator^{TM} Argus X-12 Kit (Zamora [11], Valencia [12], and Balearic Islands [12]). Actually, a unique Iberian allele frequency database for the markers of the decaplex of the GHEP-ISFG is available [1]. Recently, a new 17 X-STR panel in a single PCR reaction has been developed by our group [13] that allows the creation of a more extended database for the Spanish population.

Being all the X-STRs placed on the same chromosome, it is probable that two or more markers are linked, and hence they do not segregate independently. The use of X-STRs for forensic purposes requires a precise knowledge of the genetic linkage among the X chromosome markers [14] in order to avoid the application of the product rule that is invalid for calculating the forensic parameters when markers are linked [2]. Typically, the X chromosome has been divided into the linkage groups 1-4 located at the following regions Xp22.2, Xq12, Xq26, and Xq28 [4]. However, if the physical distance among loci is very small, the recombination and crossing over might occur in case of existence of hotspots of recombination between them [15].

The present work reports the frequency data and forensic efficiency parameters for 17 X-STRs in a sample set of 593 unrelated individuals from seven different regional populations of Spain (Alicante, Aragon, the Basque Country, Andalusia, Galicia, Madrid, and Barcelona) in order to generate a global database of the Spanish population for the 17 X-STRs.

Materials and methods

Samples and DNA extraction

A total of 593 unrelated individuals (391 men and 202 women) from seven different regions of Spain were analyzed. Samples of Alicante (N= 56, XY= 56), Aragon (N= 50, XY= 50), the Basque Country (N= 75, XX=9, XY= 66), Andalusia (N= 123, XX= 59, XY=64), Galicia (N= 74, XX= 41, XY= 33), Madrid (N= 103, XX= 35, XY= 68), and Barcelona (N= 112, XX= 58, XY= 54) were obtained from volunteer donors under informed consent, following the ethical standards of the Helsinki Declaration. Samples from Alicante and Aragon were provided by the University Miguel Hernández and University of Zaragoza, respectively. Samples from Andalusia, Madrid, Galicia, and Barcelona were provided by the Banco Nacional de ADN Carlos III (Spanish national DNA bank) (BNADN Ref. 12/0031). Favorable ethical reports were obtained (Faculty of Pharmacy UPV/EHU, September 26th, 2008; CEISH/119/2012) for the Basque Country population sample set. Male and female individuals were sampled for each population whenever possible. DNA extraction of individuals from Alicante was performed from buccal swabs by a phenol-chloroform-isoamyl alcohol method, while DNA of individuals from Aragon and the Basque Country was extracted from bloodstains and mouthwashes, respectively, by using the same method. DNA was quantified by using the Scientific NanoDrop^{TM} 1000 Spectrophotometer (ThermoFisher Scientific Inc., Wilmington, DE).

Amplification and data analysis

The amplification was performed using the 17 X-STRs system (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS6801, DXS7423, DXS6809, DXS6799, DXS7132, DXS9902, DXS6800, DXS6789, DXS10075, DXS10079, DXS6807, and DXS6803) as described in Prieto-Fernández *et al* (2016) [13]. Capillary electrophoresis of PCR products was conducted on an ABI PRISM 3130 Genetic Analyzer (Thermofisher Scientific, Waltham, MA, USA) and allele designation was performed by using GeneMapper ID software version 4.0 (Thermofisher Scientific, Waltham, MA, USA).

Confirmation of previously undescribed allele variants

New alleles previously not detected during the development and validation of the 17 X-STR panel [13] were checked by size comparison with other samples that presented neighboring alleles by blending the amplification products and analyzing the mixture by capillary electrophoresis. Additionally, to confirm a previously undescribed allele variant observed in a heterozygous genotype that did not present neighboring alleles to compare with, sequencing was performed from the DNA bands of each allele separated by electrophoresis in a 3% agarose gel. The DNA bands were cut with a scalpel from the gel, extracted from the agarose by using the QIAEX II Gel Extraction Kit (Qiagen GmbH, Hilden, Germany), and purified with the MinElute PCR Purification Kit (Qiagen GmbH, Hilden, Germany). Thereafter the reamplification was performed, and sequencing was carried out as described in Cardoso et al (2010) [16].

Linkage disequilibrium analyses

We used the physical localization of all the loci to download the HapMap SNP genotype data of the regions among consecutive markers (HapMap Data Rel 28 PhaseII+III, August10, on NCBI B36 assembly dbSNP b126) for the Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) from the HapMap website (http://www.hapmap.org). SNPs located upstream and downstream of the studied neighboring markers in each case were selected to delimit the flanking regions of interest [15]. HaploView LD plots were used to analyze recombination and linkage disequilibrium of the 17 X-STRs, and pairwise linkage disequilibrium tests were also carried out. The Haploview LD plots and pairwise linkage disequilibrium analyses were performed through the Haploview 4.2 [17] and Arlequin v3.5.1.2 [18] programs, respectively. To calculate the pairwise linkage disequilibrium, only the male subpopulation was considered.

Population comparisons

Pairwise F_{ST} genetic distances among the seven Spanish populations studied herein were estimated by using the Arlequin software v.3.5.1.2 [18]. The global Spanish population was also compared with other worldwide populations analyzed with the 17 X-STRs [13] using the same program and visualized with a multidimensional scaling (MDS) analysis, based on F_{ST} distances, using PAST software v.3.04 [19].

Population and forensic analyses

Allele frequencies and gene diversity (GD) for the seven different populations of Spain were calculated from both male and female samples. Hardy Weinberg equilibrium (HWE) was tested in the female subpopulation sample. The same parameters were also calculated for the global Spanish population. Haplotype frequencies for the cluster DXS7132-DXS10075-DXS10079 were calculated from male samples for the global Spanish population. All aforementioned parameters were estimated by using the Arlequin software v.3.5.1.2 [18]. In addition, paternity exclusion index in duos (MEC_D) and trios (MEC_T) [20], as well as, power of discrimination for males (PD_M) and females (PD_F) were calculated for the global Spanish population by using the online tool of the Forensic ChrX Research database (http://www.chrx-str.org).

Results and discussion

Genetic profiles for 593 individuals from Alicante, Aragon, the Basque Country, Andalusia, Galicia, Madrid, and Barcelona are presented in Supplementary Table S1. All the profiles were unique in the different populations. New alleles previously not detected during the development and validation of the 17 X-STR panel [13] were found in the following markers: DXS9898 (allele 9), DXS7133 (allele 7), GATA172D05 (allele 13), DXS6801 (alleles 7 and 15), DXS6799 (allele 15), DXS9902 (alleles 11.1, 13.1, and 15), DXS6789 (allele 23.1), DXS10075 (allele 12), and DXS6803 (allele 16.3). All these new alleles were checked by comparison with other samples that presented neighboring alleles by blending the amplification products and analyzing the mixture by capillary electrophoresis. Additionally, the sample BCN147 was sequenced to confirm the genotype 11/16.3 for the DXS6803 marker. This procedure was performed due to the lack of neighboring alleles for the comparison of the fragment size of the new variant.

Allele frequencies for each marker, HWE, and GD were determined for the analyzed seven Spanish populations (Supplementary Table S2). No deviation from HWE was observed for any of the analyzed seven Spanish populations after Bonferroni correction ($p > 0.0029$). The pairwise p-values of linkage disequilibrium for the seven Spanish populations analyzed herein are shown in Supplementary Table S3.

No statistically significant differences were observed among the seven populations after Bonferroni correction ($p > 0.0024$) (Supplementary Table S4, Figure 1), dismissing a possible genetic substructure. Therefore, all the individuals were included in a global Spanish population sample.

Pairwise F_{ST} genetic distances between the global Spanish population and other three populations analyzed with the 17 X-STRs (Asians from Thailand, Hispanics from Colombia, and Africans) [13] were calculated. The global Spanish population was clearly differentiated from the Asians from Thailand, Hispanics from Colombia, and African populations. This genetic scenario, as well as the results among the seven Spanish populations, is shown in the two-dimensional MDS plots in Fig. 1.

Allele frequencies and genetic parameters were determined for the first time for the Spanish population for this newly developed panel of 17 X-STRs [13]. For each marker, allele frequencies, HWE, GD, and statistical parameters of forensic interest (MEC_D, MEC_T, PD_M, and PD_F) are shown in Supplementary Table S5. No deviations from HWE were observed for any of the analyzed loci in female samples for a significance level of 0.0029 (after Bonferroni correction).

The markers included in the 17 X-STR panel [13] have been gathered into different clusters throughout the literature, such as the cluster DXS6807-DXS8378-DXS9902 in the region Xp22 [14]. In the region Xq21, the markers DXS6801, DXS6809, and DXS6789 have been considered linked by the majority of the authors [3,21,22] and additionally, DXS6799 has also been included in this cluster [14]. Finally, the markers DXS7132, DXS10079, and DXS10074 have typically been considered as a haplotype in the Xq12 region [21,22,23] and some authors have also considered the locus DXS10075 [14,24] as part of it.

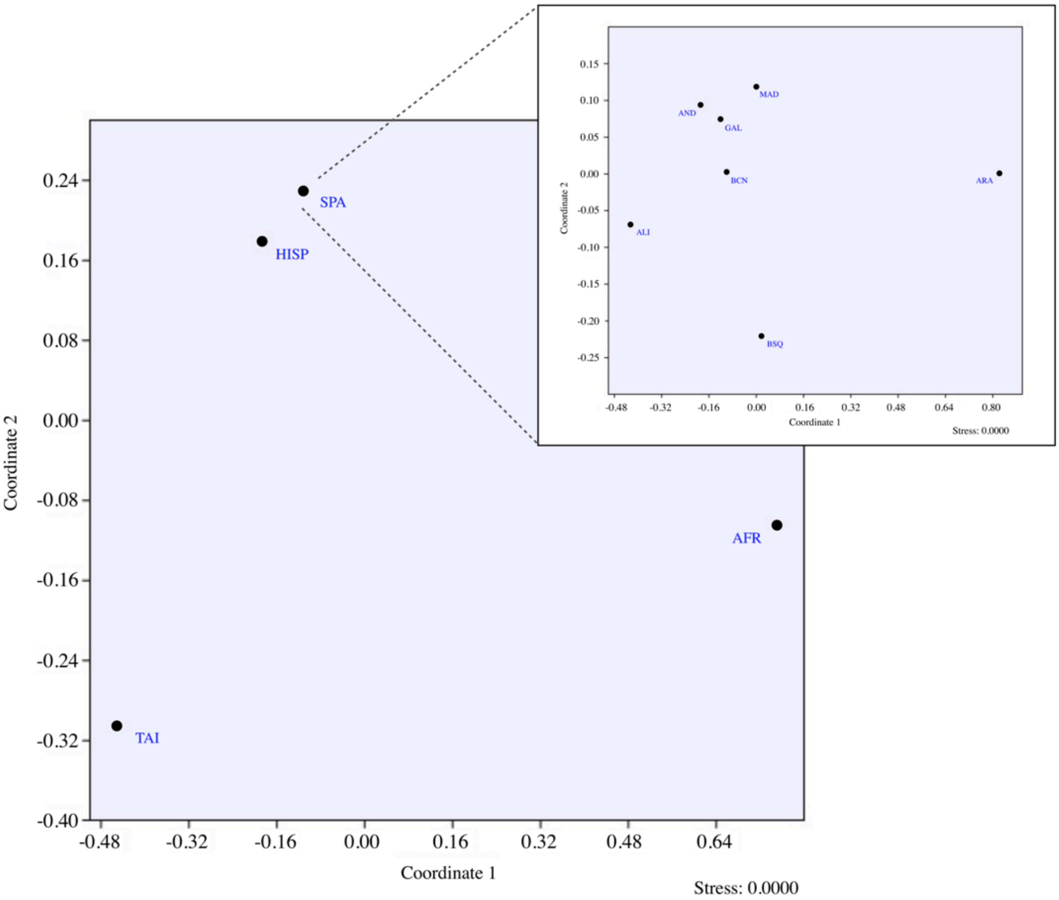


Fig. 1. MDS based on pairwise F_{ST} genetic distances calculated between the global Spanish population and other three populations analyzed with the 17 X-STRs. The result among the seven Spanish populations is also shown. Studied populations: Alicante (ALI), Aragon (ARA), Basque Country (BSQ), Andalusia (AND), Galicia (GAL), Madrid (MAD), Barcelona (BCN), and global Spanish population (SPA). Other populations included for comparisons [13]: Asians from Thailand (TAI), Hispanics from Colombia (HISP), and Africans (AFR).

In Supplementary Figure S1, the HaploView LD Plot analysis for the region 92,397,587 bp (rs1589306) - 97,265,948 bp (rs2123628) (NCBI B36), which covered the loci DXS6801, DXS6809, DXS6789, and DXS6799 is shown. In Supplementary Figure S2, the region 78,565,893 bp (rs5912942) - 87,685,115 bp (rs5942237), which covered the loci DXS6800, DXS6803, and DXS9898 (NCBI B36) is represented. Both regions showed numerous hotspots of recombination between the SNPs located upstream and downstream of the studied X-STR markers in each case and according to these results, we cannot consider any of these loci linked. On the other hand, the analysis of the region between the markers DXS10079-DXS10075 has not showed hotspots of recombination and could be considered genetically linked. The Haploview LD plot that covered the region between DXS7132 and DXS10079 was incomplete due to insufficient information about the SNPs there located for the CEU collection.

We have performed the pairwise linkage disequilibrium test with the 391 males of the global Spanish population (Supplementary Table S3) and after Bonferroni correction (p= 0.0004), only significant association was found between the loci DXS10075 and DXS10079 (p< 0.0000). These results are in concordance with the HaploView LD Plots derived from HapMap for these two markers. On the contrary, pairwise linkage disequilibrium tests do not result statistically significant neither for the pair DXS7132 and DXS10075, nor for DXS7132 and DXS10079, in the analyzed population. Nonetheless, these results do not ensure the absence of linkage, since LD or allelic association only means that alleles at different loci occur together in an individual more often than expected by chance [25]. Taking into account these results, we rather be conservative and maintain the loci DXS7132, DXS10075, and DXS10079 as a haplotype. In order to assure this cluster, family studies would be recommended. Accordingly, to calculate the combined forensic parameters with the 17 X-STR panel (Supplementary Table S5), the cluster DXS7132-DXS10075-DXS10079 was considered as a haplotype. Haplotype frequencies for this cluster are also shown in Supplementary Table S6.

Combined forensic parameters were calculated and compared with those based on the frequencies from the Iberian allele frequency database for 10 X-STRs [1]. The combined values improved considerably for the 17 X-STRs [13]. The combined MEC values increased from 0.9998 to 0.9999994 for duos ($cMEC_D$) and from 0.999995 to 0.999999998 in trios ($cMEC_T$). Likewise, the combined PD values increased from 0.9999989 to 0.9999999998 in males (cPD_M) and from 0.9999999998 to 0.99999999999999994 in females (cPD_F). Thus, this set of 17 X-STRs proved to be highly discriminative using the global Spanish database.

Conclusion

In conclusion, the 17 X-STR panel offers a highly discriminative tool for forensic identification and kinship testing. It is necessary to define the linkage of the X chromosome in order to define the haplotypes and their corresponding frequencies. We have considered the cluster DXS7132-DXS10075-DXS10079 among all the 17 X-STRs after the analyses of LD plots derived from HapMap and pairwise linkage disequilibrium tests. The lack of significant differences among the studied Spanish populations supports the use of the allele and haplotype frequency database presented herein as a global Spanish population sample for statistical evaluation of the results in forensic casework.

Acknowledgements

Funds were provided by the Basque Government (Grupo Consolidado IT833-13). The authors are grateful to the Spanish National DNA Bank for the Spanish sample collection and to PhD Maite Álvarez for her technical and human support provided by the DNA Bank Service (SGIker) of the University of the Basque Country (UPV/EHU) and European funding (ERDF and ESF).

Conflict of interest

The authors have declared no conflict of interest.

Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.fsigen.2016.06.016.

References

[1] M. Baeta, M.J. Illescas, L. García, C. Núñez, E. Prieto-Fernández, S. Jiménez-Moreno, et al., Iberian allele frequency database for 10 X-STRs, Forensic Sci. Int. Genet. 19 (2015) 76–78.

[2] A.O. Tillmar, T. Egeland, B. Lindblom, G. Holmlund, P. Mostad, Using X-chromosomal markers in relationship testing: Calculation of likelihood ratios taking both linkage and linkage disequilibrium into account, Forensic Sci. Int. Genet. 5 (2011) 506–511.

[3] R. Szibor, S. Hering, E. Kuhlisch, I. Plate, S. Demberger, M. Krawczak, et al., Haplotyping of STR cluster DXS6801-DXS6809-DXS6789 on Xq21 provides a powerful tool for kinship testing, Int. J. Legal Med. 119 (2005) 363–369.

[4] R. Szibor, X-chromosomal markers: Past, present and future, Forensic Sci. Int. Genet. 1 (2007) 93–99.

[5] L. Gusmão, C. Alves, P. Sánchez-Diz, M.T. Zarrabeitia, M.A. Abovich, I. Aragón, et al., Results of the GEP-ISFG collaborative study on an X-STR Decaplex, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 677–679.

[6] L. Gusmão, P. Sánchez-Diz, C. Alves, I. Gomes, M.T. Zarrabeitia, M. Abovich, et al., A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations, Int. J. Legal Med. 123 (2009) 227–234.

[7] M.J. Illescas, J.M. Aznar, S. Cardoso, A. López-Oceja, D. Gamarra, J.F. Sánchez-Romera, et al., Genetic characterization of ten X-STRs in a population from the Spanish Levant, Forensic Sci. Int. Genet. 6 (2012) e180–e181.

[8] M.J. Illescas, J.M. Aznar, S. Cardoso, A. López-Oceja, D. Gamarra, J.F. Sánchez-Romera, et al., Genetic diversity of 10 X-STR markers in a sample population from the region of Murcia in Spain, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) e437–e438.

