Characterization of STR sequence variants on the X- and Y- chromosomes

Mariana Pereira Tavares de Almeida Alegria
Mestrado em Genética Forense
Departamento de Biologia
2022

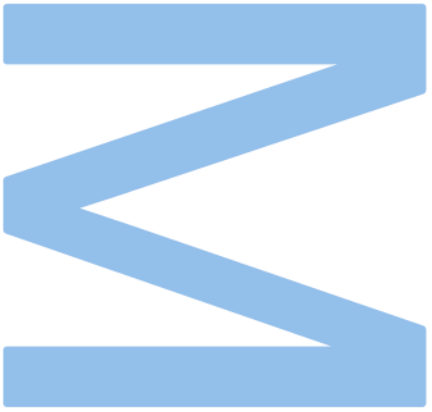
Orientadora
Iva Gomes, Investigadora Auxiliar, i3S/IPATIMUP

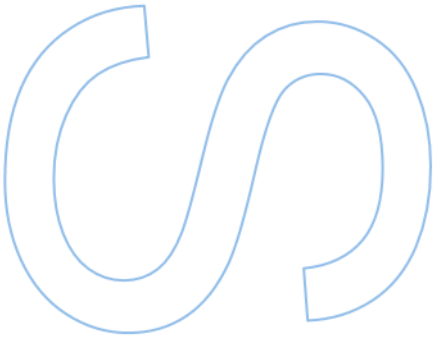
Coorientadora
Nádia Pinto, Investigadora Auxiliar, i3S/IPATIMUP/CMUP

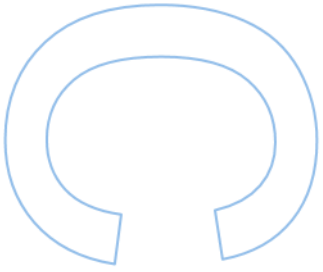

INSTITUTO DE INVESTIGAÇÃO
E INOVAÇÃO EM SAÚDE
UNIVERSIDADE DO PORTO

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO







Agradecimentos

Queria agradecer à minha orientadora, Iva Gomes, por me ter orientado e acolhido no grupo, e por todo o apoio e paciência. Agradeço também todo o acompanhamento e conhecimento que me proporcionou. É difícil encontrar uma orientadora tão compreensiva e disponível, e eu tive uma sorte enorme.

À minha coorientadora, Nádia Pinto, pela dedicação e pelos valiosos conhecimentos que partilhou comigo e que foram uma grande ajuda na elaboração desta dissertação.

À minha supervisora, Marisa Faustino, por me ter acompanhado durante todo o trabalho e me ter mostrado por onde começar, bem como por todo o apoio e disponibilidade.

A todos os membros do grupo de Genética Populacional e Evolução, por estarem sempre disponíveis para ajudar e pelos risos à hora de almoço.

Ao Professor António Amorim e à Professora Maria João Prata, por me terem recebido no grupo e pelos ensinamentos valiosos ao longo da minha licenciatura e mestrado.

O desenvolvimento desta dissertação não seria possível sem todo o apoio emocional que recebi. Quero agradecer aos meus amigos que passaram o teste dos 7 anos de amizade, pelo apoio constante apesar de as nossas vidas serem tão diferentes atualmente.

À Cientuna, a melhor tuna do mundo, pelas lembranças que para sempre irei guardar e pela distração nos momentos mais stressantes. Às amigas que fiz na tuna, com quem pude partilhar o meu percurso e que estiveram sempre lá para me apoiar. À Inês, pelas palavras fofinhas e pela confiança incondicional em mim.

Às amizades que fiz em Biologia, especialmente às BioUrsas por serem o melhor ano e por todos os momentos hilariantes e caóticos. Um agradecimento especial à Ana Sofia, Carolina, João e Marta, pelas fofocas, lanches estéticos e pelos conselhos descompensados. Adoro-vos e não faço ideia do que seria viver a faculdade sem vocês.

Ao Miguel, obrigada por seres o meu porto de abrigo e por estares disponível em todos os bons e maus momentos. Sei que fui difícil de aturar este ano, e não há palavras que cheguem para te agradecer. Aos meus pais, por todo o amor e apoio ao longo dos anos, e por acreditarem sempre em mim mesmo quando eu própria não acreditava. Obrigada por estarem sempre presentes.

Resumo

A variação genética manifesta-se na forma de polimorfismos nas moléculas de DNA. Testes de DNA são uma ferramenta poderosa de discriminação entre indivíduos, e DNA pode ser obtido da maior parte de amostras que contêm material biológico (sangue, saliva, sémen, etc.). O genoma humano consiste em 23 pares de cromossomas, sendo um deles o par de cromossomas sexuais, X e Y. Normalmente, as mulheres têm dois cromossomas X, enquanto os homens possuem um cromossoma X e um Y. Os heterossomas originaram de um par de autossomas e o Y evoluiu a partir do X. O X é um cromossoma grande, maioritariamente conservado entre mamíferos, com um grande número de marcadores descritos. O cromossoma Y é um cromossoma pequeno e determinante do sexo masculino, transmitido de maneira relativamente preservada (excluindo mutações) entre homens, sendo assim muito útil para determinar linhagens masculinas. Aneuploidias são relativamente frequentes nestes cromossomas, estando algumas delas bem-descritas.

STRs são o marcador molecular mais usado em forense, sendo altamente polimórficos, fidedignos, e, de um ponto de vista técnico, simples de usar. No contexto da genética forense, STRs dos cromossomas sexuais podem ser bastante eficientes na resolução de casos complexos. X-STRs são, em geral, pouco explorados e são maioritariamente usados como complemento para marcadores autossómicos. Alguns X-STRs, como o DXS10146, são muito polimórficos. Y-STRs são principalmente usados em misturas quando existe uma componente masculina e feminina, visto que permitem obter a linhagem paterna do homem. Mutação é a maior fonte de variação genética e, em análise de STRs, a sua compreensão é fundamental. Variação em microssatélites pode ser principalmente encontrada no tamanho, considerando as unidades variáveis do motivo repetitivo. Alelos variantes e INDELs nas regiões flanqueantes são marcadores valiosos, que são importantes de conhecer e estudar. Sequenciar STRs tem várias vantagens, tais como descobrir microvariantes novas ou confirmar os resultados de mutação obtidos por eletroforese capilar. Atualmente, o tradicional método automatizado, sequenciação de Sanger, (ainda) é o mais usado para investigação com objetivos simples.

Para compreender totalmente STRs, é necessário estudar o seu comportamento mutacional e variação genética analisando microvariantes. Sequenciar marcadores dos cromossomas sexuais, tanto em indivíduos humanos como não-humanos, pode fornecer informação importante que permite estabelecer nomenclaturas corretas. Este estudo pretende analisar a variação da sequência de DNA do X-STR DXS10146 e a

estabilidade e especificidade deste marcador ao sequenciá-lo em indivíduos de diferentes espécies, e confirmar uma trissomia X por genotipagem. Relativamente ao cromossoma Y, o objetivo é confirmar mutações em duos pai-filho sequenciando Y-STRs.

Para cumprir os objetivos estabelecidos, uma amostra de indivíduos portugueses foi selecionada e analisada. Amostras não-humanas foram também usadas para analisar a estrutura da sequência do STR ao compará-la com genomas ancestrais e para testar a especificidade humana do DXS10146. As amostras em que apenas um marcador foi analisado foram amplificadas por PCR em singleplex, observadas num gel de poliacrilamida, e as bandas dos heterozigóticos extraídas do gel e reamplificadas. Estes fragmentos foram depois sequenciados e analisados. O DNA da filha com uma possível trissomia X e dos seus pais foi amplificado com os kits multiplex Investigator® Argus X-12 e o GHEP-ISFG Decaplex, e posteriormente genotipado.

O locus DXS10146 mostrou variação no número de repetições do motivo repetitivo e foram encontrados alguns polimorfismos na região flanqueante. Os alelos sequenciados seguiram a nomenclatura estabelecida. A causa do possível alelo *off-ladder* 39.1 pode ser um isoalelo específico que causa uma mobilidade eletroforética diferente em comparação com os alelos adjacentes.

Para este locus, sequências de referência de humanos e primatas foram alinhadas e comparadas. As amostras de chimpanzé mostraram consistência com uma nomenclatura previamente descrita e foi encontrada variação no número de blocos repetitivos numa das amostras. DXS10146 provou ser um marcador específico para primatas, visto que não foi observada amplificação em amostras de outros mamíferos. Foi também possível confirmar um padrão trissómico no cromossoma X numa amostra feminina com dois multiplexes, apesar de terem ocorrido algumas dificuldades técnicas na genotipagem.

Relativamente aos Y-STRs, mutações *de novo* em duos pai-filho foram confirmados por sequenciação. Apesar de não ter sido possível sequenciar alguns duos devido a obstáculos técnicos, a sequência obtida correspondeu à genotipagem em todos os duos sequenciados, e foi possível identificar variações específicas no número de repetições.

As análises de polimorfismos genéticos, particularmente STRs, revelaram aplicações importantes em várias áreas distintas, mas com especial incidência nos estudos genéticos populacionais e forenses. Esta pesquisa destaca a importância de sequenciar STRs para revelar variação genética que pode causar problemas sérios durante

genotipagens, particularmente em investigações forenses. Sequenciar STRs em primatas também permite uma comparação direta com sequências de DNA humano, possibilitando inferir nomenclaturas mais precisas. Estes dados também enfatizam o DXS10146 como um marcador polimórfico e específico que também é maioritariamente conservado entre primatas e humanos.

Palavras-chave: cromossoma X, X-STR, DXS10146, cromossoma Y, Y-STR, polimorfismos, sequenciação

Abstract

Genetic variation is present as polymorphisms in DNA molecules. DNA testing is a powerful tool for discrimination between individuals, and DNA can be obtained from most sources containing biological material (blood, saliva, semen, etc.). The human genome consists of 23 pairs of chromosomes, one of which are the sex chromosomes, X and Y. Regularly, females have two X chromosomes, while males possess one X and one Y chromosome. The heterosomes originated from a pair of autosomes and the Y evolved from the X chromosome. The X is a large chromosome, mostly conserved between mammals, with a high number of markers described. The Y chromosome is a small, male-determining chromosome, passed down in a relatively preserved manner (barring mutations) between men, therefore very useful to determine male lineages. Aneuploidies are relatively frequent in these chromosomes, some of them being well-described.

STRs are the most used molecular marker in forensics, being highly polymorphic, reliable, and simple to use from a technical point of view. In forensic genetic settings, sex chromosome STRs may be greatly efficient at solving complex cases. X-STRs are generally underexplored and are mainly used as a complement to autosomal markers. Some of these, like DXS10146, are highly polymorphic. Y-STRs are mostly used in mixtures when there is a male and female component, as it allows to obtain the paternal lineage of the male. Mutation is the major source of genetic variation, and, in STR analysis, its understanding is fundamental. Variation in microsatellites is mostly found in length, considering variable units of the repetitive motif. Variant alleles and flanking region INDELs are valuable markers, which are important to acknowledge and study. Sequencing STRs has several advantages such as discovering novel microvariants or confirming the mutation results obtained with capillary electrophoresis. Currently, the automated and traditional Sanger sequencing method is (still) the most used for basic research purposes.

To fully understand STRs, it is necessary to study mutational behavior and genetic variation by analyzing microvariants. Sequencing sex chromosome markers, both in human and non-human subjects, can provide significant information, allowing to establish correct nomenclatures. This study aims to analyze the DNA sequence variation of the X-STR DXS10146 and the marker stability and specificity by sequencing it in individuals from different species, and confirm an X trisomy by genotyping. Relatively to the Y chromosome, the objective is to confirm mutations in father-son duos by sequencing Y-STRs.

To accomplish the aims established, a sample of Portuguese individuals was selected and analyzed. Samples from non-human subjects were also used to analyze the STR sequence structure by comparing with ancestral genomes and also to test human specificity of DXS10146. The samples for which only a marker was analyzed were PCR amplified in singleplex, observed on a polyacrylamide gel, and the bands of heterozygotes extracted from the gel and reamplified. These fragments were then sequenced and analyzed. The DNA from the daughter with a putative X trisomy and their parents was amplified with both the multiplex kits Investigator® Argus X-12 and GHEP-ISFG Decaplex, and then genotyped.

The DXS10146 locus showed variation in the number of repeats of the repetitive motif and some polymorphisms in the flanking region were also found. The sequenced alleles followed the established nomenclature. The cause of the putative off-ladder allele 39.1 might be a specific isoallele causing a different electrophoretic mobility compared to adjacent alleles.

For this locus, primate and human reference sequences were aligned and compared. The chimpanzee samples were consistent with a previously described nomenclature and variation in the number of repeat blocks was found in one of the samples. DXS10146 also proved to be a primate specific marker, since no amplification was observed in samples from other mammals. It was also possible to confirm a trisomic pattern in the X chromosome in a female sample with two multiplexes despite some technical difficulties in genotyping.

Regarding Y-STRs, *de novo* mutations in father-son duos were confirmed by sequencing. Despite sequencing not being possible in some duos due to technical obstacles, the sequence obtained matched the genotyping in all sequenced duos, and variations on the number of repeats were able to be specifically identified.

The analyses of genetic polymorphisms, in particular, STRs, have uncovered important and useful applications in many different fields but with special focuses in population and forensic genetic studies. This research highlights the importance of sequencing STRs to uncover genetic variation that might pose serious problems during genotyping assays, in particular, during forensic investigations. Sequencing of STRs in primates also allows a direct comparison with human DNA sequences, enabling the inference of more accurate allelic nomenclatures. This data also emphasizes DXS10146 as a polymorphic and specific marker that is also mostly conserved between primates and humans.

Keywords: X chromosome, X-STR, DXS10146, Y chromosome, Y-STR, polymorphisms, sequencing

Table of Contents

List of Tables

List of Figures

List of Abbreviations

APS	Ammonium Persulfate
bp	Base Pair
CE	Capillary Electrophoresis
ddNTP	2',3'-dideoxynucleotide
DNA	Deoxyribonucleic Acid
dNTP	2'-deoxynucleotide
GHEP-ISFG	Spanish and Portuguese Speaking Working Group of the International Society for Forensic Genetics
INDEL	Insertion-Deletion
KS	Klinefelter Syndrome
LD	Linkage Disequilibrium
LG	Linkage Group
NGS	Next-Generation Sequencing
NIST	National Institute of Standards and Technology
NRY	Non-recombining Region of the Y chromosome
OL	Off-Ladder
PAGE	Polyacrylamide Gel Electrophoresis
PAR	Pseudoautosomal Region
PCR	Polymerase Chain Reaction
PZM	Post-zygotic Mutation
RM	Rapidly Mutating
SRY	Sex-determining Region of the Y chromosome
SSR	Simple Sequence Repeat
STR	Short Tandem Repeat
TD-PCR	Touchdown Polymerase Chain Reaction
TEMED	1,2-Bis(dimethylamino)ethane
Tris	Tris(hydroxymethyl)aminomethane
TS	Turner Syndrome
VNTR	Variable Number Tandem Repeat
X-STR	X chromosome Short Tandem Repeat
XCI	X chromosome Inactivation
Y-STR	Y chromosome Short Tandem Repeat
YHRD	Y chromosome Haplotype Reference Database

1. Introduction

1.1. Forensic genetics

The human genome is made up of sequences encoded as deoxyribonucleic acid (DNA) molecules, located in the nucleus of the cells and in the mitochondria. These molecules comprise individuals' genetic information and variation between them is presented as polymorphisms in the DNA sequence. As such, every individual is expected to have a unique genetic profile, making individual human identification possible.

Back in the 1980s, forensic DNA testing was introduced by Gill, Jeffreys and Werrett (1985). This study showed that DNA fingerprints obtained from bloodstains, semen stains, and hair roots were as specific to individuals as whole blood and semen samples (Gill et al., 1985). Since then, forensic DNA testing has evolved into forensic casework routine, aiding the conviction of the guilty and exoneration of the innocent (Jobling & Gill, 2004), as well as being used in the analysis of paternity and complex kinship testing, missing person investigations, and mass casualty incident victim identifications (Ziętkiewicz et al., 2012).

DNA typing is a very powerful tool, as it enables high-level discrimination between individuals (Chakraborty & Kidd, 1991), and can aid investigations in the identification of suspects of committing crimes such as murder and rape (Butler, 2015).

Forensic DNA testing was facilitated by the discovery of hypervariable loci known as minisatellites (Jeffreys et al., 1985), a type of variable number tandem repeats (VNTR), which segregate according to Mendelian principles. The discovery of amplification by polymerase-chain reaction (PCR) (Mullis et al., 1986) also provided an increase in DNA testing sensitivity, allowing small amounts of DNA to be analyzed (Butler, 2015). Both techniques along with automated sequencing technology led to the current powerful systems for individual identification, providing sensitivity and discrimination power in a short time (Jobling & Gill, 2004).

In standard identification problems, DNA testing aims to obtain a matching profile between a sample with genetic material and an individual (or sample of reference), considering as an alternative the individual not being related to any contributor to the sample. If the profiles do not match then the evidence may be declared exculpatory; if there is a match then it is necessary to determine the likelihood of a match occurring by chance considering the population under analysis (Chakraborty & Kidd, 1991). DNA typing can display inconclusive results, usually due to insufficient DNA in the sample or to technical issues in the analysis. However, if the typing is done correctly and carefully,

and a set of markers properly designed for forensic purposes and a population-specific database are used, a match is significant evidence of the source of the DNA sample (Chakraborty & Kidd, 1991). Thanks to technological advances, cold cases can be solved much time after investigations have begun by analyzing degraded DNA, sometimes exonerating prisoners convicted due to errors (Jobling & Gill, 2004).

Beyond identification problems, kinship analyses represent a major fraction of the cases analyzed in forensic genetics routine. For example, in investigations of missing persons and victims of mass disasters, DNA typing has allowed many identifications by linking not only reference samples, but also relatives to recovered remains, even when in an advanced state of decay (Clayton et al., 1995). Indeed, since half of an individual's nuclear genetic code comes from the mother and the other half comes from the father, close biological relatives can be used as a reference if an individual is unavailable for genetic testing as a specified degree of genetic similarity is expected and can be included in the computations for weighing the evidence (Butler, 2015).

Any type of biological remains can be used as a source of genetic material (Chakraborty & Kidd, 1991; Ziętkiewicz et al., 2012). The quality of the genetic material used is strongly affected by both the time since death and environmental conditions and also depends on the tissue being analyzed (Ziętkiewicz et al., 2012). The source material does not need to be human, as analysis of both animal and plant DNA can be associated with a crime scene and provide evidence (the so-called silent witnesses), as well as being used in investigations of illegal trade of endangered species and other crimes (Jobling & Gill, 2004).

In the last 20 years, there has been a rapid growth in DNA testing and the tools available mostly due to technological developments. Therefore, there is now access to more rapid DNA testing, higher sensitivity in methodologies, better software, and greater depth of information (Butler, 2015), which has allowed the analysis of increasingly complex problems.

1.2. Sex chromosomes

Humans have a diploid nuclear genome, and a normal human cell contains forty-six chromosomes organized in twenty-three pairs. Twenty-two of these chromosomal pairs are autosomes, which contain the same genes in the same order along their chromosomal arm (Fuhrmann & Vogel, 1969). These are homologous chromosomes, one inherited from the mother and the other from the father.

Autosomes are different from the twenty-third pair of chromosomes that gathers the sex chromosomes, also called heterosomes: the X and Y chromosomes. These chromosomes are involved in sex determination since most females have two X chromosomes in their cells while most males have one X and one Y – exceptions to this rely on rare chromosomal aneuploidies (such as Triple X and Turner syndromes in females, and Klinefelter syndrome in males) that are compatible with life (section 1.2.4.). As such, females have twenty-three homologous chromosome pairs while males have twenty-two.

1.2.1 Origin and evolution

Both genetically and morphologically, the heterosomes X and Y are highly distinct from each other (J. A. M. Graves, 1995; Charlesworth, 2002). They do not cross over for most of their length, except on the pseudoautosomal regions (PAR) located at the tips of both chromosomes (Lahn & Page, 1999; Charlesworth, 2002; Ross et al., 2005). However, most genes on the Y chromosome have homologs on the X chromosome, showing that the heterosomes originated from a homologous pair of autosomes in a vertebrate ancestor (Ohno, 1967), and, as such, each Y gene evolved from its X counterpart (J. A. M. Graves, 1995; Lahn & Page, 1999).

Comparing sequences of X-linked genes and their Y equivalent provides evidence for various steps in the loss of homology in the sex chromosomes, as the Y was progressively deleted and inactivated (Graves & Foster, 1994; Lahn & Page, 1999; Ross et al., 2005), eventually showing a small number of active genes (Charlesworth, 2002). It is believed that several chromosomal inversions eventually led to the suppression of recombination between most of the two chromosomes (Lahn & Page, 1999).

The small PAR and the several X-Y shared genes outside this region are relics of homology and make it possible to discover more about human sex chromosome evolution (Graves & Foster, 1994).

1.2.2. X chromosome

In the human genome, the X chromosome has distinctive characteristics, specifically a haplodiploid mode of transmission. While females inherit an X chromosome from both parents, likewise autosomes, males only inherit one X chromosome from their mother, making them hemizygous for most X chromosome genes. As such, selection for and

against new alleles of a gene is more effective for X chromosomal genes than for those present in autosomes (Graves et al., 2002).

The X is one of the largest chromosomes in the human genome, with a size estimated to be around 163 Mb (Lander et al., 2001). Analyzing the X chromosome sequence shows a high number of interspersed repeats and a low number of genes (Ross et al., 2005). Genes on this chromosome have been linked to important functions in human neurodevelopment and intelligence (Skuse, 2005; Mallard et al., 2021) as well as in sex and reproduction (Rice, 1984; Saifl & Chandra, 1999).

One of the female X chromosomes has its gene expression silenced by X-chromosome inactivation (XCI), which compensates for the differences in male and female gene dosage (Graves, 1995), being reactivated in meiosis allowing the recombination with the second X chromosome. X-linked genes only spend one-third of their time in males and two-thirds in females, and since spermatogenesis undergoes more cell divisions than oogenesis, the mutation rate in the female germline is lower than in the male germline (Haldane, 1935).

In the process of separation from the ancestral autosome, the original elements of the X chromosome have been conserved, as this chromosome shows significant conservation in content within eutherian mammals (J. A. M. Graves, 1995; Ross et al., 2005).

1.2.3. Y chromosome

The Y chromosome is one of the smallest in the human genome, having about 51 Mb (Lander et al., 2001). It contains a sex-determining region Y (SRY) gene, which causes the Y chromosome to be male-determining (J. A. M. Graves, 1995).

This chromosome is divided into different regions. First, the PAR1 and PAR2, each at the terminus of the short (Yp) and long arm (Yq), respectively. These are the regions that allow the exchange of genetic material with the PAR of the X chromosome during male meiosis, and the genes located here are inherited in the same way as autosomes. There is also the euchromatic region, which consists of the centromere and the paracentromeric regions in both chromosomal arms (Quintana-Murci & Fellous, 2001), and contains genes responsible for important biological functions, like spermatogenesis, as well as a great quantity of repeating sequences (Graves, 1995). Finally, the heterochromatic region encompasses distal Yq and is a region without genes that may display variance in length between populations (Quintana-Murci & Fellous, 2001).

While both autosomes and X chromosomes have multiple ancestors due to recombination (lesser however in the latter case, as recombination in the specific zone only occurs in females), Y chromosomes have a single paternal ancestor (Jobling & Tyler-Smith, 1995) and the only cause of genetic variation is the occurrence of mutations on the non-recombining region of the Y chromosome (NRY) during meiosis (Claerhout et al., 2018). The NRY comprises the majority of the Y chromosome and since it is transmitted without recombination it holds a record of the detectable mutational events across male lineages (Quintana-Murci & Fellous, 2001). Forward mutations alter the wild-type phenotype, while reverse or backward mutations restore the wild-type phenotype in mutant cells (Rosenberg, 2013). These may occur at the same time and offset each other, turning them undetectable. As such, the Y is passed down through the patrilineal line in a relatively preserved manner and is a great tool for studying human populations and evolution.

1.2.4. Aneuploidies

Aneuploidy is defined as an irregularity in the number of chromosomes in the cell. Somatic human cells contain forty-six chromosomes and, as such, any other number of chromosomes is considered an aneuploidy. As such, monosomy refers to the loss of one chromosome and trisomy refers to the occurrence of three copies of a chromosome, while tetrasomy and pentasomy indicate, respectively, the occurrence of four or five copies of a chromosome.

The most well-described sex chromosomes aneuploidies are 45,X; 47,XXX; 47,XXY and 47,XYY. Although very rare, tetrasomy and pentasomy have also been reported in sex chromosomes (Linden et al., 1995). It is estimated that 1 in 426 children possess a sex chromosome abnormality (Nielsen & Wohlert, 1991). Non-mosaic aneuploidies occur due to either meiotic or post-zygotic mitotic (PZM) errors, while mosaic aneuploidies result from PZM errors (MacDonald et al., 1994; Álvarez-Nava & Lanes, 2018; Gruhn & Hoffmann, 2022).

The only chromosomal monosomy compatible with life postnatally is Turner syndrome (TS), with an incidence of approximately 1 in 2,500 liveborn girls (Nielsen & Wohlert, 1991; Haltrich et al., 2015; Vrtel et al., 2022). It was first described as 45,X in 1959 (Ford et al., 1959). A large proportion of individuals affected with TS are mosaics or possess structural aberrations of the X chromosome (Hall et al., 2006; Vrtel et al., 2022). In this syndrome, and depending on the relatives analyzed, it may be possible to determine which parent caused the meiotic error that gave rise to monosomy X, but not at which

stage of meiosis (Hall et al., 2006). Notwithstanding, the determination of the parental origin of the error may not be possible, depending on the individuals' genotypes and the relationship existing between the analyzed individuals as shown in an unpublished study (under review) (Faustino et al., 2023). The origin of the only X chromosome from the female with TS is mostly maternal (Ogata & Matsuo, 1995; Hall et al., 2006; Zhang et al., 2021; Vrtel et al., 2022).

The incidence of trisomy X is around 1 per 1,000 liveborn females, but most cases are not detected (Nielsen & Wohlert, 1991; Hall et al., 2006; Tartaglia et al., 2010). This condition, also known as 47,XXX or Triple X, was first described in 1959 (Jacobs et al., 1959). These individuals possess an extra X chromosome, mostly of maternal origin and due to an error in meiosis I (MacDonald et al., 1994; Hall et al., 2006). In females with trisomy X, inactivation occurs in two of the three X chromosomes, but genes in the PAR regions and genes that escape X-inactivation are expressed for all X chromosomes (Tartaglia et al., 2010). Trisomy X can occur in many mosaic combinations such as 46,XX/47,XXX or 47,XXX/48,XXXX, or in combinations including Turner syndrome cell lines (Nielsen & Wohlert, 1991).

About 1 in 600 liveborn males are affected with Klinefelter syndrome (KS) (Nielsen & Wohlert, 1991; Bojesen et al., 2003). This syndrome was first reported in 1959 and concerns males with an extra X chromosome (47, XXY) (Jacobs & Strong, 1959). Additional X chromosomes can be present, as well as mosaicism (Bojesen et al., 2003). KS is severely undiagnosed, as is estimated that only around one-fourth of adult males with KS are diagnosed (Bojesen et al., 2003). Among patients with KS, the parental origin of the extra X chromosome seems to be balanced and to happen mostly due to disjunction errors in meiosis I (MacDonald et al., 1994; Hall et al., 2006; Miki et al., 2017). A case has even been reported of two brothers with KS that had chromosomes with different parental origins (Kim et al., 2019).

The 47,XYY karyotype is fairly common, being estimated to affect 1 in 1,000 male births (Nielsen & Wohlert, 1991; Robinson & Jacobs, 1999). The trisomy 47,XYY was also described and first reported in 1961 (Sandberg et al., 1961). This condition is unique because it forcibly involves a paternal error as an embryo with two Y chromosomes can only be produced due to Y chromosome nondisjunction during paternal meiosis II or the early cleavage divisions of the embryo (Robinson & Jacobs, 1999; Hall et al., 2006).

1.3. Short tandem repeats (STRs)

A molecular marker is a gene or noncoding region of the genome that allows the detection of variations or polymorphisms among individuals in a population (Khlestkina, 2014). In forensics analysis, one type of molecular marker – the so-called short tandem repeats (STRs) – has become the standard due to its highly polymorphic nature and ability to produce reliable and robust results in a variety of biological materials, even for samples with a small DNA amount. Also, genotyping methods and nomenclatures are well standardized which allows the comparison of results from different laboratories.

STRs, also known as microsatellites or simple sequence repeats (SSRs), are DNA regions with repeating units that are 2 to 6 base pairs (bp) in length (Butler, 2004) and segregate according to Mendelian principles. These markers are present on all chromosomes and constitute about 3% of the human genome (Lander et al., 2001), appearing on average every 10,000 nucleotides (Subramanian et al., 2003).

STRs were first described as powerful tools for DNA typing in the early 1990s (Edwards et al., 1991) and have since then become popular markers due to their simple amplification using PCR, easy identification, and power of discrimination (Gill, 2002; Butler, 2004). The relatively short PCR product sizes generated in STR amplification and the fact that PCR is able to recover information from small amounts of genetic material make STR typing compatible with the analysis of degraded DNA (Butler, 2007).

Amplifying multiple STR loci at the same time is called multiplexing and it enables a high power of discrimination in a single analysis without consuming too much DNA. Commercial multiplex kits are widely available nowadays, being preferred in most laboratories to standardize procedures between different working groups (Butler, 2007).

The number of repeats in STR markers has a high degree of variation among individuals, making these markers effective for human (and non-human) identification and kinship analyses (Butler, 2004). Genetic variation in STR loci is characterized by high heterozygosity and multiple alleles (Ellegren, 2004).

STR loci can be found within the coding, genic, and extragenic regions of the genome (Edwards et al., 1991). While most STRs are not pathogenic, nearly 30 hereditary disorders have been linked to an increase in the number of copies of simple repeats in DNA, even when the repeat expansions are in the non-coding region of their resident genes (Mirkin, 2007). Both the repeat sequences and the proteins responsible for replication can play a role in copy number mutation (Richards & Sutherland, 1996).

1.3.1. X-chromosome STRs, kits, and population databases

As for all the other chromosomes, STRs can be observed at the X chromosome, being called X chromosome STRs (X-STRs). X-STRs were first reported in the 90s (Edwards et al., 1991), but the bloom in their usage only happened in the 2000s, (I. Gomes et al., 2020).

X-STRs are powerful in forensic genetics, especially when using haplotypes of closely linked markers (Szibor, 2007), which can be more discriminatory than Y chromosome STRs (Y-STRs) for some populations (Curtis et al., 2022). X-STRs can complement autosomal markers in paternity and complex kinship testing (Hering et al., 2015), whose success would depend on the amplification of a high number of autosomal markers (Szibor et al., 2005).

