

Polymorphisms and mutation rate estimates of Y- STRs and X-STRs in the Portuguese population

Marisa Mariano Faustino

Mestrado em Genética Forense
Departamento de Biologia
2021

Orientador

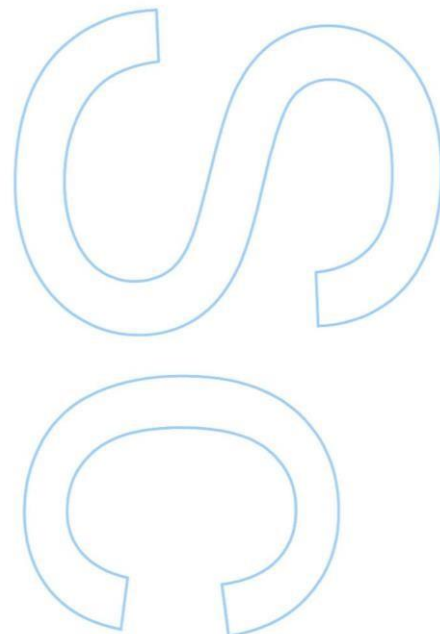
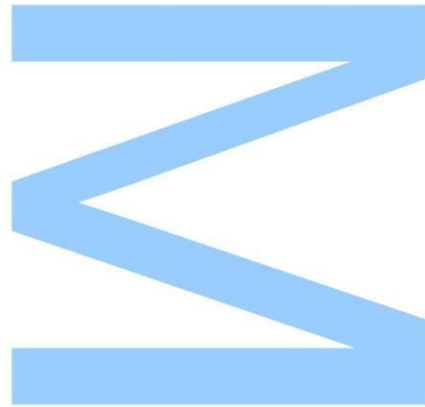
Doutora Nádia Pinto
Institute of Molecular Pathology and Immunology of the University of Porto
(IPATIMUP), Porto, Portugal
Institute of Research and Innovation in Health, I3S, University of Porto, Portugal
Center for Mathematics, Faculty of Sciences of the University of Porto (FCUP),
Portugal

Coorientador

Doutora Leonor Gusmão
DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ),
Brazil

Supervisor

Mestre Sofia Antão Sousa
Institute of Molecular Pathology and Immunology of the University of Porto
(IPATIMUP), Porto, Portugal
Institute of Research and Innovation in Health, I3S, University of Porto, Portugal
Faculty of Sciences of the University of Porto (FCUP), Portugal
DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ),
Brazil

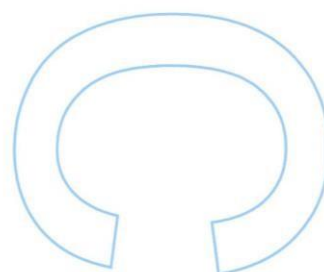
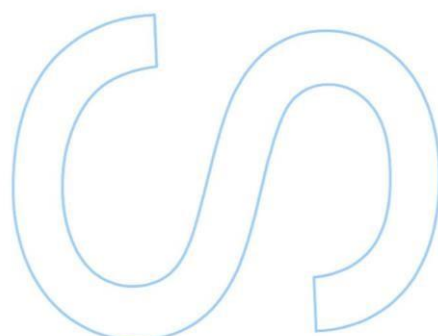
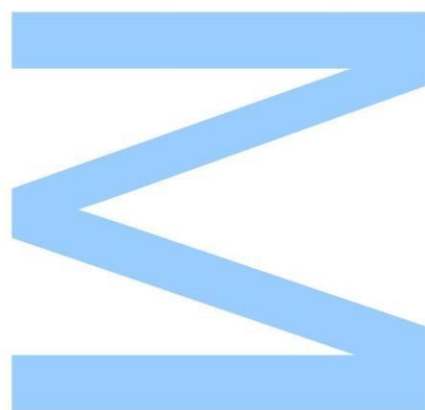




Todas as correções determinadas
pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____ / ____ / ____



Agradecimentos

Queria agradecer à minha orientadora, a professora Nádía Pinto, por me ter aceitado orientar, acolhido no grupo, pelo seu enorme apoio, paciência e dedicação necessários para me ajudar a concretizar este desafio, principalmente durante os últimos momentos!

À minha coorientadora Sofia Sousa por me ter ensinado por onde começar, ter acompanhado durante as experiências laboratoriais e por todo o seu apoio, bem como ter conseguido arranjar bastantes voluntários para este estudo.

À minha coorientadora, a professora Leonor Gusmão que apesar da distância esteve sempre disponível para ajudar, e pela enorme paciência que teve durante as longas reuniões de dúvidas.

À professora Iva Gomes pela sua disponibilidade em esclarecer algumas dúvidas.

À Mariana Nascimento por me ter ajudado também durante algumas experiências.

Ao Rob Mensink e à Ana Mafalda Rocha por realizarem as eletroforeses capilares das amostras.

Ao professor António Amorim por me ter aceitado no mestrado bem como no grupo, por tudo o que aprendi nas suas aulas, e pela generosa oferta do seu livro “A espécie das origens”.

E um enorme obrigado a todos os voluntários que participaram neste estudo, mesmo com um ano marcado pela pandemia que dificultou a logística da recolha das amostras.

Um enorme agradecimento à minha família e a todos os meus amigos que me apoiaram ao longo destes anos, em especial à minha mãe pelo apoio incondicional, e aos melhores Sara Mariano e Pedro Alves.

Um enorme obrigado a todos!

Abstract

Heterosomes have unique properties and characteristic inheritance modes, which make them crucial to solve complex kinship cases, identify disaster victims and analyse mixtures, when autosomes cannot be tested, or when they do not provide additional information. The markers mostly used for forensic purposes are the short tandem repeats (STR) due to their high discriminatory power; those are also employed for Y chromosome (Y-STR) and X chromosome (X-STR) analysis.

The lack of recombination of the Y chromosome in the male-specific region means that Y-STRs are linked and must be treated as single haplotypes, as those are transmitted from father to son intact (without recombination) barring mutations. The X chromosome only recombines in female meiosis, and it exhibits four linkage groups as a result the X-STRs within them are linked and must be treated also as haplotypes. Therefore, several data is needed to obtain a representation of the haplotypic and genetic diversity in the population, and also to estimate mutation rates, to aid the interpretation of the results and strength their statistical analysis.

In the present study were analysed 202 father-son duos, and 135 mother-father-daughter trios from the Portuguese population. The duos and trios were genotyped with the PowerPlex® Y23 System (Promega), and Investigator® Argus X-12 QS (Qiagen) kits, respectively, currently used in the forensic field. The haplotype and gene frequencies, as well as mutation rates, were estimated for 23 Y-STRs and 12 X-STRs. Genotyping analysis revealed the observation of microvariant alleles, duplications and triplications of loci, as well as a silent allele. Interestingly, a *de novo* deletion of 3 adjacent markers (DYS448, DYS549 and DYS392) on the Y chromosome was also observed, however, these findings should be confirmed by sequencing.

From the samples, 200 Y-STR haplotypes (from 200 unrelated sons) and 258 X-STR haplotypes for each linkage group (from 129 unrelated daughters) were selected for population analysis. Those revealed no significant differences when compared with the data already published for the Portuguese population. The X-STR data available for this population was joined and tested for linkage disequilibrium and Hardy-Weinberg analysis. The former revealed associations between markers within each linkage group; and no significant deviations were found from the observed and expected heterozygosity.

Taken together these results, the data was grouped with the literature on both Y-STRs and X-STRs for haplotype and genetic diversity analysis. The Portuguese population exhibits a high gene variety for 23 Y-STRs and 12 X-STRs, where higher gene diversities were observed for DXS10135 on the X chromosome, while for DYS385ab (multi-copy) and DYS570 on the Y chromosome. A total of 450 Y-STR haplotypes were analysed, with a haplotype diversity of 1.0000 ± 0.0001 . Despite 17 haplotypes were found twice, the remaining 433 were unique. A high haplotype diversity (≥ 0.9996) was also obtained for each linkage group on the X chromosome, out of the 830 haplotypes were observed 453 different haplotypes for linkage group 1, 248 for linkage group 2, 228 for linkage group 3, and 358 for linkage group 4.

In mutation analyses were observed 15 mutations from the 202 meioses for 23 Y-STRs, and 15 mutations from the 270 meioses for the 12 X-STRs. These results were gathered with other data from the literature to obtain more accurate marker specific mutation rate estimates. The mutation rate per X-STR locus ranged from 7.9×10^{-4} (at DXS7423) to 9.9×10^{-3} (at DXS10135), with an average of 3.6×10^{-3} (95% CI $3.1 - 4.2 \times 10^{-3}$) across the 12 X-STR loci. With a mutation rate near 5.4 times higher for fathers than for mothers. While per Y-STR the mutation rates ranged from 0 (at DYS643) to 1.6×10^{-2} (at DYS576), with an average of 2.6×10^{-3} (95% CI $2.4 - 2.7 \times 10^{-3}$). Most of the mutations were of one step, thus either compatible with the gain or loss of one repeat unit.

This research contributes to the knowledge about the genetic polymorphisms (allele and haplotype frequencies) and mutation rates estimates observed for both Y-STRs and X-STRs for the Portuguese population, which are very useful for kinship and forensic genetics analysis.

Keywords: Y chromosome, Y-STR, X chromosome, X-STR, mutation rates, Portugal

Resumo

Os heterossomas possuem propriedades únicas e modos característicos de transmissão, que os tornam cruciais para resolver casos complexos de parentesco, identificar vítimas de desastres em massa e analisar misturas quando os autossomas não podem ser testados, ou quando estes não fornecem informação adicional. Devido ao seu modo uniparental (paternal) de transmissão genética, os marcadores do cromossoma Y são também bastante úteis em estudos demográficos populacionais. Os marcadores autossómicos mais usados para aplicações forenses são os de repetições curtas em tandem (STRs, do inglês short-tandem repeats) devido ao seu alto poder discriminatório; esses são também os marcadores de eleição para a análise do cromossoma Y (Y-STR) e do cromossoma X (X-STR).

Devido à ausência de recombinação na região específica do cromossoma Y, os Y-STRs não são transmitidos de forma independente, tendo de ser tratados como um único haplótipo. De facto, a menos que ocorra uma mutação, a zona específica do cromossoma Y é transmitida intacta de pai para filho. Por sua vez, o cromossoma X recombina apenas durante a gametogénese nas mulheres, exibindo quatro grupos de ligação. Assim os X-STRs dentro de cada grupo de ligação estão associados e também têm de ser tratados como haplótipos. Por conseguinte, são necessários vários dados para obter uma boa representatividade dos haplótipos e da diversidade genética existentes na população. Estes são também essenciais para estimar taxas de mutação e assim suportar a interpretação dos resultados e melhorar a sua análise estatística.

Neste estudo foram analisados 202 duos pai-filho e 135 trios mãe-pai-filha da população Portuguesa. Os duos e trios foram genotipados com os kits PowerPlex® Y23 System (Promega), e Investigator® Argus X-12 QS (Qiagen), respetivamente, utilizados atualmente na área forense. As frequências dos haplótipos e dos alelos, assim como as taxas de mutação, foram estimadas para 23 Y-STRs e 12 X-STRs. A análise da genotipagem revelou a observação de microvariantes, duplicações e triplicações, bem como a presença de um alelo silencioso. Curiosamente, foi também observada uma deleção *de novo* de 3 marcadores adjacentes (DYS448, DYS549 e DYS392), sendo, no entanto necessário confirmar estes resultados por sequenciação.

Para a análise populacional das amostras relativas a duos pai-filho foram selecionados 200 Y-STR haplótipos de filhos não relacionados e 258 haplótipos X-STR (de 129 filhas não relacionadas, para as quais foi inferida a fase haplotípica). Os dados

obtidos não revelaram diferenças significativas quando comparados com os anteriormente publicados para os mesmos marcadores e população Portuguesa. Os dados X-STR disponíveis para esta população foram analisados para avaliar a presença de associação entre os marcadores, e o equilíbrio de Hardy-Weinberg. Foram observadas associações entre marcadores dentro de cada grupo de ligação; e não foram encontrados desvios significativos relativamente às heterozigotias observadas e esperadas.

Tendo em conta estes resultados, os dados obtidos foram agrupados com os disponíveis na literatura tanto para os Y-STRs como para os X-STRs, tendo sido analisada a respetiva diversidade génica e haplotípica. Para a população Portuguesa foi observada uma elevada diversidade génica para os 23 Y-STRs e 12 X-STRs, tendo sido as mais elevadas verificadas para o marcador DXS10135 (no cromossoma X) e para os marcadores DYS385ab (multi-cópia) e DYS570 (no cromossoma Y). Foram analisados 450 haplótipos Y-STR, que exibiram uma diversidade haplotípica de 1.0000 ± 0.0001 . Apesar de 17 haplótipos terem sido observados duas vezes, os restantes 433 eram únicos. Uma elevada diversidade haplotípica (≥ 0.9996) foi também obtida para cada grupo de ligação no cromossoma X, tendo sido contabilizados 830 haplótipos diferentes: 453 haplótipos diferentes para o grupo de ligação 1, 248 para o grupo de ligação 2, 228 para o grupo de ligação 3, e 358 para o grupo de ligação 4.