[9] M.J. Illescas, A. Pérez, J.M. Aznar, L. Valverde, S. Cardoso, J. Algorta, et al., Population genetic data for 10 X-STR loci in autochthonous Basques from Navarre (Spain), Forensic Sci. Int. Genet. 6 (2012) e146–e148.

[10] B. García, M. Crespillo, M. Paredes, J.L. Valverde, Population data for 10 X-chromosome STRs from north-east of Spain, Forensic Sci. Int. Genet. 6 (2012) e13–e15.

[11] J.C. Pinto, V. Pereira, S.L. Marques, A. Amorim, L. Alvarez, M.J. Prata, Mirandese language and genetic differentiation in Iberia: a study using X chromosome markers, Ann. Hum. Biol. 42 (2015) 20–25.

[12] J.F. Ferragut, K. Bentayebi, J.A. Castro, C. Ramon, A. Picornell, Genetic analysis of 12 X-chromosome STRs in Western Mediterranean populations, Int. J. Legal Med. 129 (2015) 253–255.

[13] E. Prieto-Fernández, M. Baeta, C. Núñez, M.T. Zarrabeitia, R.J. Herrera, J.J. Builes, et al., Development of a new highly efficient 17 X-STR multiplex for forensic purposes, Electrophoresis 37 (2016) 1651-1658.

[14] Q.L. Liu, J.Z. Wang, L. Quan, H. Zhao, Y. Da Wu, X.L. Huang, et al., Allele and Haplotype Diversity of 26 X-STR Loci in Four Nationality Populations from China, PLoS One. 8 (2013) e65570.

[15] H.B. Luo, Y. Ye, Y.Y. Wang, W.B. Liang, L.B. Yun, M. Liao, et al., Characteristics of eight X-STR loci for forensic purposes in the Chinese population, Int. J. Legal Med. 125 (2011) 127–131.

[16] S. Cardoso, M.T. Zarrabeitia, L. Valverde, A. Odriozola, M.A. Alfonso-Sanchez, M.M. de Pancorbo, Variability of the entire mitochondrial DNA control region in a human isolate from the Pas Valley (northern Spain), J. Forensic Sci. 55 (2010) 1196-1201.

[17] J.C. Barrett, B. Fry, J. Maller, M.J. Daly, Haploview: analysis and visualization of LD and haplotype maps, Bioinformatics. 21 (2005) 263–265.

[18] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): An integrated software package for population genetics data analysis, Evol. Bioinform. Online. 1 (2005) 47–50.

[19] Ø. Hammer, D.A.T. Harper, P.D. Ryan, Past: Paleontological statistics software package for education and data analysis, Palaeontol. Electron. 4 (2001) 9–17.

[20] D. Desmarais, Y. Zhong, R. Chakraborty, C. Perreault, L. Busque, Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), J. Forensic Sci. 43 (1998) 1046–1049.

[21] S. Pasino, S. Caratti, M. Del Pero, A. Santovito, C. Torre, C. Robino, Allele and haplotype diversity of X-chromosomal STRs in Ivory Coast, Int. J. Legal Med. 125 (2011) 749–752.

[22] S. Inturri, S. Menegon, A. Amoroso, C. Torre, C. Robino, Linkage and linkage disequilibrium analysis of X-STRs in Italian families, Forensic Sci. Int. Genet. 5 (2011) 152–154.

[23] L. Cainé, S. Costa, M.F. Pinheiro, Population data of 12 X-STR loci in a North of Portugal sample, Int. J. Legal Med. 127 (2013) 63–64.

[24] S. Hering, C. Augustin, J. Edelmann, M. Heidel, J. Dressler, H. Rodig, et al., DXS10079, DXS10074 and DXS10075 are STRs located within a 280-kb region of Xq12 and provide stable haplotypes useful for complex kinship cases, Int. J. Legal Med. 120 (2006) 337–345.

[25] M. Slatkin, Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. 9 (2008) 477–85.

Study number 4

‘A genetic overview of Atlantic coastal populations from Europe and North-West Africa based on a 17 X-STR panel’

The fourth study of the present work corresponds to the attainment of the second part of the objective 3: *Performing allele and haplotype frequency databases from Spain and other populations located on the Atlantic coast of Europe and North-West Africa by using the previously developed 17 X-STR panel that will allow its application in forensic casework*.

The establishment of suitable haplotype frequency databases requires the analysis of thousands of individuals from a certain population. However, the available number of individuals is not always large enough and therefore, a great number of haplotypes are not represented in the performed databases. Nevertheless, by merging two or more population samples that present similar allele and haplotype frequency distributions, this problem may be solved.

The objective of the present work is to analyze if some populations located on the Atlantic coast of Europe and North-West Africa actually share alike allele and haplotype frequency distributions since they have experienced genetic exchanges throughout history. With this purpose, 513 unrelated individuals from Brittany (France), Ireland, northern Portugal, and Casablanca (Morocco) have been studied with the 17 X-STR panel. Pairwise F_{ST} genetic distances between the four populations studied herein and other Spanish coastal populations previously analyzed with the 17 X-STR panel were calculated, i.e. Galicia, autochthonous Basque Country, and resident Basque Country.

Our results suggest that certain neighboring populations located on the European Atlantic coast could have experienced episodes of genetic interchange as they have shown genetic similarities between them. On the contrary, other populations, such as Casablanca, autochthonous Basques, and Brittany, have presented significant differences with other populations. In view of these differences, genetic exchanges along the Atlantic coast seem to be insufficient to totally homogenize the distribution of the X-chromosome allele and haplotype frequencies along the seashore. Therefore, the use of individual databases for each population, instead of a global database, would be more appropriate.

This study has resulted in an international publication in the journal *Forensic Science International: Genetics* in November 2016 under the heading ‘*A genetic overview of Atlantic coastal populations from Europe and North-West Africa based on a 17 X-STR panel’.* Q1, IP: 4.988. The publication is shown below.

Forensic Science International: Genetics 27 (2017) 167–171

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Short communication

A genetic overview of Atlantic coastal populations from Europe and North-West Africa based on a 17 X-STR panel

Endika Prieto-Fernández [a], Ana Díaz-de Usera [a], Miriam Baeta [a], Carolina Núñez [a], Faiza Chbel [b], Sellama Nadifi [c], Karen Rouault [d], Claude Férec [d], Orla Hardiman [e, f], Fátima Pinheiro [g], Marian M. de Pancorbo [a,*]

[a] BIOMICs Research Group, Lascaray Research Center, University of the Basque Country UPV/EHU. Avda. Miguel de Unamuno, 3. 01006 Vitoria-Gasteiz (Spain).

[b] Laboratoire National De Référence, Université Mohammed IV des Sciences de la Santé, Boulevard Mohamed Taieb Naciri, Hay El Hassani, BP 82403, Casa Oumrabii, Casablanca.

[c] Laboratoire de Génétique Humaine, Faculté de Médecine et de Pharmacie, Université Hassan II, Casablanca, Morocco.

[d] Inserm, UMR 1078, Brest, France, Université de Bretagne Occidentale, Brest, France, Etablissement Français du Sang – Bretagne, Brest, France, CHRU Brest, Hôpital Morvan, Laboratoire de Génétique Moléculaire et d'Histocompatibilité, Brest, France.

[e] Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Dublin, 2, Ireland, Department of Neurology, Beaumont Hospital, Dublin, 9, Ireland.

[f] Delegação do Norte do Instituto Nacional de Medicina Legal, Jardim Carrilho Videira, 4050-167 Porto, Portugal.

* Corresponding author

Received 23 September 2016, Revised 3 November 2016, Accepted 30 November 2016, Available online 30 November 2016.

Abstract

The forensic use of X-STRs requires the creation of allele and haplotype frequency databases in the populations where they are going to be used. Recently, an updated Spanish allele and haplotype frequency database for the new 17 X-STR panel has been created, being the only database available up to now for this new multiplex. In order to broaden the forensic applicability of the 17 X-STR panel, 513 individuals from four different populations located on the Atlantic Coast of Europe and North–West Africa have been studied, i.e. Brittany (France), Ireland, northern Portugal, and Casablanca (Morocco).

Allele and haplotype frequency databases, as well as parameters of forensic interest for these populations are presented. The obtained results showed that the 17 X-STR panel constitutes a highly discriminative tool for forensic identification and kinship testing in the studied populations. Furthermore, we aimed to study if these populations located on the Atlantic coast actually share alike allele and haplotype frequency distributions since they have experienced genetic exchanges throughout history. This would allow creating larger forensic databases that include several genetically similar populations for its use in forensic casework. For this purpose, pairwise F_{ST} genetic distances between the analyzed populations and others from the Atlantic Coast previously studied with the 17 X-STR panel or the ten coincident markers included in the decaplex of the GHEP-ISFG were estimated. Our results suggest that certain nearby populations located on the European Atlantic coast could have underwent episodes of genetic interchange as they have not shown statistically significant differentiation between them. However, the population of Casablanca showed significant differentiation with the majority of the European populations. Likewise, the autochthonous Basque Country and Brittany populations have shown distinctive allele frequency distributions between them. Therefore, these findings seem to support that the use of independent allele and haplotype frequency databases for each population instead of a global database would be more appropriate for forensic purposes.

Keywords: 17 X-STRs, forensics, database, Atlantic Coast

Introduction

The X-chromosomal short tandem repeats (X-STRs) have been widely studied over the last years by the forensic community [1]. These markers have been established as the perfect complement to the autosomal STRs (AS-STRs) and other lineage markers, such as Y-chromosomal STRs (Y-STRs) and mitochondrial DNA, when solving complex kinship cases [2]. Additionally, the X-STRs are of great utility in the identification of war victims and historical cases where the majority of the profiles to compare correspond to second or third degree relatives, such as grandparents-grandchildren, maternal uncles-nephews, etc. [3].

The forensic application of X-STRs requires the creation of allele and haplotype frequency databases in the populations where they are going to be used. Since the establishment of the decaplex of the GHEP-ISFG [4, 5] and the Investigator^{TM} Argus X-12 Kit (Qiagen

GmbH, Hilden, Germany), many populations have been studied using these two multiplexes [1]. Recently, a new 17 X-STR panel in a single PCR reaction has been developed and validated [6] and, up to now, an updated Spanish allele and haplotype frequency database for this panel has been published [7]. However, the application of the new multiplex in other populations requires the creation of new databases.

The objective of this work was to broaden the forensic applicability of the 17 X-STR panel to other populations. For this purpose, allele and haplotype frequency distributions of four populations located on the Atlantic Coast of Europe and North-West (NW) Africa have been studied in order to enlarge the current X-STR databases for their application in forensic casework.

Materials and methods

Population samples

A total of 513 unrelated individuals (500 men and 13 women) were selected for this study. This sample set comprised individuals from four different regional populations located on the Atlantic Coast of Europe and NW Africa, i.e. Brittany (N= 179; XY= 179), Ireland (N= 100; XY= 100), northern Portugal (N= 79; XY= 79), and Casablanca (N= 155; XX= 13; XY= 142) (Supplementary Figure S1). Samples from Brittany were provided by the Université de Bretagne Occidentale (Brest, France), samples from Ireland by the National Neuroscience Centre (Beaumont Hospital, Dublin, Ireland), samples from northern Portugal by the National Institute of Legal Medicine and Forensic Sciences (Porto, Portugal), and samples from Casablanca by the University of Hassan II (Casablanca, Morocco). All the samples were obtained from volunteer donors under informed consent, following the ethical standards of Helsinki Declaration.

PCR amplification, electrophoresis and data analysis

The 17 X-STR markers (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS6801, DXS7423, DXS6809, DXS6799, DXS7132, DXS9902, DXS6800, DXS6789, DXS10075, DXS10079, DXS6807, and DXS6803) were amplified in a single PCR reaction as described in [6]. Amplification of samples and performance of the PCR reaction were verified by conventional agarose gel electrophoresis. Capillary electrophoresis of PCR products was conducted on an ABI PRISM 3130 Genetic Analyzer with the POP7 polymer (Thermofisher Scientific, Waltham, MA, USA). Electrophoresis data were analyzed with GeneMapper ID software version 4.0 (Thermofisher Scientific, Waltham, MA, USA).

Confirmation of new allele variants

Previously not detected allele variants were checked by size comparison with other samples that presented neighboring alleles. With this purpose, genomic DNA of the compared samples was blended, amplified together in a single PCR reaction and checked by capillary electrophoresis.

Forensic parameters and statistical analysis

Allele frequencies and gene diversity (GD) were calculated for the four different populations analyzed herein from both male and female samples, when applicable. On the other hand, male samples were used to calculate haplotype frequencies for the cluster DXS7132-DXS10075-DXS10079 (as established in a previous study for the 17 X-STR panel [7]) and to perform the pairwise linkage disequilibrium (LD) tests. All aforementioned parameters were estimated by using the Arlequin software v.3.5.1.2 [8]. In addition, paternity exclusion index in duos (MEC_D) and trios (MEC_T) [9], as well as power of discrimination in males (PD_M) and females (PD_F) were calculated for the analyzed populations by using the online tool of the Forensic ChrX Research database (http://www.chrx-str.org).

Population comparisons

Pairwise F_{ST} genetic distances between the four populations studied and other populations previously studied with the 17 X-STR panel located on the Atlantic Coast from northern Iberian Peninsula were estimated. For this comparison, genetic profiles of Galicia and the Basque Country were taken into account [6, 7] (Supplementary Figure 1).

Samples from the Basque Country population were classified in two groups according to previous studies that have detected intrapopulation differentiation [10-14]: 1) autochthonous Basque population, that includes individuals with ancestor’s surnames that support maternal and paternal Basque ancestry for at least three generations; and 2) resident Basque population, that correspond to individuals living in the Basque Country. The criteria for this classification were based on the information collected from each donor. The genetic distance between these two groups was calculated to analyze if there are enough significant differences between them to consider them as independent.

Additionally, the pairwise F_{ST} genetic distances among the above-mentioned populations and six more collected from the bibliography, typed with the decaplex of the GHEP-ISFG [4, 5], were calculated. The following populations were selected by its nearby location to the Atlantic Coast in the Iberian Peninsula: northern and central regions of Portugal [5], Asturias [15], Cantabria [5], Pas Valley (Cantabria) [16], and autochthonous Basques from Navarre [17]. For this calculation, only the coincident markers between the decaplex of the GHEP-ISFG [4, 5] and the 17 X-STR panel [6] (DXS8378, DXS9902, DXS7132, DXS9898, DXS6809, DXS6789, DXS7133, GATA172D05, GATA31E08, and DXS7423) were considered.

The genetic distances based on F_{ST} were calculated with Arlequin software v.3.5.1.2 [8] from the male and female samples. In order to obtain a representation of the genetic distances, a 3D-nonmetric multidimensional scaling (NMDS) analysis was carried out using PAST software v.3.04 [18] and the x-y-z coordinates were represented using the rgl package [19] for R software [20].

Results and discussion

Population genetic profiles for 513 individuals from four different regional populations of the Atlantic Coast of Europe and NW Africa, i.e. Brittany (France), Ireland, northern Portugal, and Casablanca (Morocco), are presented in Supplementary Table S1. All the profiles were unique in the different populations. New variants previously not detected [1, 6, 7] were found and checked for the marker DXS10079 (alleles 13, 18.3, and 21.1). Allele frequencies and GD values for each locus were also calculated and presented in Supplementary Table S2.

For this panel, the cluster DXS7132-DXS10075-DXS10079 was considered as determined in [7], where analyses of LD plots derived from HapMap data and pairwise LD tests were carried out. Therefore, haplotype frequencies for this cluster were calculated in the four population samples studied in this work and are presented in Supplementary Table S3.