Knowledge of the allele and haplotype population frequencies, as well as of the mutation and recombination rates, is required to consider X-STRs both in routine and research settings. Not much research has been made on the mutation rates of the more commonly used X-STRs, and more studies on mutation rates, haplotype frequency, and linkage disequilibrium (LD) are needed to establish representative human reference databases for X-STRs (I. Gomes et al., 2020). X-STRs using short amplicons are a good option for DNA typing, especially if the samples used are difficult to analyze (D.-P. Chen et al., 2009).

The characterization of the different populations across the globe is far from being similarly distributed. Indeed, there is still X-STR data missing for many countries and some geographical areas, while there is a large quantity of X-STR data in some populations such as China (I. Gomes et al., 2020). Characterizing different regions or subpopulations of the same country is also relevant to studying population stratification, especially in the case of a high diversity of ethnicities (I. Gomes et al., 2020).

The main obstacles in X-STRs analyses lie in the hybrid (haplodiploid) nature of its genetic model of inheritance and in the practically unavoidable need to consider linked markers, i.e. markers that are not transmitted independently from one generation to the next (I. Gomes et al., 2020). Some issues that are still neglected in the research of X-STRs include accuracy in sequence variation, repeat structure and nomenclature (I. Gomes et al., 2020).

The X-chromosomal markers still have a lot of unused potential, and as such not many X-STR multiplex kits have become commercially available and validated. Two of the most used X-STR kits are the Investigator® Argus X-12 QS kit (QIAGEN, Germany) and

the GHEP-ISFG Decaplex (Gusmão et al., 2009). The Investigator® Argus X-12 QS is a multiplex kit containing primers for 12 X-STRs, intended for use in forensic casework, human identification, and kinship testing (*Investigator® Argus X-12 QS Handbook*, 2022). The GHEP-ISFG Decaplex (Gusmão et al., 2009) contains 10 X-STRs, is a non-commercial product, and has proven to be informative and useful in forensic practice (Gusmão et al., 2009). Both are highly sensitive and discriminatory multiplexes.

Currently, there is not any up-to-date X-STR database. There was an attempt to create a database for X-STR population data (http://www.chrx-str.org/), containing data on X chromosome markers used for forensic purposes and genetic research (Szibor et al., 2006) that lacks, however, updated data. This online database summarized some STR data regarding location, repeat motifs, allele nomenclature, and mutation rates. However, this database has not been very successful as it stopped being updated and even its availability has been intermittent. There is also the NIST STRBase (currently at http://strbase.nist.gov/), a general database for the forensic DNA typing community that includes X-STR information. This database consolidates the information available on STR markers, such as observed alleles and sequences, along with STR analysis technologies, primer sequences, and validation studies (Ruitberg et al., 2001). In any case, a centralized, curated, and updated database gathering X-STRs data is still lacking.

1.3.1.1. X-STR: DXS10146

DXS10146, first described by Edelmann and co-workers in 2008, is an STR marker with a complex structure located on the long arm of the X chromosome, more precisely at the Xq28 region (Edelmann et al., 2008). The repetitive motif was first described as $(TTCC)_x$-T-$(TTCC)_4$-TCCCTTCC-$(TCCC)_2$-TTCTTCTTTC-$(TTCC)_2$-TTTCTT-$(CTTT)_y$-T-$(CTTT)_2$, in which $(TTCC)_x$ and $(CTTT)_y$ are the common variable repeat blocks. In some alleles the $(CTTT)_y$ repeat block may contain a 2 bp insertion (TT or CT) (Edelmann et al., 2008). Several alleles with the same length but different structure – the so-called isoalleles, have been observed in this marker (Edelmann et al., 2008).

One of the most widely used multiplex kits contains the DXS10146 marker: the Investigator® Argus X-12 QS kit (QIAGEN, Germany). This X-STR has also been included in other in-house multiplexes (e.g., Huang et al., 2015; Prieto-Fernández et al., 2016; Deng et al., 2017).

The mutation rate (considering the rate between the number of mendelian incompatibilities detected and the number of allele transfers) for this locus has been estimated to be around $2.2x10^{-2}$ when analyzing father-daughter duos and around $6x10^{-3}$ in trios with a daughter (I. Gomes et al., 2020). LD has been observed between DXS10146 and other markers of its linkage group (LG4), as expected (Pasino et al., 2011; Samejima et al., 2012; Crnjac et al., 2016; I. Gomes et al., 2017; Robino et al., 2018; Mršić et al., 2018; Veselinović et al., 2018; Flores-Espinoza et al., 2021), but also with markers from other linkage groups such as DXS10135 (LG1) (Xing et al., 2019), DXS8378 (LG1) (Flores-Espinoza et al., 2021) and DXS10079 (LG2) (Bini et al., 2021; Flores-Espinoza et al., 2021).

This marker is highly polymorphic and informative in many populations (e.g., Sim et al., 2010; Bentayebi et al., 2012; Elakkary et al., 2014; Gomes et al., 2017; Messaoudi et al., 2021), therefore has the potential to solve complex kinship cases (Edelmann et al., 2008). In 2020, DXS10146 had been considered in at least 52 published articles containing human population data for X-STRs, concerning 64 populations with a varied geographical distribution (I. Gomes et al., 2020), and even more have been published since then (e.g., Bini et al., 2021; Ferragut et al., 2021; Flores-Espinoza et al., 2021; Bottinelli et al., 2022).

As is typical for highly polymorphic markers, several off-ladder alleles have been reported (Czarnogórska et al., 2010; Zhang et al., 2012; Bottinelli et al., 2022). One of these examples, the non-consensus allele 39.1 (Diegoli et al., 2011), has been referred to in the Argus X-12 handbook. This allele shows slightly different electrophoretic mobility compared to the adjacent allele 39.2 in the provided allelic ladder, which leads to a size difference of less than 1 bp between the two variants (*Investigator® Argus X-12 QS Handbook*, 2022).

Variation has also been found in the flanking regions of the core repetitive motif (Sim et al., 2010; Nagai & Bunai, 2011; Nagai et al., 2013; Pinto et al., 2020). The deletion of some insertion-deletion (INDELs) in the 3' flanking region is associated with six repeats in the variable $(TTCC)_x$ repeat block (Nagai et al., 2013). High frequency of null alleles was reported for African populations (Tomas et al., 2012; Elakkary et al., 2014; Afonso Costa et al., 2014; Kling et al., 2014; Gomes et al., 2017; Bini et al., 2021; Haddish et al., 2022), that when sequenced revealed that the non-amplification was due to one or more nucleotide variations in the primer binding sites (Tomas et al., 2012; Elakkary et al., 2014; Gomes et al., 2017; Haddish et al., 2022).

1.3.2. Y-chromosome STRs, kits, and population databases

Y-chromosome-specific STRs can be important markers in forensic genetics. Those are inherited through any male lineage and as such are not subjected to recombination, being transmitted an entire haplotype as a single marker and not allowing for individual identification (Gusmão et al., 2006). Besides being highly informative and easy to handle, Y-STRs fulfill the need for short amplification products and efficient sex determination of samples (Roewer & Epplen, 1992). These markers have a moderate number of polymorphic loci (von Wurmb-Schwark et al., 2003). The most polymorphic loci, associated with a high mutation rate (the so-called rapidly mutating or RM Y-STRs) have more usage in lineage identification and may help distinguish between male individuals, as RM markers may hamper the statistical analyses in paternity and kinship cases (Kayser, 2017; Syndercombe Court, 2021).

Some of the most used multiplex kits for Y-STR typing are different versions of the Yfiler™ (Applied Biosystems, USA) and PowerPlex® (Promega, USA) systems, in particular the AmpFLSTR™ Yfiler™ (Mulero et al., 2006), the Yfiler™ Plus (Gopinath et al., 2016), the PowerPlex® Y (Krenke et al., 2005) and the PowerPlex® Y23 (Thompson et al., 2013). These contain primers for the simultaneous genotyping of 17, 27, 12, and 23 Y-STRs, respectively. All of these have shown to be reliable and robust multiplexes suitable for genotyping forensic DNA samples, and the considered markers are included in the Y-Chromosome Haplotype Reference Database (YHRD).

YHRD (http://yhrd.org/) is a database built by submissions of population data from individual laboratories. This database contains information on frequency estimates for Y-STR and Y-SNP haplotypes useful for calculating matches in forensic and kinship cases and drawing conclusions about the history of human populations (Roewer et al., 2001). Such as for X-STRs, NIST STRBase (http://strbase.nist.gov/) also includes information on Y-STRs.

1.3.3. Mutations

The occurrence of mutations is the major source of genetic variation and its understanding is fundamental to study evolution and map genetic diseases. Mutational events can play a substantial role in parentage testing, kinship analyses, and victim identifications (Butler, 2006) and be helpful in the development of improved reagents for STR typing, such as better-designed primers (Heinrich et al., 2005; Gettings et al., 2015).

While the average mutation rate per nucleotide site has been estimated to be about $2.5x10^{-8}$ (Nachman & Crowell, 2000; Sun et al., 2012), the STRs have a much higher mutation rate estimate, from 10^{-3} to 10^{-4} (Weber & Wong, 1993; Ballantyne et al., 2010; Sun et al., 2012). The mutation rates differ significantly among loci and alleles, and, as such, it is difficult to estimate a uniform mutation rate for all STRs (Ellegren, 2004; Ballantyne et al., 2010). No differences have been observed in the rates of mutations between autosomal and heterosomal STRs (Ellegren, 2004; Szibor, 2007).

Recent studies have reported the most polymorphic type of microsatellite to be tetranucleotides (Sun et al., 2012; Willems et al., 2016), contradicting the previous belief that for loci without disease implications, the mutation rate seemed to be inversely related to the repeat lengths (Chakraborty et al., 1997). The usage of different methodologies in the study of mutation in STRs may show different results, demonstrating the obstacles of comparing results obtained from various data sets and different strategies in data collection.

More mutations have been found in males than in females, and it has been generally accepted that most of the variation at STR loci can be assumed to be caused by males (Weber & Wong, 1993; Brinkmann et al., 1998; Lander et al., 2001; Sun et al., 2012; García et al., 2017; Pinto et al., 2020; Syndercombe Court, 2021). Nevertheless, it should be noticed that the probability of maternal mutations not being detected – the so-called hidden mutations, is higher than for the paternal ones when X-STRs are analyzed (Antão-Sousa et al., 2022b). In mutation studies, the use of duos instead of trios increases the possibility of some mutations not being identified (Pinto et al., 2020), which likelihood is also dependent on the population analyzed (Antão-Sousa et al., 2022a).

Variation in microsatellites is mostly found in the length variation due to the number of repeats (Ellegren, 2004; Ballantyne et al., 2010), but also in sequence variations due to point mutations (Brinkmann et al., 1998). Length polymorphisms have been attributed to frequent slippage by DNA polymerase during replication (Levinson & Gutman, 1987; Lander et al., 2001; Ellegren, 2004), and an increase in repeat number has shown to be more common than loss of repeats (Weber & Wong, 1993). Long alleles are also most susceptible to mutation events (Brinkmann et al., 1998; Ellegren, 2004).

Sequence variation on repeat core and flanking regions can have an impact on the true allele nomenclature (Guo, 2017) and, in some cases, might contain more ancestry information than length-based alleles, as well as enable better resolution of cases (Gettings et al., 2015).

Mutations in the repetitive motif can lead to rare alleles, which differ from common allele variants in tested STRs. Mutations can be either insertions, deletions, or single nucleotide substitutions (Butler, 2004). The occurrence of rare alleles can lead to misjudgments in testing, highlighting the need for representative data collection to generate more appropriate ladders and to improve the knowledge of loci that might be used for individual identification and kinship testing (Yang et al., 2015).

Alleles that have some sequence variation are usually called microvariants. Some of these microvariants are different from the consensus alleles in the supplied allelic ladder used to calibrate PCR product sizes, and as such can be referred to as off-ladder (OL) alleles (Butler, 2004). In paternity testing, misreading OL alleles can lead to false conclusions and if an OL allele is found among alleles of other neighboring loci in electropherograms, they may cause false genotyping at both loci (W. Chen et al., 2014). Variant alleles are valuable tools and can strengthen the genetic evidence in troublesome parentage cases (Allor et al., 2005).

In the Y chromosome, a class of STRs known as RM STRs have higher mutation rates than other Y-STRs (Ballantyne et al., 2010). This feature makes mutations more likely to occur in those loci and potentially allows forensic DNA analysis to differentiate between male relatives (Ballantyne et al., 2010; Claerhout et al., 2018).

Nucleotide variations in primer binding sites can potentially create null alleles, which may lead to false homozygous results and discordant genotypes (Boutrand et al., 2001; Lazaruk et al., 2001; Heinrich et al., 2004; Rolf et al., 2011).

While most polymorphisms tend to be related to the tandem repeats and not the flanking regions (Weber & Wong, 1993), flanking region INDELs may exist in some alleles and are important to acknowledge and study (Gettings et al., 2015; Wendt et al., 2017).

1.3.4. Applications of X- and Y-STRs in forensic genetics

Markers from the sex chromosomes are particularly efficient at solving deficiency kinship cases, in which a determined biological sample is unavailable and a close relative must be used for testing (Szibor, 2007). X- and Y-STRs have also helped to determine the authenticity of historical artifacts along with autosomal STRs and other anthropological identification methods (Piniewska et al., 2017).

Genetic markers on the X chromosome are mostly used to solve deficiency cases (Szibor et al., 2003; Schmidtke et al., 2004; Szibor, 2007; Chen et al., 2009; Pinto et al., 2011)

and, in some specific situations, can be greatly efficient and be used as the only way of inferring kinship, increasing significantly the power of exclusion of the set of markers used (Bobillo et al., 2008; Auler-Bittencourt et al., 2015; Tillmar et al., 2017).

The genetic information provided by X chromosome markers is useful as a complement to autosomal markers, being recommended both in simple relationships such as mother-son and father-daughter (Szibor et al., 2003, Szibor, 2007; Chen et al., 2009), and in some complex pedigrees that are not distinguishable by only using autosomal markers (Szibor, 2007; Bobillo et al., 2008; Chen et al., 2009; Cainé et al., 2011; Pinto et al., 2011; Gomes et al., 2012; Tillmar et al., 2017; Zhang, Yu, et al., 2021; Jusic et al., 2022). X chromosome markers might provide higher statistical power than autosomal ones in cases involving at least one female (Schmidtke et al., 2004; Gomes et al., 2012; Hering et al., 2015; Tillmar et al., 2017; Jusic et al., 2022) or in kinship analyses involving large and incomplete pedigrees since these can assign pedigree members over long distances (Chen et al., 2009; Hering et al., 2015). These markers can also be useful for solving incest cases involving a female child (Tillmar et al., 2017).

X-STRs have also been used in a clinical context to infer the parental origin of the extra X chromosome in patients with KS (Miki et al., 2017; Kim et al., 2019), even if only one parent is available for testing (Kim et al., 2019), and the parental origin of the single X chromosome in patients with TS (Haltrich et al., 2015; Zhang et al., 2021; Vrtel et al., 2022). In the forensic setting, X-STRs were also already used to confirm KS in a male victim of sexual assault (Spitzer et al., 2021).

In the analysis of cases involving mixture samples, while X-STRs are more efficient at identifying female traces than autosomal STRs, it is not advised to use X chromosomal markers to test male traces where there is female contamination (Szibor et al., 2003). Instead, Y-STRs are more advised in mixtures of female and male DNA, commonly seen in stains from cases of sexual assault (von Wurmb-Schwark et al., 2003; Kayser, 2017; Syndercombe Court, 2021), since these can detect small traces of male DNA on a strong female DNA background (Roewer, 2019; Syndercombe Court, 2021). Y-STRs may also aid in investigations even when there is DNA from two or more males (except for paternally related males) or if only mixed DNA (female and male) is available for testing and the male suspect is unknown (von Wurmb-Schwark et al., 2003).

In forensic investigations, the use of Y-STRs is advised when male DNA left by a culprit is expected but not detected by autosomal screening due to strong female DNA background (Roewer, 2019). Y-STR profiling is effective in exclusion cases and, like X-STR profiling, is particularly effective in deficiency cases, as long as any male paternally

related to the unavailable male can be analyzed (Kayser, 2017; Syndercombe Court, 2021). Haplotyping Y-STRs is suitable for solving paternity disputes and paternal kinship questions, especially when using loci with lower mutation rates (Butler, 2006; Kayser, 2017); however, the analysis of the Y chromosome can only be informative on lineages and not on individuals (Syndercombe Court, 2021). Therefore, it should be interpreted in context with other available information.

Since there is no recombination in the specific zone of the Y chromosome, its markers are suitable for inferring bio-geographical ancestry, aiding in male identification cases such as missing persons, disaster victim identification, and identification of bodies in mass graves (Kayser, 2017; Syndercombe Court, 2021). Y-STRs can help establish the ethnic group and region of the paternal ancestors as long as there is representative reference data (Roewer, 2019).

1.3.5. Sequencing methodologies in STR analysis

DNA sequencing refers to the process of uncovering the sequence of nucleic acids in DNA and includes any approach used to determine the nucleotide order.

When applied to STRs, sequencing is a useful approach for validating the STR loci chosen, establishing a nomenclature for new loci, or examining species specificity of allele sequences (Lazaruk et al., 2001). Sequencing STR alleles may also aid in resolving complex mixtures and kinship analyses (Gettings et al., 2015; Wendt et al., 2017), as well as in characterizing internal sequence variation in same-size alleles that cannot be distinguished simply by genotyping (Nagai & Bunai, 2011; Butler, 2015). Sequencing silent alleles, off-ladder alleles, and microvariants is pivotal to revealing molecular variation and allowing the design of alternative primers to investigate allele dropout and sequence analysis of primer-binding nucleotide variations (Vanderheyden et al., 2007; Kline et al., 2011; Gettings et al., 2015).

When sequencing STRs, the samples are first PCR-amplified with forward and reverse primers, ideally far away enough from the repetitive zone to identify variation in the flanking regions. The fragments obtained are then sequenced with a primer (not necessarily one of the primers used for PCR amplification), obtaining sequences that are then aligned and compared to a reference. In the case of the studied locus being heterozygous, the alleles need to be separated before sequencing (Kline et al., 2011).

In 1977 a simple and accurate method for determining nucleotide sequences in DNA was described (Sanger et al., 1977). This method has since been known as “Sanger

sequencing" and exploits the DNA polymerase-dependent synthesis of a complementary strand in the presence of 2'-deoxynucleotides (dNTPs) and fluorescently labeled (to allow for nucleotide detection) 2',3'-dideoxynucleotides (ddNTPs) serving as synthesis terminators (Sanger et al., 1977). The DNA synthesis reaction is terminated whenever a ddNTP is added to the oligonucleotide chain, resulting in shortened products of different lengths with a ddNTP at the end. Nowadays, the products of the reaction are submitted to capillary electrophoresis (CE) (Karger & Guttman, 2009), whereas before the products were separated using polyacrylamide gel electrophoresis (PAGE), and the terminal ddNTPs were used to reveal the DNA sequence (Gaastra, 1985).

Currently, Sanger sequencing and next-generation sequencing (NGS) are the most used methods for sequencing. While NGS has shown to be a cost-effective technique with many applications (Morozova & Marra, 2008; Dahui, 2019), it is challenging to genotype STRs using this method, due to difficulty in aligning STR variations to a reference sample, as well as to the possibility of stutter noise in the number of repeats and low number of informative reads (Gymrek, 2017). The Sanger sequencing method remains a widely used method for basic and clinical research purposes, including STR analyses.

2. Aims

The unique properties of STRs make them useful markers for various purposes, such as kinship investigations, forensic casework, or phylogenetic and population genetic studies. When studying STRs, it is fundamental to understand the mutational behavior as well as the structure of (new) detected allelic sequence variations. In particular, X- and Y-STRs are mainly used as a complement to autosomal markers and are not yet completely understood, so it becomes notably important to identify potential genetic variations at the DNA sequence level. This is also important to establish correct allele nomenclatures by comparing human STR sequence structures to their primate counterparts. Due to sequencing methods such as the traditional Sanger sequencing which enables the knowledge of the precise order of nucleic acids in a DNA sequence, it is possible to identify the existence of mutations and discover or confirm novel sequence variations.

The genetic characterization of a Portuguese population sample for X- and Y-STRs through fragment length determination methodologies lead to the observation of alleles and DNA profiles that carried doubts. These observations contributed to the main aim of this study which was the analysis of the DNA sequence structures of some of the X- and Y-STR loci included in commonly used identification kits. With this main aim defined, specific objectives were proposed for this work:

1) To analyze the DNA sequence structure of all of the alleles identified as 39.1 and 39.2 at the DXS10146 locus in a Portuguese population sample set by sequencing the repeat motif and flanking regions;
2) To confirm the presence of a tri-allelic pattern in an X-STR profile in a female individual;
3) To analyze the general structure and stability of the STR sequence of DXS10146 in humans by comparison with the ancestral genomes of chimpanzee, orangutan and gorilla and to confirm human specificity by checking amplification in common species;
4) To infer and contribute to a more accurate nomenclature for DXS10146 and this way support better communication, data exchange and data comparison among different laboratories;
5) To confirm *de novo* mutations in father-son duos by analyzing the DNA sequence of Y-STRs in these cases.

3. Materials and Methods

3.1. Samples

3.1.1. Human

3.1.1.1. X-chromosome analyses

For the amplification of the X-STR DXS10146, the human samples used were collected in the scope of a previous work (Faustino, 2021) with the purpose of characterizing the Portuguese population and studying microsatellites' mutations. Therefore, DNA extracts from father-mother-daughter trios were already available for use. Blood samples were collected using a sterile, single-use lancing device (Accu-Chek, Roche, Germany) from healthy volunteers, under informed consent following the procedure (Faustino, 2021). The DNA extracts were then amplified using the commercial multiplex kit Investigator® Argus X-12 QS (QIAGEN, Germany) (Faustino, 2021).

The chosen samples for studying sequence variation in DXS10146 had at least one allele identified previously as either 39.1 or 39.2, using electrophoresis methodology. As such, 31 samples from Portugal (26 females and 5 males; 31 alleles previously identified as either 39.1 or 39.2 and 24 other alleles) were used for DNA sequencing (Attachments – Supplementary Table 1).

A trio (father-mother-daughter constellation), in which the daughter exhibited a trisomic pattern on the X chromosome was also selected to confirm a case of Triple X. The proximity of the trio's alleles in certain loci would difficult the process of separating and sequencing the samples, and so the chosen method for confirming X trisomy was genotyping.

Even though most samples were not extracted in this study, the methodology used to extract the DNA from the blood human samples is provided. The chosen methodology (detailed in the work of Faustino, 2021) depended on the card type used for the blood stain collection and preservation and the cards' sample age and dryness level. Two different card types were used: 1) for the Whatman® FTA® cards, DNA was extracted using a standard Chelex method (Walsh et al., 1991) or a modified version of this protocol for dry and old bloodstains in the case of older samples (Walsh et al., 1991); 2) for the GenReleaz™ cards an Elute methodology (*Fast and Efficient Recovery of DNA*, 2018) was used.

In the trio used to confirm a trisomic pattern on the X chromosome, two of the samples were re-extracted from the original Whatman® FTA® cards using the Chelex modified

version for DNA extraction of dry bloodstains to eliminate the possibility of contamination that would be present if the samples used were already extracted. There were no doubts about the genetic profile of the other sample and as such that was not re-extracted.

3.1.1.2. Y-chromosome analyses

For the Y-STR mutation duos analysis, father-son duos were used. These samples were collected in the previously described study and with the same methodologies regarding the collection and DNA extraction (Faustino, 2021). These samples had been amplified using the PowerPlex® Y23 System (Promega, USA) (Faustino, 2021).

Eighteen father-son duos were selected in which a *de novo* mutation was observed in one of the following markers: DYS19, DYS385 a/b, DYS389 I, DYS389 II, DYS390, DYS456, DYS458, DYS481, DYS549, DYS576, DYS635, and Y-GATA-H4 (Attachments – Supplementary Table 2). In most of these duos, the observed mutations were compatible with changes in the number of repeats. The duo with mutations on the DYS385 a/b was ultimately not analyzed due to the nature of the marker, as it amplifies two regions of the Y chromosome which are very similar in size and therefore difficult to separate and extract from the gel in order to sequence.

3.1.2. Non-human (primates and others)

Non-human samples were also used for studying the X-STR DXS10146. These samples had been collected previously for other studies and as such were already extracted. Therefore, there was uncertainty in which specific species were used and so only the common name of the non-human samples is mentioned here.

For sequence variation testing, primate samples were used, which consisted of a female gorilla, a male orangutan, and four male chimpanzees.

For testing marker specificity in other species, samples from domestic animals such as mouse, cat, dog, bull, horse and pig were used. These samples were supplied in order to perform routine studies.

3.2. Amplifications

3.2.1. Singleplex

3.2.1.1. DXS10146

Amplification of the X-STR DXS10146 was done through a standard endpoint-PCR reaction. Both primers were adapted from previous studies: the forward primer from Moutinho (2022), and the reverse primer from a study by Gomes and colleagues (2017). Useful information such as primer sequences, fragment size and melting temperature is shown in Table 1. The amplicons of DXS10146 generated with the primers used (Table 1) are longer than the amplicons obtained with the Investigator® Argus X-12 QS (around 280 bp at max), ensuring that the flanking region could be sequenced as well.

Table 1. Primer sequences (5'-3') and additional information about the DXS10146 locus. The melting temperature and fragment size were obtained from UCSC In-Silico PCR, using the Dec 2013 GRCh38/hg38 genome assembly as a reference.

X-STR	Primer sequences (5'-3')	T_m (ºC)	Primer reference	Fragment size (bp)
DXS10146	F - GATCTGCCTTGCCCTTCCTA	61.6	Moutinho (2022)	327
	R - TGGGGTGACAGACTGAGAGA	59.4	Gomes et al. (2017)	

T_m = melting temperature (annealing temperature)

Initially, a test was performed with few samples using the PCR conditions presented in Table 2 and including both negative and positive controls to test for possible contamination and to see if the reaction was carried out, respectively. Primer stocks were used at 5µM in a final 10µL PCR reaction volume, therefore having a final concentration of 0.4µM for each PCR reaction.

Table 2. Reagents used in the amplification of samples using DXS10146.

Reagents	Volumes (µL)
Master Mix*	5
Primer forward	0.8
Primer reverse	0.8
RNAse-free water	0.9
Template DNA	2.5
Total	**10**

* HotStarTaq Master Mix or Taq PCR Master Mix (QIAGEN, Germany).

Two different Master Mixes* were tested, one being the HotStarTaq Master Mix (QIAGEN, Germany) and the other being Taq PCR Master Mix (QIAGEN, Germany). The thermocycling conditions used are listed in Table 3.

Table 3. Thermocycling conditions for amplifying the DXS10146 loci.

Phase	Temperature (ºC)	Time	Cycles
Initial denaturation	95	15 minutes	-
Denaturation	94	30 seconds	30x
Annealing	60	90 seconds	
Extension	72	60 seconds	
Final extension	72	30 minutes	-

PCR products were visualized in a polyacrylamide gel electrophoresis (PAGE) and silver stained (procedure described in 3.5 section). The HotStarTaq Master Mix (QIAGEN, Germany) ended up performing better and, as such, was used for the rest of the amplifications.

Since some non-specific bands were detected, a Touchdown (TD) PCR using the same reagents but different thermocycling conditions (Table 4), was performed and compared with the standard PCR. The principle in TD-PCR is to begin with an annealing temperature above the projected melting temperature and transition to a lower temperature over 10 to 15 cycles, which is expected to produce more specific results compared to a standard PCR (Don et al., 1991). The standard PCR yielded stronger results, and as such was the chosen method.

Table 4. Thermocycling conditions of TD-PCR.

Phase	Temperature (ºC)	Time	Cycles
Initial denaturation	95	15 minutes	-
Denaturation	95	30 seconds	10x
Annealing	70	45 seconds	
Extension	72	60 seconds	
Denaturation	95	30 seconds	20x
Annealing	60	45 seconds	
Extension	72	60 seconds	
Final extension	72	30 minutes	-

3.2.1.2. Y-STRs

A review of the literature (Attachments – Supplementary Table 3) was made to find suitable primers in order to perform PCR amplification of the Y-STRs where *de novo* mutations were detected. While some primer pairs were adapted from the literature,

others had to be newly designed. Two primer pairs were already available in the lab, for the markers DYS458 and Y-GATA-H4, which were tested with a 10ng/µL positive 2800M Control DNA (Promega, USA) and a negative control before amplifying the samples.

A primer pair for the marker DYS456 was also available but was unsuitable since one of the primer sequences encapsulated part of the repeat motif, which made it impossible to sequence the entirety of the repeat. Additionally, the difference between the annealing temperature of both primers was 11.1ºC and the forward primer was not specific.

The newly designed primers were designed in Primer-BLAST to assure specificity (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). When that was not possible, the pairs were generated using Primer3 software (Untergasser et al., 2012) and the specificity was then confirmed in the BLAST web-based tool. The genome assembly used as reference was Dec 2013 GRCh38/hg38.

The selected primer pair sequences (Attachments – Supplementary Figure 1) followed certain general preferential primer selection criteria: not a big difference in melting temperatures among pairs, a fragment size as small as possible and a low score (assessed by Primer3 software) to amplify other regions of the genome. The adapted primer pairs that were described on the R strand were converted to the F strand.

The information about each primer, described in Table 5, was obtained from UCSC In-Silico PCR (http://genome.ucsc.edu). Before usage in PCR, the lyophilized primers were centrifuged, mixed with RNAse-free water, vortexed and incubated at 56ºC for 1 minute.

Table 5. Information about the used Y-STR primers: oligo sequences, melting temperatures (T_m), fragment sizes and respective reference. The DYS389 I/II primer pair amplifies two PCR products, in which the smaller one is contained within the sequence of the bigger one. Dec 2013 GRCh38/hg38 genome assembly.

Y-STR	Primer sequences (5'-3')	T_m (ºC)	Fragment size (bp)	Reference
DYS19	F - GAAGACAAGGACTCAGGAATTTG	59.3	380	This study
	R - TTGACAAGCCCAAAGTTCTTAACAT	62.3		
DYS389 I/II	F - CCTGAGTAGCAGAAGAATGTCATA	57.7	268 and 148	Bosch et al. (2002)
	R - CCAACTCTCATCTGTATTATCTATGT	55.1		
DYS390	F – GTATCCGCCATGGTAGCATAAT	60.1	259	This study
	R – TTTACACATTTTTGGGCCCTGC	65.1		
DYS456	F – TCTGTTGTGGGACCTTGTGA	60.1	154	This study
	R – TCAACTCAGCCCAAAACTTCT	59.0		
DYS458	F – GCAACAGGAATGAAACTCCAAT	60.4	144	Already available
	R - GTTCTGGCATTACAAGCATGAG	59.8		
DYS481	F – AGGAATGTGGCTAACGCTGT	59.8	128	Kayser et al. (2004)
	R – ACAGCTCACCAGAAGGTTGC	60.5		
DYS549	F - GTAGGTAAAGAGGAAGATGATAGATG	56.1	181	Núñez et al. (2017)
	R - GGTGTTCAGAATAGTCTCTAAAGGTTT	58.9		
DYS576	F – TCTCAGCCAAGCAACATAGC	59.2	146	Yuan et al. (2019)
	R - AGGAGATGGGAGTAATAAGCGT	58.3		
DYS635	F – GCAGCAAAATTCACAGTTGGA	61.2	214	This study
	R – TGCCCAATGGAATGCTCTCT	63.0		
Y-GATA-H4	F - GTTATGCTGAGGAGAATTTCCAA	59.6	284	Already available
	R - CCTCTGATGGTGAAGTAATGGAATTAGA	63.7		

For the amplification of the father-son samples to be submitted to Sanger sequencing, the same Master Mix and conditions described in Tables 2 and 3 were used adapting only the annealing temperatures of the different primer sequences. The annealing temperatures used were 58ºC (DYS389 I/II), 60ºC (DYS456, DYS458, DYS481, DYS549, DYS576), 61ºC (Y-GATA-H4) and 62ºC (DYS19, DYS390, DYS635). The primers available in the lab were first tested at 57ºC, but since some unspecific bands appeared, the annealing temperature was raised to 60ºC for DYS458 and 61ºC for Y-GATA-H4.