Relativamente à análise de mutações, foram observadas 15 mutações tendo em conta as 202 meioses relativas aos 23 Y-STRs, e 15 mutações para as 270 meioses dos 12 X-STRs. Estes resultados foram agrupados com outros dados da literatura para obter estimativas mais precisas das taxas de mutação específicas por marcador. A taxa de mutação específica por X-STR variou entre 7.9×10^{-4} (no DXS7423) e 9.9×10^{-3} (no DXS10135), com uma média de 3.6×10^{-3} (95% CI $3.1 - 4.2 \times 10^{-3}$) para os 12 marcadores analisados. Foi ainda observada uma taxa de mutação paterna aproximadamente 5.4 vezes mais elevada que a materna. Considerando os Y-STR, as taxas de mutação variaram entre 0 (no DYS643) e 1.6×10^{-2} (no DYS576), com uma média de 2.6×10^{-3} (95% CI $2.4 - 2.7 \times 10^{-3}$) para os 23 marcadores analisados. A maioria das incompatibilidades observadas revelaram-se explicáveis com a ocorrência de mutações de um só passo, e por isso compatíveis com o ganho ou a perda de uma única unidade repetitiva.

Este trabalho de investigação contribui para uma melhor caracterização da população Portuguesa no que diz respeito a polimorfismos genéticos Y-STR e X-STR (frequências alélicas e haplotípicas), bem como para uma melhor estimativa das

respetivas taxas de mutação, aumentando o poder estatístico da análise genética de parentescos e outros casos forense, bem como de estudos demográficos.

Palavras-chave: cromossoma Y, Y-STR, cromossoma X, X-STR, taxas de mutação, Portugal

Table of contents

Agradecimentos	i
Abstract.....	ii
Resumo.....	iv
List of Figures.....	ix
List of Tables	xi
List of Abbreviations	xiv
1. Introduction.....	1
1.1. Population genetics	1
1.1.1. Modes of genetic inheritance.....	3
1.1.1.1. Overview.....	3
1.1.1.2. Heterosomes.....	5
1.1.1.2.1. Y chromosome	6
1.1.1.2.2. X chromosome	7
1.1.2. Types of genetic polymorphisms	9
1.1.3. Population and forensic parameters	11
1.1.3.1. Hardy and Weinberg and Linkage Equilibrium.....	11
1.1.3.2. Gene diversity	12
1.1.3.3. Population comparisons	13
1.1.3.4. Forensic Parameters.....	13
1.1.4. Portuguese population	15
1.2. Heterosomal short tandem repeat markers	16
1.2.1. Y Chromosomal STRs: kits, databases, and interpretation guidelines	16
1.2.2. X Chromosomal STRs: kits, databases, and interpretation guidelines	19
1.2.3. X and Y Chromosome STRs: Portuguese population	20
1.3. STR mutations.....	20
1.3.1. Estimation of STR mutation rates	22

1.3.2.	Portuguese data in heterosomal STR mutation studies	23
2.	Aims	24
3.	Materials and Methods.....	25
3.1.	Sampling	25
3.2.	DNA Extraction	25
3.3.	Genotyping	27
3.3.1.	DNA amplification	27
3.3.2.	Capillary Electrophoresis (CE).....	30
3.4.	Data analysis.....	30
4.	Results and Discussion.....	32
4.1.	Y Chromosome.....	32
4.1.1.	Genotyping analysis	32
4.1.2.	Population genetic analysis	35
4.1.3.	Mutations analysis	38
4.1.3.1.	Mutations and father's age	39
4.1.3.2.	Locus-specific mutation rates	40
4.2.	X Chromosome.....	43
4.2.1.	Genotyping analysis	43
4.2.2.	Population genetic analysis	46
4.2.3.	Linkage disequilibrium	47
4.2.4.	Hardy-Weinberg equilibrium	47
4.2.5.	Genetic diversity and forensic parameters	48
4.2.6.	Mutations analysis	52
4.2.6.1.	Mutations and parents' age	53
4.2.6.2.	Locus-specific mutation rates	54
5.	Conclusions	57
6.	References	60
7.	Appendix.....	69

List of Figures

Figure 1- Schematic representation of the different modes of genetic inheritance for autosomes, heterosomes (X and Y chromosomes) and mitochondrial DNA. Source: Pereira and Gusmão, (2016).....	4
Figure 2- Schematic representation of the Y chromosome structure. Source: adapted from Jobling and Tyler-Smith, (2017).	6
Figure 3- Schematic representation of the X-chromosome and the physical location of the 12 X-STR loci provided in the Investigator Argus X-12 QS Kit (Qiagen). Source: Investigator® Argus X-12 QS Handbook, Qiagen (2015).	8
Figure 4- Schematic representation of a paternity case where the alleged father is the father of the mother (Hypothesis 1- H1) or her brother (Hypothesis 2- H2). The daughter will exhibit for one X chromosome one of the markers A, B or C inherited from her mother. While her other X chromosome (X ?) will either exhibit the marker A (thus the DNA evidence supports H1) or markers B or C (DNA evidence supports H2). Source: Butler, (2012a).	9
Figure 5- Punnet square of the distributions of the expected genotype probabilities for a locus containing 2 alleles (A1 and A2), with frequencies p and q, for a population under the Hardy-Weinberg equilibrium.	11
Figure 6- Schematic representation of 4 possible haplotypes for the 2 loci, one with alleles A1, A2 and another with B1, B2.	12
Figure 7- Representation of the geographical location of Portugal. Source: https://commons.wikimedia.org/wiki/File:Portugal_in_Europe_(-rivers_-mini_map).svg	15
Figure 8- Schematic representation of the duplication and divergence model for Y-STRs that can originate multiple alleles, and their correspondent electropherograms after being amplified with locus specific primer pair. Source: Butler <i>et al.</i> , (2005).....	17
Figure 9- Schematic representation on how multiple PCR primer binding sites originate multi-copy PCR products for DYS385a/b (a), and for DYS389I/II (b). F primer – forward primer; R primer – reverse primer. Source: (Butler, 2012a).	17
Figure 10- Representative structure of a short tandem repeat locus. Source: (Triggs, Buckleton and Walsh, 2004).	20
Figure 11- Electropherograms showing di-allelic (A) and tri-allelic (B) patterns at the DYS439 and DYS19 markers, respectively, considering 2 father (left) – son (right) duos.	32

Figure 12- Electropherograms from samples: father on top, and son at the bottom, for each set of markers coloured in blue (A), green (B), black (C), and red (D). The black box highlights the absence of amplification for the DYS448, DYS549 and DYS392 markers in the son's sample.	35
Figure 13- Distribution of the father's age at time of the son's birth, and the father's age that transmitted a mutation.	40
Figure 14- Electropherograms showing the DXS10079 marker from the samples of the mother (A), father (B), and daughter (C), exhibiting the tri-allelic pattern in the mother's sample.	43
Figure 15- Electropherograms showing the off-ladder between DXS10148 and D21S11, highlighted by the black box, from samples of the mother on top, father at the centre, and daughter at the bottom.	44
Figure 16- Electropherograms showing the off-ladder allele in DXS10146 marker, highlighted by the black box, from a person's sample (A), and from the Allelic ladder Argus X-12 QS (B).	45
Figure 17- Electropherograms showing the markers coloured in blue for the samples of the mother (A), father(B), and daughter (C), showing the absence of amplification for the marker DXS10101 on the father's sample, and of one allele in the daughter's sample, highlighted by the black box.	45
Figure 18- Distribution of the parental age at daughter's birth, for mothers, fathers, and father's that transmitted a mutation. Only mutations for which a most parsimonious parental origin was possible to determine were considered.	53

List of Tables

Table 1- Summary of the sequence, repeat motif, and their correspondent cluster of the Y-STR markers from PowerPlex® Y23 System (Promega).	27
Table 2- PCR reaction setup (left table) with each reagent volume and cycling conditions (right table) used for amplification of 23 Y-STRs using the PowerPlex® Y23 System.....	28
Table 3- Summary of the repeat motif sequence of the reference allele for the 12 X-STRs from the Investigator Argus X-12 QS Kit (Qiagen). * [AGAT] is the most common repeat structure; for variations see the NIST. Source: adapted from the Investigator® Argus X-12 QS Handbook, Qiagen, 2015.	29
Table 4- PCR reaction setup (left table) and cycling conditions (right table) used for amplification of 12 X-STRs using the Investigator Argus X-12 QS Kit.	30
Table 5- Pairwise genetic distances (Fst) (lower diagonal) between the 2 groups: This work (N=200) and Salazar (2013) (N=250), and corresponding non-differentiation P-values (upper diagonal), obtained after 10100 permutations.	35
Table 6- Allele frequencies calculated for 23 Y-STRs with data from 450 haplotypes from this work (N=200) and from Salazar (2013) (N=250).....	37
Table 7-Genetic diversity values obtained for 23 Y-STR loci, with data from 450 haplotypes, collected in this work (N=200) and from Salazar (2013) (N=250). The values are sorted in a descending order.....	38
Table 8- Father's and son's alleles from the duos where a mutation was observed.	39
Table 9- Average specific mutation rates obtained for each 23 Y-STR from the PowerPlex Y 23 kit, after pooling the data from this work and the literature (Antão-Sousa, 2017; Yang <i>et al.</i> , 2018; Ambrosio <i>et al.</i> , 2020; Lin <i>et al.</i> , 2020).{+1} - gain of 1 repeat ; >{+1} - gain of more than 1 repeat; {-1} - loss of 1 repeat; >{-1} -loss of more than one repeat.NI- non-integer unit variation.....	41
Table 10- Pairwise genetic distances (Fst) (lower diagonal) between the 3 groups (This work, Pinto et al., (2020), and Cainé, Costa and Pinheiro, (2013)) and corresponding non-differentiation P-values (upper diagonal), for the 3 X-STRs included in linkage group 1 (A), linkage group 2 (B), linkage group 3 (C) and linkage group 4 (D) obtained after 1023 permutations.....	46
Table 11- Pairwise linkage disequilibrium test performed in a sample containing 830 haplotypes from the Portuguese population (from This work, Pinto et al., (2020), and Cainé, Costa and Pinheiro, (2013)) between 3 X-STR inside each linkage groups from 1	

to 4. In red are the significant p-values for a significance level of 0.00417 ($\alpha=0.05/12$) after applying Bonferroni correction, for 100172 permutations.	47
Table 12-Results on the observed heterozygosity (H_{obs}), and expected heterozygosity (H_{exp}), and p-value of the exact test of Hardy-Weinberg equilibrium for 12 X-STRs, for 1001000 permutations.	48
Table 13- Forensic parameters calculated for the each of the 4 linkage groups, from a sample containing 830 haplotypes from the Portuguese population (from This work (N=258), Pinto et al., (2020) (N=422), and Cainé, Costa and Pinheiro, (2013) (N=150), namely the power of discrimination (PD) for males and females, the mean exclusion chance (MEC) for father-daughter duos and father-mother-daughter trios and paternal half-sisters (PHS). LG1- linkage group 1, LG2- linkage group 2, LG3 linkage group 3, LG4-linkage group 4.	48
Table 14- Allele frequencies of the 3 X-STR markers within the linkage groups 1 and 2, from a sample containing 976 gene copies from the Portuguese population (from this work (N=258), Pinto et al., (2020) (N=422), and Cainé, Costa and Pinheiro, (2013)(N=296)).	50
Table 15- Allele frequencies for the 3 X-STR markers within the linkage groups 3 and 4, from a sample containing 976 gene copies from the Portuguese population (from this work (N=258), Pinto et al., (2020) (N=422), and Cainé, Costa and Pinheiro, (2013) (N=296)). *During genotyping some called alleles 39.2 from the DXS10146 showed differences in their mobility (<1 bp) comparatively to the allele 39.2 from the allelic ladder.	51
Table 16- Expected heterozygosity and forensic parameters calculated for the 12 X-STR markers, namely the power of discrimination (PD) for males and females, the mean exclusion chance (MEC) for duos, trios and paternal-half-sisters (PHS).	52
Table 17- Genotypic configurations of each trio that showed an incompatibility, and the most parsimonious parental origin.	53
Table 18- Average specific mutation rates obtained for each 12 X-STR from the Investigator Argus X-12 QS kit, after pooling the data from this work (N=135 trios) and the literature (Athanasiadou <i>et al.</i> , 2003; Turrina and De Leo, 2004; Tabbada <i>et al.</i> , 2005; Turrina, Atzei and De Leo, 2007; Becker <i>et al.</i> , 2008; Tariq <i>et al.</i> , 2008; Liu <i>et al.</i> , 2008, 2012; Pico <i>et al.</i> , 2008; Tetzlaff, Wegener and Lindner, 2012; Nishi <i>et al.</i> , 2013; Diegoli <i>et al.</i> , 2014; Chen <i>et al.</i> , 2020; Pinto <i>et al.</i> , 2020).	55
Table 19- Summary of the number of mutations counted for each parsimonious parental origin and the type of mutations observed in this study (N=135 trios) and the literature (Athanasiadou <i>et al.</i> , 2003; Turrina and De Leo, 2004; Tabbada <i>et al.</i> , 2005; Turrina, Atzei and De Leo, 2007; Becker <i>et al.</i> , 2008; Tariq <i>et al.</i> , 2008; Liu <i>et al.</i> , 2008,	

2012; Pico *et al.*, 2008; Tetzlaff, Wegener and Lindner, 2012; Nishi *et al.*, 2013; Diegoli *et al.*, 2014; Chen *et al.*, 2020; Pinto *et al.*, 2020). Type of mutation: {+1} represents the gain of one repeat unit, >{+1}- gain of more than one repeat, {-1}- loss of a repeat, <{-1} the loss of more than one repeat, {+1}V{-1} mutation explained either by the gain or loss of one repeat, NI- non integer repeat unit variation. 56

List of Abbreviations

bp - base pair
CE - Capillary Electrophoresis
DL - Discrete Laplace
DNA - Deoxyribonucleic Acid
GD - Gene diversity
H - Heterozygosity
HD - Haplotype diversity
Hexp - Expected heterozygosity
Hobs - Observed heterozygosity
InDel - Insertion or Deletion
ISFG - International Society for Forensic Genetics
LD - Linkage Disequilibrium
LR - Likelihood Ratio
MEC - Mean Exclusion Chance
MP - Matching probability
MSY- Male-specific region of the Y chromosome
mtDNA - Mitochondrial DNA
Ne - Effective population size
NRY - Non-recombining region of the Y chromosome
PAR - Pseudoautosomal Region
PCR - Polymerase Chain reaction
PD - Power of Discrimination
PHS - Paternal half-sisters
PPE - Probability of Paternity Exclusion
RM - Rapidly Mutating
SMM - Stepwise Mutation Model
SNP - Single Nucleotide Polymorphism
SRY – Sex determining region of the Y chromosome
STR - Short Tandem Repeat
SWGAM - Scientific Working Group on DNA Analysis Methods
VNTR - Variable Number Tandem Repeat
XDR - X-Degenerate region

XTR - X-Transposed region

YHRD - Y-STR Haplotype Reference Database

1. Introduction

1.1. Population genetics

The genetic composition of a population results from the genetic inheritance of individuals' ancestors, their relations, history, and other factors. This is studied by population genetics, as it investigates the genetic composition within and among populations of species over time and space.