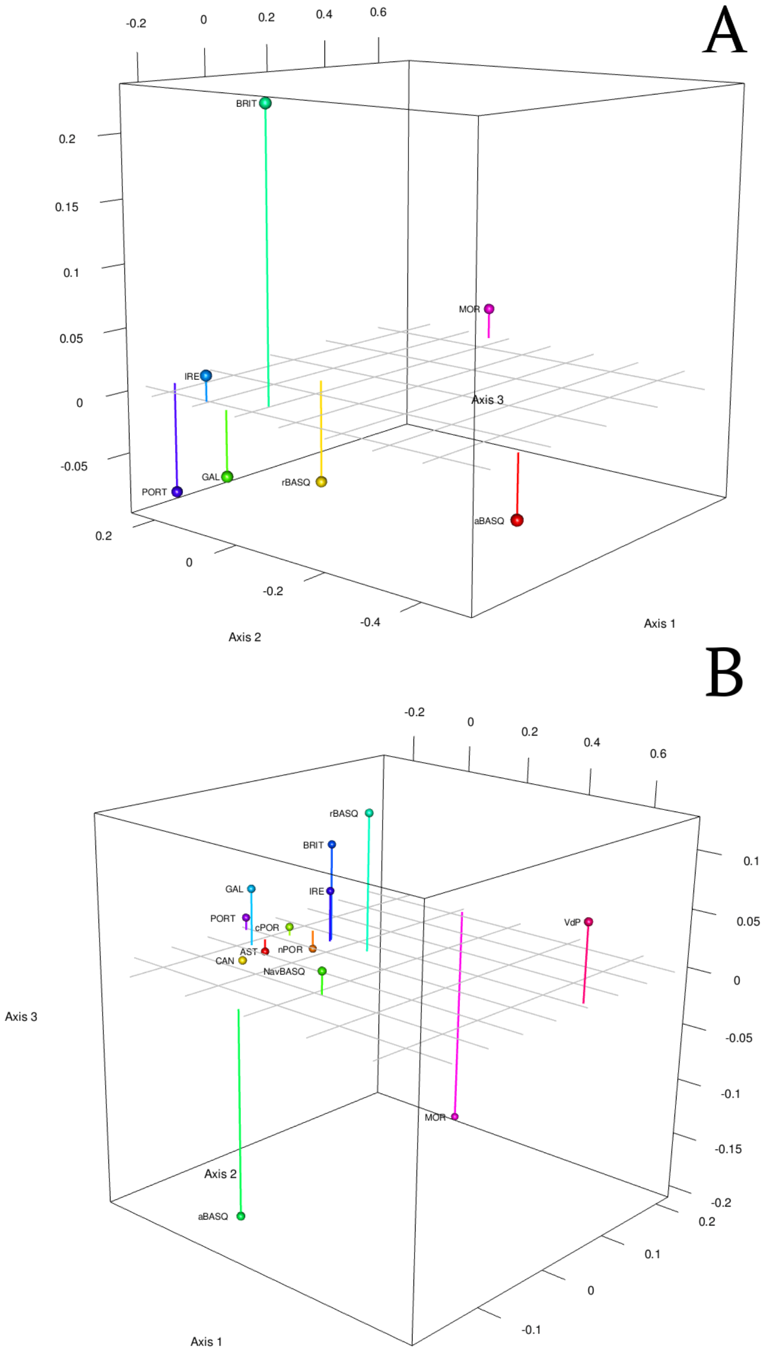


Fig. 1. (A) A 3D-nonmetric multidimensional scaling (NMDS) representation of the genetic distances based on F_{ST} among the four populations studied herein and other three populations from the Atlantic Coast from northern Iberian Peninsula previously studied with the 17 X-STR panel. Stress: 0.0000; (B) A 3D-nonmetric multidimensional scaling (NMDS) representation of the pairwise F_{ST} genetic distances among the four populations studied and nine more that are located close to the Atlantic Coast in the Iberian Peninsula previously studied with the decaplex of the GHEP-ISFG. Stress: 0.0397; Studied populations: Brittany (BRIT), Ireland (IRE), northern Portugal (PORT), and Casablanca (MOR). Other populations included for comparisons: northern and central regions of Portugal (nPOR and cPOR, respectively) [5], Galicia (GAL) [7], Asturias (AST) [15], Cantabria (CAN) [5], Pas Valley (VdP) [16], autochthonous and resident Basque populations (aBASQ and rBASQ, respectively) [6,7], and autochthonous Basques from Navarre (NavBASQ) [17].

Additionally, the pairwise p-values of LD tests were calculated for each population. No detectable evidence of LD was found between loci after Bonferroni correction ($p > 0.0004$), except for the following pairs: DXS6801–DXS6789 in northern Portugal ($p < 0.0000$), DXS6789–DXS6807 in Brittany ($p < 0.0000$), and DXS6801–DXS6803 in Casablanca ($p < 0.0000$) (Supplementary Table S4). Since no evidence of LD was detected for the same pairs of loci in all populations, the obtained results in LD tests may be attributed to spurious effects or may be population-specific [21]. Although the test for pairwise LD has low power to detect LD if the sample size is small, the LD detected among the loci included in the 17 X-STR panel was previously tested in a larger sample size in [7] and LD was not found between the above-mentioned loci.

Statistical parameters of forensic interest (PD_M, PD_F, MEC_D, and MEC_T) were also calculated for each marker included in the 17 X-STR panel, as well as for the cluster DXS7132-DXS10075-DXS10079 (Supplementary Table S5). Combined values of all the aforementioned forensic parameters (cPD_M, cPD_F, $cMEC_D$, and $cMEC_T$), as well as the average GD for each population, are shown in Table 1. The studied populations showed similar GD values for the set of 17 X-STRs, being the population of Casablanca the one that showed the highest GD (0.7453 ± 0.0757). High values for forensic parameters were obtained, with PD values ranging from 0.9999999995 to 0.99999999991 in males and from 0.9999999999999998 to 0.999999999999999991 in females. The highest values of cPD_M and cPD_F correspond to the Casablanca population, where they reached, at least, one order of magnitude higher than in European populations. In the same way, $cMEC_D$ (0.9999990–0.9999998) and $cMEC_T$ (0.999999996–0.9999999994) values are higher in the Casablanca population than in the European populations being, in this case, one order of magnitude higher only in $cMEC_T$ (Table 1). Therefore, these results have proved that this set of markers represents a highly discriminative tool for forensic identification and kinship testing in the studied populations.

Significant differentiation was observed between the autochthonous and resident Basque population groups under a p-value threshold of 0.05 ($F_{ST} = 0.0077$, $p = 0.0219$) and therefore, they have been considered as two independent groups.

The genetic distances based on F_{ST} were calculated among the four populations studied herein and other three populations located on the Atlantic Coast from northern Iberian Peninsula (Galicia, autochthonous Basques and resident Basques) that had previously been studied with the 17 X-STR panel [6, 7]. No statistically significant differentiation was observed after Bonferroni correction (p> 0.0024) between the populations of the Atlantic Coast of Europe, except for Brittany and the autochthonous Basque populations (F_{ST}= 0.0088, p= 0.0003) (Table 2). This differentiation could be due to the particular genetic pool that both Brittany [22, 23] and the autochthonous Basque populations [10-14] display. On the other hand, the population of Casablanca, located in NW Africa, showed significant p-values when comparing it with the rest of the studied populations located on the Atlantic Coast of Europe, except for northern Portugal (F_{ST}= 0.0042, p= 0.0157) and the resident Basque population (F_{ST}= 0.0070, p= 0.0030) (Table 2). The lack of differentiation between northern Portugal and Casablanca populations is in concordance with previous studies of autosomal STRs, which have suggested that among the Iberians, Portuguese are one of the genetically closest to NW Africa [24]. Furthermore, the lack of statistical significance between Casablanca and the resident Basque populations may be due to the effect of a small sample size bias. The low genetic differentiation between Brittany and Ireland populations (F_{ST}= 0.0001, p= 0.4428) may be due to the well-established historical migrations between them, especially during the sixth and seventh centuries [23] (Table 2). A 3D-nonmetric multidimensional scaling (NMDS) representation of the pairwise F_{ST} genetic distances among the populations of the Atlantic Coast is displayed in Fig. 1A.

Table 1. Average gene diversity (GD) and the following combined values of forensic parameters for each population: combined power of discrimination in males (cPD_M) and females (cPD_F), and combined mean exclusion chance in father/daughter duos ($cMEC_D$) and trios involving daughters ($cMEC_T$). Studied populations: Brittany (BRIT), Ireland (IRE), northern Portugal (PORT), and Casablanca (MOR). N= number of X chromosomes.

	BRIT (N= 179)	*IRE (N= 100)*	*PORT (N= 79)*	*MOR (N= 168)*
GD	0.7313 ± 0.0565	0.7377 ± 0.0657	0.7301 ± 0.0701	0.7453 ± 0.0757
cPD_F	0.99999999999999992	0.99999999999999993	0.9999999999999998	0.999999999999999991
cPD_M	0.9999999997	0.9999999998	0.9999999995	0.99999999991
$cMEC_T$	0.999999998	0.999999998	0.999999996	0.9999999994
$cMEC_D$	0.9999993	0.9999994	0.9999990	0.9999998

Additionally, the pairwise F_{ST} genetic distances among the above-mentioned populations and six more located close to the Atlantic Coast, i.e. northern and central regions of Portugal [5], Asturias [15], Cantabria [5], Pas Valley [16], and autochthonous Basques from Navarre [17], were calculated considering the ten coincident markers included in the decaplex of the GHEP-ISFG [4, 5]. With these X-STRs, the populations of the Iberian Peninsula are clustered together, except for the Pas Valley and the three Basque populations (Supplementary Table S6 and Fig. 1B). Compared to the results obtained from 17 X-STRs, the differentiation between Brittany and the autochthonous Basque populations can still be observed (F_{ST}= 0.0126, p= 0.0002) but it cannot be detected between Galicia and Casablanca (F_{ST}= 0.0052, p= 0.0074) after Bonferroni correction (p> 0.0006). These results show the ability of the 17 X-STR panel to detect genetic differentiation between populations more accurately than when only ten X-STR markers are considered. Finally, the lack of significant genetic differences between Casablanca and northern Portugal (F_{ST}= 0.0016, p= 0.2283) has also been detected as it happens among Casablanca and the northern and central region of Portugal populations extracted from the bibliography (F_{ST}= 0.0042, p= 0.0007; and F_{ST}= 0.0044, p= 0.0019, respectively). However, it would be interesting to analyze further populations located on the southern area of Portugal and the Iberian Peninsula to evaluate if the geographical proximity to the African Coast also means a lack of significant genetic differences or if, on the contrary, the cultural barriers act as obstacles to genetic exchanges even between nearby populations communicated by seaways.

Table 2. Pairwise genetic distances (F_{ST}) (below the diagonal) and p-values (above the diagonal; Bonferroni adjusted threshold significance of p= 0.0024) among the studied Atlantic coastal populations. Significant p-values are underlined. Included populations: autochthonous Basque population (aBASQ), resident Basque population (rBASQ), Galicia (GAL), Brittany (BRIT), Ireland (IRE), northern Portugal (PORT), and Casablanca (MOR). N= number of X chromosomes.

	aBASQ	*rBASQ*	*GAL*	*BRIT*	*IRE*	*PORT*	*MOR*
	(N= 72)	*(N= 54)*	*(N= 115)*	*(N= 179)*	*(N= 100)*	*(N= 79)*	*(N= 168)*
aBASQ	*	0.0202 ± 0.0014	0.1301 ± 0.0033	0.0003 ± 0.0002	0.0223 ± 0.0015	0.0275 ± 0.0015	<0.00001 ± 0.0000
rBASQ	0.0077	*	0.0852 ± 0.0029	0.0102 ± 0.0009	0.1505 ± 0.0035	0.2413 ± 0.0043	0.0030 ± 0.0005
GAL	0.0026	0.0039	*	0.0876 ± 0.0034	0.3115 ± 0.0053	0.7218 ± 0.0042	<0.00001 ± 0.0000
BRIT	0.0088	0.0062	0.0019	*	0.4428 ± 0.0049	0.1469 ± 0.0036	<0.00001 ± 0.0000
IRE	0.0052	0.0028	0.0008	0.0001	*	0.6335 ± 0.0044	0.0001 ± 0.0001
PORT	0.0059	0.0020	-0.0014	0.0018	-0.0009	*	0.0157 ± 0.0011
MOR	0.0150	0.0070	0.0084	0.0120	0.0088	0.0042	*

Conclusion

In conclusion, four different populations of the Atlantic Coast of Europe and NW Africa (Brittany, Ireland, northern Portugal, and Casablanca) have been studied with the new 17 X-STR panel and the corresponding allele and haplotype databases have been created. Additionally the results obtained for the forensic parameters reveal that this panel is a highly discriminative tool for forensic identification and kinship testing.

Our results suggest that certain neighboring populations located on the European Atlantic coast could have experienced episodes of genetic interchange as they have shown genetic similarities between them. On the other hand, the population of Casablanca showed significant differentiation with the majority of the European populations, except for the northern and central regions of Portugal. In the same way, the autochthonous Basque Country and Brittany populations have not shown similar allele and haplotype frequency distributions due to its particular genetic pools. In view of these differences, genetic exchanges along the Atlantic coast seem to be insufficient to have homogenized the distribution of the X chromosome markers between the studied populations. Therefore, the use of independent allele and haplotype frequency databases for each population instead of a global database would be more appropriate.

Acknowledgements

Funds were provided by the Basque Government (Grupo Consolidado IT833-13). The authors are grateful to PhD Maite Álvarez for her technical and human support provided by the DNA Bank Service (SGIker) of the University of the Basque Country (UPV/EHU) and European funding (ERDF and ESF).

Conflict of interest

The authors have declared no conflict of interest.

Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.fsigen.2016.11.011.

References

[1] T.M. Diegoli, Forensic typing of short tandem repeat markers on the X and Y chromosomes, Forensic Sci. Int. Genet. 18 (2015) 140–151.

[2] R. Szibor, X-chromosomal markers: past, present and future, Forensic Sci. Int. Genet. 1 (2007) 93 – 99.

[3] N. Pinto, L. Gusmão, A. Amorim, X-chromosome markers in kinship testing: A generalisation of the IBD approach identifying situations where their contribution is crucial, Forensic Sci. Int. Genet. 5 (2011) 27 – 32.

[4] L. Gusmão, C. Alves, P. Sánchez-Diz, M.T. Zarrabeitia, M.A. Abovich, I. Aragón, et al., Results of the GEP-ISFG collaborative study on an X-STR Decaplex, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 677–679.

[5] L. Gusmão, P. Sánchez-Diz, C. Alves, I. Gomes, M.T. Zarrabeitia, M. Abovich, et al., A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations, Int. J. Legal Med. 123 (2009) 227–234.

[6] E. Prieto-Fernández, M. Baeta, C. Núñez, M.T. Zarrabeitia, R.J. Herrera, J.J. Builes, et al., Development of a new highly efficient 17 X-STR multiplex for forensic purposes, Electrophoresis 37 (2016) 1651–1658.

[7] E. Prieto-Fernández, C. Núñez, M. Baeta, S. Jiménez-Moreno, B. Martínez-Jarreta, M.M. de Pancorbo, Forensic Spanish allele and haplotype database for a 17 X-STR panel, Forensic Sci. Int. Genet. 24 (2016) 120–123.

[8] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis, Evol. Bioinform. Online 1 (2005) 47–50.

[9] D. Desmarais, Y. Zhong, R. Chakraborty, C. Perreault, L. Busque, Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), J. Forensic Sci. 43 (1998) 1046–1049.

[10] L. Valverde, S. Köhnemann, M. Rosique, S. Cardoso, M. Zarrabeitia, H. Pfeiffer, et al., 17 Y-STR haplotype data for a population sample of Residents in the Basque Country, Forensic Sci. Int. Genet. 6 (2012) e109 – e111.

[11] J.A. Peña, S. Garcia-Obregon, A.M. Perez-Miranda, M.M. de Pancorbo, M.A. Alfonso-Sanchez, Gene flow in the Iberian Peninsula determined from Y-chromosome STR loci, Am. J. Hum. Biolo. 18 (2006) 532-539.

[12] A.M. Pérez-Miranda, M.A. Alfonso-Sánchez, A. Kalantar, S. García-Obregón, M.M. de Pancorbo, J.A. Peña, et al., Microsatellite data support subpopulation structuring among Basques, J. Hum. Genet. 50 (2005) 403 – 414.

[13] M.M. de Pancorbo, M. López-Martínez, C. Martínez-Bouzas, A. Castro, I. Fernández-Fernández, G.A. de Mayolo, et al., The Basques according to polymorphic Alu insertions, Hum. Genet. 109 (2001) 224 – 233.

[14] S. Cardoso, M.J. Villanueva-Millán, L. Valverde, A. Odriozola, J.M. Aznar, S. Piñeiro-Hermida et al., Mitochondrial DNA control region variation in an autochthonous Basque population sample from the Basque Country, Forensic Sci. Int. Genet. 6 (2012) e106 – e108.

[15] M. Baeta, M.J. Illescas, L. García, C. Núñez, E. Prieto-Fernández, S. Jiménez-Moreno, et al., Iberian allele frequency database for 10 X-STRs, Forensic Sci. Int. Genet. 19 (2015) 76–78.

[16] M.T. Zarrabeitia, F. Pinheiro, M.M. de Pancorbo, L. Cainé, S. Cardoso, L. Gusmão, et al., Analysis of 10 X-linked tetranucleotide markers in mixed and isolated populations, Forensic Sci. Int. Genet. 3 (2009) 63–66.

[17] M.J. Illescas, A. Pérez, J.M. Aznar, L. Valverde, S. Cardoso, J. Algorta, et al., Population genetic data for 10 X-STR loci in autochthonous Basques from Navarre (Spain), Forensic Sci. Int. Genet. 6 (2012) e146 – e148.

[18] Ø. Hammer, D.A.T. Harper, P.D. Ryan, Past: paleontological statistics software package for education and data analysis, Palaeontol. Electron. 4 (2001) 9 – 17.

[19] D. Adler, D. Murdoch, O. Nenadic, S. Urbanek, M. Chen, A. Gebhardt, et al., rgl: 3D Visualization Using OpenGL, R package version 0.96.0 (2016). http://cran.r-project.org/package=rgl (accessed 10.09.16).

[20] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2013). http://R-project.org (accessed 29.08.16).

[21] R. Szibor, S. Hering, E. Kuhlisch, I. Plate, S. Demberger, M. Krawczak, et al., Haplotyping of STR cluster DXS6801-DXS6809-DXS6789 on Xq21 provides a powerful tool for kinship testing, Int. J. Legal Med. 119 (2005) 363 – 369.

[22] M. Karakachoff, N. Duforet-Frebourg, F. Simonet, S. Le Scouarnec, N. Pellen, S. Lecointe et al., Fine-scale human genetic structure in Western France, Eur. J. Hum. Genet. 23 (2015), 831 – 836.

[23] V. Dubut, L. Chollet, P. Murail, F. Cartault, E. Béraud-Colomb, M. Serre, et al., mtDNA polymorphisms in five French groups: importance of regional sampling, Eur. J. Hum. Genet. 12 (2004) 293 – 300.

[24] E. Bosch, F. Calafell, A. Pérez-Lezaun, J. Clarimón, D. Comas, E. Mateu, et al., Genetic structure of north-west Africa revealed by STR analysis, Eur. J. Hum. Genet. 8 (2000) 360 – 366.

Study number 5

‘Evaluation of the forensic efficiency of the tri- and tetrallelic SNPs located on the X-chromosome’

Finally, the fifth study corresponds to the attainment of the objective 4: *Studying the efficiency of the tri- and tetrallelic SNPs located on the X-chromosome with the aim of evaluating their utility as a complement for the developed X-STR panel.*

In the present work the forensic efficiency of the tri- and tetrallelic SNPs located along the X-chromosome (X-SNPs) has been evaluated. A search through the Ensembl genome browser was carried out and 375 tri- and tetrallelic X-SNPs were identified. However, only a handful of them (22) met the criteria herein established. Therefore, the main objective of this work has been to carry out the *in silico* evaluation of the forensic efficiency of each candidate marker for X-SNP multiplexing. Parameters of forensic interest (PD_F, PD_M, MEC_T, and MEC_D) were estimated for each candidate X-SNP in the following major groups: African, American, East Asian, European, and South Asian populations.