Some samples showed low yield and were repeated with different volumes, as described in Table 6, and were tested with the previous thermocycling conditions, with 30 and 32 cycles. The used primers were at 5μM, so in a final 10μL PCR reaction volume, their final concentration was 0.4μM for each PCR reaction.

Table 6. Reagents used in a PCR used to amplify samples with very faint bands in a polyacrylamide gel.

Reagents	Volumes (µL)
Master Mix	5
Primer forward	0.8
Primer reverse	0.8
Template DNA	3.4
Total	**10**

3.2.2. Multiplex

In order to genotype the trio in which the daughter potentially has Trisomy X, two multiplexes were used: the investigator® Argus X-12 kit (QIAGEN, Germany) and the GHEP-ISFG Decaplex (Gusmão et al., 2009).

3.2.2.1. Investigator® Argus X-12 kit

The version of the Investigator® Argus X-12 kit (QIAGEN, Germany) from 2013 was used due to availability of stock in the lab. In this version, Reaction Mix A and the Multi Taq2 are separate, contrary to the most recent versions in which the Taq polymerase is already in the reaction mix included with the kit. Relevant information about the X-STRs in this kit is described in Table 7.

Table 7. Information about the X-STRs in Investigator® Argus X-12 kit (QIAGEN, Germany), taken from NIST STRBase (http://strbase.nist.gov/). The genome assembly used as reference is the GRCh38.p12.

Marker	Linkage group	Cytogenetic localization	Physical localization (Mb)
DXS10148	I	Xp 22.31	9.271
DXS10135	I	Xp 22.31	9.338
DXS8378	I	Xp 22.31	9.402
DXS7132	II	Xq 12	65.435
DXS10079	II	Xq 12	67.496
DXS10074	II	Xq 12	67.757
DXS10103	III	Xq 26.2	134.285
HPRTB	III	Xq 26.2	134.481
DXS10101	III	Xq 26.3	134.520
DXS10146	IV	Xq 28	150.404
DXS10134	IV	Xq 28	150.482
DXS7423	IV	Xq 28	150.542

The initial conditions in which the samples were amplified were taken directly from the recommendations in the Investigator® Argus X-12 handbook (*Investigator® Argus X-12 Handbook*, 2013) and are described in Tables 8 and 9.

Table 8. Initial volumes of reagents used for the amplification of samples using the Investigator® Argus X-12 kit (QIAGEN, Germany).

Reagents	Volumes (µL)
Reaction Mix A	2
Primer Mix	1
Multi Taq2	0.24
RNAse-free water	5.76
Template DNA	1
Total	**10**

Table 9. Initial thermocycling conditions used for the amplification of samples using the Investigator® Argus X-12 kit (QIAGEN, Germany).

Phase	Temperature (ºC)	Time	Cycles
Initial denaturation	94	4 minutes	-
Denaturation	96	30 seconds	5x
Annealing	63	120 seconds	
Extension	72	75 seconds	
Denaturation	94	30 seconds	25x
Annealing	60	120 seconds	
Extension	72	75 seconds	
Final extension	68	60 minutes	-

In this first amplification, double peaks appeared, and as such the PCR conditions were slightly modified (Tables 10 and 11).

Table 10. Reagent and volumes used in the PCR for amplification of samples using the Investigator® Argus X-12 kit (QIAGEN, Germany).

Reagents	Volumes (µL)
Reaction Mix A	1.5
Primer Mix	0.75
Multi Taq2	0.18
RNAse-free water	3.945
Template DNA	1.125
Total	**7.5**

Table 11. Thermocycling conditions used for PCR for amplification of samples with the Investigator® Argus X-12 kit (QIAGEN, Germany).

Phase	Temperature (ºC)	Time	Cycles
Initial denaturation	94	4 minutes	-
Denaturation	96	30 seconds	5x
Annealing	63	120 seconds	
Extension	72	75 seconds	
Denaturation	94	30 seconds	27x
Annealing	60	120 seconds	
Extension	72	75 seconds	
Final extension	72	60 minutes	-

In order to test the quality of the Taq polymerase due to PCR failure, a PCR was done using a Multiplex PCR Master Mix (QIAGEN, Germany) instead of the Reaction Mix A and Multi Taq2, with quantities and thermocycling conditions adapted from the PCR Master Mix (QIAGEN, Germany) handbook (*QIAGEN® Multiplex PCR Handbook*, 2010) and described in Tables 12 and 13.

Table 12. Reagents used in a slightly modified PCR to test the quality of the Taq polymerase.

Reagents	Volumes (µL)
Master Mix	5
Primer Mix	1
RNAse-free water	3
Template DNA	1
Total	**10**

Table 13. Thermocycling conditions of a PCR to test the quality of the Taq polymerase.

Phase	Temperature (ºC)	Time	Cycles
Initial denaturation	95	15 minutes	-
Denaturation	94	30 seconds	5x
Annealing	63	90 seconds	
Extension	72	90 seconds	
Denaturation	94	30 seconds	25x
Annealing	60	90 seconds	
Extension	72	90 seconds	
Final extension	72	30 minutes	-

Some split peaks were observed in the samples. One of the causes could be a high quantity of DNA (“overblown” samples), since the DNA polymerase is unable to complete

the extension for all amplicons. This means that the adenine addition on all PCR products results in what is usually referred to as split peaks (+A/-A peaks – a peak that contains that adenosine and a peak that does not). To test if this was the case, a PCR was performed with less volume of DNA than the others (Table 14) and with the same thermocycling conditions described in Table 9.

Table 14. Thermocycling conditions used in a PCR with the goal of eliminating split peaks.

Reagents	Volumes (µL)
Reaction Mix A	1.5
Primer Mix	0.75
Multi Taq2	0.18
RNAse-free water	4.695
Template DNA	0.375
Total	**7.5**

3.2.2.2. GHEP-ISFG Decaplex

The GHEP-ISFG Decaplex (Gusmão et al., 2009) amplifies ten X-STR markers (described in Table 15), three of which are common to the Investigator® Argus X-12 kit (QIAGEN, Germany).

Table 15. Information about the X-STRs in the GHEP-ISFG Decaplex (Gusmão et al., 2009). The information was taken from ChrX-STR (http://chrx-str.org/) using the NCBI36 genome assembly, except for the markers DXS8378, DXS7132 and DXS7423, taken from NIST STRBase (http://strbase.nist.gov/) which uses the GRCh38.p12 genome assembly as reference. These three markers are common to both multiplexes and as such are marked with *.

Marker	Cytogenetic localization	Physical localization (Mb)
DXS8378*	Xp 22.31	9.402
DXS9902	Xp 22.2	15.234
DXS7132*	Xq 12	65.435
DXS9898	Xq 21.31	87.682
DXS6809	Xq 21.33	94.825
DXS6789	Xq 21.33	95.336
DXS7133	Xq 22.3	108.928
GATA172D05	Xq 23	113.061
GATA31E08	Xq 27.1	140.062
DXS7423*	Xq 28	150.542

The conditions of the PCR amplification (Tables 12 and 16) were adapted from a previous study using this multiplex (C. Gomes, 2019). The HotStarTaq Master Mix

(QIAGEN, Germany) and the Decaplex primer mix which were available in the lab were tested.

Table 16. Thermocycling conditions for amplifying the X-STR markers included in the GHEP-ISFG Decaplex (Gusmão et al., 2009).

Phase	Temperature (ºC)	Time	Cycles
Initial denaturation	95	15 minutes	-
Denaturation	94	30 seconds	10x
Annealing	60	90 seconds	
Extension	72	60 seconds	
Denaturation	94	30 seconds	20x
Annealing	58	90 seconds	
Extension	72	60 seconds	
Final extension	60	75 minutes	-

3.3. Genotyping

In order to genotype the samples using the Investigator® Argus X-12 kit (QIAGEN, Germany), 9.6µL of Hi-Di formamide and 0.4µL of DNA Size Standard 550 (BTO) (QIAGEN, Germany) were added to each PCR product. When using the GHEP-ISFG Decaplex (Gusmão et al., 2009), 9.5µL of Hi-Di formamide and 1.5µL of GeneScan™ 500 LIZ™ Size Standard (Applied Biosystems, Thermo Fisher Scientific, USA) were mixed with the PCR product before genotyping.

Samples were separated and detected using capillary electrophoresis (CE), in which the gel and sample move inside a capillary and each fragment is identified by a detector. This step was performed in an ABI 3500 Genetic Analyzer (Applied Biosystems, Thermo Fisher Scientific, USA) at the Genomics i3S Scientific Platform, available in-house.

The fragment sizes were determined and observed using the GeneMapper™ Software 5 (Applied Biosystems, Thermo Fisher Scientific, USA) and the recommended panels and bins were used for the Investigator X-STR kit.

3.4. Polyacrylamide Gel Electrophoresis (PAGE)

In order to confirm the amplification of the amplicons, polyacrylamide gel electrophoresis was performed.

The acrylamide solution combined 25mL of Resolving Gel Buffer 1.5M Tris-HCl, pH8.8 (Bio-Rad, USA), 20mL of AccuGel 40% Acrylamide/Bis Solution, 19:1 (National Diagnostics, USA), 43mL of RNAse-free water and 7mL of glycerol. The gel solution for a small gel frame, around 5x10 cm, was prepared in a fume hood using 3mL of

acrylamide solution, 170µL of APS, and 7µL of TEMED. These quantities were doubled when using a medium gel frame (around 11x10 cm) and tripled for a large gel frame (around 23x11 cm). This solution was then poured out into a frame where it was left to polymerize. GEL-FIX™ for Polyacrylamide Gels (SERVA Electrophoresis, Germany) was used as a supporting film for the casting of the polyacrylamide gel.

After polymerizing and removing the gel from the frame, 1.3µL of each sample was loaded into the wells or 3µL when using the biggest gel frame. Two ladders were tested, one being the GRS Ladder 100bp (GRiSP, Portugal) in a 1:4 dilution and the other being O'GeneRuler 100bp DNA Ladder, ready-to-use (Thermo Fisher Scientific, USA), the latter in both a 1:1 and a 1:4 dilution.

The O'GeneRuler ladder (Thermo Fisher Scientific, USA) was not observed in any polyacrylamide gel, therefore only the GRS Ladder 100bp (GRiSP, Portugal) was used in the following tests.

The electrophoresis was then performed using a Multitemp III thermostatic circulator (Amersham Biosciences, UK) at 4ºC, a consort EV 243 power supply and a 0.125M Tris/Glycine, pH8.8 gel running buffer. Different voltages were used depending on the purpose of the gel, ranging from 90V to 195V. The gel ran at 45Ma and 20W.

PAGE was also tested with a tighter mesh gel (12% polyacrylamide) in order to observe better resolution of bands with small differences in length. However, this gel did not separate the bands so only the looser mesh was used.

The gel was stained using the silver staining method to reveal amplified products. First, the gel was covered with ethanol (10%) for 10 minutes, followed by nitric acid (1%) for 5 minutes while in a shaker, doing a quick wash with distilled water after. Next, the gel was moved to a dark recipient and put in silver nitrate (0.2%) for 20 minutes, in a shaker, and then washed with distilled water. A sodium carbonate 0.28M and formaldehyde 0.02% solution is then prepared by mixing 3g of sodium carbonate, 100mL of distilled water, and 1mL of formaldehyde and poured over the gel a little at a time in order to visualize the amplified fragments. These quantities were doubled for a medium gel frame and tripled for a large gel frame. The gel then stays in acetic acid (10%) for 2 minutes to stop the reaction. The final step is to leave the gel covered in distilled water overnight and let it dry the next day.

3.5. Band extraction from polyacrylamide gel

For sequencing of alleles that were not homozygous, separation of bands was attempted as following: the bands were extracted from the polyacrylamide gel using a sanitized

scalpel, with all the due caution inherent to the process to avoid cross-contamination. For the X-STR DXS10146, both the putative 39.2 band and the other band were extracted. For Y-STR DYS389 I/II, the larger band was extracted since it encapsulates the sequence of the other band. The fragments were then placed in 50µL TE buffer to aid the extraction of DNA from the band. The solution was then heated to 56ºC for 15 minutes and snap-cooled in the freezer for 30 minutes. This step was repeated two more times. The final step was an amplification through PCR. In this PCR (described in Table 17) only 1µL of template DNA was used since it had already been subjected to amplification once before; as such, there was no need to use as much volume of DNA as before. However, if the bands were weak the volumes of the reagents were the same as in Table 2. The used primers were at 5µM.

Table 17. Reagents used in the reamplification of the extracted bands retrieved from the polyacrylamide gel.

Reagents	Volumes (µL)
Master Mix	5
Primer forward	0.8
Primer reverse	0.8
RNAse-free water	2.4
Template DNA	1
Total	**10**

The thermocycling conditions of the PCR were the ones described in Table 3, with the melting temperatures adapted to each primer pair used. In some cases, 32 cycles were used, if 30 cycles were not enough to amplify a substantial quantity of product.

3.6. Sanger sequencing

The PCR products were submitted to a traditional Sanger DNA sequencing protocol. The complete sequencing methodology, including purification of PCR and sequencing products, was performed at the Genomics i3S Scientific Platform. Sequencing was performed using the BigDye Terminator v3.1 Cycle Sequencing kit following the manufacturer's recommendations (Applied Biosystems, Thermo Fisher Scientific, USA), and using the genetic analyzers ABI 3130xl Genetic Analyzer (Applied Biosystems, Thermo Fisher Scientific, USA) or ABI 3500 Genetic Analyzer (Applied Biosystems, Thermo Fisher Scientific, USA).

The sequencing results were observed in Sequencing Analysis Software v6.0 (Applied Biosystems, Thermo Fisher Scientific, USA). When applicable, the sequences were aligned using EMBL-EBI EMBOSS Needle (Madeira et al., 2022) for pairwise alignments and T-Coffee (Notredame et al., 2000) for multiple sequence alignments, following the ClustalW method.

4. Results and Discussion

4.1. X-chromosome analyses

4.1.1. DNA sequence structure of alleles 39.1 and 39.2 at the DXS10146 locus in a Portuguese population sample

Due to a high observation of 39.1 and 39.2 alleles in a previous work, aimed at studying the polymorphisms and mutation rates of X- and Y-STRs in a Portuguese population sample (Faustino, 2021), it was important to study the DNA sequence of these alleles. The main aim was to confirm if the genotyping results obtained by fragment sizes were correct (both 39.1 and 39.2 alleles are present in the Portuguese population) or if these are the same allele but with different electrophoretic mobility due to possible different allele/STR structures.

Of the 55 selected alleles, only 22 were sequenced (Attachments – Supplementary Table 1) due to various technical problems (Figure 1): either a. the bands were very close on the polyacrylamide gel and consequently very difficult to separate, or b. unspecific bands appeared on the reamplification, or c. the samples did not amplify at all. Of the 22 sequenced alleles, 14 had been assessed either as 39.1 or 39.2 (8 refer to the other alleles). Only 5 of these 14 alleles were assessed exclusively as 39.2 (Attachments – Supplementary Table 1), while the other 9 caused assessment doubts between 39.1 and 39.2.

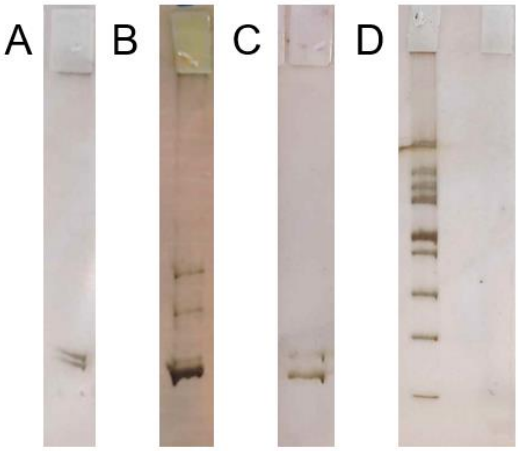


Figure 1. Examples of some technical problems that hindered the sequencing of the selected alleles in some samples, in which: the alleles had similar sizes and the bands on the gel were too close to separate (A), showed unspecific bands on the reamplification of the allele (B), the separation was not successful and both the alleles showed on the gel (C), there were no bands in the sample well (D), where the first lane is the 100 bp ladder.

DXS10146 was described as having a compound structure containing two variable repeat blocks: $(TTCC)_x$ and $(CTTT)_y$, and 10 non-variable repeats of 4 bp (Edelmann et al., 2008). This X-STR shows more variation in longer alleles, particularly in the length of the first variable repeat block and a 2 bp insertion (TT or CT) in the second repeat block (Edelmann et al., 2008). In the mentioned article the nomenclature established for this marker is $(TTCC)_x$-T-$(TTCC)_4$-TCCCTTCC-$(TCCC)_2$-TTCTTCTTTC-$(TTCC)_2$-TTTCTT-$(CTTT)_y$-T-$(CTTT)_2$. This is the most used nomenclature for DXS10146 and will be the one used in this study for analyzing variation and comparing the human and primate genomes. The expected amplified sequence using the reference genome sequence for primer design is shown in Figure 2 as an example.

GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCT**TTCCTTCCTTCCTTTCCTTCCTTCCTTC**
CTCCCTTCCTCCCTCCCTTCTTCTTTCTTCCTTCCTTTCTTCTTTCTTTCTTTCTTTCTTTCTT
TCTTTCTTTCTTTCTTTCTTTCTTTCTTTTCTTTCTTTCTCTGTCTTTCTTTCTTTTTCTTTCTTCTT
TCTTTTTCTTTCTTCTTTCTTTCCTTTCTTTCTTCTTTCTTTCTTTCCTTCTCTTTCTCTCTCTCT
CTCTCTTTCTTTCTTTCTTCCTTTCTTCCTTTCTTTCTCTCAGTCTGTCACCCCA

Figure 2. Sequence amplified with the primer pair used for DXS10146. This corresponds to the positions chrX:150,403,779-150,404,105 on the reverse complement of the GRCh38.p14 genome assembly. The forward and reverse primers are underlined, and the repeat motif is in bold.

In the first approach, the samples were sequenced with both the forward and reverse primers. The forward primer, however, showed a lot of background and, as such, only the reverse primer was used to sequence the following samples as it showed the best results. A comparison of the two can be observed in Figure 3. In some samples, the background made it impossible to read the flanking sequence and in one case the whole sample, which was not included in further analyses.

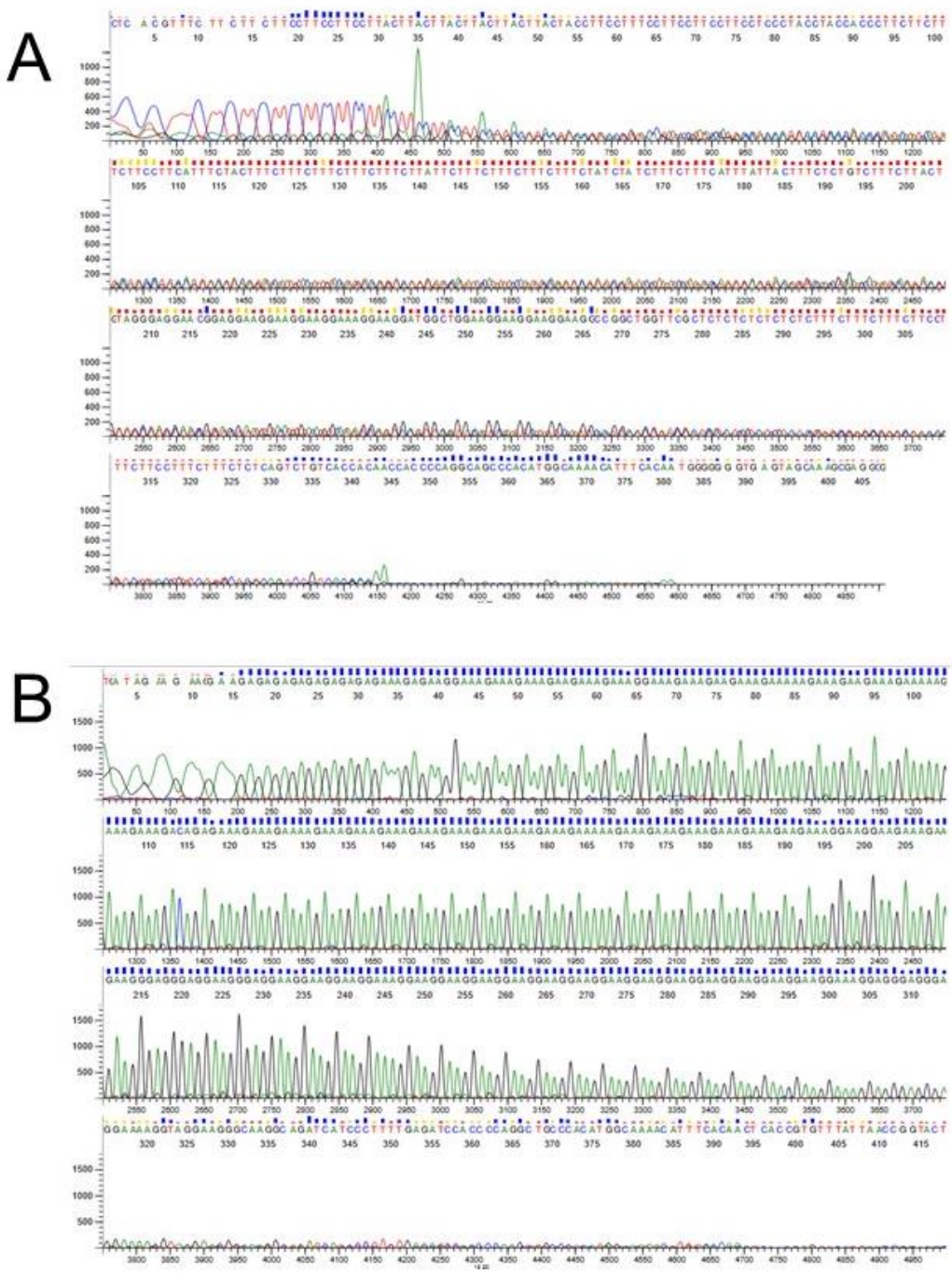


Figure 3. A sample amplified with the forward (A) and the reverse primer (B). In general, the reverse primer showed stronger and clearer results.

A gel with a tighter mesh (12% polyacrylamide) was tested. This gel, however, did not separate the bands well – these were very close to each other, and smearing could be observed. As such, only the looser, standard mesh gel was used.

Besides variation in the number of repeats in the $(TTCC)_x$ and $(CTTT)_y$ repeat blocks, no other variation was found in the core repeat motif. The sequenced alleles followed the established and most used nomenclature (Tables 18 and 19).

Table 18. Repeat motif of the sequenced alleles. N refers to the number of observations. The variable repeat blocks are in bold. The genotyping column refers to the previous genotyping assessments (Attachments – Supplementary Table 1).

Genotyping	Allele seq	Repeat motif	N
25	25	**$(TTCC)_3$**-T-$(TTCC)_4$-TCCCTTCC-$(TCCC)_2$-TTCTTCTTTC-$(TTCC)_2$-TTTCTT-**$(CTTT)_{12}$**-T-$(CTTT)_2$	1
26	26	**$(TTCC)_3$**-T-$(TTCC)_4$-TCCCTTCC-$(TCCC)_2$-TTCTTCTTTC-$(TTCC)_2$-TTTCTT-**$(CTTT)_{13}$**-T-$(CTTT)_2$	1
27	27	**$(TTCC)_3$**-T-$(TTCC)_4$-TCCCTTCC-$(TCCC)_2$-TTCTTCTTTC-$(TTCC)_2$-TTTCTT-**$(CTTT)_{14}$**-T-$(CTTT)_2$	2
28	28	**$(TTCC)_3$**-T-$(TTCC)_4$-TCCCTTCC-$(TCCC)_2$-TTCTTCTTTC-$(TTCC)_2$-TTTCTT-**$(CTTT)_{15}$**-T-$(CTTT)_2$	2
29	29	**$(TTCC)_3$**-T-$(TTCC)_4$-TCCCTTCC-$(TCCC)_2$-TTCTTCTTTC-$(TTCC)_2$-TTTCTT-**$(CTTT)_{16}$**-T-$(CTTT)_2$	2
39.1 or 39.2	39.2	**$(TTCC)_{11}$**-T-$(TTCC)_4$-TCCCTTCC-$(TCCC)_2$-TTCTTCTTTC-$(TTCC)_2$-TTTCTT-**$(CTTT)_7$**TT**$(CTTT)_{11}$**-T-$(CTTT)_2$	9
39.2	39.2	**$(TTCC)_{12}$**-T-$(TTCC)_4$-TCCCTTCC-$(TCCC)_2$-TTCTTCTTTC-$(TTCC)_2$-TTTCTT-**$(CTTT)_7$**TT**$(CTTT)_{10}$**-T-$(CTTT)_2$	1
39.2	39.2	(**$(TTCC)_{13}$**-T-$(TTCC)_4$-TCCCTTCC-$(TCCC)_2$-TTCTTCTTTC-$(TTCC)_2$-TTTCTT-**$(CTTT)_6$**TT**$(CTTT)_{10}$**-T-$(CTTT)_2$	1
39.2	39.2	**$(TTCC)_{14}$**-T-$(TTCC)_4$-TCCCTTCC-$(TCCC)_2$-TTCTTCTTTC-$(TTCC)_2$-TTTCTT-**$(CTTT)_6$**TT**$(CTTT)_9$**-T-$(CTTT)_2$	3

seq=sequenced

Table 19. Sequenced alleles and number of repeats on each variable repeat block. N refers to the number of observations.

Allele	$(TTCC)_x$ block	$(CTTT)_y$ block	N
25	$(TTCC)_3$	$(CTTT)_{12}$	1
26	$(TTCC)_3$	$(CTTT)_{13}$	1
27	$(TTCC)_3$	$(CTTT)_{14}$	2
28	$(TTCC)_3$	$(CTTT)_{15}$	2
29	$(TTCC)_3$	$(CTTT)_{16}$	2
*39.2	$(TTCC)_{11}$	$(CTTT)_7TT(CTTT)_{11}$	9
39.2	$(TTCC)_{12}$	$(CTTT)_7TT(CTTT)_{10}$	1
39.2	$(TTCC)_{13}$	$(CTTT)_6TT(CTTT)_{10}$	1
39.2	$(TTCC)_{14}$	$(CTTT)_6TT(CTTT)_9$	3

The same allelic structure (*allele 39.2, Table 19) was found between mother and daughter in two mother-daughter duos, showing no mutation.

However, a variation in the flanking regions was found. In two samples with alleles 25 and 26, a duplication of a 17 bp INDEL 35 bp downstream of the core repeat unit, previously described (Sim et al., 2010), was found (Figure 4). The existence of two copies of this INDEL had been previously observed in Caucasian and African populations (Nagai et al., 2013).

```
Reference CTCTGTCTTTCTTTCTTTTTCTTTCTTCTTTCTTT-----------------TTCTTTCT
INDEL     CTCTGTCTTTCTTTCTTTTTCTTTCTTCTTTCTTTTTCTTTCTTCTTTCTTTTTCTTTCT

Reference TCTTTCTTTCCTTTCTTTCTTCTTTCTTTCTTTCCTTCTCTTTCTCTCTCTCTCTCTC
INDEL     TCTTTCTTTCCTTTCTTTCTTCTTTCTTTCTTTCCTTCTCTTTCTCTCTCTCTCTCTC

Reference TTTCTTTCTTTCTTCCTTTCTTCCTTTCTT
INDEL     TTTCTTTCTTTCTTCCTTTCTTCCTTTCTT
```

Figure 4. 3' flanking region of the two samples that contained a duplication of a 17 bp insertion. The reference shown corresponds to the GRCh38.p14 genome assembly.

An additional TC insertion was observed 86 bp downstream of the core repeat unit in a 39.2 allele (Figure 5). To the best of our knowledge, this insertion has not been described previously.

```
Reference CTCTGTCTTTCTTTCTTTTTCTTTCTTCTTTCTTTTTCTTTCTTCTTTCTTTCCTTTCTT
INDEL     CTCTGTCTTTCTTTCTTTTTCTTTCTTCTTTCTTTTTCTTTCTTCTTTCTTTCCTTTCTT

Reference TCTTCTTTCTTTCTTTCCTTCTCTT--TCTCTCTCTCTCTCTCTTTCTTTCTTTCTTC
INDEL     TCTTCTTTCTTTCTTTCCTTCTCTTTCTCTCTCTCTCTCTCTCTTTCTTTCTTTCTTC

Reference CTTTCTTCCTTTCTT
INDEL     CTTTCTTCCTTTCTT
```

Figure 5. Sequence of an allele with a possibly novel INDEL on the DXS10146 3' flanking region. The region shown corresponds to the positions chrX:150,403,799-150,403,931 on the GRCh38.p14 genome assembly.

In the more recent version of the commercial kit Investigator® Argus X-12 QS handbook (QIAGEN, Germany), a putative OL allele is mentioned for the marker DXS10146 (*Investigator® Argus X-12 QS Handbook*, 2022). This allele, 39.1, is described as showing slightly different electrophoretic mobility compared to adjacent allele of the ladder 39.2. In this study, alleles that had been previously assessed as 39.1 or 39.2 were

sequenced in order to discover if the different electrophoretic mobility was due to a sequence variation. While no sequence variation was observed in the core repeat unit, the sequenced alleles that were previously assessed as 39.1 were all in fact 39.2 alleles in which the $(TTCC)_x$ block had 11 repeats and the $(CTTT)_y$ block had a complex pattern of 18 repeats and 2 extra bps. The alleles that had previously been assessed as 39.2 all had different patterns on these blocks, observed in Table 19. It would be necessary to sequence more samples, from unrelated individuals and from different populations to confirm if the different electrophoretic mobility is caused (or correlated) by this allelic structure or if it is a coincidence.

The sequence structure of the samples analyzed in this study was consistent with the findings of another study (Moutinho, 2022), in which sequence data of Portuguese samples were compared with African one. The sequence structure of these populations was in accordance with the established nomenclature. The referenced study shows the difficulty of sequencing STRs and highlights the lack of sequencing data for commonly used X-STRs.