The amount of genetic variation between populations (genetic distance) can unravel some of their demographic history (Jobling *et al.*, 2013). For example, populations geographically and historically distant from each other tend to exhibit a higher genetic divergence than those that are closer or with a common history. As a population begins to diverge, subpopulations are formed, which accumulate differences (mutations) over time. Usually, the higher the number of mutations accumulated in those sub-populations the older is their common ancestor. Thus, estimating mutation rates can help to date historical events in fields such as archaeology, which presents limitations in the study of fossils due to their conservation, and low resolution of dating techniques (Pereira and Ribeiro, 2009).

Population genetics also encloses the theoretical background needed in forensic genetics to evaluate the statistical weight of the evidence in problems related to the identification of individuals and/or samples, and with the existing relationship between these (Lewontin and Hartl, 1991; Butler, 2010). For example, it allows evaluating DNA evidence for problems such as “how likely it is the observation of specific DNA profile, assuming that it is from population A? And from population B?”. Kinship problems such as “how probable is this child’s DNA profile assuming that the alleged father is the true father? And if they are unrelated?” can also be investigated.

Genetic diversity of a population represents the genetic variation of alleles, genotypes, or haplotypes of the respective individuals; a higher diversity generally equates to a higher number of variants. The change in the genetic frequencies (alleles or genotypes) over time defines the evolution of species (Amorim, 2009) which are influenced by mechanisms of evolution: mutation and meiotic recombination (sexual species), genetic drift, selection, and migration (Jobling *et al.*, 2013).

Mutations are modifications at the sequence or length of the DNA. These can occur by different mechanisms, such as environmental exposures (e.g. UV light,

chemotherapeutic agents), or during DNA replication when cell repair mechanisms do not correct polymerase errors (Piraino and Furney, 2016). Mutations can be germinal (occur during meiosis) or somatic (occur during mitosis). Germinal mutations are liable of being transmitted to the offspring, whereas somatic ones are not inherited. At the allele level, mutations may originate new alleles, never observed in the population before (increasing the genetic diversity), or alleles previously observed in the population, altering its frequency (Jobling *et al.*, 2013).

In sexual species, meiotic recombination - the exchange of DNA segments between homologous chromosomes, occurs during meiosis, resulting in new chromosome combinations and thus new haplotypes. The haplotypes may contain advantageous alleles for an individual's adaptation to the environment. However, the opposite may also be observed due to genetic hitchhiking (an allele under a positive selection "hitchhikes" a deleterious allele from the same gametic phase, due to their close location) (Jobling *et al.*, 2013). The origin of new alleles and haplotypes due to mutation and recombination, respectively, contribute to the increase of genetic diversity.

On the other hand, genetic drift reduces the genetic diversity in a population through the stochastic sampling of the set of individuals that will originate the next generation. The number of members in a population that generate the offspring is designated by "effective" population size (N_e) (Wright-Fisher model). The genetic drift effect is pronounced in smaller populations due to the smaller effective population size already present before sampling, as opposed to larger populations (Jobling *et al.*, 2013). There are two main ways for population sampling that may lead to population drift:

- 1) Bottleneck effect - the population is severely reduced by an event such as a natural disaster. The population sampled (e.g.: survivors) are the ones that contribute to the genetic structure of the offspring of the next generation.
- 2) Founder effect - the members from a population are separated, contributing separately to the genetic diversity of the next generation. This isolation may happen due to physical barriers or even cultural behaviours (e.g. different religious beliefs).

Migration refers to the movement of individuals from a population (e.g. A) to another (e.g. B), population B changes its genetic diversity, depending on the migrants from population A as well as their proportion and initial gene frequencies of population B. This concept is different from colonization, which refers to the occupation of an uninhabited land by members of a population (Jobling *et al.*, 2013).

Natural selection refers to the “differential reproduction of individuals of different genotypes in sequential generations”, first elaborated by Darwin and adapted by Fisher (Jobling *et al.*, 2013). The fitness of an individual conditions its reproductive success, as it is subjected to selection at several stages such as its survival until the generation of offspring, attracting a mate (sexual selection), be fertile (fertility and gamete selection), or the number of progeny (fecundity) (Jobling *et al.*, 2013).

Mutations that reduce or improve the fitness of the individual are subjected to either a negative/purifying selection, or a positive/diversifying selection, respectively. However this effect is difficult to evaluate, and most mutations are neutral, neither advantageous nor prejudicial, and thus not subjected to selection (Theory of Neutral Evolution (Kimura, 1968)). Hence natural selection can either increase, decrease, or keep the genetic diversity in the population (Jobling *et al.*, 2013).

The interplay between the different evolutionary mechanisms is what drives the evolution within species (microevolution). Mutation (when new alleles are formed), recombination (new haplotypes) and migration contribute to an increase in diversity, while genetic drift reduces it, and selection can cause both or be neutral (Jobling *et al.*, 2013).

1.1.1. Modes of genetic inheritance

1.1.1.1. Overview

Each human body consists of trillions of cells ($\sim 10^{13}$) (Sender, Fuchs and Milo, 2016), most contain the individual's genome: the nuclear (inside the nucleus), and the mitochondrial (inside the mitochondria).

The nuclear DNA is highly compacted (chromatin) with histones to form the chromosomes. The nuclear genome is comprised of 23 pairs of chromosomes, 22 of which are pairs of homologous chromosomes designated by autosomes, and one is a pair of non-homologous chromosomes (X and Y), termed as sex chromosomes or heterosomes. Females possess a pair of X chromosomes, while males have one X chromosome and one Y chromosome.

In diploid cells, for each pair of autosomes, one chromosome is inherited (randomly) from the mother of the individual, and the other from the father, obeying to a Mendelian, biparental, inheritance. On the other hand, human heterosomes have a different inheritance mode depending on the sex of the individuals. Females possess a pair of X chromosomes, each one inherited from their mother and father, while males inherit one

X chromosome from their mother and one Y chromosome from their father. Females transmit randomly one of the two X chromosomes they have to any child, either female or male. While males transmit their unique X chromosome to their female offspring, and the Y chromosome to their male offspring. Hence, the X chromosome exhibits a haplodiploid inheritance, biparental for females and uniparental for males, while the Y chromosome shows a uniparental (paternal) inheritance.

The mitochondrial DNA (mtDNA) is transmitted in a haploid fashion by mothers to their offspring (either female or male), while mtDNA from the fathers is destroyed after fertilization. Thus, mtDNA shows a uniparental (maternal) mode of inheritance.

The nuclear DNA transmitted from the parents to the offspring undergoes meiotic recombination for autosomes and maternal X chromosomes, in contrast with the paternal X and Y-chromosomes, which are transmitted in block (without recombination, except for the pseudoautosomal regions). The different modes of genetic inheritance are represented in Figure 1.

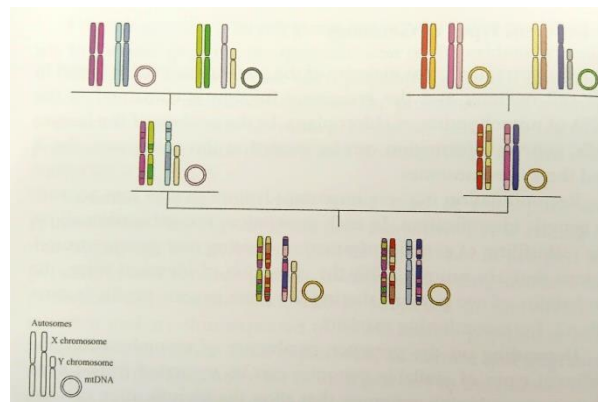


Figure 1- Schematic representation of the different modes of genetic inheritance for autosomes, heterosomes (X and Y chromosomes) and mitochondrial DNA. Source: Pereira and Gusmão, (2016).

The analysis of autosomes for identification purposes exhibits advantages compared with heterosomes and mtDNA, as they possess higher diversity due to their higher recombination rates and are less affected by genetic drift (Schaffner, 2004). The autosomal markers analysed can be spread through the 22 chromosomes which increases the hypothesis of being independently transmitted. However, the meiotic recombination in autosomes precludes the establishment of a population history (Amorim, 2009), as well as the presence of recombination in female X chromosomes, as opposed to uniparental markers (Y chromosome and mtDNA), that don't recombine during meiosis (Schaffner, 2004).

Indeed, both Y chromosome and mitochondrial DNA are uniparentally transmitted, which allows the establishment of paternal and maternal lineages, respectively, providing valuable insights into the history of humanity, their migrations, past social structures and different reproductive behaviours (Batini and Jobling, 2017; Hallast *et al.*, 2021). Both have a smaller effective size ($1/4$) relatively to the autosomes, thus are more affected by genetic drift, creating geographical distributions for both lineages. The analysis of the Y chromosome offers even higher detail over mtDNA due to their longer length and due to the patterns of human migrations (Hallast *et al.*, 2021). It is of great interest to establish paternal and maternal lineages as it allows the investigation of genetic genealogies, to solve problems of forensic genetics (further discussed), which can also aid the research of historians, archaeologists, and anthropologists.

Mitochondrial DNA (mtDNA) is a circular DNA with 16,569 base pairs (bp) located inside mitochondria, those contain about 1 to 11 mtDNA copies each (Cavelier, Johannisson and Gyllensten, 2000). Also, most cells have several, more than 1000 mitochondria (Taanman, 1999). Its uniparental (maternal) transmission enables the establishment of maternal lineages since mother and offspring share the same mtDNA, excepting mutations (Pakendorf and Stoneking, 2005). It shows higher mutation rates, 5 to 10 times more than nuclear DNA. The high copy number, maternal inheritance and higher mutation rate make mitochondrial DNA a valuable tool in several contexts. For example, it facilitates the examination of old samples which contain highly degraded DNA, and in other situations where nuclear DNA is not possible to examine (e.g. skeletal remains) (Amorim, Fernandes and Taveira, 2019). Notwithstanding, it offers a low discriminatory power when compared with autosomes as it only provides access to the maternal lineage. In kinship testing, it may be useful in cases where maternity is under analysis, as is the case of alleged (maternal) brothers. In disaster victim identification (DVI), maternal relatives can provide reference samples (e.g. identification of Tsar Nicholas II (Ivanov *et al.*, 1996)), and it also allows to estimate the geographical origin of the maternal ancestry of a sample (Pakendorf and Stoneking, 2005).

Next, attention will be focused on heterosomes as this thesis is centred on them.

1.1.1.2. Heterosomes

Ohno (1967) proposed that the heterosomes X and Y have diverged from a pair of autosomes approximately 180 million years ago. The ancestral autosomes recombined during gametogenesis, but after the proto-Y-chromosome acquired the gene SRY (sex-determining region of the Y), several inversions, and gene losses in this chromosome over time, halted the recombination between the two (except for the pseudoautosomal

regions) (Lahn and Page, 1999). This lead, to the present different sizes observed - while the X chromosome has approximately 155 Mega base pairs (Mbp) in length, the Y chromosome possesses only 60 Mbp (Skaletsky *et al.*, 2003; Ross *et al.*, 2005).

1.1.1.2.1. Y chromosome

The Y chromosome has pseudoautosomal regions 1 and 2 (PAR1 and PAR2), which constitute 5% of the chromosome length, located at the tips of the chromosome. These are the only regions that recombine with their homologous in the X chromosome during meiosis, as opposed to the remaining 95%, designated as male-specific region (MSY) or non-recombining region (NRY) of the Y chromosome. The MSY is uniparentally (paternally) transmitted, in block, from father to sons (excepting mutations), which allows the establishment of male lineages. The MSY has near 23 Mbp of euchromatin and 30 Mbp of heterochromatin, comprised of highly repetitive sequences that challenged their physical mapping and sequencing (Bachtrog and Charlesworth, 2001). Euchromatin can be divided into 3 sequence classes: a) X-transposed region (XTR), b) X-degenerate region (XDR) and c) ampliconic, further described (Skaletsky *et al.*, 2003) (Figure 2):

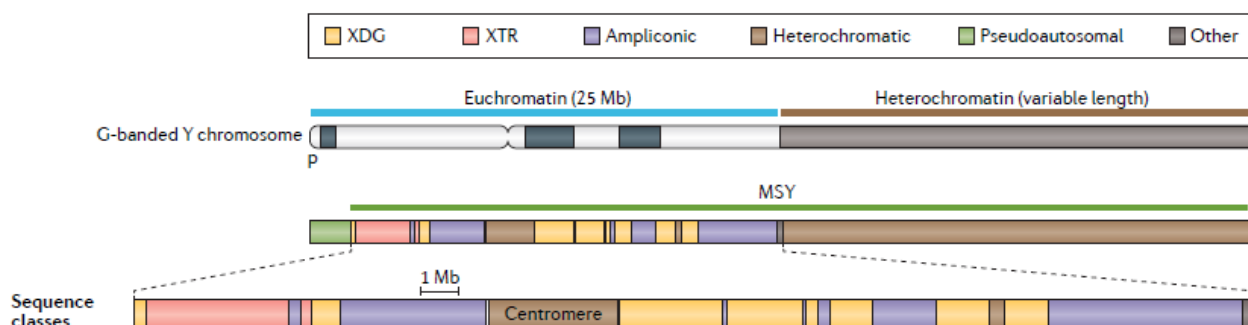


Figure 2- Schematic representation of the Y chromosome structure. Source: adapted from Jobling and Tyler-Smith, (2017).

a) The X-transposed region (XTR) is a sequence translocated from the X to the Y chromosome, of 3.4 Mbp. An inversion in this region broke the sequence into 2 blocks, precluding recombination (Page *et al.*, 1984).

b) The X-degenerate region (XDR) diverged from the proto-X-chromosome but still exhibits 96% identity with the X chromosome, it is composed of 8 blocks located in both arms of the Y chromosome, 8.6 Mbp of total length.

c) Ampliconic region is composed mostly of sequences with a higher complementarity with themselves (99%), some are large palindromes, which enhances the formation of secondary structures and subsequent rearrangements through non-

allelic homologous recombination (Jobling, 2008). They constitute 7 blocks located on both arms, 10.2 Mbp of sequence length (Skaletsky *et al.*, 2003; Jobling and Tyler-Smith, 2017).