The combined parameters of forensic interest showed that the tri- and tetrallelic X-SNP sets herein selected for each major population are not more efficient than the current biallelic X-SNP multiplexes. This may be due to the fact that only a reduced number of markers were taken into account, i.e. <7. Nonetheless, the rs12395589, rs7886527, rs6640415, rs11095509, rs5018229, and rs5980187 markers were highly discriminative and therefore they could be potential candidates for being included in new X-SNP multiplexes or in MPS kits with forensic purposes.

This study has been submitted to the journal *Forensic Science International: Genetics* in March 2017 under the heading ‘*Evaluation of the forensic efficiency of the tri- and tetrallelic SNPs located on the X-chromosome’’.* Q1, IP: 4.988. The submitted version of the manuscript is shown below.

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Short communication

Evaluation of the forensic efficiency of the tri- and tetrallelic SNPs located on the X-chromosome

Endika Prieto-Fernández [a], Tamara Kleinbielen [a], Miriam Baeta [a], Marian M. de Pancorbo [a*]

[a] BIOMICs Research Group, Lascaray Research Center, University of the Basque Country UPV/EHU. Avda. Miguel de Unamuno, 3. 01006 Vitoria-Gasteiz (Spain).

* Corresponding author

Submitted to journal: 27 March 2017

Abstract

DNA samples in forensic casework are often highly degraded and consequently, the amplification of STRs may fail. In this context, SNPs may help to obtain reliable results from challenging samples. Due to its particular inheritance pattern, SNPs located on the X chromosome (X-SNPs) may complement the results obtained through the autosomal markers. Up to now, different biallelic X-SNP sets have been evaluated. However, the forensic efficiency of tri- and tetrallelic SNPs located on the X-chromosome have not been studied yet. Therefore, we have performed an *in silico* evaluation of these markers. A search of polymorphic X-SNPs that displayed a frequency for the second allele ≥0.01 was conducted through the Ensembl genome browser. A total of 375 loci were classified as tri- or tetrallelic X-SNPs, but only 22 showed allele frequencies ≥0.01 for the third or fourth allele in at least one population. Therefore, they were considered as candidate X-SNPs. Parameters of forensic interest were calculated in each major population, i.e. African, American, East Asian, European, and South Asian, only for those markers that displayed frequencies ≥0.01 for at least three alleles. The cPD values ranged from 75.3 to 99.9 in females and from 59.5 to 98.7 in males. Moreover, the $cMEC_T$ and $cMEC_D$ ranged from 49.3 to 96.6 and 33.2 to 87.3, respectively. These low values are due to the few markers considered in each case, i.e. <7. Interestingly, the rs12395589, rs7886527, rs6640415, rs11095509, rs5018229, and rs5980187 markers were highly discriminative and therefore

they may be potential candidates for being included in new X-SNP multiplexes or in MPS kits with forensic purposes.

Keywords: Single nucleotide polymorphisms (SNPs); X-chromosome SNPs (X-SNPs); multiple-allele SNPs; tri- and tetrallelic X-SNPs; *in silico* evaluation.

Introduction

Single Nucleotide Polymorphisms (SNPs) are single base-pair differences that occur at a specific position in the genome. This kind of markers present some advantages that other polymorphic variants lacked, such as low mutation rates and the ability to obtain results from low copy number [1-3]. Regarding their usefulness in Forensic Genetics, SNPs can be divided into: 1) identity testing SNPs; 2) lineage informative SNPs; 3) ancestry informative SNPs; and 4) phenotype informative SNPs [4].

DNA samples in forensic casework are often highly degraded and therefore, in certain cases, the amplification of Short Tandem Repeats (STRs) may fail or be incomplete. In this context, SNP markers may help to obtain reliable results from extremely degraded samples as only 60-80 bp fragments in length are necessary for PCR amplification [4]. Therefore, SNPs are of great interest when dealing with highly degraded remains such as those exhumed from mass graves and/or historical skeletal remnants. Additionally, and because of their low mutation rates, they are of great utility in paternity testing. Due to their biallelic condition, the power of discrimination (PD) that can be reached by analyzing SNPs is usually lower than that obtained with STRs. Nevertheless, this drawback may be solved by considering a great number of discriminative SNPs. In this context, the analysis of about 60 biallelic SNPs provides PD values similar to those obtained by applying the current X-STR multiplexes [3].

SNP typing technologies have undergone a rapid development in the last few years. As a result, a great variety of different protocols have become available, being the minisequencing based on SNaPshot^{TM} technology (Thermofisher Scientific, Waltham, MA, USA) one of the most utilized commercial methodologies in Forensic Genetics [3,5,6]. This method, which is performed on an automatic capillary electrophoresis instrument, enables multiplexing a high number of SNPs reducing the cost and time of the analysis. Other minisequencing approach that allow analyzing a high number of SNPs in a single PCR reaction is based on MALDI-TOF Mass Spectrometry (MS) [3]. However, this methodology is not commonly used in forensics since a mass spectrometer is required (Table 1). More recently, Massive Parallel Sequencing (MPS) has allowed the simultaneous analysis of a larger battery of markers, far exceeding the capacity of other current technologies. Therefore, more forensically relevant genetic information can be obtained from highly degraded, low copy number or mixed DNA samples [7]. Hence, there is little doubt that this methodology will be implemented in forensic laboratories in the near future [8]. The search for new markers to be incorporated to MPS kits is a task that may improve the efficiency of MPS technology in the forensic field.

Currently, commercial MPS assays that analyze different kind of markers in a single PCR reaction are available [8]. However, most of the markers included in MPS assays are autosomal or Y-chromosomal, and in contrast very few loci located on the X-chromosome are being currently taken into account. Nevertheless, the interest in studying genetic markers located on the X-chromosome has rapidly grown during the last decade [9-11]. The potential of this chromosome is mainly due to its particular inheritance pattern and the recombination restriction in males. Consequently, females receive the same non-recombinant X chromosome from their fathers that facilitates the analysis of diverse relationships [10]. SNPs located on the X chromosome may complement the results of both autosomal SNP and STR markers. In this sense, the combination of STRs and SNPs markers may help to solve certain complex kinship cases that autosomal markers cannot e.g. sisters in motherless paternity cases [1]. In other words, they may play an important role in forensic casework for increasing the probability of parentage and PD in complex kinship cases, family reconstructions and/or human identification [4].

To assure the inclusion of the most efficient SNP markers in the current panels, extensive research has been performed by using the public whole-genome sequence data [12,13]. Up to now, some biallelic X-SNPs have been studied and evaluated in several populations [1,14-17] (Table 1). However, the forensic efficiency of the X-SNPs that display more than two alleles has not been reported yet. In view of this, the aim of the present study is to carry out the *in silico* evaluation of the efficiency of tri- and tetrallelic X-SNPs for their application in forensic casework.

Materials and methods

The search for polymorphic bi-, tri- and tetrallelic X-SNPs was performed through Ensembl genome browser (http://www.ensembl.org/biomart) based on the human genome assembly GRCh38.p5 of the Genome Reference Consortium (GRC). Two criteria were laid down in order to ensure that the selected X-SNPs displayed heterozygosity in all the described alleles. In the first stage we chose those X-SNP markers that displayed a 1000 Genomes Global MAF (Minor Allele Frequency) ≥ 0.01. Currently, the Single Nucleotide Polymorphism Database (dbSNP) (https://www.ncbi.nlm.nih.gov/SNP) reports the MAF as the frequency value of the second most common allele, in order to distinguish common polymorphism from rare variants. Furthermore, to select the final candidate X-SNPs a second criterion based on allele frequencies ≥ 0.01 for the third or fourth allele in at least one population was established. This second criterion was applied to ensure that the selected polymorphic X-SNPs displayed three or four variants. Only those markers that met the above-mentioned criteria were considered as candidate X-SNPs in this study.

Allele frequencies of each selected marker were compiled through the Ensembl genome browser for both the overall population and the following five major populations: African, American, East Asian, European, and South Asian. Only those tri- and tetrallelic markers that met the second criterion in each major population were considered. Parameters of forensic interest, i.e. the power of discrimination in females (PD_F) and males (PD_M), as well as the mean exclusion chance in trios (MEC_T) and duos (MEC_D) were calculated by using the Forensic ChrX Research database (http://www.chrx-str.org).

Table 1. Combined values of forensic parameters for different sets of X-SNPs extracted from the literature. Abbreviations: REF= bibliography reference; MET= typing methodologies; POP= population identifier; #= number of X-SNPs considered for each population; cPD_F = combined power of discrimination in females, cPD_M = combined power of discrimination in males, $cMEC_T$ = combined paternity exclusion index in trios; and $cMEC_D$ = combined paternity exclusion index in duos. Typing methodologies: MALDI-TOF MS= PCR amplification followed by MALDI-TOF mass spectrometry; SNaPShot^{TM}= minisequencing based on SNaPShot^{TM} technology; and Taqman®= Taqman® SNP genotyping assay. Populations: HAN= Chinese Han population; NEA= Four native North Eurasian populations (Buryat, Kazah, Khakas, and Khanty); 13P= 13 populations from Valencia, Majorca, Ibiza, Catanzaro, Cosenza, Reggio Calabria, Sicily, Tunisia, Morocco, Turkey, Iraq, Denmark, and Somalia; CAN= coastal area of Cantabria (Spain); and PV= Pas Valley (Spain).

REF	*MET*	*POP*	*#*	*cPD_F*	*cPD_M*	*$cMEC_T$*	*$cMEC_D$*
[14]	MALDI-TOF MS	HAN	*67*	±99.9999999999999	±99.9999999999999	±99.9999	±99.9999
[15]	MALDI-TOF MS	NEA	62	>99.9999999999999999999996	>0.99999999999998	-	-
[1]	SNaPShot^{TM}	13P	25	99.9999992	99.9998	99.9980	99.7800
[16]	Taqman®	HAN	14	99.9998	99.9899	99.8300	97.8800
[17]	Taqman®	CAN	10	99.9924	99.8519	99.4405	93.5023
[17]	Taqman®	PV	10	99.9933	99.8514	98.9952	93.8330

Moreover, a total of five sets of markers were built and an *in silico* evaluation was performed to assess their forensic efficiency in each major population. For that end the combined parameters of forensic interest ($cMEC_D$, $cMEC_T$, cPD_M, and cPD_F) were calculated.

Results and discussion

A total of 38,559 bi-, tri-, and tetrallelic SNPs with a 1000 Genomes Global MAF ≥ 0.01 were found along the X-chromosome. Of them, 375 were categorized as tri- or tetrallelic X-SNPs but only 22 showed allele frequencies ≥ 0.01 for the third or fourth allele in at least one population. Therefore, they were considered as polymorphic candidate X-SNPs. The selected X-SNPs and their allele frequencies in the overall population are shown in Supplementary Table S1. Additionally, genome variation data for the 22 candidate X-SNPs was compiled for all the major populations, i.e. African, American, European, East Asian, and South Asian, since allele distribution is population-specific (Supplementary Table S2). As expected, not all of the selected X-SNPs displayed three or four alleles with frequencies > 0.01 in all the studied populations. Therefore, statistical parameters of forensic interest (PD_F, PD_M, MEC_T, and MEC_D) were calculated for only those tri- and tetrallelic markers that met the second criterion in each major population (allele frequencies ≥ 0.01 for the third or fourth allele in at least one population).

Interestingly, the rs12395589, rs7886527, rs6640415, rs11095509, rs5018229, and rs5980187 markers were highly discriminative (Figure 1, Supplementary Table S2). Therefore, they could be potential candidates for being included in new X-SNP multiplexes or in MPS kits with forensic purposes.

Table 2. Combined values of forensic parameters obtained for the candidate tri- and tetrallelic X-SNPs in each major population. Populations: AFR= African; AMR= American; EAS= East Asian; EUR= European; and SAS= South Asian. # indicates the number of X-SNPs considered in each case.

POP	#	cPD_F	cPD_M	$cMEC_T$	$cMEC_D$
AFR	7	99.9	98.7	96.6	87.3
AMR	5	95.8	82.6	78.4	58.2
EAS	2	75.3	59.5	49.3	33.2
EUR	4	97.3	88.5	83.3	66.1
SAS	5	95.4	82.9	77.6	57.5

Additionally, the combined parameters of forensic interest were calculated taking into account the set of X-SNPs that met the second criterion herein established in each major population. The cPD values ranged from 75.3 to 99.9 in females (cPD$_F$) and from 59.5 to 98.7 in males (cPD$_M$). Moreover, the cMEC$_T$ and cMEC$_D$ ranged from 49.3 to 96.6 and 33.2 to 87.3, respectively (Table 2). Our results showed that the evaluated sets of tri- and tetrallelic X-SNPs are not more efficient than the current biallelic X-SNP multiplexes (Table 1), mainly due to the low number of markers considered in each case, i.e. <7.

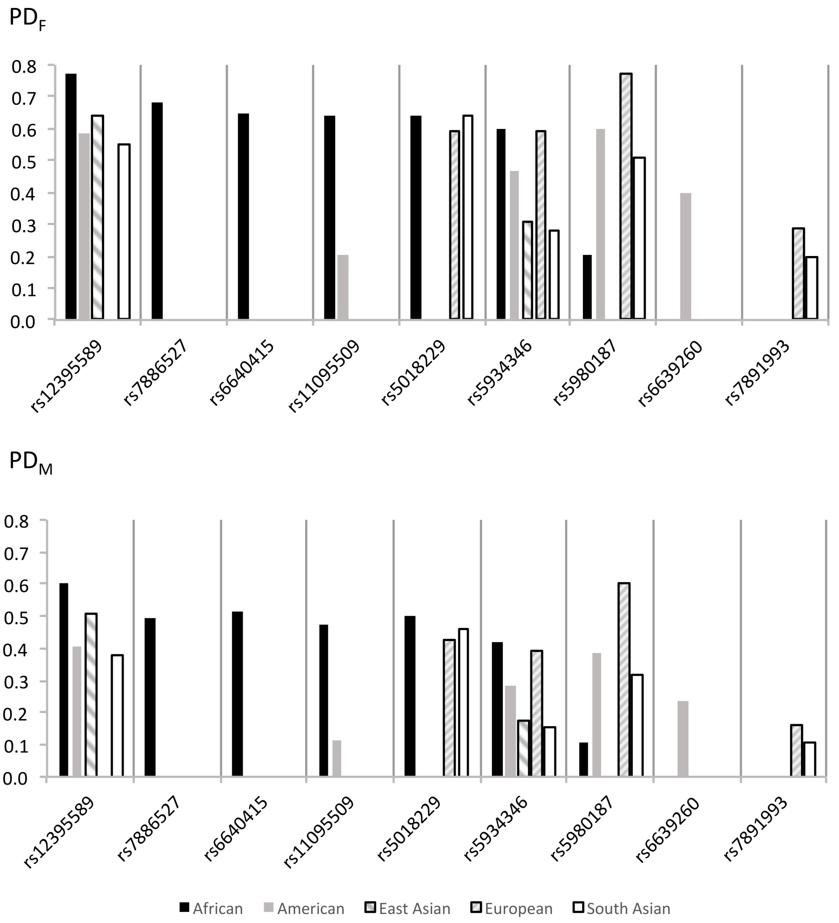


Figure 1. Representation of the values of power of discrimination in females (PD$_F$) and males (PD$_M$) for the polymorphic candidate tri- and tetrallelic X-SNPs in the African, American, East Asian, European, and South Asian populations.

Conclusion

The search through Ensembl genome browser showed only 38,559 X-SNP markers that displayed allele frequencies ≥0.01 for the second most common allele. Of these, only 22 X-SNPs met the criteria herein established for at least a third allele. Therefore, they were considered of potential interest in Forensic Genetics. The combined parameters of forensic interest showed that the sets of tri- and tetrallelic X-SNPs are not more efficient than the current biallelic X-SNP multiplexes. These low values are due to the few markers considered in each case, i.e. <7. Interestingly, the rs12395589, rs7886527, rs6640415, rs11095509, rs5018229, and rs5980187 markers did display high discrimination values and therefore, they should be taken into account for developing new X-SNP multiplexes or to be included in MPS kits with forensic purposes.

Conflict of interest

The authors have declared no conflict of interest.

References

[1] C. Tomas, J.J. Sanchez, J.A. Castro, C. Børsting, N. Morling, Forensic usefulness of a 25 X-chromosome single-nucleotide polymorphism marker set, Transfusion 50 (2010) 2258–2265.

[2] C. Børsting, N. Morling, Reinvestigations of six unusual paternity cases by typing of autosomal single-nucleotide polymorphisms, Transfusion 52 (2012) 425–430.

[3] B. Sobrino, M. Brión, A. Carracedo, SNPs in forensic genetics: A review on SNP typing methodologies, Forensic Sci. Int. 154 (2005) 181–194.

[4] B. Budowle, A. Van Daal, Forensically relevant SNP classes, Biotechniques 44 (2008) 603–610.

[5] J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, et al., A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724.

[6] C. Phillips, A. Salas, J.J. Sánchez, M. Fondevila, A. Gómez-Tato, J. Álvarez-Dios, et al., Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs, Forensic Sci. Int. Genet. 1 (2007) 273–280.

[7] A. Ambers, J. Churchill, J. King, M. Stoljarova, H. Gill-King, M. Assidi, et al., More comprehensive forensic genetic marker analyses for accurate human remains identification using massively parallel DNA sequencing, BMC Genomics 17 (2016) 21.