The frequency of 39.2 alleles in the Portuguese population (Faustino, 2021) was compared with other worldwide populations (Attachments – Supplementary Table 4). The comparison of the frequency data for these populations is shown in Figure 6. The frequency of these alleles in Portuguese samples is similar to other European populations such as Lithuania, Spain (Southeastern), Switzerland, Croatia and Czech Republic, as well as African populations such as Ethiopia and Cape Verde. This review of population data also highlights the need for an updated database of X-STR markers in order to standardize data across different working groups.

To our knowledge, only one study separated the 39.1 allele from the 39.2 one in frequency data (Diegoli et al., 2011). In this study, 39.1 was only present in African Americans, with a frequency of 0.8% of all the African American alleles. The 39.2 was present in all studied populations except U.S. Asian, with a frequency of 1.1%, 3.9% and 2.9%, in African American, U.S. Caucasian and U.S. Hispanic populations.

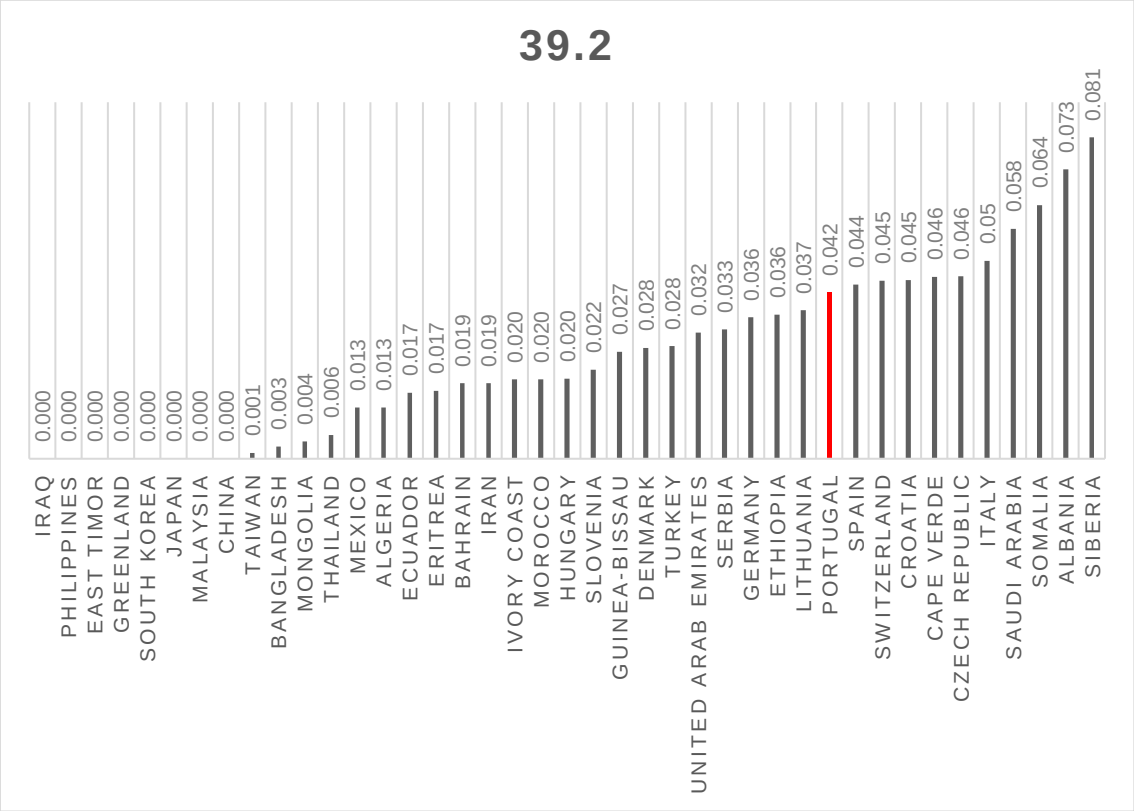


Figure 6. Comparison of the frequency of the 39.2 allele for several worldwide populations. The frequency data of Portuguese samples is highlighted in red. When more than one article was available for the same population, only one was chosen.

Since this allele is frequent worldwide, it becomes even more important to sequence it in various populations to confirm the cause of the different electrophoretic mobility, improving the sequencing data available for X-STRs.

4.1.2. An X chromosome trisomy case

In this study, a trio in which the daughter exhibited a putative trisomic pattern on the X chromosome was genotyped. The samples of this family were not sequenced for X-STRs, due to the proximity of the alleles in certain loci, which would difficult the process of extracting and sequencing each allele. As such, the chosen method for confirming X trisomy was genotyping.

In the first genotyping, incomplete adenylation could be observed (Figures 7 to 9). However, it was still possible to confirm the trisomy and obtain genotypes for all samples (Table 20). In the daughter's genotype, when only two alleles were present, one of the peaks was much larger than the other. In these cases, it was assumed that one of the alleles was present twice.

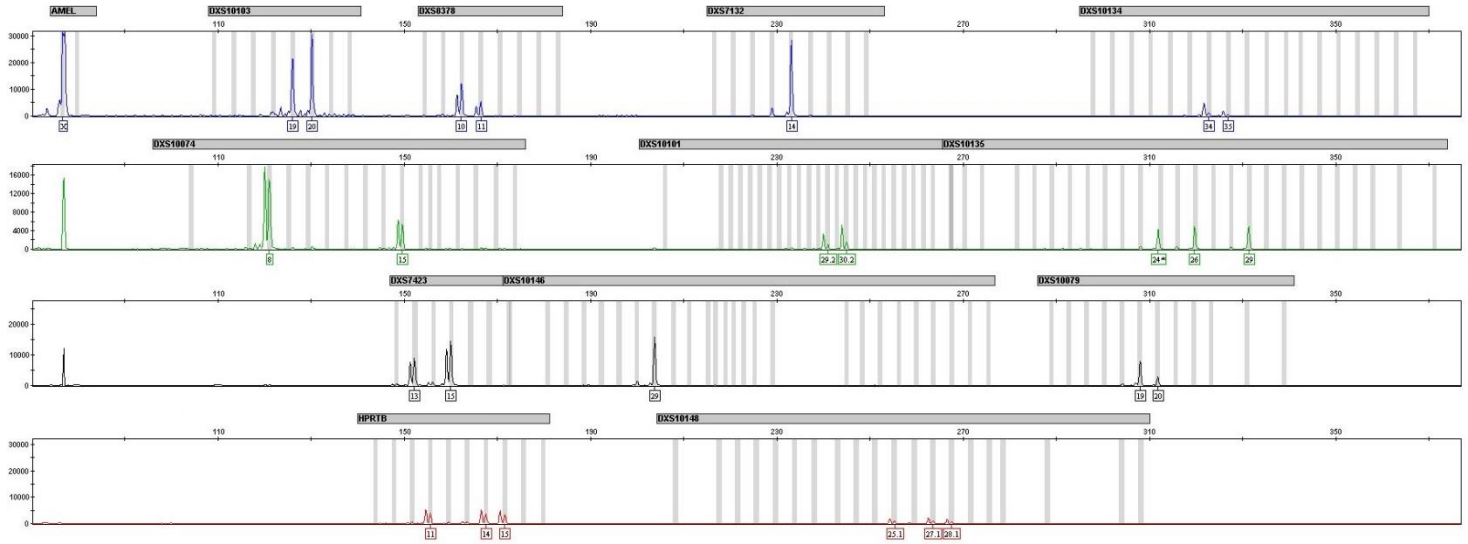


Figure 7. Electropherogram of the genetic profile of the daughter obtained from amplifying the sample with Investigator® Argus X-12 (QIAGEN, Germany). Some split peaks can be observed.

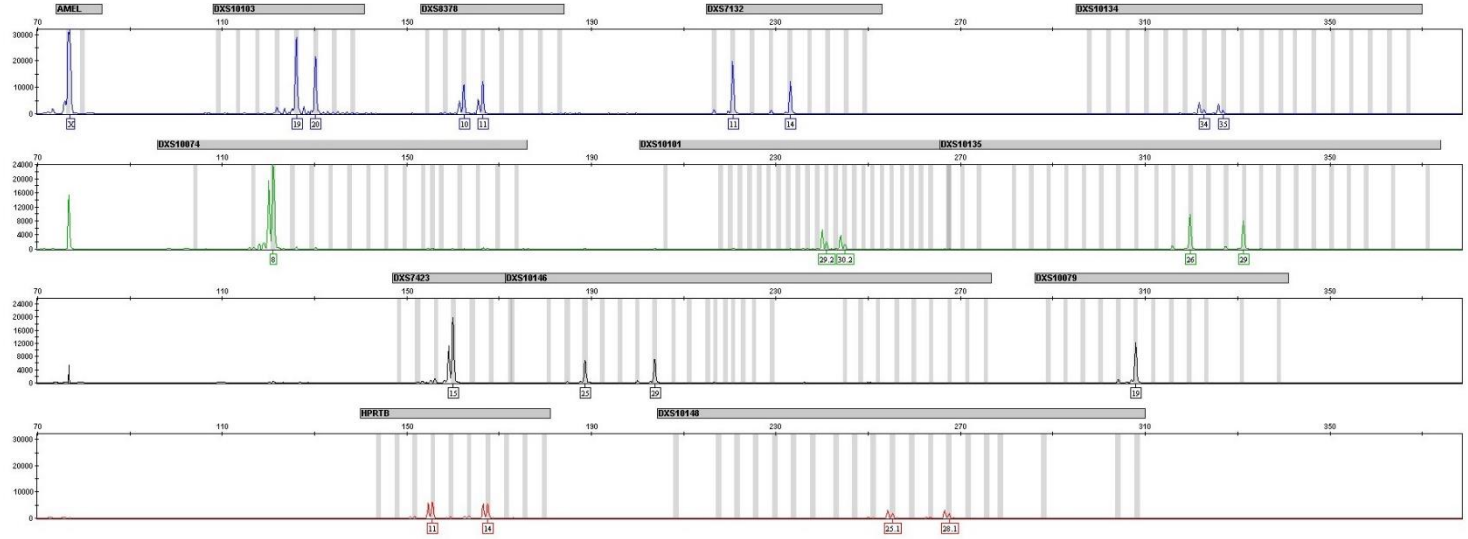


Figure 8. Electropherogram of the genetic profile of the mother obtained from amplifying the sample with Investigator® Argus X-12 (QIAGEN, Germany). Some split peaks can be observed.

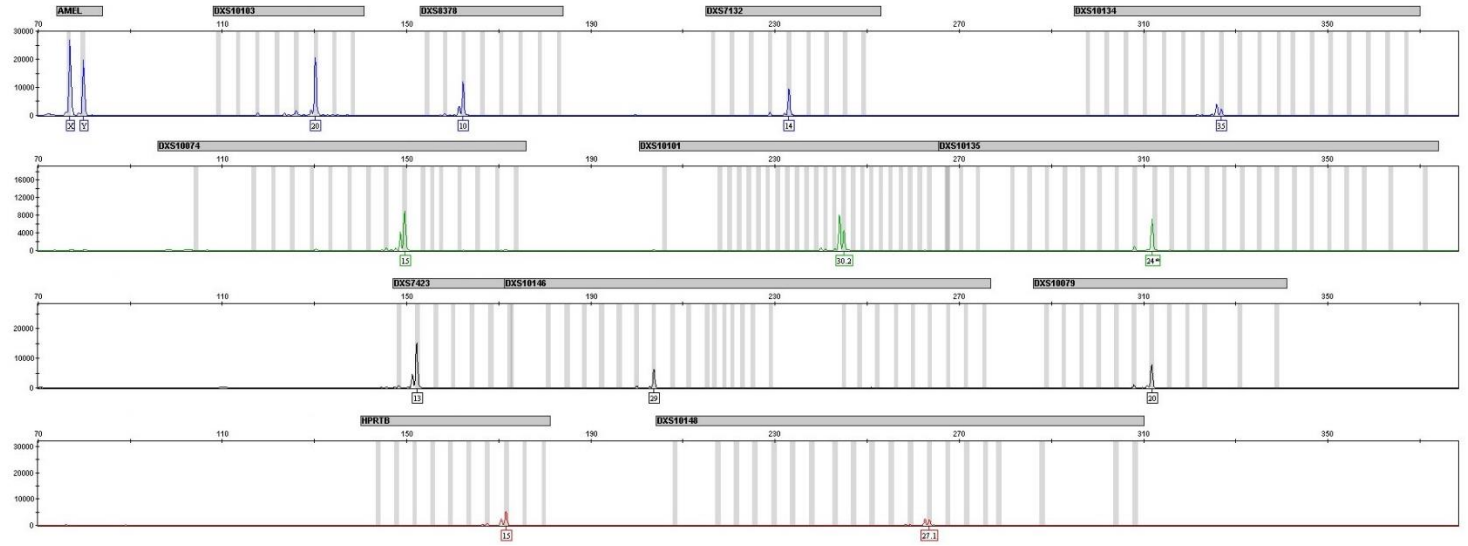


Figure 9. Electropherogram of the genetic profile of the father obtained from amplifying the sample with Investigator® Argus X-12 (QIAGEN, Germany). Some split peaks can be observed.

Table 20. Genotypes obtained with the Investigator® Argus X-12 kit (QIAGEN, Germany). The allele marked as 24* in DXS10135 was an OL.

X-STR	Daughter	Mother	Father
Amelogenin	XXX	XX	XY
DXS10103	19,20,20	19,20	20
DXS8378	10,10,11	10,11	10
DXS7132	14,14,14	11,14	14
DXS10134	34,34,35	34,35	35
DXS10074	8,8,15	8,8	15
DXS10101	29.2,30.2,30.2	29.2,30.2	30.2
DXS10135	24*,26,29	26,29	24*
DXS7423	13,15,15	15,15	13
DXS10146	29,29,29	25,29	29
DXS10079	19,19,20	19,19	20
HPRTB	11,14,15	11,14	15
DXS10148	25.1,27.1,28.1	25.1,28.1	27.1

When repeating the process with different conditions in order to remove the split peaks, the electropherograms obtained started to show unexpected peaks (Figure 10). A PCR using a Master Mix that already included Taq polymerase was performed, in order to test if the results obtained were due to Taq2 polymerase degradation. The results, however, were still similar to the previous ones and unexpected peaks could still be observed. Different reagent quantities were also used, but the same results were obtained.

In order to confirm that these results were not due to degradation of the Hi-Di formamide in the fridge, a genotyping was done immediately after mixing the Hi-Di formamide and size standard with the samples. The fact that the same previous results were obtained was evidence that the cause of the unexpected peaks could have been a degradation of the primer mix.

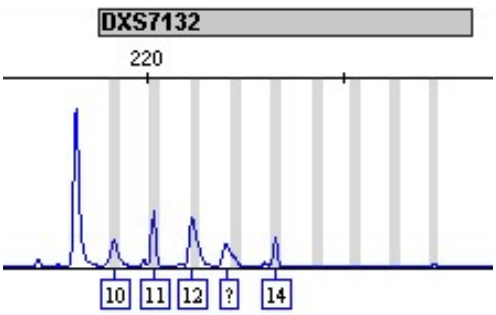


Figure 10. Example of the unexpected peaks observed in the samples.

The other multiplex used for genotyping this family was the GHEP-ISFG Decaplex (Gusmão et al., 2009) (Figures 11 to 13). This genotyping was successful, confirming the X trisomy with another multiplex and obtaining genotypes for all samples (Table 21). For one of the markers (DXS7133) no peaks were observed. The three markers common to both kits (DXS8378, DXS7132 and DXS7423) showed the same results.

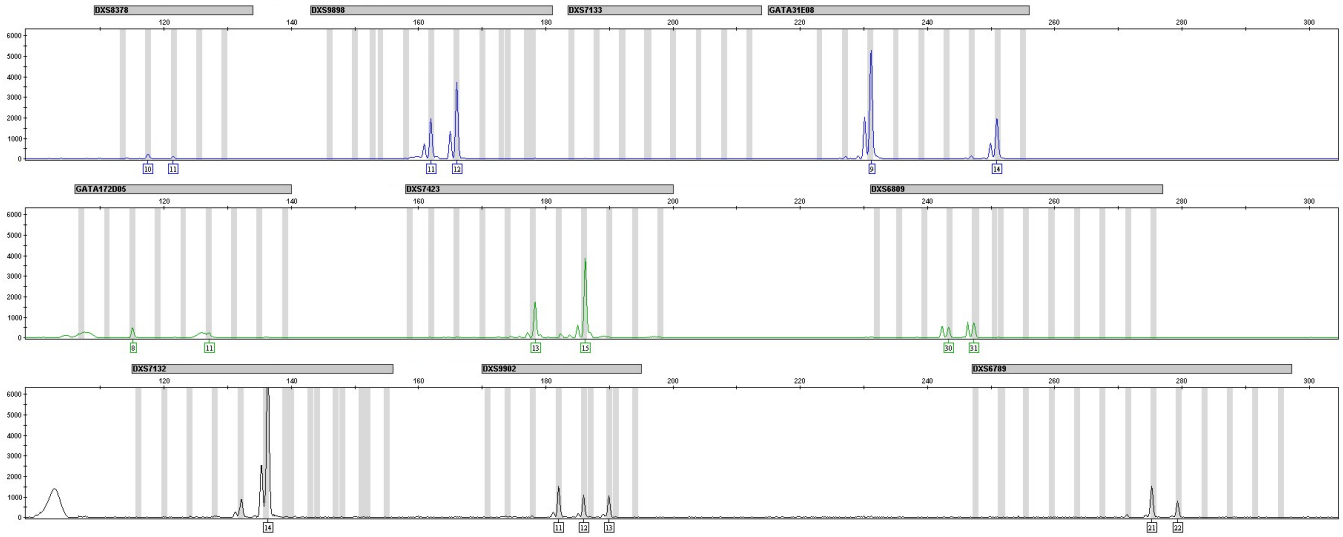


Figure 11. Electropherogram of the genetic profile of the daughter obtained from amplifying the sample with the GHEP-ISFG Decaplex (Gusmão et al., 2009). Some split peaks can be observed.

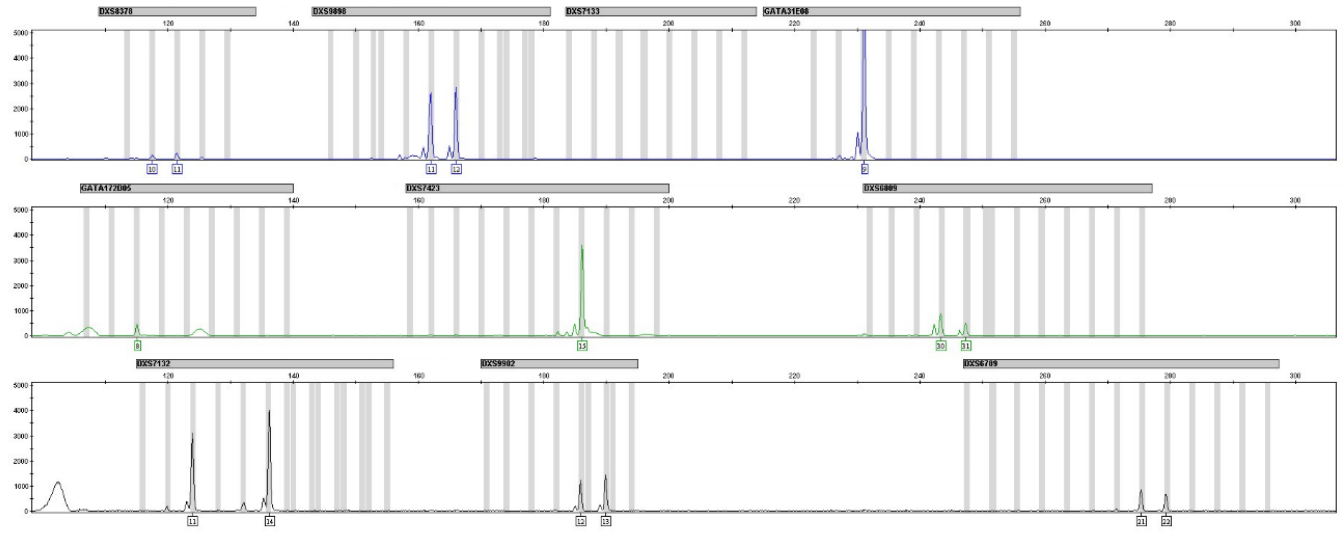


Figure 12. Electropherogram of the genetic profile of the mother obtained from amplifying the sample with the GHEP-ISFG Decaplex (Gusmão et al., 2009).

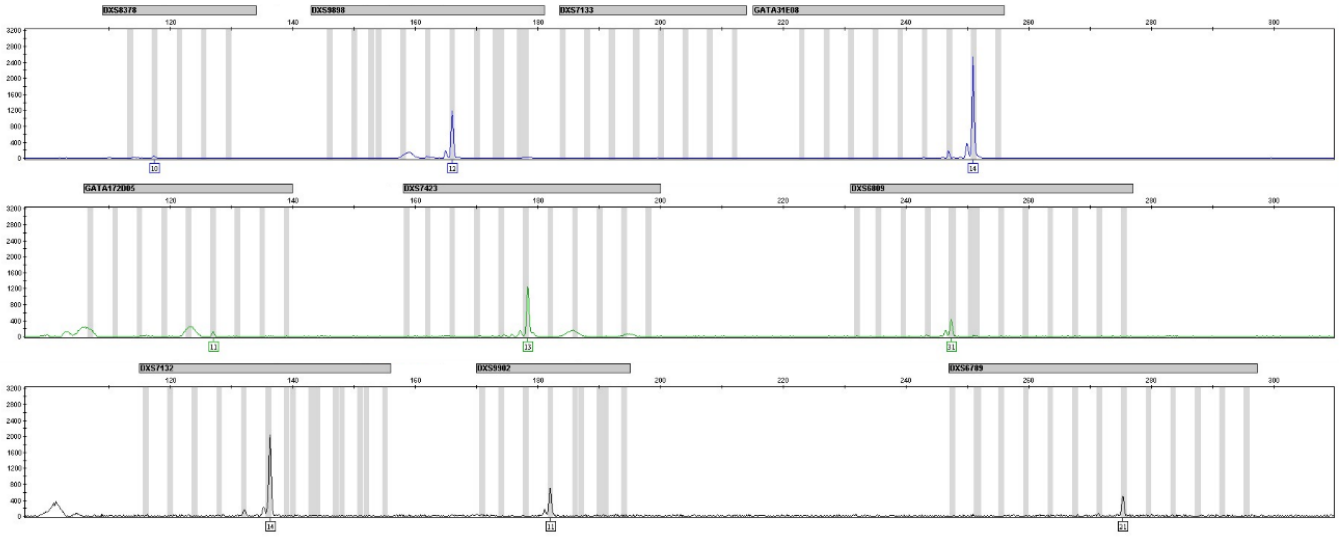


Figure 13. Electropherogram of the genetic profile of the father obtained from amplifying the sample with the GHEP-ISFG Decaplex (Gusmão et al., 2009).

Table 21. Genotypes obtained with the GHEP-ISFG Decaplex (Gusmão et al., 2009). *Markers also included in the Investigator® Argus X-12 kit.

X-STR	Daughter	Mother	Father
DXS8378*	10,10,11	10,11	10
DXS9898	11,12,12	11,12	12
DXS7133	-	-	-
GATA31E08	9,9,14	9,9	14
GATA172D05	8,8,11	8,8	11
DXS7423*	13,15,15	15,15	13
DXS6809	30,31,31	30,31	31
DXS7132*	14,14,14	11,14	14
DXS9902	11,12,13	12,13	11
DXS6789	21,21,22	21,22	21

The results seem to suggest that maternal nondisjunction occurred, as the daughter's profile shares the same alleles as the X chromosome of the father and both X chromosomes of the mother, except for three markers (DXS7132, DXS10134 and DXS10146).

4.1.3. Comparison of the DXS10146 locus with primates (STR structure inference) and other non-human samples (specificity analysis)

Six primate samples were amplified and sequenced successfully with DXS10146 (Figure 14). These samples consisted of a female gorilla, a male orangutan, and four chimpanzees, and the same primers used for humans were able to successfully amplify the corresponding DNA.

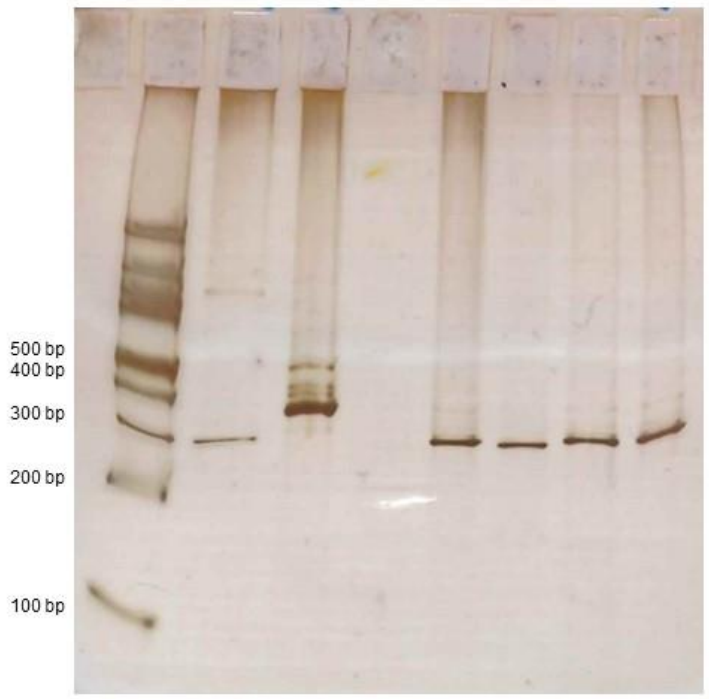


Figure 14. Polyacrylamide gel showing the successful amplification of the primate samples. From left to right: ladder, female gorilla, male orangutan, negative control, chimpanzee 108, chimpanzee 110, chimpanzee 121, chimpanzee 123.

An alignment of the reference sequences of DXS10146 locus in orangutans, gorillas, chimpanzees, and humans reveals a fairly conserved sequence (Figure 15). Some large deletions can be observed by comparing the orangutan sequence to the other primates. Some base substitutions can also be observed (primarily transitions from C to T and vice-versa), as well as some small insertions. One of the base pairs of the reverse primer does not match the reference sequence of the gorilla. The gorilla sample, however, was still amplified normally.

```
Orangutan       GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCCTCCTTTCCTTCCTTCCTT-CCTTCTTTTCTTCC-TTCCTT
Gorilla         GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCT----CCCTCCTTTCCTTCCTTCCTT-CCTTCCTTTCTTCT-TTC---
Chimpanzee      GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCT----CCCTCCTTTCCTTCCTTCCTT-CCTTCCTTCCTTCCTTTC---
Human           GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCT----CCCTCCTTTCCTTCCTTCCTTTCCTTCCTTCCTTC--------
                ********************************    ********************* ***** ** ****

Orangutan       CCTCCCTTCCTCCCTCCCTTCTTTCTTCCTTCCTTCCTTTCTTCTT---TCTTTCTTTCTTTCTCTGTCTTTCTTTCTCT
Gorilla         -CTTCCTTCCTTCCTCCCTTC-------CTCCCTTCCTCCCTCCCTTCTTCTTTCTTCCTTCC------TTTATTTC---
Chimpanzee      -CTTCCTTCCTTCCTCCCTTC-------CTCCCTTCCTCCCTCCCT---TCTTTCTTCCTTCC------TTTCT-TC---
Human           -CTCCCTTCCTCCCTCCCTTCTT-CTTTCTTCCTTCCTTTCTTCTT---TCTTTCTTTCTTTC------TTTCTTTC---
                 ** ******* *********       ** *******  ** * *   ******** *** *      *** * **

Orangutan       TTCTCTTTCTT-TCTCTTTCTTCTTTCTTGCTTTCTTTTTCTTTCTTCTTTCTTTCTTCTTTCTTTCTTTCTTTCTT
Gorilla         -----TTTCTT-TTC-----TTCTT----------------------TCTTTCTTTC-------TTTC------TCTTTCT
Chimpanzee      -----TTTCTTTTTC-----TTCTT----------------------TCTTTCTTTC-------CTTC--------TTTCT
Human           -----TTTCTT-TCT-----TTCTT----------------------TCTTTCTTTC-------TTTCTT---TTCTTTCT
                     ****** *       *****                     **********       ***       *****

Orangutan       TTCTTTCTCTTTCTTTCTTTTTCTTTCTTCTTTCTTTCCTTTCTTTCTTCTTTCTTTCCTTTCTTTCTTCTTTCTTTCTT
Gorilla         TTCTT--TCTTTCTTTCTTTTTCTTTCTTCTTTCTTT-----CTTTC--CTTTCTTTCTTCTCTCT-TTCTTTC----TT
Chimpanzee      TTCTTTCTCTGTCTTTCTTTTTCTTTCTTCTTTCTTT---TTCTTTCCTCTTTCTTTCCTTTCTTTCTTCTTTCTTTCTT
Human           TTCTCTGTCTTTCTTTCTTTTTCTTTCTTCTTTCTTT---TTCTTTCTTCTTTCTTTCCTTTCTTTCTTCTTTCTTTCTT
                ****   *** *************************     *****  ********* * *** * *******    **

Orangutan       TCTCTCTCTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTGTCTTTCTTTCTCTCTT
Gorilla         TCTTTCC--TTCTCTTT----CTTTCTCTCTCTCTCTCTT----TCTTTCTTCCTTT--------------------CTT
Chimpanzee      TCTTTCC--TTCTCTTT----CTTTCTCTCTCTCTCTCTTTCTTTCTTTCTTTCTTC--------------------CTT
Human           TCCT------TCTCTTT----CTCTCTCTCTCTCTC----TCTTTCTTTCTTTCTTC--------------------CTT
                **        *******    ** *** *** ***         ******** ***                     ***

Orangutan       TCTCTCTTTCTTTCTCTCAGTCTGTCACCCCA
Gorilla         TCTTTCTTTCTTTTTCTCAGTCTGTCACCCCA
Chimpanzee      TCTTCCTTTCTTTCTCTCAGTCTGTCACCCCA
Human           TCTTCCTTTCTTTCTCTCAGTCTGTCACCCCA
                ***  ******** ******************
```

Figure 15. Alignment of the reference sequences of orangutan, gorilla, chimpanzee and human. These sequences are located in chrX:156868749-156869174, chrX:143993022-143993338, chrX:147910978-147911305 and chrX:150403779-150404105, respectively.