The high similarity between the heterosomes X and Y challenges the design of primers to amplify specific sequences in the MSY region (Butler, 2012b).

The characteristics of the Y chromosome and its mode of transmission make it useful in population and forensic genetics research. For example, in biological mixtures analysis (as in most sexual assault cases) it may be useful to amplify only the male component. In this specific case, Y chromosomal markers may be amplified when: a. the mixture contains a higher proportion of female cells comparatively to male ones, b. sperm cannot be separated from the female cells by differential isolation, c. sperm is absent due to azoospermic or vasectomized males, and d. only digital penetration occurred. In mixture analysis (e.g. gang rape), the haploid nature of the Y chromosome makes the identification of the number of contributors easier than an autosome analysis (Butler, 2012a).

Since male lineages share the same Y chromosome, patrilineal relatives can provide reference samples to investigate a. missing males, b. male genealogies, or c. deficient paternity cases where the alleged father's DNA is unavailable (e.g. president Thomas Jefferson case (Foster *et al.*, 1999)) (Kayser and Sajantila, 2001; Butler, 2010). Patrilineal relatives (e.g. brothers) are practically indistinguishable through Y chromosomal markers, despite the identification of the rapidly mutating (RM) Y short tandem repeat markers, which may aid the differentiation when a great number of markers is analysed (Ballantyne, Keerl, Wollstein, Choi, Zuniga, Ralf, Vermeulen, De Knijff, *et al.*, 2012; Ralf *et al.*, 2020).

Resorting to markers with low mutation rates, population haplogroups may be defined, allowing the construction of an evolutionary tree of the Y chromosome (phylogeny) and studies of their geographic distribution (phylogeography). In a forensic context, this can be useful to estimate the geographical origin of the paternal lineage of a sample (Jobling and Tyler-Smith, 2017). Older haplogroups tend to show more diversity than recent ones as they had more time to accumulate mutations.

1.1.1.2.2. X chromosome

The X chromosome has unique features compared to the other chromosomes. It spent 2/3 of its history in females (females possess XX, while males XY as the 23rd pair of chromosomes) and they only recombine in females. This may lead that some markers

are not independently transmitted, having been established in 4 linkage groups located at Xp22.2, Xq12, Xq26, and Xq28 (Schaffner, 2004; Szibor, 2007) (Figure 3). The X chromosome has an effective population (N_e) size of $3/4$ relative to autosomes, but greater than uniparental markers ($1/4$), and thus are less affected by genetic drift than the latter.

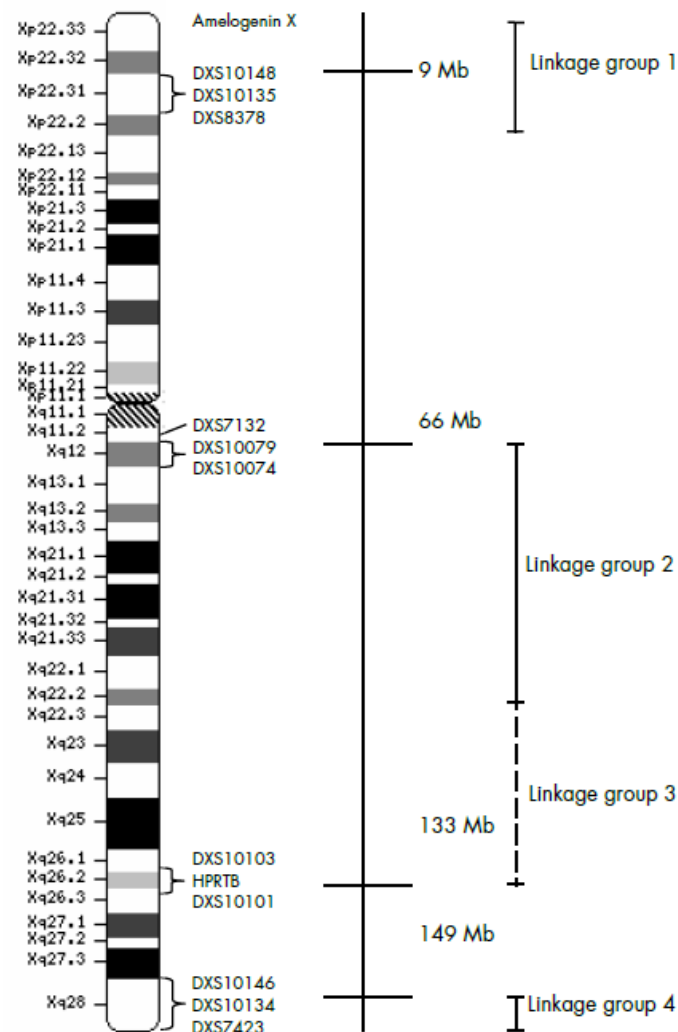


Figure 3- Schematic representation of the X-chromosome and the physical location of the 12 X-STR loci provided in the Investigator Argus X-12 QS Kit (Qiagen). Source: Investigator® Argus X-12 QS Handbook, Qiagen (2015).

In forensic genetics, the analysis of X chromosome markers may be useful in identity testing to complement autosomal profiling or analyse mixed samples (Diegoli, 2015). The power of X-chromosomal analysis in kinship investigations may be greater than the one of autosomes (Pinto *et al.*, 2013), for example in cases with daughters where the alleged father is not available for testing, contrarily to his undoubted mother or daughter. Applications also arise in incest cases (Figure 4) or when pedigrees belong to the same

autosomal class are at stake. Indeed, certain pedigrees (e.g.: aunt/niece) can be distinguished (from e.g.: half-sisters or grandmother/granddaughter) using linked X markers that were otherwise undistinguishable with unlinked autosomal markers (Pinto, Gusmão and Amorim, 2011; Gomes *et al.*, 2012). It can also be useful in missing person investigations if it is only possible to analyse reference samples from distant relatives (Diegoli and Coble, 2015).

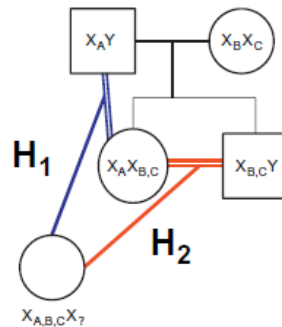


Figure 4- Schematic representation of a paternity case where the alleged father is the father of the mother (Hypothesis 1- H1) or her brother (Hypothesis 2- H2). The daughter will exhibit for one X chromosome one of the markers A, B or C inherited from her mother. While her other X chromosome (X ?) will either exhibit the marker A (thus the DNA evidence supports H1) or markers B or C (DNA evidence supports H2). Source: Butler, (2012a).

The X chromosome can also aid the study of population and evolutionary genetics, complementing the uniparental markers (Amster *et al.*, 2020). It may allow to study differences between sexes in mutation rates, recombination patterns, and selection (recessive alleles are more vulnerable to selection on males in their single X chromosome than in autosomes) (Schaffner, 2004). It can be helpful to study the human history on their global characteristics (e.g., Ne, population expansions, bottlenecks, or its stratifications), their large-scale history (e.g., the origin of non-African populations). It can also provide insights on small-scale studies, to overcome the disadvantages of uniparental markers, since those are more susceptible to genetic drift, and consequently to lose the genetic record of relationships among populations. Thus, linked markers from the X chromosome can be used to build phylogenies while the unlinked can provide higher resolution in local studies (Schaffner, 2004).

1.1.2. Types of genetic polymorphisms

Humans share approximately 99.9% of their DNA sequence, constituting the remaining differences in the variability/diversity existent between individuals. The proportion of genetic variability is higher among individuals from the same population

(intra-population), rather than between populations (inter-population) (Lewontin, 1972; Barbujani *et al.*, 1997). This variability is evaluated to characterize populations, identify individuals, and evaluate the kinship degree between them. To that purpose, several genetic markers began to be used in forensic and population genetics applications, in non-coding regions and not under selective pressures. A genetic marker is a DNA sequence present in a location of the genome ('*locus*', plural: '*loci*'), mutations to this sequence originate different forms (variants) of the initial alleles (Butler, 2010). Further will be categorized the following polymorphism types: VNTRs, STRs, INDELs, and SNPs.

Variable number tandem repeats (VNTRs) or minisatellites are sequences of approximately 10-100 base pairs repeated in tandem over 10 to more than 1000 times (Jobling and Tyler-Smith, 2017). Jeffreys, Wilson and Thein (1985) marked the beginning of forensic genetics with the introduction of "DNA fingerprinting" methodology using VNTRs, but their analysis requires large quantities of non-degraded DNA, a requirement difficult to fulfil in most of the collected samples in forensic casework (Goodwin, Linacre and Hadi, 2010). They were then replaced by shorter tandem repeats (STRs), the gold standard markers in forensic genetics analysis, given their characteristics.

Short tandem repeats (STRs), or microsatellites, are short sequences of 1-6 base pairs (per repeat unit) repeated over usually less than 50 times, spread through the genome (Fan and Chu, 2007). They are highly polymorphic, exhibiting a high number of alleles due to their high mutation rate ($\sim 10^{-3}$ per locus, per meiosis) (Weber and Wong, 1993). STRs have been widely used in forensic genetics as they grant a high power of discrimination using multiple loci (multiplexing), they are easily analysed by PCR and further separated by capillary electrophoresis (CE) (Goodwin, Linacre and Hadi, 2010). Those characteristics promoted the inclusion of several STRs from different autosomes and heterosomes in multiplex kits, analysis that has been the standard practice in forensic genetics problems (Hares, 2015).

STRs can be classified as mono-, di-, tri-, tetra-, (...), nucleotide depending on the length of the repetitive motif (Fan and Chu, 2007), the tetra is most used in forensic genetics (e.g. [GATA]_n) (Goodwin, Linacre and Hadi, 2010). On the other hand, the marker sequence may be classified into simple, compound or complex. Simple STRs are composed of a single repeat sequence (e.g. [TAT]_n), while the compound ones are composed of two or more adjacent simple repeats (e.g. [TCTG]_q[TCTA]_r). Finally, complex sequences are comprised of repeat units variable in length and may contain intervening sequences (e.g. [TAGA]_nATGGATAGATTA [GATG]_pAA [TAGA]_q) (Urquhart *et al.*, 1994).

Single nucleotide repeats (SNPs) are sequences that differ only in one nucleotide, and hence mostly bi-allelic. This characteristic is advantageous to analyse degraded samples, as we only look at a single nucleotide per locus, (Kowalczyk *et al.*, 2018) or to infer the geographical ancestry of a sample (Pereira *et al.*, 2017). They exhibit a low mutation rate (10^{-8} mutations per nucleotide site) and have a lower power of discrimination (PD) when compared with STRs, the PD of 15 STRs is only achieved with approximately 40-60 SNPs (Nachman and Crowell, 2000; Butler, Coble and Vallone, 2007).

InDels are insertions or deletions of sequences in the DNA that can range from one to hundreds of nucleotides, therefore are bi-allelic. As the above markers, they exhibit a low mutation rate ($\sim 10^{-8}$) and have a lower power of discrimination than the STRs. However, they show some of the STR advantages, such as the analysis of several loci by PCR (multiplex PCR) and CE. Their characteristics are useful for identification (Pereira *et al.*, 2009), ancestry estimation (Pereira *et al.*, 2012), and analysis of degraded samples (with short INDELs).

1.1.3. Population and forensic parameters

1.1.3.1. Hardy and Weinberg and Linkage Equilibrium

The principle of Hardy-Weinberg relates the frequencies of 2 alleles with the respective genotype frequency (only for diploid markers). This can be illustrated assuming a locus with 2 alleles: A1 and A2, where p and q represent the frequencies of those alleles in the population, respectively ($p + q = 1$). After one generation of random mating, the offspring exhibits the following genotype configurations: A1A2 or A2A1, A1A1, and A2A2. Whose genotype frequencies are given by the following binomial theorem and illustrated in the Punnett square (Figure 5).

$$(p + q)^2 = p^2 + q^2 + 2pq = 1$$

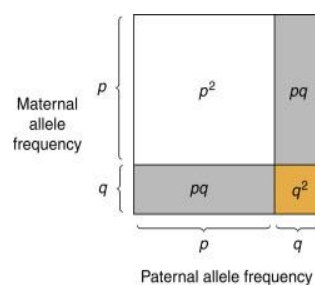


Figure 5- Punnet square of the distributions of the expected genotype probabilities for a locus containing 2 alleles (A1 and A2), with frequencies p and q , for a population under the Hardy-Weinberg equilibrium.