[8] C. Børsting, N. Morling, Next generation sequencing and its applications in forensic genetics, Forensic Sci. Int. Genet. 18 (2015) 78–89.

[9] H. Ellegren, Microsatellites: simple sequences with complex evolution, Nat. Rev. 5 (2004) 435–445.

[10] T.M. Diegoli, Forensic typing of short tandem repeat markers on the X and Y chromosomes, Forensic Sci. Int. Genet. 18 (2015) 140–151.

[11] R. Szibor, M. Krawczak, S. Hering, J. Edelmann, E. Kuhlisch, D. Krause, Use of X-linked markers for forensic purposes, Int. J. Legal Med. 117 (2003) 67–74.

[12] X. Zeng, R. Chakraborty, J. King, B. LaRue, R. Moura-Neto, B. Budowle, Selection of highly informative SNP markers for population affiliation of major US populations, Int. J. Leg. Med. 130 (2016) 341–352.

[13] C. Phillips, J. Amigo, A. Carracedo, M.V. Lareu, Tetra-allelic SNPs: Informative forensic markers compiled from public whole-genome sequence data, Forensic Sci. Int. Genet. 19 (2015) 100–106.

[14] L. Li, Y. Liu, Y. Lin, Typing of 67 SNP Loci on X Chromosome by PCR and MALDI-TOF MS, Res. Genet. (2015) Article 374688.

[15] V. Stepanov, K. Vagaitseva, V. Kharkov, A. Cherednichenko, A. Bocharova, G. Berezina, et al., Forensic and population genetic characteristics of 62 X chromosome SNPs revealed by multiplex PCR and MALDI-TOF mass spectrometry genotyping in 4 North Eurasian populations, Leg. Med. 18 (2016) 66–71.

[16] L. Li, C. Li, S. Zhang, S. Zhao, Y. Liu, Y. Lin, Analysis of 14 highly informative SNP markers on X chromosome by TaqMan® SNP genotyping assay, Forensic Sci. Int. Genet. 4 (2010) 2–5.

[17] M.T. Zarrabeitia, V. Mijares, J.A. Riancho, Forensic efficiency of microsatellites and single nucleotide polymorphisms on the X chromosome, Int. J. Legal Med. 121 (2007) 433–437.

5. Discussion

Evaluation of the New 17 X-STR Panel

Efficiency

In the present doctoral thesis work a new 17 X-STR panel has been developed [176] (Figure 11) and its forensic efficiency has been evaluated and compared with that of the most usual X-STR multiplex systems in kinship testing and human identification.

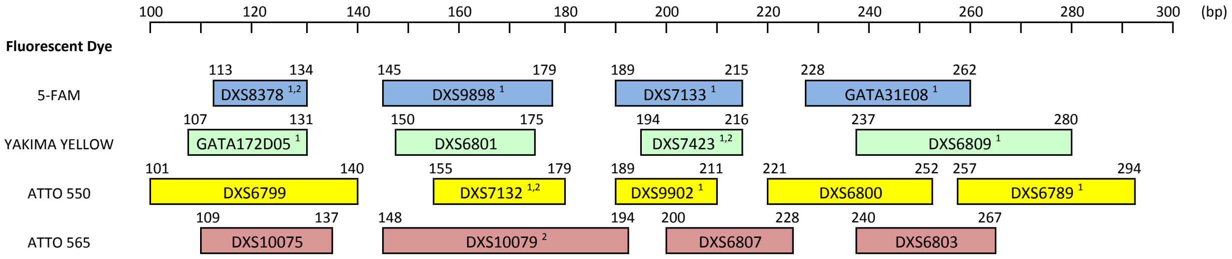


Figure 11. Definitive panel distribution of the developed X-STR panel. The boxes represent the expected fragment sizes for each locus. The scale at the top represents the size in bp. The fluorescent dye for the corresponding loci is shown on the left. Markers included in the decaplex of the GHEP-ISFG are indicated with ^{1} while the ones included in Argus X-12 Kit with ^{2}.

The new panel proved to be more efficient than the commonly used multiplexes in forensic casework in terms of X-chromosome analysis. The values obtained for the combined parameters of forensic interest, i.e. cPD$_{F}$, cPD$_{M}$, cMEC$_{T}$, cMEC$_{D}$, improved considerably when compared with those obtained by using the decaplex of the GHEP-ISFG [113,114]. For example, in the Spanish population, the combined MEC values increased from 0.9998 to 0.9999994 in duos (cMEC$_{D}$) and from 0.999995 to 0.999999998 in trios (cMEC$_{T}$). In the same context, the combined PD values raised from 0.9999990 to 0.9999999998 in males (cPD$_{M}$) and from 0.9999999998 to 0.99999999999999994 in females (cPD$_{F}$) [177]. This improvement was even more noticeable in the most heterogeneous populations studied herein, i.e. Malawi and Equatorial Guinea, as well as Casablanca populations [176,178]. Additionally, the efficiency of the 17 X-STRs has also been compared with that of the Investigator® Argus X-12 Kit (Qiagen, Valencia, CA, USA) extracted from the literature [179]. An increase of two orders of magnitude in cPD$_{F}$, cPD$_{M}$, cMEC$_{T}$, and cMEC$_{D}$ values was identified (Table 9). Additionally, the 17 X-STR panel has demonstrated to be more efficient in detecting genetic differentiation between certain populations than the decaplex of the GHEP-ISFG

[178]. This efficiency increase in terms of both genetic differentiation between populations and parameters of forensic interest may be attributed to the fact that a higher number of discriminative markers are considered in the analysis.

Table 9. Maximum values of combined parameters of forensic interest obtained with the following X-STR panels: the decaplex of the GHEP-ISFG [113,114], the Investigator® Argus X-12 Kit (Qiagen, Valencia, CA, USA), and the 17 X-STR panel [176]. Abbreviations: POP= population, cPD_F= combined power of discrimination in females, cPD_M= combined power of discrimination in males, $cMEC_T$= combined mean exclusion chance in trios, and $cMEC_D$= combined mean exclusion chance in duos.

Study	*Multiplex*	*POP*	*cPD_F*	*cPD_M*	*$cMEC_T$*	*$cMEC_D$*
[180]	decaplex-GHEP	Spain	0.9999999998	0.999999	0.999995	0.9998
[179]	Argus X-12 Kit	Spain	0.999999999999998 [A]	0.999999999 [B]	>0.99999[A, B]	>0.99999[A, B]
[177]	17 X-STR panel	Spain	0.99999999999999994	0.9999999998	0.999999998	0.9999994
[176]	decaplex-GHEP	African [C]	0.99999999997	0.9999995	0.999998	0.9999
[178]	17 X-STR panel	African [D]	0.999999999999999991	0.99999999991	0.9999999994	0.9999998
[176]	17 X-STR panel	African [C]	0.999999999999999999	0.99999999997	0.9999999998	0.9999999

[A] Indicates the value obtained in the population of Mallorca while [B] the one obtained in the population of Valencia.

[C] The values correspond to the African populations of Malawi and Equatorial Guinea.

[D] The value was obtained in the Moroccan population of Casablanca.

Applicability

In forensic casework, DNA is not always in optimal conditions and therefore, the PCR amplification may fail giving rise to incomplete or null genetic profiles. The main factors affecting the amplification are: 1) DNA degradation, 2) low copy number, and 3) presence of PCR inhibitors [72–74]. With that in mind, the new panel was carefully designed to obtain short amplicons. Of the 17 X-STRs, six markers were designed following a miniSTR approach (<200 bp) and the remaining were in midiSTR format (<300 bp). This favors the amplification of samples with highly fragmented DNA and/or with low copy number. Having conducted the corresponding sensitivity studies, the ability of this panel to obtain reliable results from low amounts of DNA (≈100 pg) has been proven. The robustness of the 17 X-STR panel has been tested by adding inhibitors to the PCR reaction. Our results showed complete amplification even in presence of high concentrations of humic acid (≤ 250 ng/µl) and hematin (≤ 300 µM). In view of this, the newly developed 17 X-STR panel is a suitable alternative to the current X-STR multiplexes in Forensic Genetics as its efficiency and applicability in both well-preserved and challenging samples has been proven [176].

Application of the new multiplex to real cases

Markers located on the X-chromosome present some advantages that others lacked, mainly due to their particular inheritance pattern [35,99]. Therefore, they may increase the resolution power and complement the results of autosomal markers in certain cases. Hence, the new 17 X-STR panel, which has demonstrate to be a highly discriminative and efficient multiplex [176–178], has recently been incorporated to the Diagnostic Service of the Biological Parentage and Genetic Identification of the University of the Basque Country UPV/EHU. In addition, it has also been applied in genetic identification of skeletal remains exhumed from mass graves of the Spanish Civil War and posterior dictatorship by our group.

Linkage and LD along the X-chromosome

As mentioned before, linkage can be defined as the co-segregation of closely located loci within a family or pedigree [99]. In other words, two markers that are closely located on the same chromosome will be transmitted to the offspring together unless there are hotspots of recombination between them. These hotspots are regions where the frequency and rate of recombination (θ) are most favorable. However, the existence of hotspots of recombination does not assure the genetic exchange. A measure of the linkage between two markers located on the same chromosome is given by θ. The mathematical relationship between θ and genetic distances is described by mapping functions being the Kosambi's function the most extended approach. The cM genetic map distance provides the expected number of crossover events per sex-averaged generation between two markers [181]. It is assumed that two loci that show a θ of 0.01 (1%) are separated by a genetic distance of 1 cM.

On the other hand, LD is simply a non-random association between two or more alleles. It means that these alleles appear together at rates that differ from what would be expected under independence [45,100]. In this context, two linked markers tend to show significant LD [134]. Moreover, there exists a strong and reverse correlation between LD and recombination rate in the human genetic map [182]. However, LD does not ensure linkage since the alleles of two markers that are located even in different chromosomes may be in linkage disequilibrium. LD throughout the genome reflects the population history, the breeding system and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution [100].

Due to their higher mutation rate, STRs tend to show less LD than SNPs [128]. In this way, closely linked polymorphic SNPs, which mutate at such low rates, tend to be in strong LD with one another and therefore, may be used to determine LD blocks within a certain region on the chromosome [100]. Most of the multiplexes for analyzing autosomal markers include loci that are located in different chromosomes and accordingly, they are transmitted separately. However, X-STRs are located within a single chromosome and then, two or more loci may be linked depending on both their physical localization and genetic distance between them.

For practical reasons, the X-chromosome was divided into four linkage groups located on the following regions Xq22.2, Xq12, Xq26, and Xq28. This division has been based on their physical location along the chromosome. Nonetheless, a precise knowledge of both, linkage and recombination rates, is required in order to perform accurate forensic calculations. With the aim of determining the linkage condition of the markers included in the 17 X-STR panel, SNP genotype data of the regions within two consecutive loci were obtained from HapMap and the corresponding Haploview LD plots were performed. The 17 X-STRs included have been gathered into different linkage groups throughout the literature, such as the cluster DXS6807-DXS8378-DXS9902 in the region Xp22 [183]. In the region Xq21, the markers DXS6801, DXS6809, and DXS6789 have been considered linked by the majority of the authors [91,137,138] and additionally, DXS6799 has also been included in this cluster [183]. Finally, the markers DXS7132, DXS10079, and DXS10074 have typically been considered as a haplotype in the Xq12 region [137,138,181] and some authors have also considered the locus DXS10075 [141,183] as part of it. Having performed the corresponding Haploview LD plots, several hotspots of recombination were detected between the above-mentioned clusters, i.e. DXS6801-DXS6809-DXS6789-DXS6799 and DXS6800-DXS6803-DXS9898 (Figures 12 and 13, respectively). On the other hand, the analysis of the region between the markers DXS10079-DXS10075 did not displayed hotspots of recombination. Thus, they could be considered genetically linked. The Haploview LD plot that covered the region between DXS7132 and DXS10079 was incomplete due to insufficient information about SNP genotype data available for the CEU collection. Taking into account these results, we rather be conservative and maintain the three loci DXS7132-DXS10075-DXS10079 as a linkage group. Pairwise LD analyses have also been performed with the population data derived herein with the 17 X-STR panel. LD has been detected for the markers DXS10075 and DXS10079 (studies number 2 and 3). In addition, significant p-values after Bonferroni correction were observed between DXS7133-DXS9902, DXS9898-DXS6803/DXS6807, DXS6801-DXS6789/DXS6803, and DXS6789-DXS6807. However, microsatellite markers are less able to detect LD compared with SNPs, mainly due to their high mutation rate.

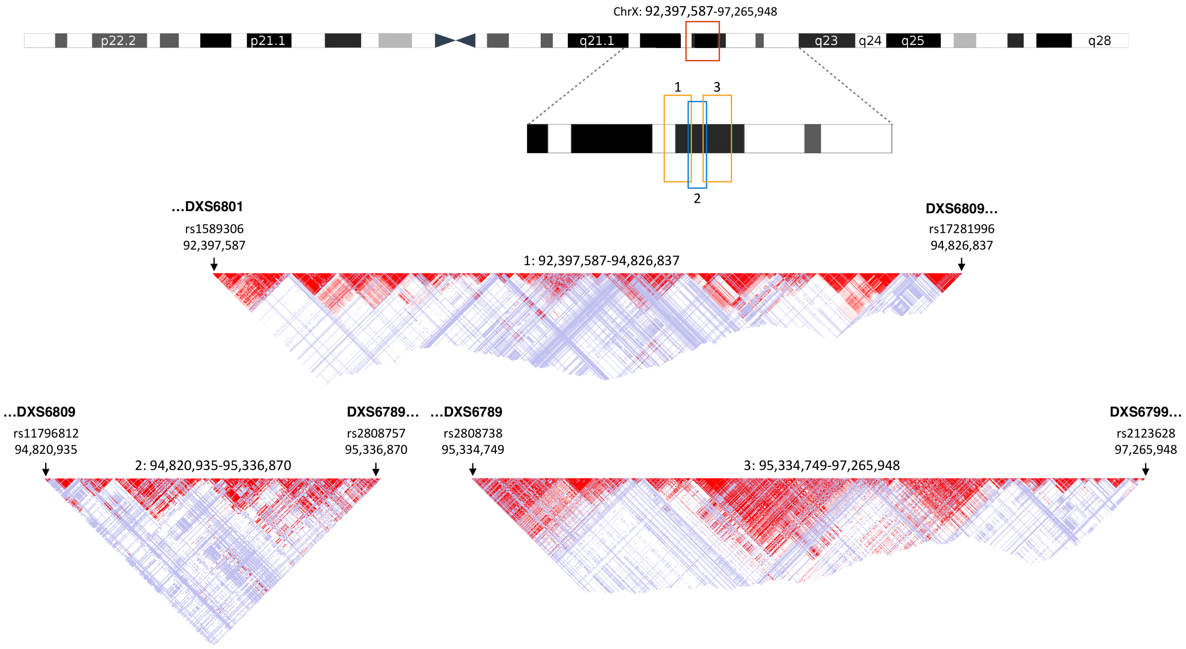


Figure 12. HapMap LD plots of DXS6801-DXS6809 (1), DXS6809-DXS6789 (2), and DXS6789-DXS6799 (3). Data source: HapMap Data Rel 28 PhaseII+III, August10, on NCBI B36 assembly dbSNP b126 for the Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) from the HapMap website (http://www.hapmap.org).

The above-mentioned linkage and LD issues are based on mathematical estimations based on gene mapping and SNP genotype data derived for a single population. Consequently, the only way to both determine realistic recombination rates and describe the linkage state of each locus along the X-chromosome is to carry out family studies composed by grandfather-daughter-grandson. For that end, the collaboration between research groups might facilitate the arduous labor of sampling the high number of trios required to obtain the necessary informative meiosis in each case.

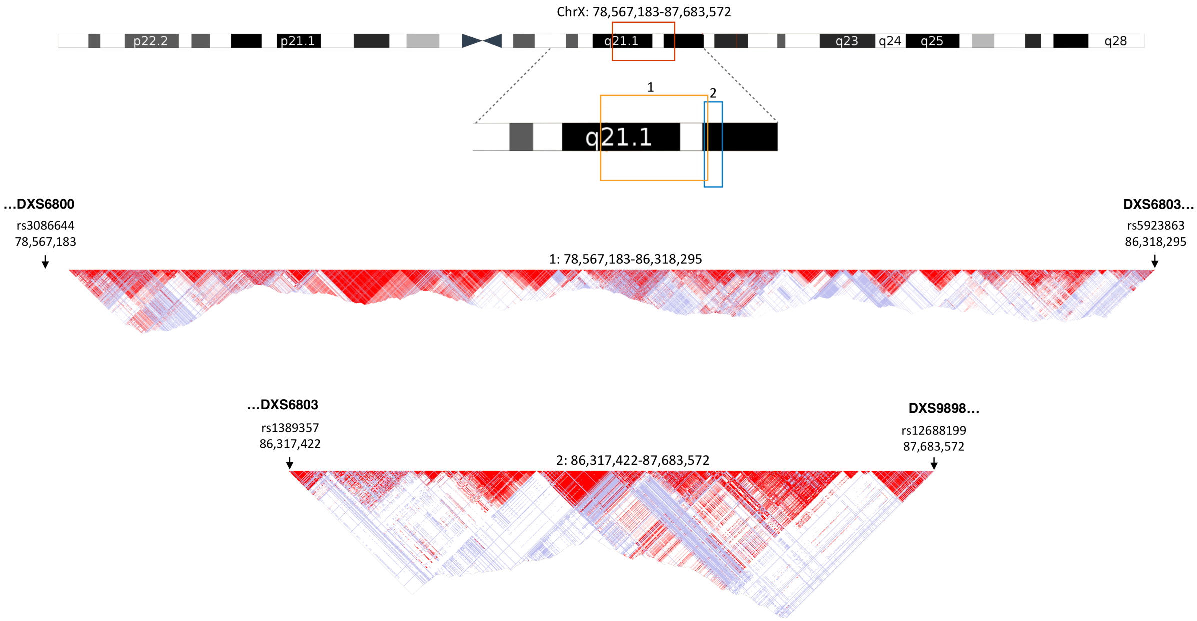


Figure 13. HapMap LD plots of DXS6800-DXS6803 (1) and DXS6803-DXS9898 (2). Data source: HapMap Data Rel 28 PhaseII+III, August10, on NCBI B36 assembly dbSNP b126 for the Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) from the HapMap website (http://www.hapmap.org).