Comparing the orangutan and the human reference, some deletions and substitutions can be observed (Figure 16). Variation can be found in the T-$(TTCC)_4$ block – a TTC deletion and a T to C substitution seem to have occurred since the structure of this zone in orangutans is TTCTTTTC-$(TTCC)_3$. Interestingly, the TCCCTTCC-$(TCCC)_2$ part is common in both sequences. TTCTTCTTTC-$(TTCC)_2$ in humans also shows a T deletion and a C to T substitution since the equivalent in orangutans is TTCTTTC-$(TTCC)_3$. The TTTCTT non-variable block and the $(CTTT)_y$ repeat block are common in both sequences, but the final T-$(CTTT)_2$ in humans cannot be found in orangutans. Differences can be found in the flanking region which do not interfere with the sequence structure or allele counting. As such, the structure for this locus in orangutans is $(TTCC)_4$-TTCTTTTC-$(TTCC)_3$-TCCCTTCC-$(TCCC)_2$-TTCTTTC-$(TTCC)_3$-TTTCTT-$(CTTT)_4$.

```
Orangutan       GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCCTCCTTTCCTTCCTTCCTTCC
Human           GATCTGCCTTGCCCTTCCTACCTTTT----CCTCCCTCCCTCCTTTCCTTCCTTC-----

Orangutan       TTCTTTTCTTCCTTCCTTCCTCCCTTCCTCCCTCCCTTCTTTCTTCCTTCCTTCCTTTCT
Human           --CTTTCCTTCCTTCCTTCCTCCCTTCCTCCCTCCCTTC-TTCTTTCTTCCTTCCTTTCT

Orangutan       TCTTTCTTTCTTTCTTTCTCTGTCTTTCTTTCTCTTTCTCTTTCTTTCTCTTTCTTCTTT
Human           TCTTTCTTTCTTTCTT------TCTTTCTTTCT--------------------TTCTTT

Orangutan       CTTGCTTTCTTTTTCTTTCTTCTTTCTTTCTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT
Human           CTTTCTTTCT---------TTCTTTCTTTCT--------------TTTCTTTCTTTCTCT

Orangutan       CTCTTTCTTTCTTTTTCTTTCTTCTTTCTTTCCTTTCTTTCTTCTTTCTTTCCTTTCTTT
Human           GTCTTTCTTTCTTTTTCTTTCTTCTTTCTT---TTTCTTTCTTCTTTCTTTCCTTTCTTT

Orangutan       CTTCTTTCTTTCTTTCTCTCTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC
Human           CTTCTTTCTTTCTTTC-CT-----TCTCTTTCTCTCTCTCTCTCTCTTTCTTTCTTTC

Orangutan       TTTCTTTCTTTCTTTCTGTCTTTCTTTCTCTCTTTCTCTCTTTCTTTCTCTCAGTCTGTC
Human           TTCCTTTCTT----------------------------CCTTTCTTTCTCTCAGTCTGTC

Orangutan       ACCCCA
Human           ACCCCA
```

Figure 16. Alignment of the orangutan and human reference sequences.

The orangutan sample showed some variation in comparison with the reference sequence, in particular TC, TT, and T deletions (Figure 17).

```
Reference       GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCCTCCTTTCCTTCCTTCCTTCC
Orangutan       GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCCTCCTTTCCTTCCTTCCTTCC

Reference       TTCTTTTCTTCCTTCCTTCCTCCCTTCCTCCCTCCCTTCTTTCTTCCTTCCTTCCTTTCT
Orangutan       TTCTTTTCTTCCTTCCTTCCTCCCTTCCTCCCTCCCTTCTTTCTTCCTTCCTTCCTTTCT

Reference       TCTTTCTTTCTTTCTTTCTCTGTCTTTCTTTCTCTTTCTCTTTCTTTCTCTTTCTTCTTT
Orangutan       TCTTTCTTTCTTTCTTTCTCTGTCTTTCTT--TCTTTCTCTTTCTCTTTCTTTCTTCTTT

Reference       CTTGCTTTCTTTTTCTTTCTTCTTTCTTTCTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT
Orangutan       CTTGCTTTCTTTTTCTTTCTTCTTTCTTTC--CTTTCTTTC-TTCTTTCTTTCTTTC--T

Reference       CTCTTTCTTTCTTTTTCTTTCTTCTTTCTTTCCTTTCTTTCTTCTTTCTTTCCTTTCTTT
Orangutan       CTCTTTCTTTCTTTTTCTTTCTTCTTTCTTTCCTTTCTTTCTTCTTTCTTTCCTTTCTTT

Reference       CTTCTTTCTTTCTTTCTCTCTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC
Orangutan       CTTCTTTCTTTCTTTCTCTCTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC

Reference       TTTCTTTCTTTCTTTCTGTCTTTCTTTCTCTCTTTCTCTCTTTCTTTCTCTCAGTCTGTC
Orangutan       TTTCTTTC----------------------------------------------------

Reference       ACCCCA
Orangutan       ------
```

Figure 17. Alignment of the orangutan sequence with the reference. Some deletions can be observed. The large section at the end is missing because the sample was only sequenced with the reverse primer and that was the part of the sequence that was obscured.

Comparing the sequences of the gorilla and human references some relevant variation can be observed (Figure 18). The equivalent to the T after the variable $(TTCC)_x$ repeat block in humans seems to be TTTCTTCT in gorillas. The TCCCTTCC non-variable block is duplicated in gorillas, while most of the sequence afterward is conserved between the two species. In the non-variable block TTTCTT on the human sequence, an A to C substitution and a T deletion can be observed, as this sequence is TTTATTT in gorillas. The $(CTTT)_y$ block shows two repeats in gorillas. In the final part of the sequence structure, a CTT deletion can be observed in humans, as well as a deletion of two repeats of the last non-variable (CTTT) block. As such, the sequence structure of this locus in gorillas is $(TTCC)_5$-TTTCTTCT-$(TTCC)_4$-$(TCCCTTCC)_2$-$(TCCC)_2$-TTCTTCTTTC-$(TTCC)_2$-TTTATTT-$(CTTT)_2$-TCTT-$(CTTT)_4$.

```
Gorilla         GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCTTTCCTTCCTTCCTTCCTTCC
Human           GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCTTTCCTTCCTTCCT---TTCC

Gorilla         TTTCTTCTTTCCTTCCTTCCTTCCTCCCTTCCTCCCTTCCTCCCTCCCTTCTTCTTTCTT
Human           TTCCTTCCTTCCTCCCTTCCTCCCTCCCTT-CTTCTTTCTTCCTTCCTTTCTTCTTTCTT

Gorilla         CCTTCCTTTATTTCTTTCT---TTTC-TTCTTTCTTTCTTTCTTTCTCTTTCTTTCTT--
Human           TCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT-TTTCTTTCTTTC

Gorilla         ----TCTTTCTTTCTTTTTCTTTCTTCTTTC--TTTCTTTC--CTTTCTTTC--TTCTCT
Human           TCTGTCTTTCTTTCTTTTTCTTTCTTCTTTCTTTTTCTTTCTTCTTTCTTTCCTTTCTTT

Gorilla         CTTTCTTTCTTTCTTTCCTTCTCTTTCTTTCTCTCTCTCTCTCTTTCTTTCTTCCTTTCT
Human           C-TTCTTTCTTTCTTTCCTTCTCTTTCTCTCTCTCTCTCTCTTTCTTTCTTTCTTCCT

Gorilla         TTCTTTCTTTCTTTTTCTCAGTCTGTCACCCCA
Human           TTCTTCCTTTCTTTCTCTCAGTCTGTCACCCCA
```

Figure 18. Alignment of the gorilla and human reference sequences.

The sequenced gorilla sample showed no variation in comparison with the reference sequence (Figure 19).

```
Reference       GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCTTTCCTTCCTTCCTTCCTTCC
Gorilla         GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCTTTCCTTCCTTCCTTCCTTCC

Reference       TTTCTTCTTTCCTTCCTTCCTTCCTCCCTTCCTCCCTTCCTCCCTCCCTTCTTCTTTCTT
Gorilla         TTTCTTCTTTCCTTCCTTCCTTCCTCCCTTCCTCCCTTCCTCCCTCCCTTCTTCTTTCTT

Reference       CCTTCCTTTATTTCTTTCTTTTCTTCTTTCTTTCTTTCTTTCTCTTTCTTTCTTTCTTTC
Gorilla         CCTTCCTTTATTTCTTTCTTTTCTTCTTTCTTTCTTTCTTTCTCTTTCTTTCTTTCTTTC

Reference       TTTCTTTTTCTTTCTTCTTTCTTTCTTTCCTTTCTTTCTTCTCTCTTTCTTTCTTTCTTT
Gorilla         TTTCTTTTTCTTTCTTCTTTCTTTCTTTCCTTTCTTTCTTCTCTCTTTCTTTCTTTCTTT

Reference       CCTTCTCTTTCTTTCTCTCTCTCTCTCTTTCTTTCTTCCTTTCTTTCTTTCTTTCTTTTT
Gorilla         CCTTCTCTTTCTTTCTCTCTCTCTCTC--------------------------------

Reference       CTCAGTCTGTCACCCCA
Gorilla         -----------------
```

Figure 19. Alignment of the gorilla sequence with the reference. The large section at the end is missing because the sample was only sequenced with the reverse primer and that was the part of the sequence that was obscured.

A previous study (Moutinho, 2022) analyzed the sequence structure of this locus in chimpanzees, coming up with the following nomenclature: $(TTCC)_7$-T-$(TTCC)_4$-$(TCCCTTCC)_2$-$(TCCC)_2$-TTCTT-TC-$(TTCC)_2$-TTTCTT-$(CTTT)_2$-TTCTT-$(CTTT)_3$-CCTT-$(CTTT)_3$.

For three of the chimpanzee samples, the results obtained show to be consistent with the aforementioned study and follow the proposed sequence structure (Figure 20). However, one sample showed some variation in the number of repeats, following the structure $(TTCC)_5$-T-$(TTCC)_5$-$(TCCCTTCC)_2$-$(TCCC)_2$-TTCTT-TC-$(TTCC)_2$-TTTCTT-$(CTTT)_2$-TTCTT-$(CTTT)_3$-CCTT-$(CTTT)_3$. This may suggest that this locus is variable in chimpanzees, and more variation may be found by sequencing more samples.

```
Reference       GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCTTTCCTTCCTTCCTTCCTTCCTTCCTTTCCTTCCTTC
Chimp108        GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCTTTCCTTCCTTCCTTCCTTCCTTCCTTTCCTTCCTTC
Chimp110        GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCTTTCCTTCCTTCCTTCCTTCC----TTTCCTTCCTTCCTTC
Chimp121        GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCTTTCCTTCCTTCCTTCCTTCCTTCCTTTCCTTCCTTC
Chimp123        GATCTGCCTTGCCCTTCCTACCTTTTCCTCCCTCCCTCCTTTCCTTCCTTCCTTCCTTCCTTCCTTTCCTTCCTTC
                ************************************************************    ** * **********

Reference       CTTCCTCCCTTCCTCCCTTCCTCCCTCCCTTCTTTCTTCCTTCCTTTCTTCTTTCTTTTTCTTCTTTCTTTCTTTCCTTC
Chimp108        CTTCCTCCCTTCCTCCCTTCCTCCCTCCCTTCTTTCTTCCTTCCTTTCTTCTTTCTTTTTCTTCTTTCTTTCTTTCCTTC
Chimp110        CTTCCTCCCTTCCTCCCTTCCTCCCTCCCTTCTTTCTTCCTTCCTTTCTTCTTTCTTTTTCTTCTTTCTTTCTTTCCTTC
Chimp121        CTTCCTCCCTTCCTCCCTTCCTCCCTCCCTTCTTTCTTCCTTCCTTTCTTCTTTCTTTTTCTTCTTTCTTTCTTTCCTTC
Chimp123        CTTCCTCCCTTCCTCCCTTCCTCCCTCCCTTCTTTCTTCCTTCCTTTCTTCTTTCTTTTTCTTCTTTCTTTCTTTCCTTC
                ********************************************************************************

Reference       TTTCTTTCTTTCTCTGTCTTTCTTTTTCTTTCTTCTTTCTTTTTCTTTCCTCTTTCTTTCCTTTCTTTCTTCTTTCTTTC
Chimp108        TTTCTTTCTTTCTCTGTCTTTCTTTTTCTTTCTTCTTTCTTTTTCTTTCCTCTTTCTTTCCTTTCTTTCTTCTTTCTTTC
Chimp110        TTTCTTTCTTTCTCTGTCTTTCTTTTTCTTTCTTCTTTCTTTTTCTTTCCTCTTTCTT-----TCTTTCTTCTTTCTTTC
Chimp121        TTTCTTTCTTTCTCTGTCTTTCTTTTTCTTTCTTCTTTCTTTTTCTTTCCTCTTTCTTTCCTTTCTTTCTTCTTTCTTTC
Chimp123        TTTCTTTCTTTCTCTGTCTTTCTTTTTCTTTCTTCTTTCTTTTTCTTTCCTCTTTCTTTCCTTTCTTTCTTCTTTCTTTC
                ************************************************************     ****************

Reference       TTTCTTTCCTTCTCTTTCTTTCTCTCTCTCTCTTTCTTTCTTTCTTTCTTCCTTTCTTCCTTTCTTTCTCTCAGTCTG
Chimp108        TTTCTTTCCTTCTCTTTCTTTCTCTCTCTCT--C----------------------TT--------------------
Chimp110        TTTCTTTCCTTCTCTTTCTTTCTCTCTCTCTC------------------------------------------------
Chimp121        TTTCTTTCCTTCTCTTTCTTTCTCTCTCTCT--C----------------------TT--------------------
Chimp123        TTTCTTTCCTTCTCTTTCTTTCTCTCTCTCTC------------------------------------------------
                ********************************

Reference       TCACCCCA
Chimp108        ---C-TTC
Chimp110        ------TC
Chimp121        ---CTTTC
Chimp123        ------TC
```

Figure 20. Comparison of the chimpanzee samples with the reference sequence. Variation can be observed in the chimpanzee 110 sequence.

The PCR amplification of the various mammal samples produced negative results (Figure 21), showing that the sequence of the DXS10146 marker in humans differs considerably from the corresponding sequence in these non-human samples, even though the X chromosome is fairly conserved between mammals. Since these samples represent DNA from domestic animals, their DNA are often found at a crime scene in certain circumstances and be an important piece of evidence in a crime investigation. The lack of amplification shows that the presence of domestic animal DNA in a sample would not be a concern in forensic investigations using this marker. The human sample showed unspecific bands. However, since the goal here was just to confirm that these primers amplified this locus in humans (it would not be sequenced), it was not important to eliminate these bands.

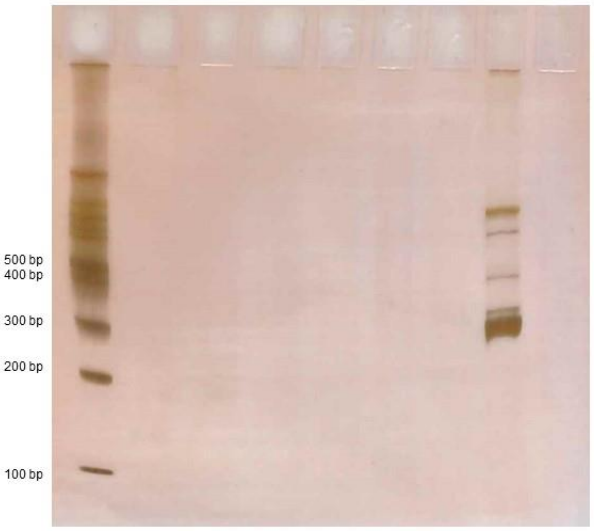


Figure 21. Polyacrylamide gel used for checking if the PCR amplification of the DXS10146 marker in non-human and domestic samples was successful. From left to right: ladder, mouse, cat, dog, bull, horse, pig, positive control (human), negative control. In the human sample, a few unspecific bands can be observed.

These results are consistent with the tested ones in the Investigator® Argus X-12 QS kit developmental validation (*Developmental Validation of the Investigator® Argus X-12 QS Kit*, 2016). As for primates, it is possible to observe the amplification of some products, some of which have the same size as the human STR products, while other non-human samples showed low-level reactivity. In the mentioned study (*Developmental Validation of the Investigator® Argus X-12 QS Kit*, 2016), however, some peaks appeared in the dog DNA electropherogram that seemed to be derived from STRs, while in this study amplification was not observed.

An accurate and common nomenclature is essential to ensure better communication, as well as comparison and exchange of data between laboratories and working groups. When proposing a nomenclature for a new STR, it is important to screen individuals from various populations as well as compare human and chimpanzee genomes. Comparing genomes allows for a better understanding of the possible mutational mechanisms involved in evolution, as well as a more precise understanding of the sequence structure of the loci. It is also important to study sequence variation thoroughly before proposing a new nomenclature, as an inaccurate nomenclature may lead to errors in the reported frequencies in population and genetic data.

4.2 Y-chromosome analyses

4.2.1 *De novo* mutations

The 11 analyzed father-son duos were sequenced first using only the forward primer. When sequencing, the sequence close to the primer had either not appeared on the electropherogram or was low in quality. For the duos amplified using DYS456, DYS481, DYS549 and one sample amplified using DYS576, the forward primer was designed too close to the core repeat sequence, and as consequence it was not possible to read the whole repeat sequence. For those cases, the reverse primer was used. The markers DYS458 and Y-GATA-H4 were only sequenced with the reverse primers since they were not marked with a fluorochrome.

The DYS389 I/II marker was ultimately not analyzed due to the difficulty in separating the bands (Figure 22). The primers used were not specific to each of the fragments in this marker. As such, even after extracting the larger band from the polyacrylamide gel the smaller band was still being amplified and so both bands were still appearing. Since it was not possible to extract the band or sequence directly, this marker was not considered for further analyses.

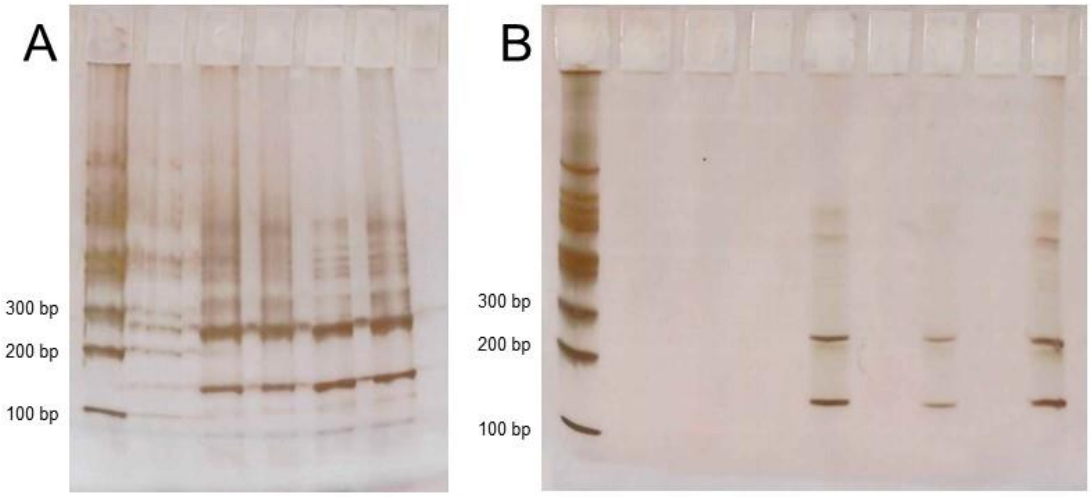


Figure 22. Polyacrylamide gel showing samples amplified with DYS389 I/II. Both bands still appeared after separation and reamplification. First many unspecific bands appeared (A), which disappeared after performing reamplification with different water and master mix aliquots (B). From left to right: ladder, negative control, N128 (father), N129 (son), N83 (father), N58 (son). One of the four samples, N128, did not show up on the second gel, probably due to a technical error since it appeared before.

For DYS481, it was also impossible to read the whole sequence with the reverse primer, as both flanking sequences were too small to sequence and the core repeat unit was only partly shown.

Four duos were amplified for the DYS576 marker. In one of these duos, the amplification of the sample of the son was unsuccessful. As such, that duo was not considered for further analyses. In the remaining duos, the mutations discovered by genotyping were confirmed with sequencing. The *de novo* length polymorphisms were successfully confirmed in the other markers: DYS19, DYS390, DYS456, DYS458, DYS549, DYS635, and Y-GATA-H4. For all of these, the sequence obtained matched the genotyping (Table 22).

Table 22. Sequences of the Y-STRs repeat motifs in father-son duos. The repeat motifs used as references were taken from the NIST STRBase (http://strbase.nist.gov/) and YHRD (http://yhrd.org/). In DYS19, the tagg block is not included in the reference repeat motif, even though it exists in the sequence. It was included in the column with the sequenced alleles in order to show a more accurate structure. The presented reference repeat motif for DYS635 is the long type.

Y-STR	Father's allele	Son's allele	Sequence of the alleles	Reference repeat motif
DYS19	14	15	$[TAGA]_3$tagg $[TAGA]_{11\rightarrow12}$	$[TAGA]_n$
DYS390	24	23	$[TCTG]_8[TCTA]_{11\rightarrow10}$ $TCTG[TCTA]_4$	$[TCTG]_8[TCTA]_n$ $[TCTG]_1[TCTA]_4$
DYS456	16	15	$[AGAT]_{16\rightarrow15}$	$[AGAT]_n$
	16	17	$[AGAT]_{16\rightarrow17}$	
DYS458	18	17	$[GAAA]_{18\rightarrow17}$	$[GAAA]_n$
DYS549	13	12	$[GATA]_{13\rightarrow12}$	$[GATA]_n$
DYS576	18	17	$[AAAG]_{18\rightarrow17}$	$[AAAG]_n$
	20	21	$[AAAG]_{20\rightarrow21}$	
	19	18	$[AAAG]_{19\rightarrow18}$	
DYS635	22	21	$[TCTA]_4[TGTA]_2$ $[TCTA]_2[TGTA]_2$ $[TCTA]_{12\rightarrow11}$	$[TCTA]_4[TGTA]_2[TCTA]_2$ $[TGTA]_2[TCTA]_2$ $[TGTA]_2[TCTA]_n$
Y-GATA-H4	12	11	$[TCTA]_{12\rightarrow11}$	$[TCTA]_n$

The DYS635 marker, also known as Y-GATA-C4, has two different possible sequence structures, a short one and a long one (Sánchez-Diz et al., 2003). The long one is referenced in Table 22, while the short one is $[TCTA]_4[TGTA]_2[TCTA]_2[TGTA]_2[TCTA]_n$. The duo sequenced with this Y-STR presented the short form of the allele. Some methodologies might lead to an inaccurate sizing of the alleles in this complex marker (Sánchez-Diz et al., 2003).

In one of the duos, the sample of the son showed a lack of signal at the DYS549, DYS392, and DYS448 markers which were observed in the father's profile (Figure 23). This duo was not used for this study, due to the inability to use individual primers for each marker and the difficulty of finding primers for such a large section of the Y chromosome. If used, the fragment size of the positive control would be around 3 kb, which would be too large to sequence using automated Sanger Sequencing and compare to the sample of the individual. Null alleles reported for these three markers had been observed previously in another study (Purps et al., 2014), and in other studies including the DYS385 locus (Beltramo et al., 2015; Silva et al., 2010).

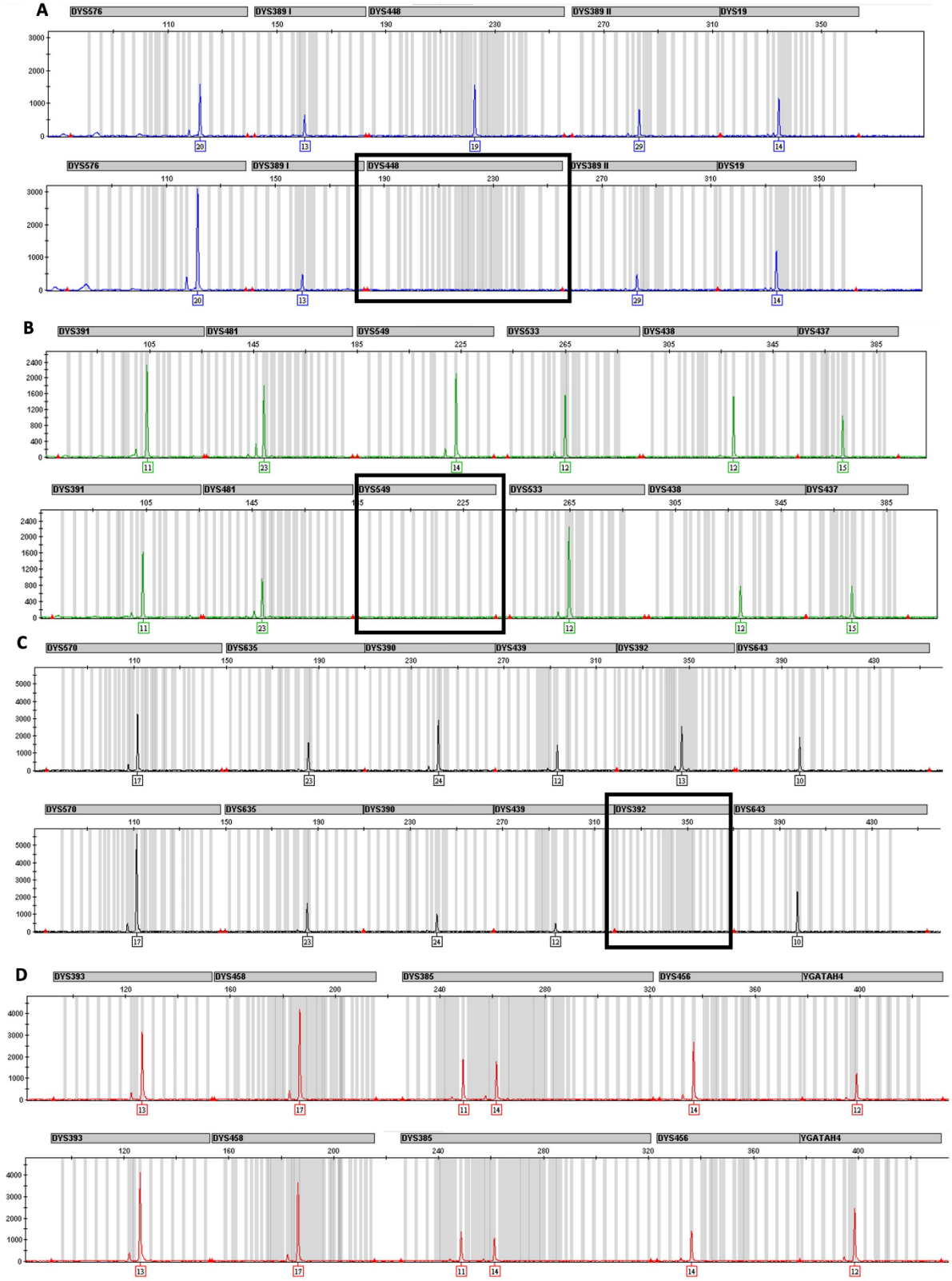


Figure 23. Electropherograms of the duo in which a microdeletion of three markers was observed in the son's sample. Top electropherograms belong to the father, and the bottom ones to the son, for each set of markers colored in blue (A), green (B), black (C), and red (D). The black box highlights the absence of amplification for the DYS549, DYS392, and DYS448 markers in the son's sample. Image taken from (Faustino, 2021).

Two of these markers, DYS549 and DYS392, are located in Yq11.222 while DYS448 is located in Yq11.223 (Hanson & Ballantyne, 2006). The azoospermia factor (AZF) is divided into three zones on the long arm of the Y chromosome, AZFa, AZFb, and AZFc, located in proximal Yq11, middle Yq11, and distal Yq11, respectively (Vogt et al., 1996), meaning that the missing markers map into the AZF region. Some authors have suggested that DYS392 and DYS448 are located in AZFb (Silva et al., 2010) while some supported that DYS448 maps to the proximal region of AZFc (Balaresque et al., 2008). AZF microdeletions have been shown to be specific for spermatogenic arrest and cause oligospermia or azoospermia (Dada et al., 2004).

5. Conclusion

5.1. X chromosome

One of the objectives of this study was to analyze DNA sequence variation, by amplifying and sequencing the repeat motif and flanking regions of an X-STR in a sample of Portuguese subjects. The marker used for this study was the X-STR DXS10146, a very polymorphic marker with a compound structure and two variable repeat blocks (Edelmann et al., 2008).

The subjects included in this study were part of a larger set of individuals previously analyzed for a different study (Faustino, 2021). The chosen samples for this work had at least one allele genotyped as 39.2, which was sequenced to confirm its length, as it was uncertain about whether this allele in some samples was actually 39.1. As such, in heterozygotic samples, both the putative 39.2 and the other allele were sequenced. Not all the selected samples were sequenced due to a variety of technical problems with the samples.

Variation was found due to variation in the repetitive blocks and not as single nucleotide changes, insertions or deletions. The alleles followed the established nomenclature. Some variation in the flanking regions could be detected since flanking region INDELs were found in three samples.

The 39.2 alleles had various isoalleles, while the shorter alleles (25, 26, 27, 28, and 29) only showed one form. This supports the hypothesis previously reported that DXS10146 shows more variation in longer alleles (Edelmann et al., 2008). It was observed in this study that all the alleles that had been previously assessed as 39.1 but are in fact 39.2 shared the same allelic DNA structure, which may justify the different electrophoretic mobility compared to adjacent alleles of the ladder (*Investigator® Argus X-12 QS Handbook*, 2022). Since no point mutations were found in the repetitive motif, the electrophoretic shift, at least in these alleles, could be due to that particular isoallele. More samples would need to be sequenced in order to confirm if this allelic structure is the cause of the different electrophoretic mobility.

A trio in which the daughter exhibited a possible triple X pattern was chosen, in order to confirm this condition using two different X-STR multiplexes, the Investigator® Argus X-12 (QIAGEN, Germany) and the GHEP-ISFG Decaplex (Gusmão et al., 2009). Despite the several technical difficulties in genotyping the samples using the Investigator® Argus

X-12, it was possible to confirm the trisomy and obtain genotypes for the trio. The GHEP-ISFG Decaplex (Gusmão et al., 2009) was also successful in genotyping the samples, although one of the markers (DXS7133) did not have any peaks. The results suggest maternal nondisjunction.

Primate samples were amplified and sequenced using primers for the X-STR DXS10146. The objective was to analyze the structure and stability of this marker's sequence by comparing the human sequence with the primates' and establish accurate nomenclature in the latter. Establishing a common nomenclature is very important in order to assure accurate comparisons and exchange of data between different working groups.

Six primate samples were used: a female gorilla, a male orangutan, and four chimpanzees. By aligning the primate and human reference sequences for this locus, it was possible to observe a relatively conserved sequence, with some bigger differences being observed in the orangutan sequence. This sequence had around 100 bp more than the other reference sequences. Large deletions, small insertions, and some base substitutions were observed in the primates. Comparing primate and human genomes helped us to perceive the evolution that this STR underwent.

In this study, it was possible to establish the sequence structure of this locus in orangutans and gorillas. Some deletions and substitutions can be observed between the reference sequences of these primates and the human.

The results obtained from sequencing the chimpanzee samples were consistent with the nomenclature established in a previous study (Moutinho, 2022), except for one sample. This sample had a similar structure but a different number of repeats as the other three, which may point to this locus being polymorphic in chimpanzees. It would be interesting to sequence more primate samples in future studies, in order to test if this locus is as variable in primates as it is in humans. In this study, DXS10146 was successfully confirmed as a specific marker in human-(non-primate) mixtures by checking for amplification in other species. The PCR amplification of various mammal samples was unsuccessful, showing that this marker does not amplify the DNA of these species. These results show that, in the presence of some domestic species DNA in a sample from a crime scene, the DXS10146 would still be useful and specific for humans.