It is expected that: the frequency of homozygotes A1A1, $f(A1A1) = p^2$, frequency of homozygotes A2A2, $f(A2A2) = q^2$, and frequency of heterozygotes A1A2 or A1A2, $f(A1A2 \text{ or } A2A1) = 2pq$. Hence a population is considered under the Hardy-Weinberg equilibrium if the observed probability of a genotype in the population doesn't deviate significantly from the one expected. This equilibrium can be reached after 1 generation of random mating in the population.

Linkage equilibrium relates the frequency of alleles at 2 different loci with their haplotype frequencies if they are transmitted independently to the offspring. This can be illustrated by Figure 6 where a locus has 2 alleles: A1 and A2, while another has: B1 and B2, with frequencies p_1 , p_2 , q_1 and q_2 , respectively. Linkage equilibrium exists when the haplotype frequencies correspond to the product of the allele frequencies, for example, $f(A1B1) = f(A1) \times f(B1) = p_1 \times q_1$.

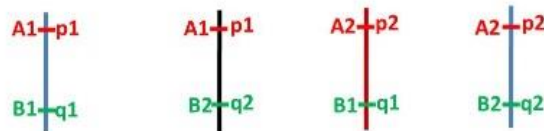


Figure 6- Schematic representation of 4 possible haplotypes for the 2 loci, one with alleles A1, A2 and another with B1, B2.

If this is not observed, markers are said to be in linkage disequilibrium or gametic association in that specific population. Linked alleles, i.e. alleles that are closely located in the chromosome, have a higher probability of being transmitted to the offspring in block (as recombination did not occur between them), which may result in gametic association in the population. However, the gametic association may also occur due to population genetic effects (e.g.: substructure or admixture).

Both the Hardy-Weinberg and Linkage equilibrium analyses make the following assumptions: a. infinite population size (no genetic drift), b. random mating, c. no mutation, d. no migration and e. no selection. The linkage equilibrium also considers an infinite number of generations. Thus deviations to these equilibriums due to the evolutionary mechanisms may alter the observed allele, genotype and haplotype frequencies in the population, which can make the statistical analysis of DNA evidence difficult (Triggs, Buckleton and Walsh, 2004).

1.1.3.2. Gene diversity

In a population, for diploid data, a higher number of heterozygotes indicates a higher variability of alleles. This variability is thus measured by gene diversity (GD) or Heterozygosity (H), which equals the number of heterozygotes expected in a population at Hardy-Weinberg equilibrium. For haploid data, this is equivalent to haplotype diversity (HD). At a locus we have the following equation for a sample (Nei, 1978):

$$H = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2 \right)$$

Where n is the number of gene copies in the sample, p_i is the frequency of the i^{th} allele (i^{th} haplotype for haploid data), k is the number of alleles (or haplotypes for haploid data).

1.1.3.3. Population comparisons

A population is often divided into groups due to cultural, religious, political reasons. Individuals belonging to these groups tend to establish relationships with each other and generate offspring, favouring the decrease of genetic diversity, since the founders only transmit a small pool of alleles to the offspring (creating genetic structure). To evaluate the genetic structure within or between population(s), the fixation index (F_{ST}) is used (Wright, 1949). It calculates the difference between the observed genetic diversity in the total population (H_T) and the expected in a subpopulation (H_S) relative to the total. It ranges from 0 to 1, and a higher value indicates a higher genetic structure.

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

1.1.3.4. Forensic Parameters

Usually, forensic parameters are estimated to evaluate the informative content of the markers for genetic analysis in a specific population. Since they are calculated *a priori*, before genetic analysis, it can help to decide which markers should be analysed to provide more information in a specific kinship test.

Matching Probability

It is often useful to evaluate the matching probability (MP) of 2 samples randomly selected using the panel of loci analysed. The MP is given by the following algebraic expression for a locus l (for diploid data, i.e, X chromosome in females)(Fisher, 1951):

$$MP_l = \sum_i^n (p_i^2)^2 + \sum_{i < j}^n (2p_i p_j)^2$$

The first part of the equation accounts for homozygous genotypes while the second the heterozygotes. Where p corresponds to the frequency of the i^{th} allele or j^{th} allele, n is the number of alleles. The equation can be applied to haploid data (i.e., X and Y chromosome markers in males):

$$MP_h = \sum_i^n p_i^2$$

Here, p corresponds to the frequency of the i^{th} haplotype.

Power of discrimination

It is also possible to determine the power of discrimination (PD) which is the probability of choosing in a specific population 2 different random genotypes from unrelated individuals. The PD for a locus or haplotype is given by the following expression:

$$PD = 1 - MP$$

A lower MP or higher power of discrimination indicates a higher probability of distinguishing 2 samples using the loci tested, and also higher genotypic diversity for the loci analysed.

Statistical parameters for X chromosome markers

Mean exclusion chance

In kinship tests, such as paternity, it is useful to evaluate the ability to exclude a random person from being the father using a panel of genetic markers. This is assessed by the mean exclusion chance (MEC) or probability of paternity exclusion (PPE), first described for autosomal markers (Ohno, Sebetan and Akaishi, 1982) and later adapted for X markers (Kishida *et al.*, 1997; Desmarais *et al.*, 1998). The equation to estimate MEC for markers in the X chromosome in father-daughter duos is given by (Desmarais *et al.*, 1998):

$$MEC \text{ duos} = 1 - 2 \sum_{i=1}^n f_i^2 + \sum_{i=1}^n f_i^3$$

To estimate MEC when the trio (mother-father-daughter) is completed for genotyping we have:

$$MEC \text{ trios} = 1 - \sum_{i=1}^n f_i^2 + \sum_{i=1}^n f_i^4 - \left(\sum_{i=1}^n f_i^2 \right)^2$$

To estimate MEC when even only paternal half-siblings are available for testing the following equation is used:

$$MEC_{PHS} = \sum_{i=1}^n f_i^2 (1 - f_i)^2 + \sum_{i=1}^n \sum_{j=i+1}^n 2f_i f_j (1 - f_i - f_j)^2$$

The parameters f refers to the frequency of the i^{th} or j^{th} allele of the analysed marker, and n the number of markers evaluated.

1.1.4. Portuguese population

Portugal is a country located in Southern Europe, on the west coast of the Iberian Peninsula (Figure 7). It is divided in 7 regions: 5 from the continent (North, Center, Lisbon Metropolitan area, Alentejo and Algarve), the remaining two belong to the islands (Azores and Madeira autonomous regions) (REGULAMENTO (UE) N. o 868/2014 DA COMISSÃO de 8 de Agosto de 2014). It is inhabited by 9,860,175 people (CENSOS 2021 preliminary results), most of the population live in the coastal area and a third is concentrated in the metropolitan areas of Lisbon and Porto.



Figure 7- Representation of the geographical location of Portugal. Source: [https://commons.wikimedia.org/wiki/File:Portugal_in_Europe_\(-rivers_-mini_map\).svg](https://commons.wikimedia.org/wiki/File:Portugal_in_Europe_(-rivers_-mini_map).svg)

The genetics of the Portuguese population, through the analysis of the Y chromosome and mtDNA, displays the pre-historic population migrations to this location and relationships that were established among them. Their genetic composition is mainly constituted by 70-80% old European lineages from the Modern Humans that migrated “Out of Africa” arriving in Europe approximately 40 000 years ago. While 10-20% belong to Middle East lineages, mainly due to Jew’s migrations, and other communities, such as the Phoenicians. They also exhibit North and Sub-Saharan African lineages. The former may result from the Moor’s invasion and their presence for several centuries in Portugal. The Sub-Saharan lineages may result from the transportation of slaves from Portuguese ex-colonies to the country (Pereira and Ribeiro, 2009).

1.2. Heterosomal short tandem repeat markers

1.2.1.Y Chromosomal STRs: kits, databases, and interpretation guidelines

Several STRs in the Y chromosome (Y-STRs) were described, but to facilitate their genotyping and comparison between laboratories, multiplexes were developed, just as for autosomes (Diegoli, 2015). Initially, a 'minimal' haplotype of 9 loci was defined but later more loci were included to provide greater discriminatory power. Currently, there are commercial kits that include a greater number of loci, as the 17 loci multiplex AmpFLSTR™ Yfiler™ (Life Technologies) (Mulero *et al.*, 2006), the 23 loci multiplex PowerPlex Y23 (Promega Corporation) (Thompson *et al.*, 2013), and more recently the 27-loci multiplex AmpFISTR Yfiler Plus (Life Technologies) (Bai *et al.*, 2016). The kits include Y-STRs with repeat units ranging from tri- to pentanucleotide, as well as simple, compound, and complex markers.

The frequent intrachromosomal rearrangements in the Y chromosome and lack of recombination with the X chromosome make the Y chromosome prone to accumulate deletions or copies of some regions (Jobling, 2008), which may hamper their analysis. When regions are deleted, the loci that they covered are not amplified, which results in their drop-out during PCR. While the copy of some regions originates a multi-copy locus, which may diverge, as explained by the duplication and divergence model (Figure 8)(Butler *et al.*, 2005). In this model, the duplication of a locus results in two alleles of equal size (a and a'), and if they diverge by length mutation, two alleles of different sizes (a and b) are observed. The patterns observed for multi-copy Y-STRs can be misinterpreted as a mixture, having some of these questions and recommendations been addressed by Butler *et al.*, (2005). Thus, during PCR, the same pair of primers generates 1 amplicon for single-copy Y-STRs, while multiple amplicons can be obtained for multi-copy Y-STRs (e.g. DYS385 a/b and DYS3891 I/II) (Figure 9)

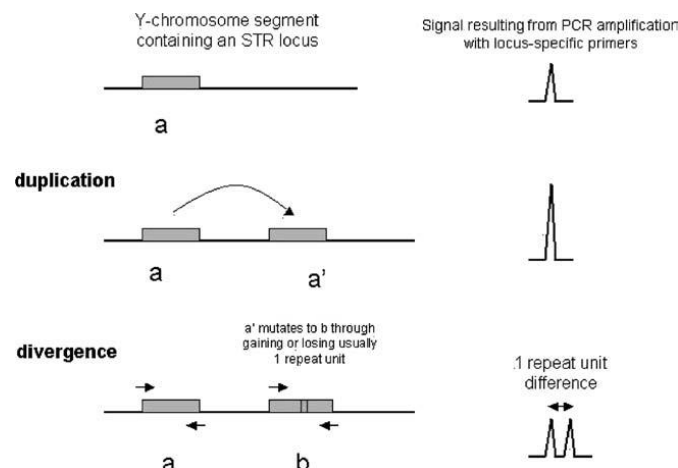


Figure 8- Schematic representation of the duplication and divergence model for Y-STRs that can originate multiple alleles, and their correspondent electropherograms after being amplified with locus specific primer pair. Source: Butler *et al.*, (2005).

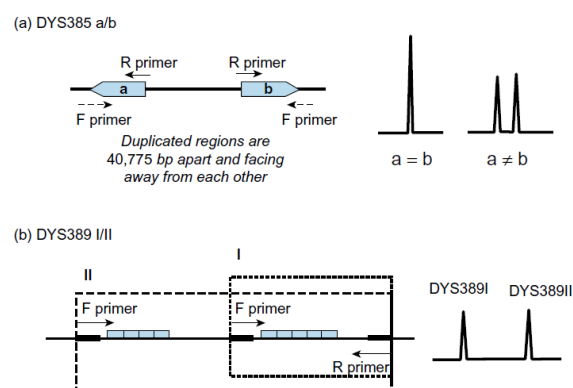


Figure 9- Schematic representation on how multiple PCR primer binding sites originate multi-copy PCR products for DYS385a/b (a), and for DYS389I/II (b). F primer – forward primer; R primer – reverse primer. Source: (Butler, 2012a).

To gather the information related to Y-STRs and facilitate its statistical analysis, some databases were created. There are mainly 2 types of genetic databases: genealogy and forensic population databases, with different purposes. The last accesses the frequency of the haplotypes from populations to evaluate the forensic evidence (Butler, 2012a).

An example of a genetic genealogical database is the Ysearch, no longer available due to its controversial misuse in the identification of persons of interest in criminal investigations (<https://dna-explained.com/category/ysearch/>).

The YHRD (Y-Chromosome Haplotype Reference Database) is a database widely used in population and forensic genetics research. It is freely available at the www.yhrd.org website. It gathers 280,337 haplotypes for the 17-loci Yfiler kit, 98,565

haplotypes for the Powerplex Y23 kit, and 98,274 for the 27-loci Yfiler Plus, along with information on Y-SNP (consulted at 19.08.2021), sampled from worldwide populations. It contains databases from 140 countries (consulted at 19.08.2021). It can provide the frequency of a haplotype for a. the total dataset gathered from worldwide populations in the database, b. the meta-population or c. the national database. Meta-populations are composed of populations that share common paternal ancestors and thus a similar pool of deep-rooting lineages (e.g. Western European) (Willuweit and Roewer, 2015).

The scientific community must correctly evaluate the Y-STRs profiles and present the statistical weight of evidence in reports. Results must also be easily interchanged between laboratories. Therefore interpretation guidelines were established by some entities such as the International Society of Forensic Genetics (ISFG) (Roewer *et al.*, 2020), and the Scientific Working Group on DNA Analysis Methods (SWGDM) and the Forensic Science Regulator in the UK, which may be adjusted according to the regulations of each country.

When DNA profiles from problem and reference samples are compared in identification analysis, when both profiles match, those may belong to the same individual or same male lineage, or even an unrelated individual. Hence reports should address this hypothesis in statistical analysis. To calculate the frequency of a match, one needs to use haplotypes frequencies estimates from the metapopulation or population according to the country recommendations.