Forensic Allele and Haplotype Frequency Databases

In the present doctoral thesis, three allele and haplotype frequency databases have been performed with both the decaplex of the GHEP-ISFG and the 17 X-STR panel developed herein.

Study number 1: 'Iberian allele frequency database for 10 X-STRs'

Until the attainment of the objective number 1 of the present work, no X-STR allele frequency databases were available for the Iberian Peninsula population. Therefore, the appliance of the decaplex of the GHEP-ISFG to real forensic cases based on realistic allele frequencies was not possible in the Spanish population.

With that in mind, 1,136 unrelated individuals of several populations, which were representative of the most populated regions of the Iberian Peninsula, i.e. Alicante, Andalusia, Asturias, Barcelona, Basque Country, Galicia, and Madrid, were studied with the 10 X-STRs. Having assured that all the populations were in HWE, no genetic differentiation was observed. The genetic distances were calculated between the seven populations studied herein and other previously studied from the Iberian Peninsula, i.e.

North Portugal, Central Portugal, Galicia, Murcia, Barcelona, Cantabria, and autochthonous from the Pas Valley (Cantabria), the Basque Country, and Navarre. In comparison with other Iberian populations, differences were observed with respect to groups such as the autochthonous groups from the Basque Country, Navarre or the Pas Valley.

This microdifferentation has been previously reported for X-STRs [184,185] and other markers, such as Y-STRs, AS-STRs, and mitochondrial DNA [186–190]. This condition may be attributed to their genetic singularity mainly derived from a geographical, linguistical, and/or cultural isolation. However, the seven Iberian populations analyzed herein did not present heterogeneity among them. Therefore, they were pooled together to form the first overall Iberian population database for the decaplex of the GHEP-ISFG, which may be safely applied for matching or kinship probabilities.

The 10 X-STRs included in the panel of the GHEP-ISFG proved to be sufficiently discriminative for its application in real cases as combined PD values ranging from 0.9999999997–0.99999999985 for females (cPD_F) and 0.999998–0.9999990 for males (cPD_M) were obtained. Furthermore, it can be also particularly useful in kinship testing, since combined mean exclusion chance values of 0.999993–0.999995 for trios ($cMEC_T$) and 0.9997–0.9998 for duos ($cMEC_D$) were observed.

Study number 3: 'Forensic Spanish allele and haplotype database for a 17 X-STR panel'

In Spain, a great number of skeletal remnants from executed and murdered individuals during the Spanish Civil War and posterior dictatorship remains yet unidentified. Their identification requires a comparison with the putative relatives, which in a high number of cases correspond to second or third degree family members [75]. In this context, the study of the X-chromosome, and especially the new 17 X-STR panel, may be of great utility in skeletal identification, mainly due to the particular inheritance pattern of the X-chromosome [176]. On the other hand, the number of kinship testing in our country is continuously growing and this panel may help to distinguish between two close relatives when other markers cannot. However, its applicability in the Spanish population was not possible until the establishment of a forensic database covering all the 17 X-STRs included.

Following the same procedure as in the study number 1, the Spanish allele and haplotype frequency database was updated and accommodated to the new 17 X-STR panel [177]. For that, 593 unrelated individuals from seven Spanish populations, i.e. Alicante, the Basque Country, Andalusia, Galicia, Madrid, and Barcelona, were analyzed with the 17 X-STR panel. Having compared all the populations, no genetic substructures were found. Therefore, all the populations could be placed together to form the first Spanish allele and haplotype frequency database for the new 17 X-STR panel.

Additionally, genetic distances based on F_{ST} between the global Spanish population and other three populations previously studied with the 17 X-STRs, i.e. Asians from Thailand, Hispanics from Colombia, and Africans from Malawi and Equatorial Guinea, were calculated. The overall Spanish population was clearly differentiated from the rest of the above-mentioned populations. Moreover, the closest population groups were the Spanish and the Hispanic population from Colombia, probably due to the historical events that both populations share. In terms of Forensic Genetics, the higher the number of markers considered, the higher the ability for a more accurate kinship determination and human identification. In this sense, the 17 X-STR panel increased the number of markers analyzed in a single PCR reaction if it is compared with the current X-STR multiplexes. The combined forensic parameters were calculated and compared with those obtained from the Iberian database previously performed with the 10 X-STRs of the GHEP-ISFG [180], showing a considerable improvement in the combined parameters of forensic interest.

Study number 4: 'A genetic overview of Atlantic coastal populations from Europe and North-West Africa based on a 17 X-STR panel'

In the last years, several population data have been provided by researchers regarding X-STRs. However, the forensic use of markers requires the performance of allele and haplotype frequency databases in the populations where they are going to be used. In this sense, when a new multiplex is developed, population data for each marker included is required to take advantage of the maximum potential of the multiplex. If not, resolution power of the considered set of markers may be underestimated in statistical calculations.

Taking this into consideration, the implement of the newly developed 17 X-STR panel requires establishing new forensic databases. When dealing with STRs that may be linked, especially with those of the X-chromosome, haplotype frequencies are needed.

The establishment of suitable haplotype frequency databases requires the analysis of thousands of individuals from a certain population. However, the available number of individuals is not always large enough. Consequently, a great number of haplotypes are not represented in the performed databases. Nevertheless, by merging two or more population samples that present similar allele and haplotype frequency distributions this problem may be solved.

Populations located on the Atlantic coast have experienced several genetic exchanges throughout history. Genetic exchanges may homogenize the allele and haplotype frequency distributions between different populations, which would allow creating larger forensic databases for its use in forensic casework. The aim of this study was to broaden the applicability of the 17 X-STR panel to the region of the European and North-African Atlantic coastline. With this purpose, 513 individuals from four different populations, i.e. Brittany, Ireland, northern Portugal, and Casablanca, were studied. Additionally, these populations were compared with other populations previously analyzed with the 17 X-STR panel, i.e. Galician, autochthonous Basque, and resident Basque populations, in search of similarities between populations, which will allow to create a global database for the Atlantic coastline of Europe and North-West Africa [178].

Population substructures have been observed within the overall Basques population with the 17 X-STR panel. Therefore, the autochthonous and resident Basques populations were considered independently. This clear differentiation between Basques has been reported by several authors for other kind of markers. On the contrary, no statistically significant differentiation was observed between the populations of the Atlantic coast of Europe studied herein, except for Brittany and the autochthonous Basque populations, which could be due to the particular genetic pool that both populations display [186–192]. Interestingly, a close genetic distance between Brittany and Ireland population has been detected that would be attributed to the well-established historical migrations between these two close populations during the sixth and seventh centuries [192]. On the other hand, the African population of Casablanca displayed differentiation with the rest of the European populations, except for autochthonous Basques and northern Portugal. The second case is in accordance with previous studies of autosomal STRs, which have suggested that among the Iberians, the Portuguese are one of the genetically closest to North-West African populations [193]. However, it would be interesting to analyze further populations located on the southern area of Portugal and the Iberian Peninsula, e.g. Andalusia; to evaluate if the geographical proximity to the African coast also means a lack of significant genetic differences or if, on the contrary, cultural barriers act as obstacles to genetic exchanges even between nearby populations communicated by seaways. On the other hand, the lack of differentiation with the resident Basques may be due to sample size bias [178].

Finally, our results suggest that certain neighboring populations located on the European Atlantic coast could have experienced episodes of genetic exchange as they have shown genetic similarities between them. On the contrary, other populations have not. In view of these results, genetic exchanges along the Atlantic coast seem to be insufficient to have totally homogenized the distribution of the X-chromosome markers and therefore, the use of independent allele and haplotype frequencies for each population would be more appropriate. The establishment of suitable haplotype frequency databases requires the analysis of a higher number of individuals for each population.

Forensic Efficiency of the Tri- and Tetrallelic X-SNPs

Up to now, the forensic efficiency of several biallelic SNPs located on the X-chromosome (X-SNPs) has been evaluated [145, 149,194–196]. However, the *in silico* evaluation of the tri- and tetrallelic X-SNP markers had not been evaluated until the achievement of the objective no. 4 of the present doctoral thesis work. A total of 38,559 bi-, tri-, and tetrallelic SNPs with a 1000 Genomes Global Minor Allele Frequency (MAF) ≥ 0.01, were found along the X-chromosome. Currently, the Single Nucleotide Polymorphism Database (dbSNP) reports the MAF as the frequency value of the second most common allele, in order to distinguish common polymorphisms from rare variants. Of them, 375 were classified as tri- and tetrallelic variants in the Ensembl genome browser. However, only 22 markers displayed allele frequencies ≥ 0.01 for the third or fourth allele in at least one population. Consequently, they were considered as polymorphic candidate tri- and tetrallelic X-SNPs in each major population, i.e. African, American, East Asian, European, and South Asian [197]. The rest of the markers that displayed allele frequencies < 0.01 for the third or fourth allele may be due to punctual mutations in the positions where the loci of interest are located.

The forensic efficiency of each candidate X-SNP was *in silico* evaluated by calculating the parameters of forensic interest, i.e. the power of discrimination in females (PD_F) and males (PD_M), as well as the mean exclusion chance in trios (MEC_T) and duos (MEC_D). The tri- and tetrallelic X-SNPs that did show the highest forensic efficiency were rs12395589, rs7886527, rs6640415, rs11095509, rs5018229, and rs5980187. Therefore, they should be taken into consideration for developing new X-SNP multiplexes or to be included in MPS kits with forensic purposes. In addition, the obtained values of cPD_F, cPD_M, $cMEC_T$, and $cMEC_D$ were calculated in each major population. Our results showed that the combined values of forensic interest were not more efficient than the current biallelic X-SNP multiplexes [145, 149,194–196]. This low values are due to the few markers considered in each case, i.e. <7 [197].

6. Conclusions

Conclusions

1. Seven Spanish populations have been analyzed with the decaplex of the GHEP-ISFG. Pairwise F_{ST} genetic distances between these populations were calculated and absence of genetic heterogeneity was observed dismissing possible genetic substructures. This supports the formation of the first global Iberian allele database with the markers included in the decaplex of the GHEP-ISFG.

2. A new multiplex system named '17 X-STR panel' that analyzes 17 microsatellite regions of the X-chromosome has been developed. Of the 17 X-STRs, six were designed following a mini X-STR approach and the remaining markers were in midi X-STR format. This allows obtaining reliable results from degraded samples and/or with low copy number. The newly developed panel has been validated following the SWGDAM Validation Guidelines for DNA Analysis Methods. Therefore, its applicability in routine forensic casework has been proved.

3. Linkage and linkage disequilibrium tests for the 17 X-STR loci were carried out by using the SNP data available in the HapMap Project and the X-STR population data derived from our studies. Multiple hotspots of recombination between the markers included in the '17 X-STR panel' were observed indicating independence among markers, except for the cluster DXS7132-DXS10075-DXS10079. This cluster was considered as a linkage group due to the very close genetic distance among markers and the lack of hotspots of recombination in this region.

4. The utility and applicability of the '17 X-STR panel' for the analysis of the X-chromosome variability has been proved. 593 unrelated individuals from seven different regional populations of Spain have been studied and forensic parameters, as well as allele and haplotype frequencies have been provided. The lack of significant differences between the analyzed populations supports the use of this database in the global Spanish population for statistical evaluation of the results in forensic casework.

5. Four different populations of the Atlantic Coast of Europe and North-West Africa were analyzed with the '17 X-STR panel' and the corresponding allele and haplotype databases have been performed. The genetic distances based on F_{ST} revealed significant differences between certain populations studied herein and others previously analyzed with the 17 X-STRs. Therefore, the genetic exchanges suffered throughout history along this seashore seem to be insufficient to have totally homogenized the distribution of the X-chromosome allele and haplotype frequencies. In view of these results, the use of individual databases for each population instead of a global database is recommended.

6. The search for X-SNPs through Ensembl genome browser, showed only a limited number of polymorphic candidate tri- and tetrallelic X-SNP markers, i.e. 22. The combined parameters of forensic interest showed that the tri- and tetrallelic X-SNP sets are not more efficient than the current biallelic X-SNP multiplexes, mainly due to the low number of markers that are polymorphic in each major population. Interestingly, the rs12395589, rs7886527, rs6640415, rs11095509, rs5018229, and rs5980187 markers did display high discrimination values and therefore, they should be taken into account when developing new X-SNP multiplexes or to be included in MPS kits with forensic purposes.

7. References

[1] J.D. Watson, F.H.C. Crick, Molecular structure of nucleic acids, *Nature* 171 (1953) 737–738.

[2] International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, *Nature* 409 (2001) 860–921.

[3] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, et al., The sequence of the human genome, *Science* 291 (2001) 1304–1351.

[4] International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome, *Nature* 431 (2004) 931–945.

[5] J.D. McPherson, M. Marra, L. Hillier, R.H. Waterston, A. Chinwalla, J. Wallis, et al., A physical map of the human genome, *Nature* 409 (2001) 934–941.

[6] The ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome, *Nature* 489 (2012) 57–74.

[7] The ENCODE Project Consortium, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, *Nature* 447 (2007) 799–816.

[8] J.M. Butler, Forensic DNA typing: Biology, Technology, and Genetics of STR markers (second ed.), *Elsevier Ltd. Academic Press, London* (2005).

[9] J.M. Butler, Advanced Topics in Forensic DNA Typing: Methodology, *Elsevier Ltd. Academic Press, San Diego* (2012).

[10] K. Landsteiner, Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe, *Zentralblatt. Bakteriol.* 27 (1900) 357–362.

[11] E. Von Dungern, L. Hirschfeld, Ueber Vererbung gruppenspezifischer Strukturen des Blutes, *Z. Immun. Forsch* 6 (1910) 284–292.

[12] E. Von Dungern, L. Hirschfeld, Ueber gruppenspezifische Strukturen des Blutes, *Z. Immun. Forsch* 8 (1911) 526–562.

[13] G. Geserick, I. Wirth, Genetic kinship investigation from blood groups to DNA markers, *Transfus. Med. Hemotherapy* 39 (2012) 163–175.

[14] A.J. Jeffreys, V. Wilson, S.L. Thein, Hypervariable “minisatellite” regions in human DNA, *Nature* 314 (1985) 67–73.

[15] Y. Nakamura, M. Leppert, P. O’Connell, R. Wolff, T. Holm, M. Culver, et al., Variable number of tandem repeat (VNTR) markers for human gene mapping, *Science* 235 (1987) 1616–1622.

[16] A.J. Jeffreys, S.D.J. Pena, A Brief Introduction to Human DNA Fingerprinting. In: DNA Fingerprinting: State of the Science (S.D.J. Pena, R. Chakraborty, J.T. Epplen, A.J. Jeffreys, eds.), *Birkhãuser Verlag, Basel* (1993) 1–20.

[17] T.S.L. Jeffreys, A. J, Wilson, V., Individual-specific “fingerprints” of human DNA, *Nature* 316 (1985) 76–79.

[18] A.J. Jeffreys, M. Turner, P. Debenham, The efficiency of multilocus DNA fingerprinting probes for individualization and establishment of family relationship, determined from extensive casework, *Am. J. Hum. Genet.* 48 (1991) 824–840.

[19] S. Ali, C.R. Muller, J.T. Epplen, DNA fingerprinting by oligonucleotide probes specific for simple repeats, *Hum. Genet.* 74 (1986) 239–243.

[20] G. Vassart, M. Georges, R. Monsieur, H. Brocas, A.S. Lequarre, D. Christophe, A sequence in MI3 phage detects hypervariable minisatellites in human and animal DNA, *Science* 235 (1987) 683–684.

[21] L. Roewer, DNA fingerprinting in forensics: past, present, future, *Investig. Genet.* 4 (2013) 22.

[22] M.L. Baird, Use of the AmpliType PM + HLA DQA1 PCR Amplification and Typing Kits for Identity Testing, *Methods Mol. Biol.* 98 (1998) 261–277.

[23] R.K. Saiki, T.L. Bugawan, G.T. Horn, K.B. Mullis, H. a Erlich, Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes, *Nature* 324 (1986) 163–166.

[24] A. Edwards, A. Civitello, H.A. Hammond, C.T. Caskey, DNA typing and genetic mapping with trimeric and tetrameric tandem repeats, *Am. J. Hum. Genet.* 49 (1991) 746–756.

[25] C. Børsting, N. Morling, Next generation sequencing and its applications in forensic genetics, *Forensic Sci. Int. Genet.* 18 (2015) 78–89.

[26] S. Anderson, A.T. Bankier, B.G. Barrell, M.H. de Bruijn, A.R. Coulson, J. Drouin, et al., Sequence and organization of the human mitochondrial genome, *Nature* 290 (1981) 457–465.

[27] B. Budowle, M.W. Allard, M.R. Wilson, R. Chakraborty, Forensics and mitochondrial DNA: applications, debates, and foundations, *Annu. Rev. Genomics Hum. Genet.* 4 (2003) 119–141.

[28] S. Lutz, H. Wittig, H.J. Weisser, J. Heizmann, A. Junge, N. Dimo-Simonin, et al., Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions?, *Forensic Sci. Int.* 113 (2000) 97–101.

[29] C. Bini, S. Ceccardi, D. Luiselli, G. Ferri, S. Pelotti, C. Colalongo, et al., Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy), *Forensic Sci. Int.* 135 (2003) 48–52.