5.2. Y chromosome

Regarding the Y chromosome, the aim of this study was to confirm *de novo* mutations genotyped in a previous study (Faustino, 2021) in father-son duos by sequencing Y-

STRs. The mutations sequenced were observed in markers DYS19, DYS390, DYS456, DYS458, DYS549, DYS576, DYS635, and Y-GATA-H4, and compatible with one-step changes, involving either the gain or the loss of a single repeat.

When sequencing, the sequence close to the primer tends to have a lower quality in the electropherogram due to technical reasons. Therefore, in STR analysis, if the primers are designed too close to the repetitive motif, the sequence may not appear in the electropherogram in its entirety. In this study, this was observed on the samples sequenced for DYS481, showing the importance of designing primers with a larger flanking region.

In this study, there were also difficulties in separating the two bands of the marker DYS389 I/II. In the future, when analyzing this marker, primers should be designed to encompass a larger sequence, instead of amplifying two bands.

The *de novo* length polymorphisms were successfully confirmed for all markers and duos, except for the cases mentioned above and for a sample that could not be amplified. The sequence obtained matched the genotyping, however, this is not always the case, as it depends on the primers used for genotyping. Some insertions in the flanking regions can cause discrepancies between allele designations and the number of repeats (Nagai & Bunai, 2011b), and as such it becomes important to sequence possible mutations.

A duo in which the son showed a *de novo* deletion in the markers DYS549, DYS392, and DYS448 was selected but not considered for further analyses. These markers map into the azoospermia factor (AZF) zone, and some authors have suggested that two of these DYS392 and DYS448 are located in AZFb (Silva et al., 2010) while some have said that DYS448 maps to the proximal region of AZFc (Balaresque et al., 2008). More studies would be needed in order to be certain.

The data obtained in this dissertation will, hopefully, improve the knowledge of sequence variation and structure of the X-STR DXS10146, as well as prove its stability and specificity. This study also confirms the importance of sequencing STRs in order to both confirm mutations and discover genetic polymorphisms, which are important in forensic genetic analysis.

References

Afonso Costa, H., Morais, P., Vieira da Silva, C., Matos, S., Marques Santos, R., Espinheira, R., Costa Santos, J., & Amorim, A. (2014). X-Chromosome STR markers data in a Cabo Verde immigrant population of Lisboa. *Molecular Biology Reports*, *41*(4), 2559–2569. https://doi.org/10.1007/s11033-014-3114-9

Allor, C., Einum, D. D., & Scarpetta, M. (2005). Identification and Characterization of Variant Alleles at CODIS STR Loci. *Journal of Forensic Sciences*, *50*(5), 1–6. https://doi.org/10.1520/JFS2005024

Álvarez-Nava, F., & Lanes, R. (2018). Epigenetics in Turner syndrome. *Clinical Epigenetics*, *10*(1), 45. https://doi.org/10.1186/s13148-018-0477-0

Antão-Sousa, S., Conde-Sousa, E., Gusmão, L., Amorim, A., & Pinto, N. (2022a). Estimations of Mutation Rates Depend on Population Allele Frequency Distribution: The Case of Autosomal Microsatellites. *Genes*, *13*(7), 1248. https://doi.org/10.3390/genes13071248

Antão-Sousa, S., Conde-Sousa, E., Gusmão, L., Amorim, A., & Pinto, N. (2022b). How often have X- and autosomal-STRs mutations equivocal parental origin been assigned? *Forensic Science International: Genetics Supplement Series*, *8*, 99–101. https://doi.org/10.1016/j.fsigss.2022.09.035

Auler-Bittencourt, E., Iwamura, E. S. M., Lima, M. J. M., da Silva, I. D. C. G., & dos Santos, S. E. B. (2015). Exploring the applicability of analysing X chromosome STRs in Brazilian admixed population. *Science & Justice*, *55*(5), 323–328. https://doi.org/10.1016/j.scijus.2015.03.004

Balaresque, P., Bowden, G. R., Parkin, E. J., Omran, G. A., Heyer, E., Quintana-Murci, L., Roewer, L., Stoneking, M., Nasidze, I., Carvalho-Silva, D. R., Tyler-Smith, C., de Knijff, P., & Jobling, M. A. (2008). Dynamic Nature of the Proximal AZFc Region of the Human Y Chromosome: Multiple Independent Deletion and Duplication Events Revealed by Microsatellite Analysis. *Human Mutation*, *29*(10), 1171–1180. https://doi.org/10.1002/humu.20757

Ballantyne, K. N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., Choi, Y., van Duijn, K., Vermeulen, M., Brauer, S., Decorte, R., Poetsch, M., von Wurmb-Schwark, N., de Knijff, P., Labuda, D., Vézina, H., Knoblauch, H., Lessig, R., Roewer, L., … Kayser, M. (2010). Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications. *The American*

Journal of Human Genetics, *87*(3), 341–353. https://doi.org/10.1016/j.ajhg.2010.08.006

Beltramo, J., Pena, M. A., & Lojo, M. M. (2015). The finding of Y-STR microdeletion involving DYS448, DYS392, DYS549 and DYS385a/b markers in a paternity case with deceased alleged father. *Forensic Science International: Genetics Supplement Series*, *5*, e141–e143. https://doi.org/10.1016/j.fsigss.2015.09.057

Bentayebi, K., Picornell, A., Bouabdeallah, M., Castro, J. A., Aboukhalid, R., Squalli, D., Misericòrdia, M., & Amzazi, S. (2012). Genetic diversity of 12 X-chromosomal short tandem repeats in the Moroccan population. *Forensic Science International: Genetics*, *6*(1), e48–e49. https://doi.org/10.1016/j.fsigen.2011.03.008

Bini, C., Sarno, S., Tangorra, E., Iuvaro, A., De Fanti, S., Tseghereda, Y. G., Pelotti, S., & Luiselli, D. (2021). Haplotype data and forensic evaluation of 23 Y-STR and 12 X-STR loci in eight ethnic groups from Eritrea. *International Journal of Legal Medicine*, *135*(2), 449–453. https://doi.org/10.1007/s00414-020-02446-2

Bobillo, C., Marino, M., Sala, A., Gusmão, L., & Corach, D. (2008). X-STRs: Relevance in complex kinship cases. *Forensic Science International: Genetics Supplement Series*, *1*(1), 496–498. https://doi.org/10.1016/j.fsigss.2007.10.080

Bojesen, A., Juul, S., & Gravholt, C. H. (2003). Prenatal and Postnatal Prevalence of Klinefelter Syndrome: A National Registry Study. *The Journal of Clinical Endocrinology & Metabolism*, *88*(2), 622–626. https://doi.org/10.1210/jc.2002-021491

Bosch, E., Lee, A. C., Calafell, F., Arroyo, E., Henneman, P., de Knijff, P., & Jobling, M. A. (2002). High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. *Forensic Science International*, *125*(1), 42–51. https://doi.org/10.1016/S0379-0738(01)00627-2

Bottinelli, M., Gouy, A., Utz, S., & Zieger, M. (2022). Population genetic analysis of 12 X-chromosomal STRs in a Swiss sample. *International Journal of Legal Medicine*, *136*(2), 561–563. https://doi.org/10.1007/s00414-021-02684-y

Boutrand, L., Egyed, B., Füredi, S., Mommers, N., Mertens, G., & Vandenberghe, A. (2001). Variations in primer sequences are the origin of allele drop-out at loci D13S317 and CD4. *International Journal of Legal Medicine*, *114*(4–5), 295–297. https://doi.org/10.1007/s004140000183

Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J., & Rolf, B. (1998). Mutation Rate in Human Microsatellites: Influence of the Structure and Length of the Tandem Repeat. *The American Journal of Human Genetics*, *62*(6), 1408–1415. https://doi.org/10.1086/301869

Butler, J. M. (2004). *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers* (2nd ed.). Elsevier Academic Press.

Butler, J. M. (2006). Genetics and Genomics of Core Short Tandem Repeat Loci Used in Human Identity Testing. *Journal of Forensic Sciences*, *51*(2), 253–265. https://doi.org/10.1111/j.1556-4029.2006.00046.x

Butler, J. M. (2007). Short tandem repeat typing technologies used in human identity testing. *BioTechniques*, *43*(4), Sii-Sv. https://doi.org/10.2144/000112582

Butler, J. M. (2015). The future of forensic DNA analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1674), 20140252. https://doi.org/10.1098/rstb.2014.0252

Cainé, L., Carvalho, R., Costa, S., Pereira, M. F., & Pinheiro, M. F. (2011). Interest of X chromosome (Argus X-12 kit) in complex kinship analysis. *Forensic Science International: Genetics Supplement Series*, *3*(1), e206–e207. https://doi.org/10.1016/j.fsigss.2011.08.103

Chakraborty, R., & Kidd, K. K. (1991). The Utility of DNA Typing in Forensic Work. *Science*, *254*(5039), 1735–1739. https://doi.org/10.1126/science.1763323

Chakraborty, R., Kimmel, M., Stivers, D. N., Davison, L. J., & Deka, R. (1997). Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proceedings of the National Academy of Sciences*, *94*(3), 1041–1046. https://doi.org/10.1073/pnas.94.3.1041

Charlesworth, B. (2002). *The Evolution of Chromosomal Sex Determination* (pp. 207–224). J. Wiley. https://doi.org/10.1002/0470868732.ch17

Chen, D.-P., Tseng, C.-P., Tsai, S.-H., Wang, M.-C., Lu, S.-C., Wu, T.-L., Chang, P.-Y., & Sun, C.-F. (2009). Use of X-linked short tandem repeats loci to confirm mutations in parentage caseworks. *Clinica Chimica Acta*, *408*(1–2), 29–33. https://doi.org/10.1016/j.cca.2009.06.042

Chen, W., Cheng, J., Tong, D., Liu, S., Zhang, Y., & Sun, H. (2014). Identification of a rare off-ladder allele of the D13S325 locus during paternity testing. *Legal Medicine*, *16*(1), 48–51. https://doi.org/10.1016/j.legalmed.2013.09.005

Claerhout, S., Vandenbosch, M., Nivelle, K., Gruyters, L., Peeters, A., Larmuseau, M. H. D., & Decorte, R. (2018). Determining Y-STR mutation rates in deep-routing genealogies: Identification of haplogroup differences. *Forensic Science International: Genetics*, *34*, 1–10. https://doi.org/10.1016/j.fsigen.2018.01.005

Clayton, T. M., Whitaker, J. P., & Maguire, C. N. (1995). Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci. *Forensic Science International*, *76*(1), 7–15. https://doi.org/10.1016/0379-0738(95)01787-9

Crnjac, J., Ozretić, P., Merkaš, S., Ratko, M., Lozančić, M., Rožić, S., Špoljarić, D., Korolija, M., Popović, M., & Mršić, G. (2016). Analysis of 12 X-chromosomal markers in the population of central Croatia. *Legal Medicine*, *21*, 77–84. https://doi.org/10.1016/j.legalmed.2016.07.001

Curtis, R., Ward, D., Taylor, D., & Henry, J. (2022). Investigation of X-STR haplotype diversity in the Australian Aboriginal population. *Australian Journal of Forensic Sciences*, 1–13. https://doi.org/10.1080/00450618.2022.2048690

Czarnogórska, M., Sanak, M., Piniewska, D., Polańska, N., Stawowiak, A., & Opolska-Bogusz, B. (2010). [Identification of rare genetic variants at DXS10074, DXS10079, DXS10146 and DXS10148 loci of investigator argus X-12 multiplex in the south Polish population]. *Archiwum Medycyny Sadowej i Kryminologii*, *60*(4), 235–242. http://www.ncbi.nlm.nih.gov/pubmed/21863730

Dada, R., Gupta, N. P., & Kucheria, K. (2004). Yq Microdeletions—Azoospermia Factor Candidate Genes and Spermatogenic Arrest. *Journal of Biomolecular Techniques*, *15*(3), 176–183. http://www.ncbi.nlm.nih.gov/pubmed/15331583

Dahui, Q. (2019). Next-generation sequencing and its clinical application. *Cancer Biology & Medicine*, *16*(1), 4–10. https://doi.org/10.20892/j.issn.2095-3941.2018.0055

Deng, C., Song, F., Li, J., Li, Y., Hou, Y., & Luo, H. (2017). Multiplex PCR for 19 X-chromosomal STRs in Chinese population. *Forensic Science International: Genetics Supplement Series*, *6*, e24–e26. https://doi.org/10.1016/j.fsigss.2017.09.016

Developmental validation of the Investigator® Argus X-12 QS Kit. (2016). QIAGEN.

Diegoli, T. M., Linacre, A., Vallone, P. M., Butler, J. M., & Coble, M. D. (2011). Allele frequency distribution of twelve X-chromosomal short tandem repeat markers in four U.S. population groups. *Forensic Science International: Genetics Supplement Series*, *3*(1), e481–e483. https://doi.org/10.1016/j.fsigss.2011.09.102

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K., & Mattick, J. S. (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research*, *19*(14), 4008–4008. https://doi.org/10.1093/nar/19.14.4008

Edelmann, J., Hering, S., Augustin, C., & Szibor, R. (2008). Characterisation of the STR markers DXS10146, DXS10134 and DXS10147 located within a 79.1kb region at Xq28. *Forensic Science International: Genetics*, *2*(1), 41–46. https://doi.org/10.1016/j.fsigen.2007.08.001

Edwards, A., Civitello, A., Hammond, H. A., & Caskey, C. T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics*, *49*(4), 746–756. http://www.ncbi.nlm.nih.gov/pubmed/1897522

Elakkary, S., Hoffmeister-Ullerich, S., Schulze, C., Seif, E., Sheta, A., Hering, S., Edelmann, J., & Augustin, C. (2014). 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 Science International: Genetics*, *11*(1), 26–30. https://doi.org/10.1016/j.fsigen.2014.02.007

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, *5*(6), 435–445. https://doi.org/10.1038/nrg1348

Fast and efficient recovery of DNA. (2018). Ahlstrom-Munksjö.

Faustino, M. (2021). *Polymorphisms and mutation rate estimates of Y-STRs and X-STRs in the Portuguese population* [Master's Degree in Forensic Genetics]. Faculty of Sciences of the University of Porto.

Faustino, M., Gusmão, L., Amorim, A., Kling, D., & Pinto, N. (2023). A Mathematical Framework for Genetic Relatedness Analysis Involving X Chromosome Aneuploidies. *Under Review*.

Ferragut, J. F., Bentayebi, K., Barbaro, A., Ramírez, M., Saguillo, A. Y., Ramon, C., & Picornell, A. (2021). Exploring the Western Mediterranean through X-chromosome.

International Journal of Legal Medicine, *135*(3), 787–790. https://doi.org/10.1007/s00414-020-02498-4

Flores-Espinoza, R., Paz-Cruz, E., Ruiz-Pozo, V. A., Lopez-Carrera, M., Cabrera-Andrade, A., Gusmão, L., & Burgos, G. (2021). Investigating genetic diversity in admixed populations from Ecuador. *American Journal of Physical Anthropology*, *176*(1), 109–119. https://doi.org/10.1002/ajpa.24341

Ford, C. E., Jones, K. W., Polani, P. E., De Almeida, J. C., & Briggs, J. H. (1959). A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *The Lancet*, *273*(7075), 711–713. https://doi.org/10.1016/S0140-6736(59)91893-8

Fuhrmann, W., & Vogel, F. (1969). The Autosomal Dominant Mode of Inheritance. In *Genetic Counseling* (pp. 16–23). Springer New York. https://doi.org/10.1007/978-1-4899-2672-2_3

Gaastra, W. (1985). Gel Electrophoretic Analysis of DNA Sequencing Products. In *Nucleic Acids* (pp. 343–350). Humana Press. https://doi.org/10.1385/0-89603-064-4:343

García, M. G., Gusmão, L., Catanesi, C. I., Penacino, G. A., & Pinto, N. (2017). Mutation rate of 12 X-STRs from investigator Argus X-12 kit in Argentine population. *Forensic Science International: Genetics Supplement Series*, *6*, e562–e564. https://doi.org/10.1016/j.fsigss.2017.09.219

Gettings, K. B., Aponte, R. A., Vallone, P. M., & Butler, J. M. (2015). STR allele sequence variation: Current knowledge and future issues. *Forensic Science International: Genetics*, *18*, 118–130. https://doi.org/10.1016/j.fsigen.2015.06.005

Gill, P. (2002). Role of Short Tandem Repeat DNA in Forensic Casework in the UK—Past, Present, and Future Perspectives. *BioTechniques*, *32*(2), 366–385. https://doi.org/10.2144/02322rv01

Gill, P., Jeffreys, A. J., & Werrett, D. J. (1985). Forensic application of DNA 'fingerprints.' *Nature*, *318*(6046), 577–579. https://doi.org/10.1038/318577a0

Gomes, C. (2019). *Genetic insight into Nigerian population groups using an X chromosome decaplex system* [Master's Degree in Forensic Genetics]. Faculty of Sciences of the University of Porto.

Gomes, C., Magalhães, M., Alves, C., Amorim, A., Pinto, N., & Gusmão, L. (2012). Comparative evaluation of alternative batteries of genetic markers to complement

autosomal STRs in kinship investigations: autosomal indels vs. X-chromosome STRs. *International Journal of Legal Medicine*, *126*(6), 917–921. https://doi.org/10.1007/s00414-012-0768-5

Gomes, I., Pereira, P. J. P., Harms, S., Oliveira, A. M., Schneider, P. M., & Brehm, A. (2017). Genetic characterization of Guinea-Bissau using a 12 X-chromosomal STR system: Inferences from a multiethnic population. *Forensic Science International: Genetics*, *31*, 89–94. https://doi.org/10.1016/j.fsigen.2017.08.016

Gomes, I., Pinto, N., Antão-Sousa, S., Gomes, V., Gusmão, L., & Amorim, A. (2020). Twenty Years Later: A Comprehensive Review of the X Chromosome Use in Forensic Genetics. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.00926

Gopinath, S., Zhong, C., Nguyen, V., Ge, J., Lagacé, R. E., Short, M. L., & Mulero, J. J. (2016). Developmental validation of the Yfiler® Plus PCR Amplification Kit: An enhanced Y-STR multiplex for casework and database applications. *Forensic Science International: Genetics*, *24*, 164–175. https://doi.org/10.1016/j.fsigen.2016.07.006

Graves, J. A. M. (1995). The origin and function of the mammalian Y chromosome and Y-borne genes - an evolving understanding. *BioEssays*, *17*(4), 311–320. https://doi.org/10.1002/bies.950170407

Graves, J. A. M., & Foster, J. W. (1994). Evolution of Mammalian Sex Chromosomes and Sex-Determining Genes. In *International Review of Cytology* (Vol. 154, Issue C, pp. 191–259). https://doi.org/10.1016/S0074-7696(08)62200-7

Graves, J. A. M., Gécz, J., & Hameister, H. (2002). Evolution of the human X – a smart and sexy chromosome that controls speciation and development. *Cytogenetic and Genome Research*, *99*(1–4), 141–145. https://doi.org/10.1159/000071585

Gruhn, J. R., & Hoffmann, E. R. (2022). Errors of the Egg: The Establishment and Progression of Human Aneuploidy Research in the Maternal Germline. *Annual Review of Genetics*, *56*(1), 369–390. https://doi.org/10.1146/annurev-genet-072820-033609

Guo, F. (2017). Population genetic data for 12 X-STR loci in the Northern Han Chinese and StatsX package as tools for population statistics on X-STR. *Forensic Science International: Genetics*, *26*, e1–e8. https://doi.org/10.1016/j.fsigen.2016.10.012

Gusmão, L., Butler, J. M., Carracedo, A., Gill, P., Kayser, M., Mayr, W. R., Morling, N., Prinz, M., Roewer, L., Tyler-Smith, C., & Schneider, P. M. (2006). DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Science International*, *157*(2–3), 187–197. https://doi.org/10.1016/j.forsciint.2005.04.002

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, Ó., … Amorim, A. (2009). A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations. *International Journal of Legal Medicine*, *123*(3), 227–234. https://doi.org/10.1007/s00414-008-0309-4

Gymrek, M. (2017). A genomic view of short tandem repeats. *Current Opinion in Genetics & Development*, *44*, 9–16. https://doi.org/10.1016/j.gde.2017.01.012

Haddish, K., Chierto, E., Di Vella, G., Lacerenza, D., Raddi, S., Aneli, S., Bogale, A. L., Kidane, E., Yizengaw, A., Getaneh, Y., Tewelemedhin, G., Kumar, H. R. S., & Robino, C. (2022). A reference database of forensic autosomal and gonosomal STR markers in the Tigray population of Ethiopia. *Forensic Science International: Genetics*, *56*, 102618. https://doi.org/10.1016/j.fsigen.2021.102618

Haldane, J. B. S. (1935). The rate of spontaneous mutation of a human gene. *Journal of Genetics*, *31*(3), 317–326. https://doi.org/10.1007/BF02982403

Hall, H., Hunt, P., & Hassold, T. (2006). Meiosis and sex chromosome aneuploidy: how meiotic errors cause aneuploidy; how aneuploidy causes meiotic errors. *Current Opinion in Genetics & Development*, *16*(3), 323–329. https://doi.org/10.1016/j.gde.2006.04.011

Haltrich, I., Pikó, H., Pamjav, H., Somogyi, A., Völgyi, A., David, D., Beke, A., Garamvölgyi, Z., Kiss, E., Karcagi, V., & Fekete, G. (2015). Complex X chromosome rearrangement associated with multiorgan autoimmunity. *Molecular Cytogenetics*, *8*(1), 51. https://doi.org/10.1186/s13039-015-0152-5

Hanson, E. K., & Ballantyne, J. (2006). Comprehensive annotated STR physical map of the human Y chromosome: Forensic implications. *Legal Medicine*, *8*(2), 110–120. https://doi.org/10.1016/j.legalmed.2005.10.001

Heinrich, M., Felske-Zech, H., Brinkmann, B., & Hohoff, C. (2005). Characterisation of variant alleles in the STR systems D2S1338, D3S1358 and D19S433. *International Journal of Legal Medicine*, *119*(5), 310–313. https://doi.org/10.1007/s00414-005-0554-8

Heinrich, M., Müller, M., Rand, S., Brinkmann, B., & Hohoff, C. (2004). Allelic drop-out in the STR system ACTBP2 (SE33) as a result of mutations in the primer binding region. *International Journal of Legal Medicine*, *118*(6), 361–363. https://doi.org/10.1007/s00414-004-0473-0

Hering, S., Edelmann, J., Haas, S., & Grasern, N. (2015). Paternity testing of two female siblings with Investigator Argus X-12 kit: A case with several rare mutation and recombination events. *Forensic Science International: Genetics Supplement Series*, *5*, e341–e343. https://doi.org/10.1016/j.fsigss.2015.09.135

Investigator® Argus X-12 Handbook. (2013). QIAGEN.

Investigator® Argus X-12 QS Handbook. (2022). QIAGEN.

Jacobs, P. A., Baikie, A. G., Court Brown, W. M., Macgregor, T. N., Maclean, N., & Harnden, D. G. (1959). Evidence for the existence of the human “super female.” *The Lancet*, *274*(7100), 423–425. https://doi.org/10.1016/S0140-6736(59)90415-5

Jacobs, P. A., & Strong, J. A. (1959). A Case of Human Intersexuality Having a Possible XXY Sex-Determining Mechanism. *Nature*, *183*(4657), 302–303. https://doi.org/10.1038/183302a0

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Hypervariable ‘minisatellite’ regions in human DNA. *Nature*, *314*(6006), 67–73. https://doi.org/10.1038/314067a0

Jobling, M. A., & Gill, P. (2004). Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics*, *5*(10), 739–751. https://doi.org/10.1038/nrg1455

Jobling, M. A., & Tyler-Smith, C. (1995). Fathers and sons: the Y chromosome and human evolution. *Trends in Genetics*, *11*(11), 449–456. https://doi.org/10.1016/S0168-9525(00)89144-1

Jusic, B., Pilav, A., Dzehverovic, M., & Cakar, J. (2022). DNA Paternıty Testıng wıth Two Mısmatches: Our Experıence. *Jordan Journal of Biological Sciences*, *15*(03), 387–394. https://doi.org/10.54319/jjbs/150307

Karger, B. L., & Guttman, A. (2009). DNA sequencing by CE. *ELECTROPHORESIS*, *30*(S1), S196–S202. https://doi.org/10.1002/elps.200900218

Kayser, M. (2017). Forensic use of Y-chromosome DNA: a general overview. *Human Genetics*, *136*(5), 621–635. https://doi.org/10.1007/s00439-017-1776-9

Kayser, M., Kittler, R., Erler, A., Hedman, M., Lee, A. C., Mohyuddin, A., Mehdi, S. Q., Rosser, Z., Stoneking, M., Jobling, M. A., Sajantila, A., & Tyler-Smith, C. (2004). A Comprehensive Survey of Human Y-Chromosomal Microsatellites. *The American Journal of Human Genetics*, *74*(6), 1183–1197. https://doi.org/10.1086/421531

Khlestkina, E. K. (2014). Molecular markers in genetic studies and breeding. *Russian Journal of Genetics: Applied Research*, *4*(3), 236–244. https://doi.org/10.1134/S2079059714030022

Kim, S.-H., Park, M.-J., Cho, E. H., Kim, S., & Yoo, S. J. (2019). Different parental origins of supernumerary X chromosomes in brothers with Klinefelter syndrome. *Medicine*, *98*(44), e17838. https://doi.org/10.1097/MD.0000000000017838

Kline, M. C., Hill, C. R., Decker, A. E., & Butler, J. M. (2011). STR sequence analysis for characterizing normal, variant, and null alleles. *Forensic Science International: Genetics*, *5*(4), 329–332. https://doi.org/10.1016/j.fsigen.2010.09.005

Kling, D., Dell'Amico, B., Haddeland, P. J. T., & Tillmar, A. O. (2014). Population genetic analysis of 12 X-STRs in a Somali population sample. *Forensic Science International: Genetics*, *11*(1), e7–e8. https://doi.org/10.1016/j.fsigen.2014.02.013

Krenke, B. E., Viculis, L., Richard, M. L., Prinz, M., Milne, S. C., Ladd, C., Gross, A. M., Gornall, T., Frappier, J. R. H., Eisenberg, A. J., Barna, C., Aranda, X. G., Adamowicz, M. S., & Budowle, B. (2005). Validation of a male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex. *Forensic Science International*, *148*(1), 1–14. https://doi.org/10.1016/j.forsciint.2004.07.008

Lahn, B. T., & Page, D. C. (1999). Four Evolutionary Strata on the Human X Chromosome. *Science*, *286*(5441), 964–967. https://doi.org/10.1126/science.286.5441.964

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., … Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Lazaruk, K., Wallin, J., Holt, C., Nguyen, T., & Walsh, P. S. (2001). Sequence variation in humans and other primates at six short tandem repeat loci used in forensic

identity testing. *Forensic Science International*, *119*(1), 1–10. https://doi.org/10.1016/S0379-0738(00)00388-1

Levinson, G., & Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*. https://doi.org/10.1093/oxfordjournals.molbev.a040442

Linden, M. C., Bender, B. G., & Robinson, A. (1995). Sex Chromosome Tetrasomy and Pentasomy. *Pediatrics*, *96*(4), 672–682. https://doi.org/10.1542/peds.96.4.672

MacDonald, M., Hassold, T., Harvey, J., Wang, L. H., Morton, N. E., & Jacobs, P. (1994). The origin of 47, XXY and 47, XXX aneuploidy: heterogeneous mechanisms and role of aberrant recombination. *Human Molecular Genetics*, *3*(8), 1365–1371. https://doi.org/10.1093/hmg/3.8.1365

Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*, *50*(W1), W276–W279. https://doi.org/10.1093/nar/gkac240

Mallard, T. T., Liu, S., Seidlitz, J., Ma, Z., Moraczewski, D., Thomas, A., & Raznahan, A. (2021). X-chromosome influences on neuroanatomical variation in humans. *Nature Neuroscience*, *24*(9), 1216–1224. https://doi.org/10.1038/s41593-021-00890-w

Messaoudi, S. A., Babu, S. R., Alsaleh, A. B., Albujja, M., Al-Snan, N. R., Chaudhary, A. R., Kassab, A. Ch., & Assidi, M. (2021). Population genetic data for 12 X-STR loci in the Central Saudi region using investigator Argus X-12 amplification kit. *Annals of Human Biology*, *48*(4), 321–326. https://doi.org/10.1080/03014460.2021.1957147

Miki, T., Nagayoshi, M., Takemoto, Y., Yamaguchi, T., Takeda, S., Watanabe, S., & Tanaka, A. (2017). Genetic risk of Klinefelter's syndrome in assisted reproductive technology. *Reproductive Medicine and Biology*, *16*(2), 188–195. https://doi.org/10.1002/rmb2.12029

Mirkin, S. M. (2007). Expandable DNA repeats and human disease. *Nature*, *447*(7147), 932–940. https://doi.org/10.1038/nature05977

Morozova, O., & Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, *92*(5), 255–264. https://doi.org/10.1016/j.ygeno.2008.07.001

Moutinho, B. (2022). *Genetic diversity of human specific X-STRs in humans and chimpanzees: a comparative study and proposal for nomenclature* [Master's Degree in Forensic Genetics]. Faculty of Sciences of the University of Porto.