As loci from the Y chromosome are linked, not independently transmitted, their frequencies cannot be estimated by independent multiplication of the alleles' frequencies. There are several methods including the counting estimators: a. the counting method, n/N where n is the number of specific haplotypes in a database, and N is the size of the database, b. the augmented counting method, $n+1/N+1$, the augmentation by 1 is due to the addition of the sample profile in question to both the observations and the database, c. the augmented method modified: $n+2/N+2$ adds also the reference sample profile (Schwarz, 1978), c. the kappa "inflation" factor which accounts the proportion of singletons in the database (profiles only observed once in the database) (Brenner, 2010). Besides those, the Discrete Laplace (LD) estimation method estimates Y-STRs frequencies based on the haplotypes on the database (Andersen, Eriksen and Morling, 2013).

While counting methods are conservative because profiles rarely match those in the database, often due to the database' size, estimators like DL, based on haplotype

diversification, retain the “high evidential power” of rare haplotypes even in databases of moderate size (Roewer *et al.*, 2020).

1.2.2.X Chromosomal STRs: kits, databases, and interpretation guidelines

As for autosomes, the X chromosome markers of choice in forensic analyses are the short tandem repeats (X-STRs), for which multiplex PCR assays were developed (Gusmão *et al.*, 2009; Elakkary *et al.*, 2014; Gao *et al.*, 2019). The first version of the most used commercial kit was the 12 loci ArgusX-12 kit, including X-STRs located in the four linkage groups (Figure 3). However it was based on individuals with European ancestry, and new and silent alleles (Kling *et al.*, 2014; Gomes *et al.*, 2020; Pinto *et al.*, 2020), and aberration patterns (Mršić *et al.*, 2018), were observed when genotyping other populations. To overcome this situation, the Investigator Argus X-12 QS Kit (Qiagen) was developed containing newly designed primers (Elakkary *et al.*, 2014). This highlights the importance of genotyping several populations to develop robust STR kits.

Unlike what occurs for autosomal (<https://strider.online/qc>), Y-chromosomal (<https://yhrd.org/>) and mtDNA (<https://empop.online/>) loci, there is not an updated, centralized, and curated worldwide population database containing X-chromosomal data. Some published X-STR data are available at the FamLinkX website (<http://famlink.se/Databases/>) and the outdated X-STR database (<http://www.chrx-str.org/>) (Szibor, Hering and Edelmann, 2006), without quality control (QC) strategies. QC procedures are essential to reduce the number of errors in scientific publications (Bodner *et al.*, 2016). Therefore, a centralized platform gathering information on X-STRs would avoid discrepancies and errors in data exchange and allow comparison among the scientific community (Szibor *et al.*, 2009; Gomes *et al.*, 2016).

The haplodiploid mode of inheritance and characteristics of the X chromosome must be considered in the statistical analysis of forensic evidence. The presence of linkage groups in the chromosome impairs the use of the product rule to estimate frequencies of haplotypes or genotypes considering linked markers. Kling, Dell’Amico and Tillmar, (2015) developed a software that uses an algorithm (Kling *et al.*, 2014) to estimate likelihoods considering linkage, linkage disequilibrium, and mutations in the X markers. ISFG also developed guidelines for the use of X chromosome markers in kinship analysis (Tillmar *et al.*, 2017).

1.2.3.X and Y Chromosome STRs: Portuguese population

YHRD holds a Portuguese database (https://yhrd.org/tools/national_database/Portugal) composed of 1,884 haplotypes (at least with minimal haplotype), 551 haplotypes for the 12 loci PowerPlex Y and 16 loci Yfiler, and only 248 haplotypes for the 23-loci PowerPlex Y23 (consulted at 19.08.2021). Only for the minimal haplotype, the samples were collected from all regions of Portugal (Fernandes et al., 2001; Carvalho et al., 2003; Alves et al., 2007; M.L. Pontes et al., 2007; Bento et al., 2009; Salazar, 2013). This population exhibits a high haplotype diversity, and intra-population analysis shows no significant genetic distances among the different regions (Beleza et al., 2006; Alves et al., 2007). Inter-population analysis revealed higher genetic distances for populations more geographically distant (Carvalho et al., 2003; Alves et al., 2007; M.L. Pontes et al., 2007).

On the other hand, genetic characterization of the Portuguese population was performed by the analysis of sets of X-STRs. Portuguese populations didn't exhibit statistically significant differences among them and its neighbour country, Spain (Pereira et al., 2007; Gusmão et al., 2009; Pinto et al., 2020). Similar results were observed for the different markers analysed in each linkage group (Pinto et al., 2020). Portuguese population samples exhibit a high genetic diversity and discriminatory power for the loci tested in each research (Pereira et al., 2007; Cainé, Costa and Pinheiro, 2013; Pinto et al., 2020).

1.3. STR mutations

Overview

The structure of an STR amplified by PCR is composed of the primer binding sites (forward and reverse), flanking regions (upstream and downstream) to the repeat units (Figure 10). Mutations can occur in any of these regions, changing either their sequence or length. Length mutations are assumed if they are compatible with the results obtained by capillary electrophoresis (CE) but with some uncertainty if the amplicons are not sequenced. For a matter of simplicity length mutations in the repeat units will be discussed.

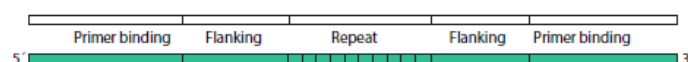


Figure 10- Representative structure of a short tandem repeat locus. Source: (Triggs, Buckleton and Walsh, 2004).

The rates at which mutations occur depend on several factors. Estimating them is critical to establish a molecular clock and help to date historical events (Pereira and Ribeiro, 2009). It is also essential to evaluate DNA results in kinship analysis, as the observation of mendelian incompatibilities may hamper the interpretation of results. For example, mutations may lead to mendelian incompatibilities between (true) parent(s)-child pairs in paternity testing, which has to be properly evaluated, to be compared with the hypothesis of non-relatedness (Ballantyne *et al.*, 2010; Kayser, 2017). It is also likelier to observe more incompatibilities by examining more markers and/or if they have higher mutation rates (Ballantyne *et al.*, 2010; Pinto, Gusmão and Amorim, 2014). This highlights the need to understand the molecular mechanisms that give origin to mutations, their rates, as well as the proper publication of the results to aid the development of theoretical models that estimate them (Amorim and Pinto, 2019).

Mutations are assumed to be germinal as the used methodology does not allow to distinguish them from somatic ones. Three mechanisms of mutations have been proposed for STRs: unequal crossing over during meiosis, retrotransposition mechanism, and strand-slippage during DNA replication (Kornberg *et al.*, 1964; Nadir *et al.*, 1996; Huang *et al.*, 2002). The last is the most accepted mechanism, which may result in the loss or gain of repeat units; the incomplete insertion or deletion of one repeat unit may originate an intermediate allele. This model is also referred to as the Loop-out theory (Levinson and Gutman, 1987; Fan and Chu, 2007).

Mutations in either autosomal, X or Y chromosome markers depend on several factors such as the allele length, repeat unit size, structure, interruptions in the STR sequence, flanking regions sequence, recombination (not applicable to the Y chromosome markers), age and sex of the individual (Fan and Chu, 2007; Ballantyne *et al.*, 2010).

Allele length seems to have a correlation with the mutation rate (Brinkmann *et al.*, 1998; Yang *et al.*, 2018) and to influence the direction of mutation. Antão-Sousa *et al.* (2019) observed that shorter alleles tend to gain repeats (expand) and the longer ones to lose them (contract), a similar result was also found by Ballantyne *et al.*, (2010).

The length of the repeat unit (e.g. di-, tri-, tetra-...) seems to exert an effect on the mutation rate. As length increases, the mutation rate seems to decrease (Chakraborty *et al.*, 1997; Lee *et al.*, 1999; Ballantyne *et al.*, 2010).

The structure of the repeat unit seems to influence the mutation rate, as well as the number of steps involved despite the modest significance of the results. Ballantyne *et al.* (2010) reported a higher mutation rate for repeats with higher differences between the number of purines and pyrimidines. Antão-Sousa (2017) observed GAAA repeats had a higher tendency towards multistep mutation when compared with GATA repeats.

The sex of the individual also affects the likelihood of mutation transmission, with men being more prone to suffer a mutation in germline cells than women, probably due to the higher number of cell divisions to form gametes in men, and this rate tends to increase as men get older (Hardy, 2010). Indeed, the rate of paternal mutations showed to be higher than maternal ones for both autosomal and X chromosome STRs (Weber and Wong, 1993; Brinkmann *et al.*, 1998; Dauber *et al.*, 2012; Tomas, Pereira and Morling, 2012; Diegoli *et al.*, 2014; Chen *et al.*, 2020; Gomes *et al.*, 2020). While in Y-STR mutation studies old fathers transmitted more mutations than younger ones (Brinkmann *et al.*, 1998; Ballantyne *et al.*, 2010; Zhao *et al.*, 2015),

Populations may differ in their mutation rates due to their allelic diversity, and cultural behaviour, which impact the age at which people become parents (Jobling and Tyler-Smith, 2017). Differences were observed for some populations (Diegoli *et al.*, 2014; Chen *et al.*, 2020). Yet, this is difficult to evaluate as several data is needed to obtain significant results.

1.3.1. Estimation of STR mutation rates

Mutations are rare events, which imply that a great amount of data is necessary to increase confidence in rate estimates. Therefore, the collaboration between research groups and laboratories is fundamental to gather data. Under this scope, the Spanish and Portuguese Speaking Working Group of the International Society for Forensic Genetics (GHEP-ISFG) organized two working groups to study mutation rates in X- and Y-chromosomal STRs.

Mutation rates are generally estimated by the direct count of the observed mutations transmitted from parent to child, through the analysis of parent(s)-child duos or trios, depending on the mode of transmission considered. Thus, the mutation rate equals the number of mutations divided by the number of allelic transmissions. STR mutation rates have been reported to range between 10^{-4} and 10^{-2} .

Mutations are better detected when it is straightforward to assign which allele from the parent was transmitted to the offspring. Hence the Y chromosome is the simplest system to study mutations, because it allows the unambiguous identification of both

origin (father) and mutated alleles. Contrarily, for autosomal and X markers, the identification of both origin (father or mother or both) and mutated allele may be ambiguous. Moreover, in these recombinant systems, “hidden mutations”, i.e. mutations that are not detected because they do not result in incompatibilities, lead to an underestimation of mutation frequency (Pinto, Gusmão and Amorim, 2014).

Moreover, estimates of mutation rates are biased depending on the mode of transmission and type of pedigree analysed, which may affect the accuracy of the statistical analysis. As for both autosomal and X chromosomal markers, mutated alleles are easier detected in trios than in duos. While for duos and for the X markers, mutations are better detected in mother-son, then in father-daughter and finally in mother-daughter pairs (Antão-Sousa, Conde-Sousa, *et al.*, 2019).

The mechanisms of mutation of autosomal STRs are similar to that in heterosomes (X and Y), hence they show similar mutation rates. Therefore, the knowledge gained using the simplest system to study mutations, the Y chromosome, can be applied to the others. Notwithstanding, the others should also be studied to account for their characteristics (including the effect of recombination).

Collaborative studies reported that mutation rates per Y-STR marker varied between 5.7×10^{-4} (DYS438) and 6.1×10^{-3} (DYS439) (Antão-Sousa *et al.*, 2017), while for X-STRs those varied between 7.5×10^{-4} (DXS7423) and 1.1×10^{-2} (DXS10135) (Pinto *et al.*, 2020).

1.3.2. Portuguese data in heterosomal STR mutation studies

Only a few studies have analysed Y-STR segregation data from the Portuguese population to estimate Y-STR mutation analysis (Gusmão *et al.*, 2005; M.L. Pontes *et al.*, 2007; Sánchez-Diz *et al.*, 2008; Vieira-Silva *et al.*, 2009; Antão-Sousa *et al.*, 2017). From those, only M. L. Pontes *et al.*, (2007) and Vieira-Silva *et al.*, (2009) analysed specifically the mutation rates for the Portuguese population, using the data from 45 and 95 father-son duos, respectively, for the 17 Y-STRs using the Yfiler kit.

Regarding X-chromosomal markers, only Pinto *et al.*, (2020) included 211 trios from the Portuguese population to analyse mutation rates of the kit Argus-X12 QS, Qiagen.

2. Aims

The knowledge of X-STR and Y-STR genetic diversities and mutation behaviour is fundamental for kinship investigations and other forensic cases, as well as for population genetics and phylogenetic studies. In this work a sample from the Portuguese population will be genotyped with either the PowerPlex® Y23 System (Promega) or the Investigator® Argus X-12 QS (Qiagen), to improve the genetic characterization of the Portuguese population on these heterosomal markers and corresponding mutation rates. To accomplish this, the following main goals were established:

- a. Provide new haplotypic data for the Portuguese population to obtain reliable estimates of allele and haplotype frequencies for both the set of 23 Y-STRs and each of the four linkage groups of the 12 X-STRs. Those will contribute to the creation of a Portuguese database for X-STRs, while Y-STRs haplotypes will be added to the Portuguese database, at YHRD;
- b. Perform pairwise comparisons between the new data gathered on this study and those from other studies previously published for the Portuguese population;
- c. Calculate forensic parameters for the gathered data and evaluate their relevance for analysis;
- d. Improve the estimates of specific mutation rates for the analyzed markers (including paternal and maternal ones in the case of X-STRs), gathering the new data with those previously published, as well as strengthen the correlation between parents' age at the time of the child's birth on the transmission of mutations.