[30] H.J. Bandelt, V. Macaulay, M. Richard, Human Mitochondrial DNA and the Evolution of Homo Sapiens, *Springer* (2006).

[31] A. Torroni, A. Achilli, V. Macaulay, M. Richards, H.J. Bandelt, Harvesting the fruit of the human mtDNA tree, *Trends Genet.* 22 (2006) 339–345.

[32] P.A. Underhill, T. Kivisild, Use of y chromosome and mitochondrial DNA population structure in tracing human migrations, *Annu. Rev. Genet.* 41 (2007) 539–564.

[33] B. Budowle, A. Van Daal, Forensically relevant SNP classes, *Biotechniques* 44 (2008) 603–610.

[34] P. Turnpenny, S. Ellard, Emery's Elements of Medical Genetics 14th edition, *Elsevier Ltd. Churchill Livingstone, Philadelphia* (2012).

[35] T.M. Diegoli, Forensic typing of short tandem repeat markers on the X and Y chromosomes, *Forensic Sci. Int. Genet.* 18 (2015) 140–151.

[36] H. Ellegren, Microsatellites: simple sequences with complex evolution, Nat. Rev. 5 (2004) 435–445.

[37] J.R. Collins, R.M. Stephens, B. Gold, B. Long, M. Dean, S.K. Burt, An exhaustive DNA micro-satellite map of the human genome using high performance computing, *Genomics* 82 (2003) 10–19.

[38] S. Subramanian, R.K. Mishra, L. Singh, Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions, *Genome Biol.* 4 (2003) R13.

[39] G. Levinson, G.A. Gutman, High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in Escherichia coli K-12, *Nucleic Acids Res.* 15 (1987) 5323–5338.

[40] P. Wiegand, E. Meyer, B. Brinkmann, Microsatellite structures in the context of human evolution, *Electrophoresis* 21 (2000) 889–895.

[41] Y. Zhu, J.E. Strassmann, D.C. Queller, Insertions, substitutions, and the origin of microsatellites, *Genet. Res.* 76 (2000) 227–236.

[42] O. Rose, D. Falush, A threshold size for microsatellite expansion, *Mol. Biol. Evol.* 15 (1998) 613–615.

[43] T. Pupko, D. Graur, Evolution of microsatellites in the yeast Saccharomyces cerevisiae: role of length and number of repeated units, *J. Mol. Evol.* 48 (1999) 313–316.

[44] R.R. Lyer, A. Pluciennik, V. Burdett, P.L. Modrich, DNA mismatch repair: Functions and mechanisms, *Chem. Rev.* 106 (2006) 302–323.

[45] T. Egeland, D. Kling, P. Mostad, Relationship Inference with Familias and R: Statistical Methods in Forensic Genetics, *Elsevier Ltd. Academic Press, London* (2016).

[46] E. Meyer, P. Wiegand, S.P. Rand, D. Kuhlmann, M. Brack, B. Brinkmann, Microsatellite polymorphisms reveal phylogenetic relationships in primates, *J. Mol. Evol.* 41 (1995) 10–14.

[47] R. Chakraborty, M. Kimmel, D.N. Stivers, L.J. Davison, R. Deka, Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci, *Proc. Natl. Acad. Sci. U. S. A.* 94 (1997) 1041–1046.

[48] Y.C. Li, A.B. Korol, T. Fahima, A. Beiles, E. Nevo, Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review, *Mol. Ecol.* 11 (2002) 2453–2465.

[49] L.D. Hurst, H. Ellegren, Sex biases in the mutation rate, *Trends Genet.* 14 (1998) 446–452.

[50] H. Ellegren, Heterogeneous mutation processes in human microsatellite DNA sequences, *Nat. Genet.* 24 (2000) 400–402.

[51] B. Brinkmann, M. Klintschar, F. Neuhuber, J. Hühne, B. Rolf, Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat, Am. J. Hum. Genet. 62 (1998) 1408–1415.

[52] F. Aşicioglu, F. Oguz-Savran, U. Ozbek, Mutation rate at commonly used forensic STR loci: paternity testing experience, *Dis. Markers* 20 (2004) 313–315.

[53] J.R. Dettman, J.W. Taylor, Mutation and evolution of microsatellite loci in neurospora, *Genetics* 168 (2004) 1231–1248.

[54] A. Urquhart, C.P. Kimpton, T.J. Downes, P. Gill, Variation in short tandem repeat sequences--a survey of twelve microsatellite loci for use as forensic identification markers, *Int. J. Legal Med.* 107 (1994) 13–20.

[55] J.M. Butler, C.R. Hill, Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis, *Forensic Sci Rev.* 4049 (2012) 15–26.

[56] T.M. Diegoli, M.D. Coble, Development and characterization of two mini-X chromosomal short tandem repeat multiplexes, *Forensic Sci. Int. Genet.* 5 (2011) 415–421.

[57] Recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms, *Forensic Sci. Int.* 55 (1992) 1–3.

[58] DNA recommendations - 1992 report concerning recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms, *Int. J. Legal Med.* 105 (1992) 63–64.

[59] DNA recommendations – 1994 report concerning further recommendations of the DNA commission of the ISFH regarding PCR-based polymozphisms in STR (short tandem repeat) systems, *Int. J. Legal Med.* 107 (1994) 159–160.

[60] P. Gill, B. Brinkmann, E. D'Aloja, J. Andersen, W. Bar, A. Carracedo, et al., Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature, *Forensic Sci. Int.* 87 (1997) 185–192.

[61] P.D. Martin, H. Schmitter, P.M. Schneider, A brief history of the formation of DNA databases in forensic science within Europe, *Forensic Sci. Int.* 119 (2001) 225–231.

[62] C.H. Asplen, S.A. Lane, International perspectives on forensic DNA databases, *Forensic Sci. Int.* 146 (2004) Suppl: S119-121.

[63] F. Corte-Real, Forensic DNA databases, *Forensic Sci. Int.* 146 (2004) Suppl: S143-144.

[64] K. Bender, P.M. Schneider, Validation and casework testing of the BioPlex-11 for STR typing of telogen hair roots, *Forensic Sci. Int.* 161 (2006) 52–59.

[65] A. Odriozola, J.M. Aznar, D. Celorrio, M.L. Bravo, J.J. Builes, J.F. Val-Bernal, et al., Development and validation of I-DNA1: A 15-Loci multiplex system for identity testing, *Int. J. Legal Med.* 125 (2011) 685–694.

[66] A. Odriozola, J.M. Aznar, D. Celorrio, M.L. Bravo, J.J. Builes, M.M. de Pancorbo, Development and validation for identity testing of I-DNADuo, a combination of I-DNA1 and a new multiplex system, I-DNA2, *Int. J. Legal Med.* 126 (2012) 167–172.

[67] J.M. Aznar, D. Celorrio, A. Odriozola, S. Köhnemann, M.L. Bravo, J.J. Builes, et al., I-DNASE21 system: Development and SWGDAM validation of a new STR 21-plex reaction, *Forensic Sci. Int. Genet.* 8 (2014) 10–19.

[68] K. Müller, G. Braunschweiger, R. Klein, E. Miltner, P. Wiegand, Validation of the Short Amplicon Multiplex Q8 Including the German DNA Database Systems, *J. Forensic Sci.* 54 (2009) 862–865.

[69] A. Odriozola, J.M. Aznar, D. Celorrio, M.M. de Pancorbo, Recent advances and considerations for the future in forensic analysis of degraded DNA by autosomic miniSTR multiplex genotyping, *Recent Pat. DNA Gene Seq.* 5 (2011) 110–116.

[70] P. Gill, L. Fereday, N. Morling, P.A.M. Schneider, New multiplexes for Europe - Amendments and clarification of strategic development, *Forensic Sci Int.* 163 (2006) 155–157.

[71] F. Guo, H. Shen, H. Tian, P. Jin, X. Jiang, Development of a 24-locus multiplex system to incorporate the core loci in the Combined DNA Index System (CODIS) and the European Standard Set (ESS), Forensic Sci. Int. Genet. 8 (2014) 44–54.

[72] Q. Hu, Y. Liu, S. Yi, D. Huang, A comparison of four methods for PCR inhibitor removal, *Forensic Sci. Int. Genet.* 16 (2015) 94–97.

[73] C.C. Tebbe, W. Vahjen, Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast, *Appl. Environ. Microbiol.* 59 (1993) 2657–2665.

[74] A. Akane, K. Matsubara, H. Nakamura, S. Takahashi, K. Kimura, Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification, *J. Forensic Sci.* 39 (1994) 362–372.

[75] M. Baeta, C. Núñez, S. Cardoso, L. Palencia-Madrid, L. Herrasti, F. Etxeberria, et al., Digging up the recent Spanish memory: Genetic identification of human remains from mass graves of the Spanish Civil War and posterior dictatorship, *Forensic Sci. Int. Genet.* 19 (2015) 272–279.

[76] A. Carracedo, M.V. Lareu, Development of new STRs for forensic casework: criteria for selection, sequencing & population data and forensic validation, Proceedings of the 9th International Symposium on Human Identification (http://www.promega.com/geneticidproc/ussymp9proc/content/21.pdf) (accessed 11.03.17).

[77] SWGDAM Validation Guidelines for DNA Analysis Methods, http://swgdam.org/SWGDAM_Validation_Guidelines_APPROVED_Dec_2012.pdf (accessed 09.08.15).

[78] P. Gill, C. Kimpton, E. D'Aloja, J.F. Andersen, W. Bar, B. Brinkmann, et al., Report of the European DNA profiling group (EDNAP) - towards standardisation of short tandem repeat (STR) loci, *Forensic Sci. Int.* 65 (1994) 51–59.

[79] J.M. Butler, Genetics and genomics of core short tandem repeat loci used in human identity testing, *J. Forensic Sci.* 51 (2006) 253–265.

[80] M.D. Coble, Capillary electrophoresis of MiniSTR markers to genotype highly degraded DNA samples, *Methods Mol. Biol.* 830 (2012) 31–42.

[81] F. Van Nieuwerburgh, D. Van Hoofstat, C. Van Neste, D. Deforce, Retrospective study of the impact of miniSTRs on forensic DNA profiling of touch DNA samples, *Sci. Justice* 54 (2014) 369–372.

[82] P.S. Walsh, N.J. Fildes, R. Reynolds, Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA, *Nucleic Acids Res.* 24 (1996) 2807–2812.

[83] A. Hosseinzadeh-Colagar, M.J. Haghighatnia, Z. Amiri, M. Mohadjerani, M. Tafrihi, Microsatellite (SSR) amplification by PCR usually led to polymorphic bands: Evidence which shows replication slippage occurs in extend or nascent DNA strands, *Mol. Biol. Res. Commun.* 5 (2016) 167–174.

[84] C.M. Ruitberg, D.J. Reeder, J.M. Butler, STRBase: a short tandem repeat DNA database for the human identity testing community, *Nucleic Acids Res.* 29 (2001) 320–322.

[85] H.M. Wallace, A.R. Jackson, J. Gruber, A.D. Thibedeau, Forensic DNA databases-Ethical and legal standards: A global review, *Egypt. J. Forensic Sci.* 4 (2014) 57–63.

[86] J.M. Butler, The future of forensic DNA analysis, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370 (2015) 577–579.

[87] P. Gill, Role of short tandem repeat DNA in forensic casework in the UK--past, present, and future perspectives, *Biotechniques* 32 (2002) 366–368, 370, 372, passim.

[88] J.M. Butler, Advanced Topics in Forensic DNA Typing: Interpretation, *Elsevier Ltd. Academic Press, San Diego* (2014).

[89] M. Nei, A.K. Roychoudhury, Sampling variances of heterozygosity and genetic distance, *Genetics* 76 (1974) 379–390.

[90] A. Edwards, H.A. Hammond, L. Jin, C.T. Caskey, R. Chakraborty, Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups, *Genomics* 12 (1992) 241–253.

[91] R. Szibor, S. Hering, E. Kuhlisch, I. Plate, S. Demberger, M. Krawczak, et al., Haplotyping of STR cluster DXS6801-DXS6809-DXS6789 on Xq21 provides a powerful tool for kinship testing, *Int. J. Legal Med.* 119 (2005) 363–369.

[92] L. Excoffier, G. Laval, D. Balding, Gametic phase estimation over large genomic regions using an adaptive window approach, *Hum. Genomics* 1 (2003) 7–19.

[93] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): An integrated software package for population genetics data analysis, *Evol. Bioinform. Online* 1 (2005) 47–50.

[94] S. Cardoso, R. Sevillano, M.J. Illescas, M.M. de Pancorbo, Analysis of 16 autosomal STRs and 17 Y-STRs in an indigenous Maya population from Guatemala, *Int. J. Legal Med.* 130 (2016) 365–366.

[95] A.M. Pérez-Miranda, M.A. Alfonso-Sánchez, A. Kalantar, J.A. Peña, M.M. de Pancorbo, R.J. Herrera, Allelic frequencies of 13 STR loci in autochthonous Basques from the province of Vizcaya (Spain), *Forensic Sci. Int.* 152 (2005) 259–262.

[96] A.M. Pérez-Miranda, M.A. Alfonso-Sánchez, J.A. Peña, M.M. de Pancorbo, R.J. Herrera, Genetic polymorphisms at 13 STR loci in autochthonous Basques from the province of Alava (Spain), *Leg. Med.* 7 (2005) 58–61.

[97] W.R. Engels, Exact tests for Hardy-Weinberg proportions, *Genetics* 183 (2009) 1431–1441.

[98] S.W. Guo, E.A. Thompson, Performing the exact test of Hardy-Weinberg proportion for multiple alleles, *Biometrics* 48 (1992) 361–372.

[99] A.O. Tillmar, T. Egeland, B. Lindblom, G. Holmlund, P. Mostad, Using X-chromosomal markers in relationship testing: Calculation of likelihood ratios taking both linkage and linkage disequilibrium into account, *Forensic Sci. Int. Genet.* 5 (2011) 506–511.

[100] M. Slatkin, Linkage disequilibrium - understanding the evolutionary past and mapping the medical future, *Nat. Rev. Genet.* 9 (2008) 477–485.

[101] K.E. Holsinger, B.S. Weir, Genetics in geographically structured populations: defining, estimating and interpreting FST, *Nat. Rev. Genet.* 10 (2009) 639–650.

[102] S. Wright, The Genetical Structure of Populations, *Ann. Eugen.* 15 (1951) 322–354.

[103] S. Wright, The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating, *Evolution* 19 (1965) 395–420.

[104] M. Slatkin, A measure of population subdivision based on microsatellite allele frequencies, *Genetics* 139 (1995) 457–462.

[105] F. Balloux, N. Lugon-Moulin, The estimation of population differentiation with microsatellite markers, *Mol. Ecol.* 11 (2002) 155–165.

[106] R.A. Fisher, Standard calculations for evaluating a blood-group system, *Heredity* 5 (1951) 95–102.

[107] D.A. Jones, Blood samples: probability of discrimination, *J. Forensic Sci. Soc.* 12 (1972) 355–359.

[108] G. Sensabaugh, Biomechanical markers of individuality. In: Forensic Science Handbook volume 3 (R. Saferstein ed.), *Prentice-Hall New York* (1982).

[109] C. Brenner, J.W. Morris, Paternity index calculations in single locus hypervariable DNA probes: validation and other studies. In International symposium on human identification 1989: proceedings, Madison, Wisconsin Promega Corp.

[110] D. Botstein, R.L. White, M. Skolnick, R.W. Davis, Construction of a genetic linkage map in man using restriction fragment length polymorphisms, *Am. J. Hum. Genet.* 32 (1980) 314–331.

[111] J. Krüger, W. Fuhrmann, K. Lichte, C. Steffens, Zur Verwendung der sauren Erythrocytenphosphatase bei der Vaterschaftsbegutachtung, *Dtsch. Z. Gerichtl. Med.* 64 (1968) 127–146.

[112] N. Pinto, L. Gusmão, A. Amorim, X-chromosome markers in kinship testing: A generalisation of the IBD approach identifying situations where their contribution is crucial, *Forensic Sci. Int. Genet.* 5 (2011) 27–32.

[113] L. Gusmão, C. Alves, P. Sánchez-Diz, M.T. Zarrabeitia, M.A. Abovich, I. Aragón, et al., Results of the GEP-ISFG collaborative study on an X-STR Decaplex, *Forensic Sci. Int. Genet. Suppl. Ser.* 1 (2008) 677–679.

[114] L. Gusmão, P. Sánchez-Diz, C. Alves, I. Gomes, M.T. Zarrabeitia, M. Abovich, et al., A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: Data on 15 Iberian and Latin American populations, *Int. J. Legal Med.* 123 (2009) 227–234.

[115] J.M. Butler, Fundamentals of Forensic DNA Typing, *Elsevier Ltd. New York* (2010).

[116] M.T. Ross, D. V Grafham, A.J. Coffey, S. Scherer, K. McLay, D. Muzny, et al., The DNA sequence of the human X chromosome, *Nature* 434 (2005) 325–337.

[117] T. Strachan, A. Read, Human Molecular Genetics 4th edition, Garland Science/*Taylor & Francis Group* New York (2010).

[118] N.A. Johnson, J. Lachance, The genetics of sex chromosomes: evolution and implications for hybrid incompatibility, *Ann. N. Y. Acad. Sci.* 1256 (2012) e1-22.

[119] Q.L. Liu, D.J. Lu, X.G. Li, H. Zhao, J.M. Zhang, Y.K. Lai, et al., Development of the nine X-STR loci typing system and genetic analysis in three nationality populations from China, *Int J Legal Med.* 125 (2011) 51–58.

[120] M. Israr, A.A. Shahid, Z. Rahman, M.S. Zar, M.S. Shahzad, T. Husnain, et al., Development and characterization of a new 12-plex ChrX miniSTR system, *Int. J. Legal Med.* 128 (2014) 595–598.