Mršić, G., Ozretić, P., Crnjac, J., Merkaš, S., Sukser, V., Račić, I., Rožić, S., Barbarić, L., Popović, M., & Korolija, M. (2018). Expanded Croatian 12 X-STR loci database with an overview of anomalous profiles. *Forensic Science International: Genetics*, *34*, 249–256. https://doi.org/10.1016/j.fsigen.2018.03.004

Mulero, J. J., Chang, C. W., Calandro, L. M., Green, R. L., Li, Y., Johnson, C. L., & Hennessy, L. K. (2006). Development and Validation of the AmpFLSTR™ Yfiler™ PCR Amplification Kit: A Male Specific, Single Amplification 17 Y-STR Multiplex System. *Journal of Forensic Sciences*, *51*(1), 64–75. https://doi.org/10.1111/j.1556-4029.2005.00016.x

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, *51*, 263–273. https://doi.org/10.1101/SQB.1986.051.01.032

Nachman, M. W., & Crowell, S. L. (2000). Estimate of the Mutation Rate per Nucleotide in Humans. *Genetics*, *156*(1), 297–304. https://doi.org/10.1093/genetics/156.1.297

Nagai, A., & Bunai, Y. (2011a). Insertion/deletion polymorphisms at the X-STR DXS10146 and DXS10147 flanking regions. *Forensic Science International: Genetics Supplement Series*, *3*(1), e129–e130. https://doi.org/10.1016/j.fsigss.2011.08.064

Nagai, A., & Bunai, Y. (2011b). Structural polymorphisms at the X-chromosomal short tandem repeat loci DXS10134, DXS10135, DXS10146 and DXS10148. *Forensic Science International: Genetics Supplement Series*, *3*(1), e343–e344. https://doi.org/10.1016/j.fsigss.2011.09.034

Nagai, A., Hara, M., Ishihara, T., Tamura, A., Kido, A., & Bunai, Y. (2013). INDEL polymorphisms at the DXS10146 flanking region in four racial populations. *Forensic Science International: Genetics Supplement Series*, *4*(1), e318–e319. https://doi.org/10.1016/j.fsigss.2013.10.162

Nielsen, J., & Wohlert, M. (1991). Chromosome abnormalities found among 34910 newborn children: results from a 13-year incidence study in Århus, Denmark. *Human Genetics*, *87*(1), 81–83. https://doi.org/10.1007/BF01213097

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment 1 1Edited by J. Thornton. *Journal of Molecular Biology*, *302*(1), 205–217. https://doi.org/10.1006/jmbi.2000.4042

Núñez, C., Baeta, M., Ibarbia, N., Ortueta, U., Jiménez-Moreno, S., Blazquez-Caeiro, J. L., Builes, J. J., Herrera, R. J., Martínez-Jarreta, B., & de Pancorbo, M. M. (2017). 17 to 23: A novel complementary mini Y-STR panel to extend the Y-STR databases from 17 to 23 markers for forensic purposes. *ELECTROPHORESIS*, *38*(7), 1016–1021. https://doi.org/10.1002/elps.201600313

Ogata, T., & Matsuo, N. (1995). Turner syndrome and female sex chromosome aberrations: deduction of the principal factors involved in the development of clinical features. *Human Genetics*, *95*(6). https://doi.org/10.1007/BF00209476

Ohno, S. (1967). *Sex Chromosomes and Sex-Linked Genes* (Vol. 1). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-88178-7

Pasino, S., Caratti, S., Del Pero, M., Santovito, A., Torre, C., & Robino, C. (2011). Allele and haplotype diversity of X-chromosomal STRs in Ivory Coast. *International Journal of Legal Medicine*, *125*(5), 749–752. https://doi.org/10.1007/s00414-011-0591-4

Piniewska, D., Sanak, M., Wojtas, M., & Polanska, N. (2017). The genetic evidence for human origin of Jivaroan shrunken heads in collections from the Polish museums. *International Journal of Legal Medicine*, *131*(3), 643–650. https://doi.org/10.1007/s00414-016-1448-7

Pinto, N., Gusmão, L., & Amorim, A. (2011). X-chromosome markers in kinship testing: A generalisation of the IBD approach identifying situations where their contribution is crucial. *Forensic Science International: Genetics*, *5*(1), 27–32. https://doi.org/10.1016/j.fsigen.2010.01.011

Pinto, N., Pereira, V., Tomas, C., Loiola, S., Carvalho, E. F., Modesti, N., Maxzud, M., Marcucci, V., Cano, H., Cicarelli, R., Januario, B., Bento, A., Brito, P., Burgos, G., Paz-Cruz, E., Díez-Juárez, L., Vannelli, S., Pontes, M. de L., Berardi, G., … Gusmão, L. (2020). Paternal and maternal mutations in X-STRs: A GHEP-ISFG collaborative study. *Forensic Science International: Genetics*, *46*, 102258. https://doi.org/10.1016/j.fsigen.2020.102258

Prieto-Fernández, E., Baeta, M., Núñez, C., Zarrabeitia, M. T., Herrera, R. J., Builes, J. J., & de Pancorbo, M. M. (2016). Development of a new highly efficient 17 X-STR

multiplex for forensic purposes. *ELECTROPHORESIS*, *37*(12), 1651–1658. https://doi.org/10.1002/elps.201500546

Purps, J., Siegert, S., Willuweit, S., Nagy, M., Alves, C., Salazar, R., Angustia, S. M. T., Santos, L. H., Anslinger, K., Bayer, B., Ayub, Q., Wei, W., Xue, Y., Tyler-Smith, C., Bafalluy, M. B., Martínez-Jarreta, B., Egyed, B., Balitzki, B., Tschumi, S., … Roewer, L. (2014). A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Science International: Genetics*, *12*, 12–23. https://doi.org/10.1016/j.fsigen.2014.04.008

QIAGEN® Multiplex PCR Handbook. (2010). QIAGEN.

Quintana-Murci, L., & Fellous, M. (2001). The Human Y Chromosome: The Biological Role of a "Functional Wasteland." *Journal of Biomedicine and Biotechnology*, *1*(1), 18–24. https://doi.org/10.1155/S1110724301000080

Rice, W. R. (1984). SEX CHROMOSOMES AND THE EVOLUTION OF SEXUAL DIMORPHISM. *Evolution*, *38*(4), 735–742. https://doi.org/10.1111/j.1558-5646.1984.tb00346.x

Richards, R. I., & Sutherland, G. R. (1996). Repeat offenders: Simple repeat sequences and complex genetic problems. *Human Mutation*, *8*(1), 1–7. https://doi.org/10.1002/(SICI)1098-1004(1996)8:1<1::AID-HUMU1>3.0.CO;2-G

Robino, C., Lacerenza, D., Aneli, S., Di Gaetano, C., Matullo, G., Robledo, R., & Calò, C. (2018). Allele and haplotype diversity of 12 X-STRs in Sardinia. *Forensic Science International: Genetics*, *33*, e1–e3. https://doi.org/10.1016/j.fsigen.2017.12.002

Robinson, D. O., & Jacobs, P. A. (1999). The Origin of the Extra Y Chromosome in Males with a 47,XYY Karyotype. *Human Molecular Genetics*, *8*(12), 2205–2209. https://doi.org/10.1093/hmg/8.12.2205

Roewer, L. (2019). Y-chromosome short tandem repeats in forensics—Sexing, profiling, and matching male DNA. *WIREs Forensic Science*, *1*(4). https://doi.org/10.1002/wfs2.1336

Roewer, L., & Epplen, J. T. (1992). Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. *Forensic Science International*, *53*(2), 163–171. https://doi.org/10.1016/0379-0738(92)90193-Z

Roewer, L., Krawczak, M., Willuweit, S., Nagy, M., Alves, C., Amorim, A., Anslinger, K., Augustin, C., Betz, A., Bosch, E., Cagliá, A., Carracedo, A., Corach, D., Dekairelle, A.-F., Dobosz, T., Dupuy, B. M., Füredi, S., Gehrig, C., Gusmaõ, L., … Kayser, M. (2001). Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Science International*, *118*(2–3), 106–113. https://doi.org/10.1016/S0379-0738(00)00478-3

Rolf, B., Bulander, N., & Wiegand, P. (2011). Insertion-/deletion polymorphisms close to the repeat region of STR loci can cause discordant genotypes with different STR kits. *Forensic Science International: Genetics*, *5*(4), 339–341. https://doi.org/10.1016/j.fsigen.2011.01.005

Rosenberg, S. M. (2013). Reverse Mutation. In *Brenner's Encyclopedia of Genetics* (pp. 220–221). Elsevier. https://doi.org/10.1016/B978-0-12-374984-0.01325-5

Ross, M. T., Grafham, D. V., Coffey, A. J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G. R., Burrows, C., Bird, C. P., Frankish, A., Lovell, F. L., Howe, K. L., Ashurst, J. L., Fulton, R. S., Sudbrak, R., Wen, G., Jones, M. C., Hurles, M. E., … Bentley, D. R. (2005). The DNA sequence of the human X chromosome. *Nature*, *434*(7031), 325–337. https://doi.org/10.1038/nature03440

Ruitberg, C. M., Reeder, D. J., & Butler, J. M. (2001). STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Research*, *29*(1), 320–322. https://doi.org/10.1093/nar/29.1.320

Saifl, G. M., & Chandra, H. S. (1999). An apparent excess of sex– and reproduction–related genes on the human X chromosome. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1415), 203–209. https://doi.org/10.1098/rspb.1999.0623

Samejima, M., Nakamura, Y., Nambiar, P., & Minaguchi, K. (2012). Genetic study of 12 X-STRs in Malay population living in and around Kuala Lumpur using Investigator Argus X-12 kit. *International Journal of Legal Medicine*, *126*(4), 677–683. https://doi.org/10.1007/s00414-012-0705-7

Sánchez-Diz, P., Gusmão, L., Beleza, S., Benítez-Páez, A., Castro, A., García, O., Solla, L. P., Geada, H., Martín, P., Martínez-Jarreta, B., Pinheiro, M. D. F., Raimondi, E., Silva De La Fuente, S. M., Vide, M. C., Whittle, M. R., Zarrabeitia, M. T., Carracedo, A., & Amorim, A. (2003). Results of the GEP-ISFG collaborative study on two Y-STRs tetraplexes: GEPY I (DYS461, GATA C4, DYS437 and DYS438) and GEPY

II (DYS460, GATA A10, GATA H4 and DYS439). *Forensic Science International*, *135*(2), 158–162. https://doi.org/10.1016/S0379-0738(03)00200-7

Sandberg, A. A., Koepf, G. F., Ishihara, T., & Hauschka, T. S. (1961). An XYY human male. *The Lancet*, *278*(7200), 488–489. https://doi.org/10.1016/S0140-6736(61)92459-X

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Schmidtke, J., Kühnau, W., Wand, D., Edelmann, J., Szibor, R., & Krawczak, M. (2004). Prenatal exclusion without involving the putative fathers of an incestuous father–daughter parenthood. *Prenatal Diagnosis*, *24*(8), 662–664. https://doi.org/10.1002/pd.929

Silva, J., Ribeiro, T., Gonçalves, J., & Geada, H. (2010). Forensic Y-STR study in Y-chromosome abnormalities. In *Acta Medicinae Legalis et Socialis* (pp. 87–90). Imprensa da Universidade de Coimbra. https://doi.org/10.14195/978-989-26-0173-1_15

Sim, J. E., Lee, H. Y., Yang, W. I., & Shin, K.-J. (2010). Population genetic study of four closely-linked X-STR trios in Koreans. *Molecular Biology Reports*, *37*(1), 333–337. https://doi.org/10.1007/s11033-009-9733-x

Skuse, D. H. (2005). X-linked genes and mental functioning. *Human Molecular Genetics*, *14*(suppl_1), R27–R32. https://doi.org/10.1093/hmg/ddi112

Spitzer, A., Elkayam Sapir, L., & Amiel, M. (2021). What is she doing here? Klinefelter syndrome in forensic casework. *Science & Justice*, *61*(4), 443–448. https://doi.org/10.1016/j.scijus.2021.03.003

Subramanian, S., Mishra, R. K., & Singh, L. (2003). Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biology*, *4*(2), R13. https://doi.org/10.1186/gb-2003-4-2-r13

Sun, J. X., Helgason, A., Masson, G., Ebenesersdóttir, S. S., Li, H., Mallick, S., Gnerre, S., Patterson, N., Kong, A., Reich, D., & Stefansson, K. (2012). A direct characterization of human mutation based on microsatellites. *Nature Genetics*, *44*(10), 1161–1165. https://doi.org/10.1038/ng.2398

Syndercombe Court, D. (2021). The Y chromosome and its use in forensic DNA analysis. *Emerging Topics in Life Sciences*, *5*(3), 427–441. https://doi.org/10.1042/ETLS20200339

Szibor, R. (2007). X-chromosomal markers: Past, present and future. *Forensic Science International: Genetics*, *1*(2), 93–99. https://doi.org/10.1016/j.fsigen.2007.03.003

Szibor, R., Hering, S., & Edelmann, J. (2006). A new Web site compiling forensic chromosome X research is now online. *International Journal of Legal Medicine*, *120*(4), 252–254. https://doi.org/10.1007/s00414-005-0029-y

Szibor, R., Hering, S., Kuhlisch, E., Plate, I., Demberger, S., Krawczak, M., & Edelmann, J. (2005). Haplotyping of STR cluster DXS6801–DXS6809–DXS6789 on Xq21 provides a powerful tool for kinship testing. *International Journal of Legal Medicine*, *119*(6), 363–369. https://doi.org/10.1007/s00414-005-0550-z

Szibor, R., Krawczak, M., Hering, S., Edelmann, J., Kuhlisch, E., & Krause, D. (2003). Use of X-linked markers for forensic purposes. *International Journal of Legal Medicine*, *117*(2), 67–74. https://doi.org/10.1007/s00414-002-0352-5

Tartaglia, N. R., Howell, S., Sutherland, A., Wilson, R., & Wilson, L. (2010). A review of trisomy X (47,XXX). *Orphanet Journal of Rare Diseases*, *5*(1), 8. https://doi.org/10.1186/1750-1172-5-8

Thompson, J. M., Ewing, M. M., Frank, W. E., Pogemiller, J. J., Nolde, C. A., Koehler, D. J., Shaffer, A. M., Rabbach, D. R., Fulmer, P. M., Sprecher, C. J., & Storts, D. R. (2013). Developmental validation of the PowerPlex® Y23 System: A single multiplex Y-STR analysis system for casework and database samples. *Forensic Science International: Genetics*, *7*(2), 240–250. https://doi.org/10.1016/j.fsigen.2012.10.013

Tillmar, A. O., Kling, D., Butler, J. M., Parson, W., Prinz, M., Schneider, P. M., Egeland, T., & Gusmão, L. (2017). DNA Commission of the International Society for Forensic Genetics (ISFG): Guidelines on the use of X-STRs in kinship analysis. *Forensic Science International: Genetics*, *29*, 269–275. https://doi.org/10.1016/j.fsigen.2017.05.005

Tomas, C., Pereira, V., & Morling, N. (2012). Analysis of 12 X-STRs in Greenlanders, Danes and Somalis using Argus X-12. *International Journal of Legal Medicine*, *126*(1), 121–128. https://doi.org/10.1007/s00414-011-0609-y

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), e115–e115. https://doi.org/10.1093/nar/gks596

Vanderheyden, N., Mai, A., Gilissen, A., Cassiman, J.-J., & Decorte, R. (2007). Identification and sequence analysis of discordant phenotypes between AmpFlSTR SGM Plus™ and PowerPlex® 16. *International Journal of Legal Medicine*, *121*(4), 297–301. https://doi.org/10.1007/s00414-007-0167-5

Veselinović, I., Vapa, D., Djan, M., Veličković, N., Veljović, T., & Petrić, G. (2018). Genetic analysis of 12 X-STR loci in the Serbian population from Vojvodina Province. *International Journal of Legal Medicine*, *132*(2), 405–408. https://doi.org/10.1007/s00414-017-1677-4

Vogt, P. H., Edelmann, A., Kirsch, S., Henegariu, O., Hirschmann, P., Kiesewetter, F., Köhn, F. M., Schill, W. B., Farah, S., Ramos, C., Hartmann, M., Hartschuh, W., Meschede, D., Behre, H. M., Castel, A., Nieschlag, E., Weidner, W., Gröne, H. J., Jung, A., … Haidl, G. (1996). Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Human Molecular Genetics*, *5*(7), 933–943. https://doi.org/10.1093/hmg/5.7.933

von Wurmb-Schwark, N., Petermann, S., & Wegener, R. (2003). Y-STR typing in forensic analysis. *International Congress Series*, *1239*, 487–490. https://doi.org/10.1016/S0531-5131(02)00591-5

Vrtel, P., Vrtel, R., Klaskova, E., Vrbicka, D., Adamova, K., Pavlicek, J., Hana, V., Hana, V., Soucek, O., Stara, V., Lebl, J., Snajdrova, M., Zapletalova, J., Furst, T., Kapralova, S., Tauber, Z., Krejcirikova, E., Routilova, M., Stellmachova, J., … Prochazka, M. (2022). Haplotype analysis of the X chromosome in patients with Turner syndrome in order to verify the possible effect of imprinting on selected symptoms. *Biomedical Papers*, *166*(1), 63–67. https://doi.org/10.5507/bp.2020.060

Walsh, P. S., Metzger, D. A., & Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, *10*(4), 506–513. http://www.ncbi.nlm.nih.gov/pubmed/1867860

Weber, J. L., & Wong, C. (1993). Mutation of human short tandem repeats. *Human Molecular Genetics*, *2*(8), 1123–1128. https://doi.org/10.1093/hmg/2.8.1123

Wendt, F. R., King, J. L., Novroski, N. M. M., Churchill, J. D., Ng, J., Oldt, R. F., McCulloh, K. L., Weise, J. A., Smith, D. G., Kanthaswamy, S., & Budowle, B. (2017). Flanking

region variation of ForenSeq™ DNA Signature Prep Kit STR and SNP loci in Yavapai Native Americans. *Forensic Science International: Genetics*, *28*, 146–154. https://doi.org/10.1016/j.fsigen.2017.02.014

Willems, T., Gymrek, M., Poznik, G. D., Tyler-Smith, C., & Erlich, Y. (2016). Population-Scale Sequencing Data Enable Precise Estimates of Y-STR Mutation Rates. *The American Journal of Human Genetics*, *98*(5), 919–933. https://doi.org/10.1016/j.ajhg.2016.04.001

Xing, J., Adnan, A., Rakha, A., Kasim, K., Noor, A., Xuan, J., Zhang, X., Yao, J., McNevin, D., & Wang, B. (2019). Genetic analysis of 12 X-STRs for forensic purposes in Liaoning Manchu population from China. *Gene*, *683*, 153–158. https://doi.org/10.1016/j.gene.2018.10.020

Yang, Y. L., Wang, J. G., Wang, D. X., Zhang, W. Y., Liu, X. J., Cao, J., & Yang, S. L. (2015). Analysis of an “off-ladder” allele at the Penta D short tandem repeat locus. *Genetics and Molecular Research*, *14*(4), 15096–15101. https://doi.org/10.4238/2015.November.24.18

Yuan, L., Chen, W., Zhao, D., Li, Y., Hao, S., Liu, Y., & Lu, D. (2019). Mutation analysis of 13 RM Y-STR loci in Han population from Beijing of China. *International Journal of Legal Medicine*, *133*(1), 59–63. https://doi.org/10.1007/s00414-018-1949-7

Zhang, S., Zhao, S., Zhu, R., & Li, C. (2012). Genetic polymorphisms of 12 X-STR for forensic purposes in Shanghai Han population from China. *Molecular Biology Reports*, *39*(5), 5705–5707. https://doi.org/10.1007/s11033-011-1379-9

Zhang, Y., Yang, Y., Li, P., & Guo, S. (2021). Origin of the X-chromosome influences the development and treatment outcomes of Turner syndrome. *PeerJ*, *9*, e12354. https://doi.org/10.7717/peerj.12354

Zhang, Y., Yu, Z., Mo, X., Zhao, X., Li, W., Liu, H., Liu, C., Wu, R., & Sun, H. (2021). Comparative evaluation of autosomal STRs and X-chromosome STRs as a complement of autosomal STRs in kinship testing in Southern Han Chinese. *Annals of Human Biology*, *48*(1), 66–69. https://doi.org/10.1080/03014460.2020.1856926

Ziętkiewicz, E., Witt, M., Daca, P., Żebracka-Gala, J., Goniewicz, M., Jarząb, B., & Witt, M. (2012). Current genetic methodologies in the identification of disaster victims and in forensic analysis. *Journal of Applied Genetics*, *53*(1), 41–60. https://doi.org/10.1007/s13353-011-0068-7

Attachments

Supplementary Table 1. The 31 father-mother-daughter trio samples selected to study the DNA sequence of 39.1 and 39.2 alleles in DXS10146. Genotyping was performed with the commercial multiplex kit Investigator® Argus X-12 QS (QIAGEN, Germany). Two different readings of the electropherogram results were done by two independent experts, referred here as “First assessment” and “Second assessment”.

Sample	Trio	Kinship	Allele of interest (First assessment)	Allele of interest (Second assessment)	Other allele (in case of heterozygosity)
N17	Trio 4	Mother	39.1	39.2	39.2 (homozygote)
N18		Daughter	39.1	39.2	28
N19	Trio 5	Father	39.2	39.2	-
N20		Mother	39.1	39.2	30
N21		Daughter	39.2	39.2	30
N47	Trio 28	Mother	39.1	39.1	27
N49		Daughter	39.1	39.1	27
N121	Trio 22	Mother	39.1	39.1	38.1
N123		Daughter	39.1	39.1	27
N178	Trio 36	Mother	39.2	39.1	28
N232	Trio 47	Mother	39.2	39.2	29
N345	Trio 66	Father	39.2	39.1	-
N346		Daughter	39.2	39.1	25
N349	Trio 67	Mother	39.2	39.1	28
N350		Daughter	39.2	39.1	28
N373	Trio 73	Mother	39.2	39.2	41.2
N375		Daughter	39.2	39.2	44.2
N397	Trio 80	Daughter	39.2	39.2	40.2
N398		Mother	39.2	39.2	29
N402	Trio 81	Mother	39.1	39.1	28
N427	Trio 84	Mother	39.2	39.1	28
Z7250	Trio 129	Father	39.2	39.2	-
Z7253		Daughter	39.2	39.2	26
Z7707	Trio 116	Daughter	39.2	39.2	31
Z7712		Father	39.2	39.2	-
Z7946	Trio 106	Mother	39.2	39.1	30
Z7947		Daughter	39.2	39.1	28
Z8288	Trio 101	Mother	39.2	39.2	42.2
Z8636	Trio 94	Father	39.2	39.2	-
Z8640		Mother	39.2	39.2	31
Z8641		Daughter	39.2	39.2	39.2 (homozygote)

Supplementary Table 2. Information about the duos chosen as samples to confirm mutations on commonly used Y-STRs. Genotyping was performed with the commercial multiplex kit PowerPlex® Y23 System (Promega, USA).

Father	Son	Y-STR	Father's genotype	Son's genotype
N229	N230	DYS19	14	15
Z6668	Z6670	DYS385	11,13	11,11
N128	N129	DYS389 I	12	13
N83	N58	DYS389 II	31	30
Z8295	Z8292	DYS390	24	23
N210	N209	DYS456	16	15
Z6746	Z7648		16	17
Z8387	Z8386	DYS458	18	17
N1	N2	DYS481	22	23
N258	N259		29	28
Z6565	Z6564	DYS549	13	12
N247	N248	DYS576	18	17
N313	N314		20	21
N442	N443		18	19
Z7251	Z7248		19	18
Z6658	Z6660	DYS635	22	21
N305	N306	Y-GATA-H4	12	11
N108	N107	DYS448	19	0
		DYS549	14	0
		DYS392	13	0

Supplementary Table 3. Reviewed primers in the literature for Y-STRs. The used primer pairs are in bold and italic.

Locus	Primer sequences 5'-3'	Reference
DYS576	F – GTTGGGCTGAGGAGTTCAATC	Alghafri, R., Goodwin, W., Ralf, A., Kayser, M., & Hadi, S. (2015). A novel multiplex assay for simultaneously analysing 13 rapidly mutating Y-STRs. Forensic Science International: Genetics, 17, 91–98. https://doi.org/10.1016/j.fsigen.2015.04.004
	R – GGCAGTCTCATTTCCTGGAG	
DYS481	F - TCTGTGAGAGTGTTGCGAGAGTTAG	Aliferi, A., Thomson, J., McDonald, A., Paynter, V. M., Ferguson, S., Vanhinsbergh, D., Syndercombe Court, D., & Ballard, D. (2018). UK and Irish Y-STR population data—A catalogue of variant alleles. Forensic Science International: Genetics, 34, e1–e6. https://doi.org/10.1016/j.fsigen.2018.02.018
	R - TGCCAGCATGTCTTGGCATACTTA	
DYS576	F - TGAGGAGTTCAATCTCAGCCAA	
	R - ATGGCAGTCTCATTTCCTGGAG	
DYS389 I/II	F - CCAACTCTCATCTGTATTATCTATG	Almeida, J. L., Hill, C. R., & Cole, K. D. (2011). Authentication of African green monkey cell lines using human short tandem repeat markers. BMC Biotechnology, 11(1), 102. https://doi.org/10.1186/1472-6750-11-102
	R - GTCTTATCTCCACCCACCAGA	
DYS481	F - AATGTGGCTAACGCTGTTCA	Asamura, H., Fujimori, S., Ota, M., Oki, T., & Fukushima, H. (2008). Evaluation of miniY-STR multiplex PCR systems for extended 16 Y-STR loci. International Journal of Legal Medicine, 122(1), 43–49. https://doi.org/10.1007/s00414-007-0193-3
	R - CCAGAAGGTTGCAAGACTCA	
DYS549	F - GGTAAAGAGGAAGATGATAGATGATTA	
	R - TCCCCTTTTCCATTTGTGA	
DYS19	F - CTACTGAGTTTCTGTTATAGT	Beleza, S., Alves, C., González-Neira, A., Lareu, M., Amorim, A., Carracedo, A., & Gusmão, L. (2003). Extending STR markers in Y chromosome haplotypes. International Journal of Legal Medicine, 117(1), 27–33. https://doi.org/10.1007/s00414-002-0317-8
	R - GGGTTAAGGAGAGTGTCACTA	
DYS19	F - ACTGAGTTTCTGTTATAGTGTTTTT	Bosch, E., Lee, A. C., Calafell, F., Arroyo, E., Henneman, P., de Knijff, P., & Jobling, M. A. (2002). High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. Forensic Science International, 125(1), 42–51. https://doi.org/10.1016/S0379-0738(01)00627-2
	R - ATGGCATGTAGTGAGGACA	
DYS389 I/II	***F - CCAACTCTCATCTGTATTATCTATGT***	
	R - CCTGAGTAGCAGAAGAATGTCATA	
DYS19	F - ACTACTGAGTTTCTGTTATAGTGTTTTT	Butler, J. M., Schoske, R., Vallone, P. M., Kline, M. C., Redd, A. J., & Hammer, M. F. (2002). A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. Forensic Science International, 129(1), 10–24. https://doi.org/10.1016/S0379-0738(02)00195-0
	R - GTCAATCTCTGCACCTGGAAAT	
DYS390	F - TATATTTTACACATTTTTGGGCCC	
	R - GTGACAGTAAAATGAAAACATTGC	
DYS481	F - GAATGTGGCTAACGCTGTTC	D'Amato, M. E., Ehrenreich, L., Cloete, K., Benjeddou, M., & Davison, S. (2010). Characterization of the highly discriminatory loci DYS449, DYS481, DYS518, DYS612, DYS626, DYS644 and DYS710. Forensic Science International: Genetics, 4(2), 104–110. https://doi.org/10.1016/j.fsigen.2009.06.011
	R - TCACCAGAAGGTTGCAAGAC	
DYS19	F - CTGAGTTTCTGTTATAGTGTTTTTTA	Edlund, H., & Allen, M. (2009). Y chromosomal STR analysis using Pyrosequencing technology. Forensic Science International: Genetics, 3(2), 119–124. https://doi.org/10.1016/j.fsigen.2008.11.010
	R - CTGGGTTAAGGAGAGTGTCA	
DYS389 I/II	F - CCAACTCTCATCTGTATTATCTATGT	
	R - GATAGATTGATAGAGGGAGGGA	
DYS389 I/II	F - GTATCCAACTCTCATCTGTATTATCTA	
	R - GATAGATTGATAGAGGGAGGGA	
DYS389 I/II	F - CAACTCTCATCTGTATTATCTATGTGTGT	
	R - GATAGATTGATAGAGGGAGGGA	
DYS456	F - CCCATCAACTCAGCCCAAAAC	Hanson, E. K., & Ballantyne, J. (2004). A highly discriminating 21 locus Y-STR "megaplex" system designed to augment the minimal haplotype loci for forensic casework. Journal of Forensic Sciences, 49(1), 40–51. http://www.ncbi.nlm.nih.gov/pubmed/14979342
	R - GGACCTTGTGATAATGTAAGATA	
DYS576	F - TTGGGCTGAGGAGTTCAATC	Javed, F., Sumbal, S., Shafique, M., Shahid, A. A., Shahzadi, A., Rani, N., Javid, H., Javed, M., Gillani, N., & Husnain, T. (2018). Male individualization using 12 rapidly

(cont.)	R - TTCCTGGAGATGAAGGAGGA	mutating Y-STRs in Araein ethnic group and shared paternal lineage of Pakistani population. International Journal of Legal Medicine, 132(6), 1621–1624. https://doi.org/10.1007/s00414-018-1851-3
DYS19	F - CTACTGAGTTTCTGTTATAGT	Kayser, M., Caglià, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., … Roewer, L. (1997). Evaluation of Y-chromosomal STRs: a multicenter study. International Journal of Legal Medicine, 110(3), 125–133. https://doi.org/10.1007/s004140050051
	R - ATGGCCATGTAGTGAGGACA	
DYS389 I/II	F - CCAACTCTCATCTGTATTATCTAT	
	R - TCTTATCTCCACCCACCAGA	
DYS481	***F - AGGAATGTGGCTAACGCTGT***	Kayser, M., Kittler, R., Erler, A., Hedman, M., Lee, A. C., Mohyuddin, A., Mehdi, S. Q., Rosser, Z., Stoneking, M., Jobling, M. A., Sajantila, A., & Tyler-Smith, C. (2004). A Comprehensive Survey of Human Y-Chromosomal Microsatellites. The American Journal of Human Genetics, 74(6), 1183–1197. https://doi.org/10.1086/421531
	R - ACAGCTCACCAGAAGGTTGC	
DYS549	F - AACCAAATTCAGGGATGTACTGA	
	R - GTCCCCTTTTCCATTTGTGA	
DYS576	F - TTGGGCTGAGGAGTTCAATC	
	R - GGCAGTCTCATTTCCTGGAG	
DYS19	F - CCATCTGGGTTAAGGAGAGTGT	Kwon, S. Y., Lee, H. Y., Kim, E. H., Lee, E. Y., & Shin, K.-J. (2016). Investigation into the sequence structure of 23 Y chromosomal STR loci using massively parallel sequencing. Forensic Science International: Genetics, 25, 132–141. https://doi.org/10.1016/j.fsigen.2016.08.010
	R - TTCACTATGACTACTGAGTTTCTGTT	
DYS389 I/II	F - CCAACTCTCATCTGTATTATCTATGTG	
	R - GATAGATTGATAGAGGGAGGGA	
DYS390	F - GTGTATACTCAGAAACAAGGAAAGA	
	R - CTGCATTTTGGTACCCCATA	
DYS456	F - CTGTTGTGGGACCTTGTGATA	
	R - ACTCAGCCCAAAACTTCTTAAA	
DYS481	F - CAGAAGGTTGCAAGACTCAAA	
	R - AGGAATGTGGCTAACGCTGT	
DYS549	F - GTCCCCTTTTCCATTTGTGA	
	R - GCAATTAGGTAGGTAAAGAGGAAGA	
DYS576	F - GCGTATTTGTCTTGGCTTTTT	
	R - CATAGCAAGACCTCATCTCTGAA	
DYS635	F - TGGCTTCTCACTTTGCATAGAA	
	R - GTGGAACCAGCCCAAATATC	
DYS481	F - AAAAGGAATGTGGCTAACGCTGTTC	Leat, N., Ehrenreich, L., Benjeddou, M., Cloete, K., & Davison, S. (2007). Properties of novel and widely studied Y-STR loci in three South African populations. Forensic Science International, 168(2–3), 154–161. https://doi.org/10.1016/j.forsciint.2006.07.009
	R - GCTCACCAGAAGGTTGCAAGACTCA	
DYS635	F - GCACTGTATTTCAGCTTGAGTGATGG	
	R - CTCTTGGCTTCTCACTTTGCATAGAAT	
DYS576	F - GCGTATTTGTCTTGGCTTTTTC	Lee, E. Y., Lee, H. Y., Kwon, S. Y., Oh, Y. N., Yang, W. I., & Shin, K.-J. (2017). A multiplex PCR system for 13 RM Y-STRs with separate amplification of two different repeat motif structures in DYF403S1a. Forensic Science International: Genetics, 26, 85–90. https://doi.org/10.1016/j.fsigen.2016.10.019
	R - TCAATCTCAGCCAAGCAACA	
DYS456	F - GCCCTCAAACATTGGACTC	Núñez, C., Baeta, M., Ibarbia, N., Ortueta, U., Jiménez-Moreno, S., Blazquez-Caeiro, J. L., Builes, J. J., Herrera, R. J., Martínez-Jarreta, B., & de Pancorbo, M. M. (2017). 17 to 23: A novel complementary mini Y-STR panel to extend the Y-STR databases from 17 to 23 markers for forensic purposes. ELECTROPHORESIS, 38(7), 1016–1021. https://doi.org/10.1002/elps.201600313
	R - ATCAACTCAGCCCAAAACTT	
DYS481	F - AGGAATGTGGCTAACGCTG	
	R - ACAGCTCACCAGAAGGTTG	
DYS549	***F - GTAGGTAAAGAGGAAGATGATAGATG***	
	R - GGTGTTCAGAATAGTCTCTAAAGGTTT	
DYS576	F - AGGAGTTCAATCTCAGCCAA	
	R - GGAGATGGGAGTAATAAGCGT	