3. Materials and Methods

3.1. Sampling

Blood samples were collected from 750 healthy individuals from the population of mainland Portugal; 434 of these samples were collected from volunteers with legally confirmed paternity, while the remaining 316 were taken from former paternity cases at IPATIMUP (whose kinship relationship was confirmed by autosomal STRs). These individuals represented a total of 202 father-son duos and 135 mother-father-daughter trios.

Both projects were approved by the Ethics and Responsible Conduct in Research Committee of the i3S – Institute of Research and Innovation in Health, University of Porto. All the participants provided written informed consent for investigation (legal representatives in case of underage), all the samples were anonymized, and minimal personal data (age and place of birth) were gathered from each participant.

Blood samples were collected from the volunteer's finger using disposable lancets (Accu-Chek) onto Whatman® FTA® cards (Merk) or GenReleaz™ cards (Ahlstrom-Munksjö), following manufacturer's protocol.

3.2. DNA Extraction

Chemical reagents were prepared for DNA extraction: saline sodium citrate buffer (SSC) 20X stock solution (3 M NaCl (Sigma-Aldrich), 300 mM Sodium citrate (May & Baker), pH 7.8); 5% (w/v) and 1% (w/v) Chelex® 100 (Bio-Rad) solutions. The final volume of each solution was adjusted with sterile H₂O.

The methodology used to extract the DNA from the blood samples depended on the card type used, the card sample's age, and dryness level. DNA was extracted from Whatman® FTA® cards using the standard Chelex method or the modified version for dry and old bloodstains (Walsh, Metzger and Higuchi, 1991). It was extracted from GenReleaz™ cards using an Elute methodology (Fast and efficient recovery of DNA, Ahlstrom-Munksjö, 2018). Firstly, approximately 3mm² of bloodstain from the card were cut and placed in a 1.5 mL sterile Eppendorf tube; between each sample cut, the scissor was disinfected with sodium hypochlorite, water, and ethanol (99%), consecutively.

In the standard Chelex method, to wash the card, 1mL of sterile water was added to the tube, incubated at room temperature for 30 minutes, and mixed occasionally by

vortexing (Snijders, model 34524). Then the mixture was centrifuged. Most of the supernatant was discarded and 200 µL of 5% (w/v) Chelex® 100 (Bio-Rad) solution (in agitation) was added. The tube was vortexed and further incubated at 56°C for 30 minutes in a Thermal Shake lite (VWR). The tube was once more vortexed, and incubated at 100°C, for 8 minutes. The mixture was vortexed and centrifuged, leaving the DNA in the solution at the top and the resin with the card at the bottom.

In the modified version of the Chelex method, 500 µL of the saline sodium citrate buffer (SSC) 1X was added to the tube containing the card, incubated at room temperature for 30 minutes and vortexed occasionally. Then it was centrifuged, and 400 µL of the supernatant was discarded. Further 500 µL of 1% (w/v) Chelex and 400 µL SSC buffer 1X were added, the mixture was vortexed and incubated at 40°C overnight, with agitation until most of the blood from the card passed to the solution. The tubes were centrifuged, and 950 µL of the supernatant was discarded. Afterwards, 950 µL of 1% Chelex solution was added and incubated at 40°C, for 1 to 2 hours. The tubes were centrifuged and 950 µL of the supernatant was discarded. Then 150 µL of 5% Chelex solution was added. The protocol continued as described for the standard Chelex method, starting from the incubating step at 56°C for 30 minutes.

For both standard and modified versions of the Chelex method, the centrifugations were performed in a microcentrifuge at 14,000 rpm for 4 minutes, and 3 minutes, respectively.

In the elute methodology used to extract DNA from GenReleaz™ cards, the bloodstains were washed with 500 µL of sterile dH₂O and vortexed 3 times for 5 seconds. The water was discarded, and the tube was centrifuged at maximum speed for 1 minute to remove the excess water. Further 50 µL of sterile H₂O was added and heated to 95°C for 25 minutes. The tube was pulse vortexed 10 times and centrifuged for 30 seconds at maximum speed to separate the matrix from the eluate. The card was removed using a sterile pipette tip and discarded. The remaining eluate contained the DNA.

The DNA samples were stored at 4°C or -20°C, if analysed within a week or longer, respectively. After analysis, all samples were stored at -20°C.

3.3. Genotyping

3.3.1. DNA amplification

Y-STRs

The extracted DNA from the father-son duos was amplified for 23 loci Y-STRs using the PowerPlex® Y23 System (Promega) by PCR. This kit allows the co-amplification of 23 Y-STRs markers: DYS576, DYS389I/II, DYS448, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS393, DYS458, DYS385a/b, DYS456 and Y-GATA-H4 (Table 1), with four colour detection dyes.

Table 1- Summary of the sequence, repeat motif, and their correspondent cluster of the Y-STR markers from PowerPlex® Y23 System (Promega).

Marker	Sequence	Repeat Motif	Cluster	Reference
DYS19	[TAGA] ₃ TAGG[TAGA] _n	TAGA	GATA	https://strbase.nist.gov/str_y19.htm
DYS385 a/b	[GAAA] _n	GAAA	GAAA	https://strbase.nist.gov/str_y385.htm
DYS389 I	[TCTG] ₃ [TCTA] _r	TCTA	GATA	https://strbase.nist.gov/str_y389.htm
DYS389 II	[TCTG] _n [TCTA] _p [TCTG] _q [TCTA] _r	-	-	https://strbase.nist.gov/str_y389.htm
DYS390	[TCTG] _n [TCTA] _m [TCTG] _p [TCTA] _q	-	-	https://strbase.nist.gov/str_y390.htm
DYS391	[TCTA] _n	TCTA	GATA	https://strbase.nist.gov/str_y391.htm
DYS392	[TAT] _n	TAT	ATT	https://strbase.nist.gov/str_y392.htm
DYS393	[AGAT] _n	AGAT	GATA	https://strbase.nist.gov/str_y393.htm
DYS437	[TCTA] _n [TCTG] ₂ [TCTA] ₄	TCTA	GATA	https://strbase.nist.gov/str_y437.htm
DYS438	[TTTTC] _n	TTTTC	GAAAA	https://strbase.nist.gov/str_y438.htm
DYS439	[GATA] _n	AGAT	GATA	https://strbase.nist.gov/str_y439.htm
DYS448	[AGAGAT] _n N ₄₂ [AGAGAT] _m	AGAGAT	AGAGAT	https://strbase.nist.gov/str_y448.htm
DYS456	[AGAT] _n	AGAT	GATA	https://strbase.nist.gov/str_y456.htm
DYS458	[GAAA] _n	GAAA	GAAA	https://strbase.nist.gov/str_y458.htm
DYS481	[CTT] _n	CTT	GAA	Amato et al. (2010)
DYS533	[ATCT] _n	ATCT	GATA	Butler et al. (2006)
DYS549	[GATA] _n	GATA	GATA	https://strbase.nist.gov/training/ISHI2014-ChrYinfo.pdf
DYS570	[TTTC] _n	TTTC	GAAA	Butler et al. (2006)
DYS576	[AAAG] _n	AAAG	GAAA	Butler et al. (2006)
DYS635	[TCTA] ₄ (TGTA) ₂ [TCTA] ₂ (TGTA) ₂ [TCTA] ₂ (TGTA) _{0,2} [TCTA] _n	TSTA compound	-	https://strbase.nist.gov/str_y635.htm
DYS643	[CTTTT] _n	CTTTT	GAAAA	Butler et al. (2006)
Y-GATA-H4	[TAGA] _n ATGGATAGATTA [GATG] _p AA [TAGA] _q	TAGA	GATA	https://strbase.nist.gov/str_yh4.htm

PCR was performed following the manufacturer's instructions (PowerPlex® Y23 System Technical Manual #TMD035, Promega, 2019), but using a reduced amount of the recommended reagent's volumes to maximize the number of samples analysed. The

PCR reaction components and cycling conditions are described in Table 2. A negative control, on which the template DNA was replaced by water was analysed to detect if contaminations occurred.

Table 2- PCR reaction setup (left table) with each reagent volume and cycling conditions (right table) used for amplification of 23 Y-STRs using the PowerPlex® Y23 System.

Component	Volume per reaction (µL)	Step	Temperature	Time	Cycles
Water, Amplification Grade	3	1. Initial denaturation	96°C	2 min	-
PowerPlex® Y23 5X Master Mix	1	2. Denaturation	94°C	10 s	28
PowerPlex® Y23 10X Primer Pair Mix	0,5	3. Annealing	61°C	1 min	
Template DNA	0,5	4. Extension	72°C	30 s	
Total	5	5. Final extension	60°C	20 min	-

After amplification, 0.5 µL of the PCR product (or 1.2 µL PowerPlex® Y23 Allelic Ladder Mix) was mixed with 9.8 µL of Hi-Di™ Formamide and 0.2 µL WEN Internal Lane Standard (ILS) 500 Y23 (Promega) and further analysed by capillary electrophoresis (CE).

X-STRs

The extracted DNA from the father-mother-daughter trios was amplified for 12 X-STRs loci using the Investigator Argus X-12 QS Kit (Qiagen) by PCR. It performs the multiplex PCR with four colour detection dyes of the 12 X-STRs: DXS7132, DXS7423, DXS8378, DXS10074, DXS10079, DXS10101, DXS10103, DXS10134, DXS10135, DXS10146, DXS10148 and HPRTB, plus an Amelogenin (AM) marker, and the D21S11 (Table 3).

Table 3- Summary of the repeat motif sequence of the reference allele for the 12 X-STRs from the Investigator Argus X-12 QS Kit (Qiagen). * [AGAT] is the most common repeat structure; for variations see the NIST. Source: adapted from the Investigator® Argus X-12 QS Handbook, Qiagen, 2015.

Locus	Repeat motif of the reference allele
Amelogenin X	–
Amelogenin Y	–
DXS7132	[TCTA] ₁₃
DXS7423	[TCCA] ₃ TCTGTCCT [TCCA] ₁₂
DXS8378	[CTAT] ₁₂
DXS10074	[AAGA] ₁₄
DXS10079	[AGAG] ₃ TGAAAGAG [AGAA] ₁₇ AGAG [AGAA] ₃
DXS10101	[AAAG] ₃ GAAAGAAG [GAAA] ₃ A [GAAA] ₄ AAGA [AAAAG] ₅ AAAAAGAA [AAAG] ₁₃ AA
DXS10103	[TAGA] ₂ CTGA [CAGA] ₁₁ [CAGA] ₄ [TAGA]
DXS10134	[GAAA] ₃ GAGA [GAAA] ₄ AA [GAAA] GAGA [GAAA] ₄ GAGA [GACAGA] ₃ [GAAA] GTAA [GAAA] ₃ AAA [GAAA] ₄ AAA [GAAA] ₁₅
DXS10135	[AAGA] ₃ GAAAG [GAAA] ₂₀
DXS10146	[TTCC] ₃ T [TTCC] ₃ TTTC CTCCCTTCC [TTCC] [TCCC] TTCTTCTTTC [TTCC] ₂ TTTCTT [CTTT] ₂ CTTC [CTTT] ₁₀ T [CTTT] ₂
DXS10148	[GGAA] ₄ [AAGA] ₁₂ [AAAG] ₄ N ₈ [AAGG] ₂
HPRTB	[AGAT] ₁₂ *
D21S11	[TCTA] ₄ [TCTG] ₆ [TCTA] ₃ TA [TCTA] ₃ TCA [TCTA] ₂ TCCATA [TCTA] ₁₁

The PCR was performed following the manufacturer's instructions (Investigator® Argus X-12 QS Handbook, Qiagen, 2015), but using a reduced amount from the recommended reagent's volumes to maximize the number of samples analysed. The PCR reaction components and cycling conditions are described in Table 4. A negative control, on which the template DNA was replaced by water was analysed to detect if contaminations occurred.

Table 4- PCR reaction setup (left table) and cycling conditions (right table) used for amplification of 12 X-STRs using the Investigator Argus X-12 QS Kit.

Component	Volume per reaction (µL)	Step	Temperature	Time	Cycles
Water, Amplification Grade	2,5	1. Initial denaturation	98°C	60 s	3
Fast Reaction Mix 2.0	1,5	2. Annealing	61°C	100 s	
Primer Mix	0,5	3. Extension	72°C	5 s	
Template DNA	0,5	4. Denaturation	96°C	10 s	28
Total	5	5. Annealing	61°C	100 s	
		6. Extension	72°C	5 s	
		7. Final extension	60°C	20 min	-

After amplification, 0.5 µL of the PCR product (or 1 µL Allelic ladder Argus X-12 QS) was mixed with 9.8 µL Hi-Di™ Formamide and 0.2 µL of DNA Size Standard 550 (BTO) and further analysed by CE.

3.3.2. Capillary Electrophoresis (CE)

The amplified DNA fragments (Y-STRs and X-STRs) were separated by length using capillary electrophoresis (CE) using the ABI 3500 Genetic Analyzer (Applied Biosystems™), with POP-7™ Polymer and a capillary length of 50 cm, using the manufacturer's recommended running conditions. They were further genotyped using GeneMapper® Software v5 using the recommended panels and bins files.

3.4. Data analysis

A total of 129 unrelated daughters and 200 unrelated sons were selected for population analysis and calculation of forensic parameters for the 12 X-STRs and 23 Y-STRs, respectively. From the females, 258 unrelated haplotypes were obtained after inferring their gametic phase by comparison with the father's haplotype and mother's genotype. In each family cluster, only one daughter was randomly selected and considered for the analysis, to avoid result biases. From males, 200 unrelated haplotypes were obtained. For DYS389II, Y-STR marker, alleles were determined by subtracting the total repeat number of the DYS389I from the DYS389II since the DYS389I sequence is included in the DYS389II (Figure 9b). The alleles from the multi-copy locus DYS385ab were considered as single alleles for the analysis.