[121] S. Turrina, R. Atzei, G. Filippini, D. de Leo, Development and forensic validation of a new multiplex PCR assay with 12 X-chromosomal short tandem repeats, *Forensic Sci. Int. Genet.* 1 (2007) 201–204.

[122] E.M. Rodrigues Ribeiro, F.P. Leite, M.H. Hutz, J. PalhaTde, A.K. Ribero dos Santos, S.E. dos Santos, A multiplex PCR for 11 X chromosome STR markers and population data from a Brazilian Amazon Region, *Forensic Sci. Int. Genet.* 2 (2008) 154–158.

[123] H.L. Hwa, Y.Y. Chang, J.C.I. Lee, H.Y. Yin, Y.H. Chen, L.H. Tseng, et al., Thirteen X-chromosomal short tandem repeat loci multiplex data from Taiwanese, *Int. J. Legal Med.* 123 (2009) 263–269.

[124] Y. Nakamura, K. Minaguchi, Sixteen X-chromosomal STRs in two octaplex PCRs in Japanese population and development of 15-locus multiplex PCR system, *Int. J. Legal Med.* 124 (2010) 405–414.

[125] Q.L. Liu, D.J. Lu, L. Quan, Y.F. Chen, M. Shen, H. Zhao, Development of multiplex PCR system with 15 X-STR loci and genetic analysis in three nationality populations from China, *Electrophoresis* 33 (2012) 1299–1305.

[126] X. Yang, W. Wu, L. Chen, C. Liu, X. Zhang, L. Chen, et al., Development of the 19 X-STR loci multiplex system and genetic analysis of a Zhejiang Han population in China, *Electrophoresis* 37 (2016) 2260–2272.

[127] R. Szibor, S. Hering, J. Edelmann, The HumARA genotype is linked to spinal and bulbar muscular dystrophy and some further disease risks and should no longer be used as a DNA marker for forensic purposes, *Int. J. Legal Med.* 119 (2005) 179–180.

[128] R. Szibor, X-chromosomal markers: Past, present and future, *Forensic Sci. Int. Genet.* 1 (2007) 93–99.

[129] M. Castañeda, A. Odriozola, J. Gómez, M.T. Zarrabeitia, Development and validation of a multiplex reaction analyzing eight miniSTRs of the X chromosome for identity and kinship testing with degraded DNA, *Int. J. Legal Med.* 127 (2013) 735–739.

[130] M. Krawczak, Kinship testing with X-chromosomal markers: Mathematical and statistical issues, *Forensic Sci. Int. Genet.* 1 (2007) 111–114.

[131] B.S. Weir, A.D. Anderson, A.B. Hepler, Genetic relatedness analysis: modern data and new challenges, *Nat. Rev. Genet.* 7 (2006) 771–780.

[132] D. Kling, A.O. Tillmar, T. Egeland, P. Mostad, A general model for likelihood computations of genetic marker data accounting for linkage, linkage disequilibrium, and mutations, *Int. J. Legal Med.* 129 (2015) 943–954.

[133] D. Kling, B. Dell'Amico, A.O. Tillmar, FamLinkX - Implementation of a general model for likelihood computations for X-chromosomal marker data, *Forensic Sci. Int. Genet.* 17 (2015) 1–7.

[134] H.B. Luo, Y. Ye, Y.Y. Wang, W.B. Liang, L.B. Yun, M. Liao, et al., Characteristics of eight X-STR loci for forensic purposes in the Chinese population, *Int. J. Legal Med.* 125 (2011) 127–131.

[135] C. Dong, L. Fu, X. Zhang, C. Ma, F. Yu, S. Li, et al., Development of three X-linked tetrameric microsatellite markers for forensic purposes, *Mol. Biol. Rep.* 41 (2014) 6429–6432.

[136] H.T. Meng, J.T. Han, Y.D. Zhang, W.J. Liu, T.J. Wang, J.W. Yan, et al., Diversity study of 12 X-chromosomal STR loci in Hui ethnic from China, *Electrophoresis* 35 (2014) 2001–2007.

[137] S. Pasino, S. Caratti, M. Del Pero, A. Santovito, C. Torre, C. Robino, Allele and haplotype diversity of X-chromosomal STRs in Ivory Coast, *Int. J. Legal Med.* 125 (2011) 749–752.

[138] S. Inturri, S. Menegon, A. Amoroso, C. Torre, C. Robino, Linkage and linkage disequilibrium analysis of X-STRs in Italian families, *Forensic Sci. Int. Genet.* 5 (2011) 152–154.

[139] P. Paul, D. Nag, S. Chakraborty, Recombination hotspots: Models and tools for detection, *DNA Repair (Amst).* 40 (2016) 47–56.

[140] C. Phillips, D. Ballard, P. Gill, D.S. Court, A. Carracedo, M.V. Lareu, The recombination landscape around forensic STRs: Accurate measurement of genetic distances between syntenic STR pairs using HapMap high density SNP data, *Forensic Sci. Int. Genet.* 6 (2012) 354–365.

[141] S. Hering, C. Augustin, J. Edelmann, M. Heidel, J. Dressler, H. Rodig, et al., DXS10079, DXS10074 and DXS10075 are STRs located within a 280-kb region of Xq12 and provide stable haplotypes useful for complex kinship cases, *Int. J. Legal Med.* 120 (2006) 337–345.

[142] D.D. Kosambi, The estimation of map distance from recombination values, *Ann. Eugen.* 12 (1944) 172–175.

[143] H.J. Muller, The mechanism of crossing over, *Am Nat.* 50 (1916) 193-221.

[144] S. Elakkary, S. Hoffmeister-Ullerich, C. Schulze, E. Seif, A. Sheta, S. Hering, et al., Genetic polymorphisms of twelve X-STRs of the investigator Argus X-12 kit and additional six X-STR centromere region loci in an Egyptian population sample, *Forensic Sci. Int. Genet.* 11 (2014) 26–30.

[145] C. Tomas, J.J. Sanchez, J.A. Castro, C. Børsting, N. Morling, Forensic usefulness of a 25 X-chromosome single-nucleotide polymorphism marker set, *Transfusion* 50 (2010) 2258–2265.

[146] C. Børsting, N. Morling, Reinvestigations of six unusual paternity cases by typing of autosomal single-nucleotide polymorphisms, *Transfusion* 52 (2012) 425–430.

[147] B. Sobrino, M. Brión, A. Carracedo, SNPs in forensic genetics: A review on SNP typing methodologies, *Forensic Sci. Int.* 154 (2005) 181–194.

[148] 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.

[149] L. Li, C. Li, S. Zhang, S. Zhao, Y. Liu, Y. Lin, Analysis of 14 highly informative SNP markers on X chromosome by TaqMan® SNP genotyping assay, *Forensic Sci. Int. Genet.* 4 (2010) 2–5.

[150] C. Tomàs, J.J. Sanchez, J.A. Castro, C. Børsting, N. Morling, Utility of X-chromosome SNPs in relationship testing, *Forensic Sci. Int. Genet. Suppl. Ser.* 1 (2008) 528–530.

[151] J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, et al., A multiplex assay with 52 single nucleotide polymorphisms for human identification, *Electrophoresis* 27 (2006) 1713–1724.

[152] A. Odriozola, J.M. Aznar, L. Valverde, S. Cardoso, M.L. Bravo, J.J. Builes, et al., SNPSTR rs59186128_D7S820 polymorphism distribution in European Caucasoid, Hispanic, and Afro-American populations, *Int. J. Legal Med.* 123 (2009) 527-533.

[153] M. Castañeda, Estudio de los microsatélites y miniSTRs del cromosoma X de aplicación forense, PhD Thesis, University of Cantabria (Spain) 2013.

[154] J. Edelmann, R. Szibor, Validation of the X-linked STR DXS6801, *Forensic Sci. Int.* 148 (2005) 219–220.

[155] J. Edelmann, S. Hering, M. Michael, R. Lessig, D. Deischel, G. Meier-Sundhausen, et al., 16 X-chromosome STR loci frequency data from a German population, *Forensic Sci. Int.* 124 (2001) 215–218.

[156] J.J. Sanchez, C. Børsting, N. Morling, Typing of Y chromosome SNPs with multiplex PCR methods, *Methods Mol. Biol.* 297 (2005) 209–228.

[157] R.M. Woodsmall, D.A. Benson, Information resources at the National Center for Biotechnology Information, *Bull. Med. Libr. Assoc.* 81 (1993) 282–284.

[158] J. Edelmann, D. Deichsel, S. Hering, I. Plate, R. Szibor, Sequence variation and allele nomenclature for the X-linked STRs DXS9895, DXS8378, DXS7132, DXS6800, DXS7133, GATA172D05, DXS7423 and DXS8377, *Forensic Sci. Int.* 129 (2002) 99–103.

[159] S. Hering, R. Szibor, Development of the X-linked tetrameric microsatellite marker DXS9898 for forensic purposes, *J. Forensic Sci.* 45 (2000) 929–931.

[160] J. Edelmann, D. Deichsel, I. Plate, M. Käser, R. Szibor, Validation of the X-chromosomal STR DXS6809, *Int. J. Legal Med.* 117 (2003) 241–244.

[161] S. Hering, E. Kuhlisch, R. Szibor, Development of the X-linked tetrameric microsatellite marker HumDXS6789 for forensic purposes, *Forensic Sci. Int.* 119 (2001) 42–46.

[162] J. Edelmann, R. Szibor, Validation of the HumDXS6807 short tandem repeat polymorphism for forensic application, *Electrophoresis* 20 (1999) 2844–2846.

[163] O.J. Marshall, PerlPrimer: Cross-platform, graphical primer design for standard, bisulphite and real-time PCR, *Bioinformatics* 20 (2004) 2471–2472.

[164] P.M. Vallone, J.M. Butler, AutoDimer: A screening tool for primer-dimer and hairpin structures, *Biotechniques* 37 (2004) 226–231.

[165] SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories, https://www.fbi.gov/about-us/lab/biometric-analysis/codis/swgdam.pdf (accessed 13.01.16).

[166] J.C. Barrett, B. Fry, J. Maller, M.J. Daly, Haploview: analysis and visualization of LD and haplotype maps, *Bioinformatics* 21 (2005) 263–265.

[167] H. Levene, On a matching problem arising in genetics, *Ann. Math. Stat.* 20 (1949) 91–94.

[168] M. Slatkin, L. Excoffier, Testing for linkage disequilibrium in genotypic data using the EM algorithm, *Heredity* 76 (1996) 377–383.

[169] L. Excoffier, M. Slatkin, Incorporating genotypes of relatives into a test of linkage disequilibrium, *Am. J. Hum. Genet.* 62 (1998) 171–180.

[170] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood estimation from incomplete data via the EM algorithm, *J. R. Stat. Assoc.* 39 (1977) 1–38.

[171] Ø. Hammer, D.A.T. Harper, P.D. Ryan, PAST: Paleontological Statistics Software Package for Education and Data Analysis, *Palaeontol. Electron.* 4 (2001) 1–9.

[172] D. Adler, D. Murdoch, O. Nenadic, S. Urbanek, M. Chen, A. Gebhardt, et al., Rgl: 3D visualization using OpenGL, R Package Version 0.96.0, 2016. http://cran.r-project.org/package=rgl (Accessed 10 September 2016).

[173] R Core Team, R: A language and environment for statistical computing, R foundation for statistical computing, Vienna, Austria, 2013 http://R-project.org/ (Accessed 10 September 2016).

[174] T. Kishida, W. Wang, M. Fukuda, Y. Tamaki, Duplex PCR of the Y-27H39 and HPRT loci with reference to japanese population data on the HPRT locus, *Japanese J. Leg. Med.* 51 (1997) 67–69.

[175] D. Desmarais, Y. Zhong, R. Chakraborty, C. Perreault, L. Busque, Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), *J. Forensic Sci.* 43 (1998) 1046–1049.

[176] E. Prieto-Fernández, M. Baeta, C. Núñez, M.T. Zarrabeitia, R.J. Herrera, J.J. Builes, et al., Development of a new highly efficient 17 X-STR multiplex for forensic purposes, *Electrophoresis* 37 (2016) 1651–1658.

[177] E. Prieto-Fernández, C. Núñez, M. Baeta, S. Jiménez-Moreno, B. Martínez-Jarreta, M.M. de Pancorbo, Forensic Spanish allele and haplotype database for a 17 X-STR panel, *Forensic Sci. Int. Genet.* 24 (2016) 120–123.

[178] E. Prieto-Fernández, A. Díaz-De Usera, M. Baeta, C. Núñez, F. Chbel, S. Nadifi, et al., A genetic overview of Atlantic coastal populations from Europe and North-West Africa based on a 17 X-STR panel, *Forensic Sci. Int. Genet.* 27 (2017) 167–171.

[179] J.F. Ferragut, K. Bentayebi, J.A. Castro, C. Ramon, A. Picornell, Genetic analysis of 12 X-chromosome STRs in Western Mediterranean populations, *Int. J. Legal Med.* 129 (2015) 253–255.

[180] M. Baeta, M.J. Illescas, L. García, C. Núñez, E. Prieto-Fernández, S. Jiménez-Moreno, et al., Iberian allele frequency database for 10 X-STRs, *Forensic Sci. Int. Genet.* 19 (2015) 76–78.

[181] L. Cainé, S. Costa, M.F. Pinheiro, Population data of 12 X-STR loci in a North of Portugal sample, *Int. J. Legal Med.* 127 (2013) 63–64.

[182] E. Dawson, G.R. Abecasis, S. Bumpstead, Y. Chen, S. Hunt, D.M. Beare, et al., A first-generation linkage disequilibrium map of human chromosome 22, *Nature* 418 (2002) 544–548.

[183] Q.L. Liu, J.Z. Wang, L. Quan, H. Zhao, Y. Da Wu, X.L. Huang, et al., Allele and Haplotype Diversity of 26 X-STR Loci in Four Nationality Populations from China, *PLoS One* 8 (2013) e65570.

[184] M.T. Zarrabeitia, F. Pinheiro, M.M. de Pancorbo, L. Cainé, S. Cardoso, L. Gusmão, et al., Analysis of 10 X-linked tetranucleotide markers in mixed and isolated populations, *Forensic Sci. Int. Genet.* 3 (2009) 63–66.

[185] M.J. Illescas, A. Pérez, J.M. Aznar, L. Valverde, S. Cardoso, J. Algorta, et al., Population genetic data for 10 X-STR loci in autochthonous Basques from Navarre (Spain), *Forensic Sci. Int. Genet.* 6 (2012) e146–148.

[186] L. Valverde, S. Köhnemann, M. Rosique, S. Cardoso, M. Zarrabeitia, H. Pfeiffer, et al., 17 Y-STR haplotype data for a population sample of Residents in the Basque Country, *Forensic Sci. Int. Genet.* 6 (2012) e109–111.

[187] J.A. Peña, S. Garcia-Obregon, A.M. Perez-Miranda, M.M. de Pancorbo, M.A. Alfonso-Sanchez, Gene flow in the Iberian Peninsula determined from Y-chromosome STR loci, *Am. J. Hum. Biol.* 18 (2006) 532–539.

[188] A.M. Pérez-Miranda, M.A. Alfonso-Sánchez, A. Kalantar, S. García-Obregón, M.M. de Pancorbo, J.A. Peña, et al., Microsatellite data support subpopulation structuring among Basques, *J. Hum. Genet.* 50 (2005) 403–414.

[189] M.M. de Pancorbo, M. López-Martínez, C. Martínez-Bouzas, A. Castro, I. Fernández-Fernández, G. Antúnez de Mayolo, et al., The Basques according to polymorphic Alu insertions, *Hum. Genet.* 109 (2001) 224–233.

[190] S. Cardoso, M.J. Villanueva-Millán, L. Valverde, A. Odriozola, J.M. Aznar, S. Piñeiro-Hermida, et al., Mitochondrial DNA control region variation in an autochthonous Basque population sample from the Basque Country, *Forensic Sci. Int. Genet.* 6 (2012) e106-108.

[191] M. Karakachoff, N. Duforet-Frebourg, F. Simonet, S. Le Scouarnec, N. Pellen, S. Lecointe, et al., Fine-scale human genetic structure in Western France, *Eur. J. Hum. Genet.* (2014) 1–6.

[192] V. Dubut, L. Chollet, P. Murail, F. Cartault, E. Béraud-Colomb, M. Serre, et al., mtDNA polymorphisms in five French groups: importance of regional sampling, *Eur. J. Hum. Genet.* 12 (2004) 293–300.

[193] E. Bosch, F. Calafell, A. Pérez-Lezaun, J. Clarimón, D. Comas, E. Mateu, et al., Genetic structure of north-west Africa revealed by STR analysis, *Eur. J. Hum. Genet.* 8 (2000) 360–366.

[194] L. Li, Y. Liu, Y. Lin, Typing of 67 SNP Loci on X Chromosome by PCR and MALDI-TOF MS, *Res. Genet.* (2015) Article ID 374688.

[195] V. Stepanov, K. Vagaitseva, V. Kharkov, A. Cherednichenko, A. Bocharova, G. Berezina, et al., Forensic and population genetic characteristics of 62 X chromosome SNPs revealed by multiplex PCR and MALDI-TOF mass spectrometry genotyping in 4 North Eurasian populations, *Leg. Med.* 18 (2016) 66–71.

[196] M.T. Zarrabeitia, V. Mijares, J.A. Riancho, Forensic efficiency of microsatellites and single nucleotide polymorphisms on the X chromosome, *Int. J. Legal Med.* 121 (2007) 433–437.

[197] E. Prieto-Fernández, T. Kleinbielen, M. Baeta, M.M. de Pancorbo, Evaluation of the forensic efficiency of the tri- and tetrallelic SNPs located on the X-chromosome, *Forensic Sci. Int. Genet.* (Submitted to publication 27th March 2017).

Por cierto, queda pendiente llamar a Paco…

eman ta zabal zazu

Universidad
del País Vasco

Euskal Herriko
Unibertsitatea