DYS481	F - CTCACCAGAAGGTTGCAAGAC	Oh, Y. N., Lee, H. Y., Lee, E. Y., Kim, E. H., Yang, W. I., & Shin, K.-J. (2015). Haplotype and mutation analysis for newly suggested Y-STRs in Korean father–son pairs. Forensic Science International: Genetics, 15, 64–68. https://doi.org/10.1016/j.fsigen.2014.09.023
	R - AGGAATGTGGCTAACGCTGT	
DYS549	F - TGTGTGCATAGAGGTGTTCAGA	
	R - AGGTAGGTAAAGAGGAAGATGA	
DYS576	F - GCGTATTTGTCTTGGCTTTTTC	
	R - GCAAGACCTCATCTCTGAATAAAA	
DYS576	F - CAATCTCAGCCAAGCAACATAG	Palha, T., Ribeiro-Rodrigues, E., Ribeiro-dos-Santos, Â., & Santos, S. (2012). Fourteen short tandem repeat loci Y chromosome haplotypes: Genetic analysis in populations from northern Brazil. Forensic Science International: Genetics, 6(3), 413–418. https://doi.org/10.1016/j.fsigen.2011.08.003
	R - GAAAAACCCAACACCATGATTC	
DYS635	F - AGTGTCTCACTTCAAGCACCAA	
	R - ACAGTTGGAAAAATGTGGAACC	
DYS390	F - CTGCATTTTGGTACCCCATA	Park, M. J., Lee, H. Y., Chung, U., Kang, S.-C., & Shin, K.-J. (2007). Y-STR analysis of degraded DNA using reduced-size amplicons. International Journal of Legal Medicine, 121(2), 152–157. https://doi.org/10.1007/s00414-006-0133-7
	R - GCAATGTGTATACTCAGAAACAAGG	
DYS635	F - GGCTTCTCACTTTGCATAGAATC	
	R - ACCAGCCCAAATATCCATCA	
DYS389 I/II	F - CCAACTCTCATCTGTATTATCTATG	Pestoni, C., Cal, M. L., Lareu, M. V., Rodríguez-Calvo, M. S., & Carracedo, A. (1998). Y chromosome STR haplotypes: genetic and sequencing data of the Galician population (NW Spain). International Journal of Legal Medicine, 112(1), 15–21. https://doi.org/10.1007/s004140050191
	R - TCTTATCTCCACCCACCAGA	
DYS456	F - GGACCTTGTGATAATGTAAGATA	Redd, A. J., Agellon, A. B., Kearney, V. A., Contreras, V. A., Karafet, T., Park, H., de Knijff, P., Butler, J. M., & Hammer, M. F. (2002). Forensic value of 14 novel STRs on the human Y chromosome. Forensic Science International, 130(2–3), 97–111. https://doi.org/10.1016/S0379-0738(02)00347-X
	R - CCCATCAACTCAGCCCAAAAC	
DYS19	F - CTACTGAGTTTCTGTTATAGT	Roewer, L., & Epplen, J. T. (1992). Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. Forensic Science International, 53(2), 163–171. https://doi.org/10.1016/0379-0738(92)90193-Z
	R - ATGGCATGTAGTGAGGACA	
DYS389 I/II	F - TCTGTATTATCTATGTGTG	Redd, A. J., Agellon, A. B., Kearney, V. A., Contreras, V. A., Karafet, T., Park, H., de Knijff, P., Butler, J. M., & Hammer, M. F. (2002). Forensic value of 14 novel STRs on the human Y chromosome. Forensic Science International, 130(2–3), 97–111. https://doi.org/10.1016/S0379-0738(02)00347-X
	R - CCAGACATTGCCAAGTGTTACTTG	
DYS389 I/II	F - TCATCTGTATTATCTATGTGTG	Rossi, E., Rolf, B., Schürenkamp, M., & Brinkmann, B. (1998). Y-Chromosome STR haplotypes in an Italian population sample. International Journal of Legal Medicine, 112(1), 78–81. https://doi.org/10.1007/s004140050206
	R - TCTTATCTCCACCCACCAGA	
DYS456	F - GGACCTTGTGATAATGTAAGATAG	Schoske, R., Vallone, P. M., Kline, M. C., Redman, J. W., & Butler, J. M. (2004). High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays. Forensic Science International, 139(2–3), 107–121. https://doi.org/10.1016/j.forsciint.2003.10.007
	R - GTAGAGGGACAGAACTAATGGAA	
DYS389 I/II	F - CCAACTCTCATCTGTATTATCTATGTG	Thomas, M. G., Bradman, N., & Flinn, H. M. (1999). High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y-chromosome. Human Genetics, 105(6), 577–581. https://doi.org/10.1007/s004399900181
	R - CCTGAGTAGCAGAAGAATGTCATA	
DYS481	F – AGGAATGTGGCTAACGCTGT	Vermeulen, M., Wollstein, A., van der Gaag, K., Lao, O., Xue, Y., Wang, Q., Roewer, L., Knoblauch, H., Tyler-Smith, C., de Knijff, P., & Kayser, M. (2009). Improving global and regional resolution of male lineage differentiation by simple single-copy Y-chromosomal short tandem repeat polymorphisms. Forensic Science International: Genetics, 3(4), 205–213. https://doi.org/10.1016/j.fsigen.2009.01.009
	R - GACAGCTCACCAGAAGGTTGC	

DYS390	F - GTACCAGTAAGTGCACGTTTAA	Watahiki, H., Fujii, K., Fukagawa, T., Mita, Y., Kitayama, T., & Mizuno, N. (2019). Polymorphisms and microvariant sequences in the Japanese population for 25 Y-STR markers and their relationships to Y-chromosome haplogroups. Forensic Science International: Genetics, 41, e1–e7. https://doi.org/10.1016/j.fsigen.2019.03.004
	R - ATGGGATGGGAAATGATGTTTC	
DYS635	F - AGTGTCTCACTTCAAGCACCAAGCAC	White, P. S., Tatum, O. L., Deaven, L. L., & Longmire, J. L. (1999). New, Male-Specific Microsatellite Markers from the Human Y Chromosome. Genomics, 57(3), 433–437. https://doi.org/10.1006/geno.1999.5782
	R - GCAGCAAAATTCACAGTTGGAAAAATGT	
DYS576	F - AGCAACATAGCAAGACCTCAT	Xiong, C., Yang, C., Wu, W., Zeng, Y., Lin, T., Chen, L., Liu, H., Liu, C., Du, W., Wang, M., Sun, H., & Liu, C. (2022). Development and validation of a multiplex typing system with 32 Y-STRs for forensic application. Forensic Science International, 339, 111409. https://doi.org/10.1016/j.forsciint.2022.111409
	R - TAATAAGCGTATTTGTCTTGGC	
DYS576	***F - TCTCAGCCAAGCAACATAGC***	Yuan, L., Chen, W., Zhao, D., Li, Y., Hao, S., Liu, Y., & Lu, D. (2019). Mutation analysis of 13 RM Y-STR loci in Han population from Beijing of China. International Journal of Legal Medicine, 133(1), 59–63. https://doi.org/10.1007/s00414-018-1949-7
	R - AGGAGATGGGAGTAATAAGCGT	
DYS19	F - AGGTATGAGATCAAATTGACTGTG	Zhang, S., Tian, H., Wang, Z., Zhao, S., Hu, Z., Li, C., & Ji, C. (2014). Development of a new 26plex Y-STRs typing system for forensic application. Forensic Science International: Genetics, 13, 112–120. https://doi.org/10.1016/j.fsigen.2014.06.015
	R - CCAGGAGTAATACTTCGGGCCAT	
DYS389 I/II	F - GTATCCAACTCTCATCTGTATTATCTATGT	
	R - ACAATTATCCCTGAGTAGCAGAAGAATG	
DYS390	F - ATTACATTCACACATATATTTTACAC	
	R - TAAAATGAAAACATTGCAATGTGT	
DYS456	F - TTGTGGGACCTTGTGATAATG	
	R - GGACAGAACTAATGGAATATCTAT	
DYS481	F - GTGTCTGTCCCTTTAAGAGGAGTCTG	
	R - AAATCAGAACACAGAGCCCCACAAC	
DYS549	F - TTAATACAACAAAAATTTGGTAATCTG	
	R - GTGATTTTGTTTATGTAGATTTTTTC	
DYS576	F - TTTTTAATGTATGAGCAAGAATATCCTAG	
	R - GGATTATGGGAGCTAGAATTCAAGATGAG	
DYS635	F - ATGCCCAATGGAATGCTCTCTTGGCT	
	R - TCTCAAACAACAAAAACACAAAAAATGAA	
DYS576	F - CTCAGCCAAGCAACATAGCA	Zhang, W., Xiao, C., Yu, J., Wei, T., Liao, F., Wei, W., & Huang, D. (2017). Multiplex assay development and mutation rate analysis for 13 RM Y-STRs in Chinese Han population. International Journal of Legal Medicine, 131(2), 345–350. https://doi.org/10.1007/s00414-016-1489-y
	R - GTTCCTGGAGATGAAGGAGGA	
DYS19	F - AATTTGCTGGTCAATCTCTGCA	Zhao, X., Ma, K., Li, H., Cao, Y., Liu, W., Zhou, H., & Ping, Y. (2015). Multiplex Y-STRs analysis using the ion torrent personal genome machine (PGM). Forensic Science International: Genetics, 19, 192–196. https://doi.org/10.1016/j.fsigen.2015.06.012
	R - ACTATGACTACTGAGTTTCTGTTATAGTGT	
DYS389 I/II	F - AATTGTTTCCAGACATTGCCAAGT	
	R - TGTATCCAACTCTCATCTGTATTATCTATGTGT	
DYS390	F - GCCCTGCATTTTGGTACCCC	
	R - TGCAATGTGTATACTCAGAAACAAGG	
DYS456	F - GTCTGTTGTGGGACCTTGTGATAATG	
	R - CTATACCTAGAAAACCCCATCAACTC	
DYS635	F - CCAGCCCAAATATCCATCAATCAATG	
	R - CAATGGAATGCTCTCTTGGCTTCTC	

A
GAAGACAAGGACTCAGGAATTTGCTGGTCAATCTCTGCACCTGGAAATAGTGGCTGGGGCACCAGGAG
TAATACTTCGGGCCATGGCCATGTAGTGAGGACAAGGAGTCCATCTGGGTTAAGGAGAGTGTCACTAT
A**TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTACCTATCTATCTATCTA**AAA
CACTATATATATATAACACTATATATATAATACTATATATATATTAAAAAACACTATAACAGAAACTC
AGTAGTCATAGTGAAATCAAAAAATAATCACAGTCAATTTGATCTCATACCTAGACTGAAATATGAAA
CTTCAAAAGAAAAGAATGTTAAGAACTTTGGGCTTGTCAA

B
CCTGAGTAGCAGAAGAATGTCATAGATAGATGATGGACTGCTAGATAAATAGATAGATTGATAGAGGG
AGGGA**TAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGA**TACATAGATAATACA
GATGAGAGTTGGATACAGAAGTAGGTATAATGA**TAGATAGATAGATAGATAGATAGATAGATAGATAG**
ATAGATAGATAGACAGACAGACAGACAGACAGACACACACATAGATAATACAGATGAGAGTTGG

C
GTATCCGCCATGGTAGCATAATAGAAATTTTATGAGTGGGAGAAATGGATGACAGTAAAATGAAAACA
TTGCAATGTGTATACTCAGAAACAAGGAAAGATAGATAGATGA**TAGATAGATAGATAGACAGATAGAT**
AGATAGATAGATAGATAGATAGATAGATAGATAGATAGACAGACAGACAGACAGACAGACAGACAGAC
AGATAGATAGAATATATTATGGGGTACCAAAATGCAGGGCCCAAAAATGTGTAAA

D
TCTGTTGTGGGACCTTGTGATAATGTA**AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**
GATAGATAGATAGATAGATATTCCATTAGTTCTGTCCCTCTAGAGAACCCTAATACATCAGTTTAAGA
AGTTTTGGGCTGAGTTGA

E
GCAACAGGAATGAAACTCCAATGAAAGAAAGAAAAGGAAG**GAAAGAAAGAAAGAAAGAAAGAAAGAAA**
GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGAGGGTGGGCGTGGTGGCTCATGCTTGTAAT
GCCAGAAC

F
AGGAATGTGGCTAACGCTGTTCAGCATGCTG**CTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC**
TTCTTCTTCTTCTTCTTCTTCTTCTTTTTTTGAGTCTTGCAACCTTCTGGTGAGCTGT

G
GTAGGTAAAGAGGAAGATGATAGATGATTAGAAAGAT**GATAGATAGATAGATAGATAGATAGATAGAT**
AGATAGATAGATAGATAGATAGAAAAAATCTACATAAACAAAATCACAAATGGAAAAGGGGACATTAC
CACTTATGCCACCAAAAAAAACCTTTAGAGACTATTCTGAACACC

H
TCTCAGCCAAGCAACATAGCAAGACCTCATCTCTGAATA**AAAGAAAGAAAGAAAGAAAGAAAGAAAGA**
AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAAAGCCAAGACAAATACGCTTATTACT
CCCATCTCCT

I
GCAGCAAAATTCACAGTTGGAAAAATGTGGAACCAGCCCAAATATCCATCAATCAATGAATGGATAAA
GAAAATGTGA**TAGATAGATAGATAGATAGATAGATAGATAGATAGATACATACATAGATAGATACATA**
CATAGATAGATACATACATAGATAGATAGATAGAGATTCTATGCAAAGTGAGAAGCCAAGAGAGCATT
CCATTGGGCA

J
CCTCTGATGGTGAAGTAATGGAATTAGAAATAAATAAGAAAATCTACAAACTACTTATGTTTAAATAT
TAGTAAAATTTTACATAGCCCACTTGTTAAACAACTTAACAGGATAAATCACCTATCTATGTATCTAT
CTATCTATCTATTCATCCATCTAATCTATCCAT**TCTATCTATCTATCTATCTATCTATCTATCTATCT**
ATCTATCTATCTACCTACCTACCTATCTATCTATAGATCTATCTATCTATCTTAAATTTGGAAATTCT
CCTCAGCATAAC

Supplementary Figure 1. Sequence amplified with the designed primer pairs for various Y-STRs: DYS19 (A), DYS389 I/II (B), DYS390 (C), DYS456 (D), DYS458 (E), DYS481 (F), DYS549 (G), DYS576 (H), DYS635 (I), Y-GATA-H4 (J). The sequence used as reference is from the GRCh38.p14 genome assembly. The forward and reverse primers are underlined, and the repeat motif is in bold.

Supplementary Table 4. Populations used for comparing frequency data for 39.2 alleles in DXS10146. Frequencies were rounded to three decimals.

Population	Frequency of allele 39.2	Reference
Albania	0.073	Poulsen, L., Tomas, C., Drobnič, K., Ivanova, V., Mogensen, H. S., Kondili, A., Miniati, P., Bunokiene, D., Jankauskiene, J., Pereira, V., & Morling, N. (2016). NGMSElectTM and Investigator ® Argus X-12 analysis in population samples from Albania, Iraq, Lithuania, Slovenia, and Turkey. Forensic Science International: Genetics, 22, 110–112. https://doi.org/10.1016/j.fsigen.2016.02.004
Algeria	0.013	Bekada, A., Benhamamouch, S., Boudjema, A., Fodil, M., Menegon, S., Torre, C., & Robino, C. (2010). Analysis of 21 X-chromosomal STRs in an Algerian population sample. International Journal of Legal Medicine, 124(4), 287–294. https://doi.org/10.1007/s00414-009-0397-9
Bahrain	0.019	Al-Snan, N. R., Messaoudi, S. A., Mansoor, L. A., & Bakhiet, M. (2021). Population Genetic Analysis of 12 X-Chromosomal Short Tandem Repeats in a Bahraini Population Sample. Forensic Genomics, 1(1), 27–37. https://doi.org/10.1089/forensic.2020.0003
Bangladesh	0.003	Sufian, A., Hosen, Md. I., Fatema, K., Hossain, T., Hasan, Md. M., Mazumder, A. K., & Akhteruzzaman, S. (2017). Genetic diversity study on 12 X-STR loci of investigator® Argus X STR kit in Bangladeshi population. International Journal of Legal Medicine, 131(4), 963–965. https://doi.org/10.1007/s00414-016-1513-2
Cape Verde	0.046	Afonso Costa, H., Morais, P., Vieira da Silva, C., Matos, S., Marques Santos, R., Espinheira, R., Costa Santos, J., & Amorim, A. (2014). X-Chromosome STR markers data in a Cabo Verde immigrant population of Lisboa. Molecular Biology Reports, 41(4), 2559–2569. https://doi.org/10.1007/s11033-014-3114-9
China	0.000	Uchigasaki, S., Tie, J., & Takahashi, D. (2013). Genetic analysis of twelve X-chromosomal STRs in Japanese and Chinese populations. Molecular Biology Reports, 40(4), 3193–3196. https://doi.org/10.1007/s11033-012-2394-1
Croatia	0.045	Mršić, G., Ozretić, P., Crnjac, J., Merkaš, S., Sukser, V., Račić, I., Rožić, S., Barbarić, L., Popović, M., & Korolija, M. (2018). Expanded Croatian 12 X-STR loci database with an overview of anomalous profiles. Forensic Science International: Genetics, 34, 249–256. https://doi.org/10.1016/j.fsigen.2018.03.004
Czech Republic	0.046	Zidkova, A., Capek, P., Horinek, A., & Coufalova, P. (2014). Investigator® Argus X-12 study on the population of Czech Republic: Comparison of linked and unlinked X-STRs for kinship analysis. ELECTROPHORESIS, 35(14), 1989–1992. https://doi.org/10.1002/elps.201400046
Denmark	0.028	Tomas, C., Pereira, V., & Morling, N. (2012). Analysis of 12 X-STRs in Greenlanders, Danes and Somalis using Argus X-12. International Journal of Legal Medicine, 126(1), 121–128. https://doi.org/10.1007/s00414-011-0609-y
East Timor	0.000	Moreira, H., Costa, H., Tavares, F., & Souto, L. (2015). Genetic variation of 12 X-chromosomal STR loci in an East Timor sample. International Journal of Legal Medicine, 129(2), 257–258. https://doi.org/10.1007/s00414-014-1126-6
Ecuador	0.017	Gaviria, A., Boada, L., Rodríguez-Pólit, C., Vela, M., Fiallos, G., Gruezo, C., Paz-y-Miño, C., & Zambrano, A. K. (2019). Genetic polymorphisms of 12 X-STRs in the Ecuadorian population. Forensic Science International: Genetics Supplement Series, 7(1), 677–679. https://doi.org/10.1016/j.fsigss.2019.10.135
Eritrea	0.017	Bini, C., Sarno, S., Tangorra, E., Iuvaro, A., De Fanti, S., Tseghereda, Y. G., Pelotti, S., & Luiselli, D. (2021). Haplotype data and forensic evaluation of 23 Y-STR and 12 X-STR loci in eight ethnic groups from Eritrea. International Journal of Legal Medicine, 135(2), 449–453. https://doi.org/10.1007/s00414-020-02446-2

Ethiopia	0.036	Edelmann, J., Hering, S., Augustin, C., & Szibor, R. (2008). Haplotypes and haplotype stability within a 126.6kb region at Xq28. Forensic Science International: Genetics Supplement Series, 1(1), 554–556. https://doi.org/10.1016/j.fsigss.2007.10.028
Germany	0.036	Edelmann, J., Lutz-Bonengel, S., Naue, J., & Hering, S. (2012). X-chromosomal haplotype frequencies of four linkage groups using the Investigator Argus X-12 Kit. Forensic Science International: Genetics, 6(1), e24–e34. https://doi.org/10.1016/j.fsigen.2011.01.001
Greenland	0.000	Tomas, C., Pereira, V., & Morling, N. (2012). Analysis of 12 X-STRs in Greenlanders, Danes and Somalis using Argus X-12. International Journal of Legal Medicine, 126(1), 121–128. https://doi.org/10.1007/s00414-011-0609-y
Guinea-Bissau	0.027	Gomes, I., Pereira, P. J. P., Harms, S., Oliveira, A. M., Schneider, P. M., & Brehm, A. (2017). Genetic characterization of Guinea-Bissau using a 12 X-chromosomal STR system: Inferences from a multiethnic population. Forensic Science International: Genetics, 31, 89–94. https://doi.org/10.1016/j.fsigen.2017.08.016
Hungary	0.020	Horváth, G., Zalán, A., Kis, Z., & Pamjav, H. (2012). A genetic study of 12 X-STR loci in the Hungarian population. Forensic Science International: Genetics, 6(1), e46–e47. https://doi.org/10.1016/j.fsigen.2011.03.007
Iran	0.019	Poulsen, L., Farzad, M. S., Børsting, C., Tomas, C., Pereira, V., & Morling, N. (2015). Population and forensic data for three sets of forensic genetic markers in four ethnic groups from Iran: Persians, Lurs, Kurds and Azeris. Forensic Science International: Genetics, 17, 43–46. https://doi.org/10.1016/j.fsigen.2015.03.010
Iraq	0.000	Poulsen, L., Tomas, C., Drobnič, K., Ivanova, V., Mogensen, H. S., Kondili, A., Miniati, P., Bunokiene, D., Jankauskiene, J., Pereira, V., & Morling, N. (2016). NGMSElectTM and Investigator ® Argus X-12 analysis in population samples from Albania, Iraq, Lithuania, Slovenia, and Turkey. Forensic Science International: Genetics, 22, 110–112. https://doi.org/10.1016/j.fsigen.2016.02.004
Italy	0.050	Bini, C., Riccardi, L. N., Ceccardi, S., Carano, F., Sarno, S., Luiselli, D., & Pelotti, S. (2015). Expanding X-chromosomal forensic haplotype frequencies database: Italian population data of four linkage groups. Forensic Science International: Genetics, 15, 127–130. https://doi.org/10.1016/j.fsigen.2014.11.008
Ivory Coast	0.020	Pasino, S., Caratti, S., Del Pero, M., Santovito, A., Torre, C., & Robino, C. (2011). Allele and haplotype diversity of X-chromosomal STRs in Ivory Coast. International Journal of Legal Medicine, 125(5), 749–752. https://doi.org/10.1007/s00414-011-0591-4
Japan	0.000	Uchigasaki, S., Tie, J., & Takahashi, D. (2013). Genetic analysis of twelve X-chromosomal STRs in Japanese and Chinese populations. Molecular Biology Reports, 40(4), 3193–3196. https://doi.org/10.1007/s11033-012-2394-1
Lithuania	0.037	Poulsen, L., Tomas, C., Drobnič, K., Ivanova, V., Mogensen, H. S., Kondili, A., Miniati, P., Bunokiene, D., Jankauskiene, J., Pereira, V., & Morling, N. (2016). NGMSElectTM and Investigator ® Argus X-12 analysis in population samples from Albania, Iraq, Lithuania, Slovenia, and Turkey. Forensic Science International: Genetics, 22, 110–112. https://doi.org/10.1016/j.fsigen.2016.02.004
Malaysia	0.000	Samejima, M., Nakamura, Y., Nambiar, P., & Minaguchi, K. (2012). Genetic study of 12 X-STRs in Malay population living in and around Kuala Lumpur using Investigator Argus X-12 kit. International Journal of Legal Medicine, 126(4), 677–683. https://doi.org/10.1007/s00414-012-0705-7
Mexico	0.013	Cortés-Trujillo, I., Zuñiga-Chiquette, F., Ramos-González, B., Chávez-Briones, M. de L., Islas-González, K. L., Betancourt-Guerra, D. A., Peralta-Coria, R., Martínez-Cortés, G., & Rangel-Villalobos, H. (2019). Allele and haplotype frequencies of 12 X-STRs in Mexican population. Forensic Science International: Genetics, 38, e11–e13. https://doi.org/10.1016/j.fsigen.2018.10.012

Mongolia	0.004	Tao, R., Zhang, J., Bian, Y., Dong, R., Liu, X., Jin, C., Zhu, R., Zhang, S., & Li, C. (2018). Investigation of 12 X-STR loci in Mongolian and Eastern Han populations of China with comparison to other populations. Scientific Reports, 8(1), 4287. https://doi.org/10.1038/s41598-018-22665-3
Morocco	0.020	Bentayebi, K., Picornell, A., Bouabdeallah, M., Castro, J. A., Aboukhalid, R., Squalli, D., Misericòrdia, M., & Amzazi, S. (2012). Genetic diversity of 12 X-chromosomal short tandem repeats in the Moroccan population. Forensic Science International: Genetics, 6(1), e48–e49. https://doi.org/10.1016/j.fsigen.2011.03.008
Philippines	0.000	Salvador, J. M., Apaga, D. L. T., Delfin, F. C., Calacal, G. C., Dennis, S. E., & De Ungria, M. C. A. (2018). Filipino DNA variation at 12 X-chromosome short tandem repeat markers. Forensic Science International: Genetics, 36, e8–e12. https://doi.org/10.1016/j.fsigen.2018.06.008
Portugal	0.042	Faustino, M. M. (2021). Polymorphisms and mutation rate estimates of Y-STRs and X-STRs in the Portuguese population [Master's Degree in Forensic Genetics]. Faculty of Sciences of the University of Porto.
Saudi Arabia	0.058	Messaoudi, S. A., Babu, S. R., Alsaleh, A. B., Albujja, M., Al-Snan, N. R., Chaudhary, A. R., Kassab, A. Ch., & Assidi, M. (2021). Population genetic data for 12 X-STR loci in the Central Saudi region using investigator Argus X-12 amplification kit. Annals of Human Biology, 48(4), 321–326. https://doi.org/10.1080/03014460.2021.1957147
Serbia	0.033	Veselinović, I., Vapa, D., Djan, M., Veličković, N., Veljović, T., & Petrić, G. (2018). Genetic analysis of 12 X-STR loci in the Serbian population from Vojvodina Province. International Journal of Legal Medicine, 132(2), 405–408. https://doi.org/10.1007/s00414-017-1677-4
Siberia	0.081	Semikhodskii, A., Krassotkin, Y., Makarova, T., Zavarin, V., Ilina, V., & Sutyagina, D. (2021). Genetic diversity and forensic parameters of 12 X-STR included in Argus X-12® marker panel in the population of the Russian Federation. Annals of Human Biology, 48(5), 430–436. https://doi.org/10.1080/03014460.2021.1980104
Slovenia	0.022	Poulsen, L., Tomas, C., Drobnič, K., Ivanova, V., Mogensen, H. S., Kondili, A., Miniati, P., Bunokiene, D., Jankauskiene, J., Pereira, V., & Morling, N. (2016). NGMSElectTM and Investigator ® Argus X-12 analysis in population samples from Albania, Iraq, Lithuania, Slovenia, and Turkey. Forensic Science International: Genetics, 22, 110–112. https://doi.org/10.1016/j.fsigen.2016.02.004
Somalia	0.064	Tomas, C., Pereira, V., & Morling, N. (2012). Analysis of 12 X-STRs in Greenlanders, Danes and Somalis using Argus X-12. International Journal of Legal Medicine, 126(1), 121–128. https://doi.org/10.1007/s00414-011-0609-y
South Korea	0.000	Sim, J. E., Lee, H. Y., Yang, W. I., & Shin, K.-J. (2010). Population genetic study of four closely-linked X-STR trios in Koreans. Molecular Biology Reports, 37(1), 333–337. https://doi.org/10.1007/s11033-009-9733-x
Spain	0.044	Susana, J.-M., Celia, G., Ester, L., Miriam, B., Eladio, B., & de Pancorbo Marian, M. (2019). Genetic analysis of 12 X-chromosomal STRs an autochthonous population of Southeast Spain. Forensic Science International: Genetics Supplement Series, 7(1), 482–484. https://doi.org/10.1016/j.fsigss.2019.10.060
Switzerland	0.045	Bottinelli, M., Gouy, A., Utz, S., & Zieger, M. (2022). Population genetic analysis of 12 X-chromosomal STRs in a Swiss sample. International Journal of Legal Medicine, 136(2), 561–563. https://doi.org/10.1007/s00414-021-02684-y
Taiwan	0.001	Chen, M.-Y., Ho, C.-W., Pu, C.-E., & Wu, F.-C. (2014). Genetic polymorphisms of 12 X-chromosomal STR loci in Taiwanese individuals and likelihood ratio calculations applied to case studies of blood relationships. ELECTROPHORESIS, 35(12–13), 1912–1920. https://doi.org/10.1002/elps.201300645

Thailand	0.006	Vongpaisarnsin, K., Boonlert, A., Rasmeepaisarn, K., & Dangkao, P. (2016). Genetic variation study of 12 X chromosomal STR in central Thailand population. International Journal of Legal Medicine, 130(6), 1497–1499. https://doi.org/10.1007/s00414-016-1363-y
Turkey	0.028	Kalaoglu, E., Ay, M., Ulubay, A., Canan, H., Serin, A., & Alper, B. (2019). Population data of 12 X-STRs in Turkey. Forensic Science International: Genetics Supplement Series, 7(1), 632–634. https://doi.org/10.1016/j.fsigss.2019.10.118
United Arab Emirates	0.032	Almarri, M. A., & Lootah, R. A. (2018). Allelic and haplotype diversity of 12 X-STRs in the United Arab Emirates. Forensic Science International: Genetics, 33, e4–e6. https://doi.org/10.1016/j.fsigen.2017.12.013