Pairwise Fst genetic distance analysis between the data gathered in the study and others (for both Y-STRs and X-STRs), as well as Hardy-Weinberg equilibrium and pairwise Linkage disequilibrium (LD) testes between X-STRs loci, were performed using software Arlequin v3.5.2.2 (Excoffier and Lischer, 2010). A significance level of 0.05 was

considered, having Bonferroni's correction been properly applied in case of multiple tests.

Allele and haplotype frequencies, as well as gene and haplotype diversities, were also determined by software Arlequin v3.5.2.2 (Excoffier and Lischer, 2010). The matching probability (MP) was estimated using the sum of the squared haplotype frequencies. The discriminatory capacity (DC) was determined by the ratio between the number of unique haplotypes by the total number of haplotypes. Other forensic parameters for X-STRs data were estimated, such as the power of discrimination (PD) for males and females; and the mean exclusion chance (MEC) of paternity exclusion in trios involving daughters, duos father-daughter, and in duos of paternal half-siblings, using the equations described in section 1.1.3.4, developed by Desmarais *et al.*, (1998) and Pinto *et al.*, (2013), implemented through in house developed algorithms.

Mutation rates analysis

A total of 135 mother-father-daughter trios and 202 father-son duos were used to estimate marker specific mutation rates for the analysed 12 X-STRs and 23 Y-STRs, respectively. The data was pooled together with corresponding data from the literature to obtain more accurate mutation rate estimates. X-STR mutation rates were estimated using data from the researches: Athanasiadou *et al.*, (2003); Turrina and De Leo, (2004); Tabbada *et al.*, (2005); Turrina, Atzei and De Leo, (2007); Becker *et al.*, (2008); Tariq *et al.*, (2008); Liu *et al.*, (2008); Liu *et al.*, (2012); Pico *et al.*, (2008); Tetzlaff, Wegener and Lindner, (2012); Nishi *et al.*, (2013); Diegoli *et al.*, (2014); Chen *et al.*, (2020); Pinto *et al.*, (2020). While Y-STRs mutation rates were estimated using the following data: Antão-Sousa, (2017), included data from several articles; Yang *et al.*, (2018); Ambrosio *et al.*, (2020); Lin *et al.*, (2020). Mutation rates were calculated considering the number of incompatibilities observed divided by the total number of allele transfers for each marker, except in the case of DYS389I/II where a mutation in DYS389I results in an incompatibility in DYS389II. Confidence intervals (CI) were estimated from the binomial standard deviation (Clopper and Pearson, 1934), which formula was implemented using Microsoft® Excel® 2016.

4. Results and Discussion

4.1. Y Chromosome

4.1.1. Genotyping analysis

During the 23 Y-STR genotyping analysis, variant alleles not exhibiting simple repetitions of the microsatellite motif unit tested were detected. Intermediate alleles i.e. exhibiting a non-integral number of repeats, were observed for DYS458 (17.2, 18.2 and 19.2) and DYS385ab (12.2). One di-allelic pattern was observed at the DYS439 locus and a tri-allelic one at DYS19 (Figure 11), these observations suggest a duplication and a triplication, respectively. These patterns were observed in two distinct duos, in both father and son. Similar observations were also reported in other studies, namely duplications at DYS439 by Diederiche *et al.*, (2005) and Soares-Vieira *et al.*, (2008); and triplications at DYS19 by Santos *et al.*, (1996) and Purps *et al.*, (2014).

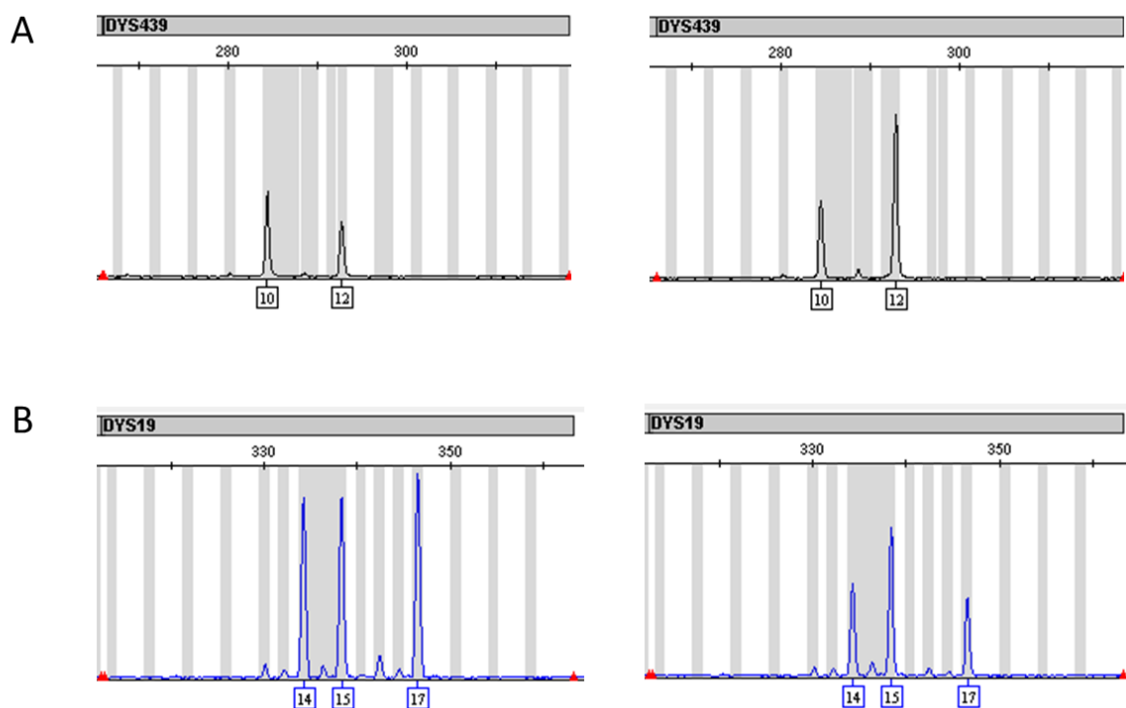
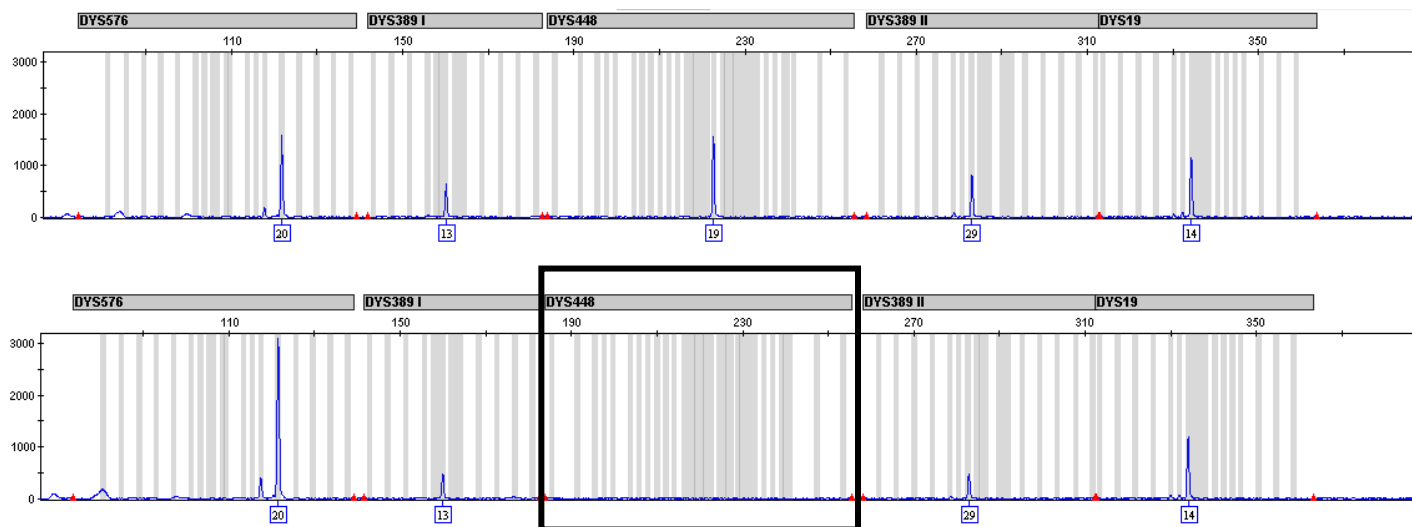


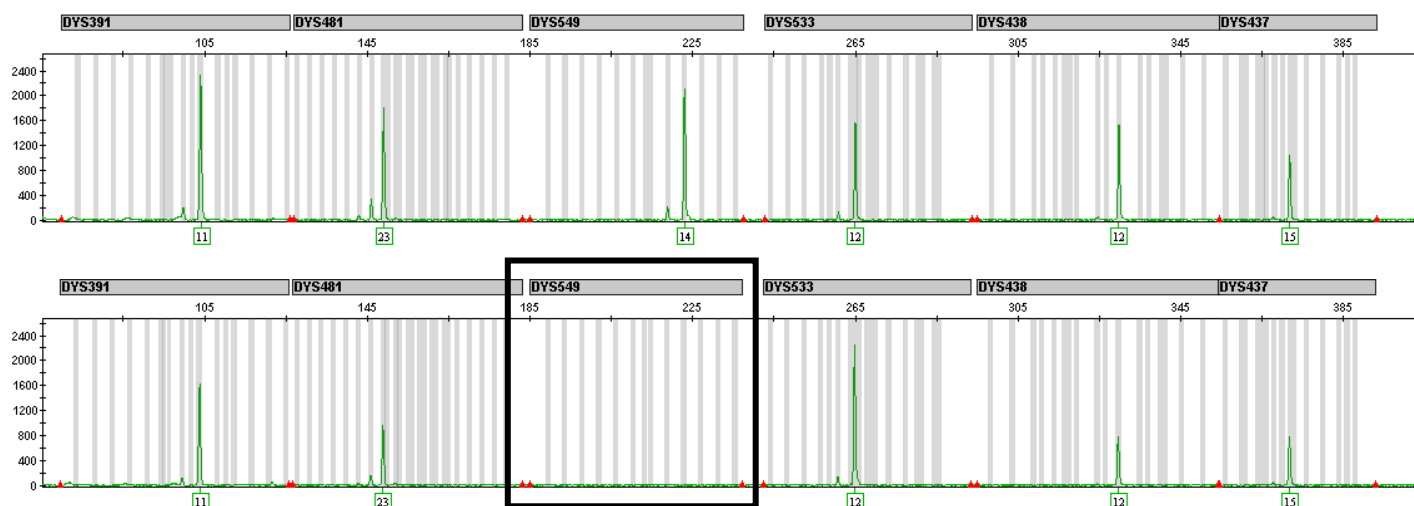
Figure 11- Electropherograms showing di-allelic (A) and tri-allelic (B) patterns at the DYS439 and DYS19 markers, respectively, considering 2 father (left) – son (right) duos.

The amplification of the adjacent markers DYS448, DYS549 and DYS392 was absent in the two times that a son's sample was analysed. As the father's haplotype exhibited those alleles (Figure 12), a *de novo* deletion from father to son seems to have occurred. The microdeletion of the referred markers and another adjacent (DYS385ab) was previously observed in Beltramo, Pena and Lojo, (2015). These findings can be explained by the intrachromosomal rearrangements on the Y chromosome through non-allelic homologous recombination, due to its highly repetitive regions. This enhances deletions or copies of some regions.

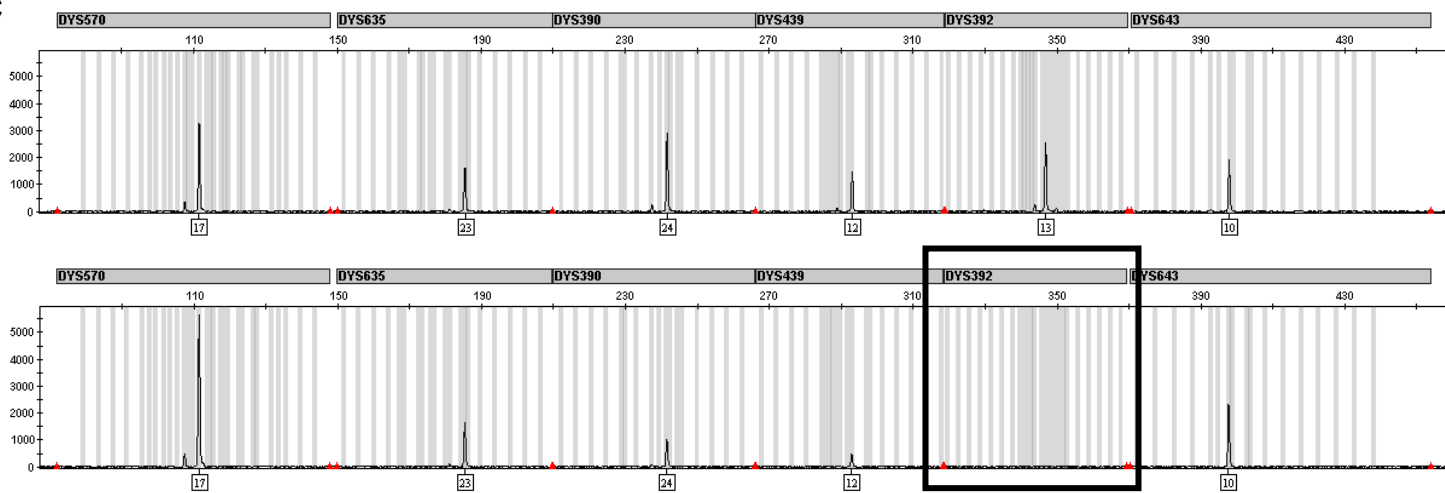
A



B



C



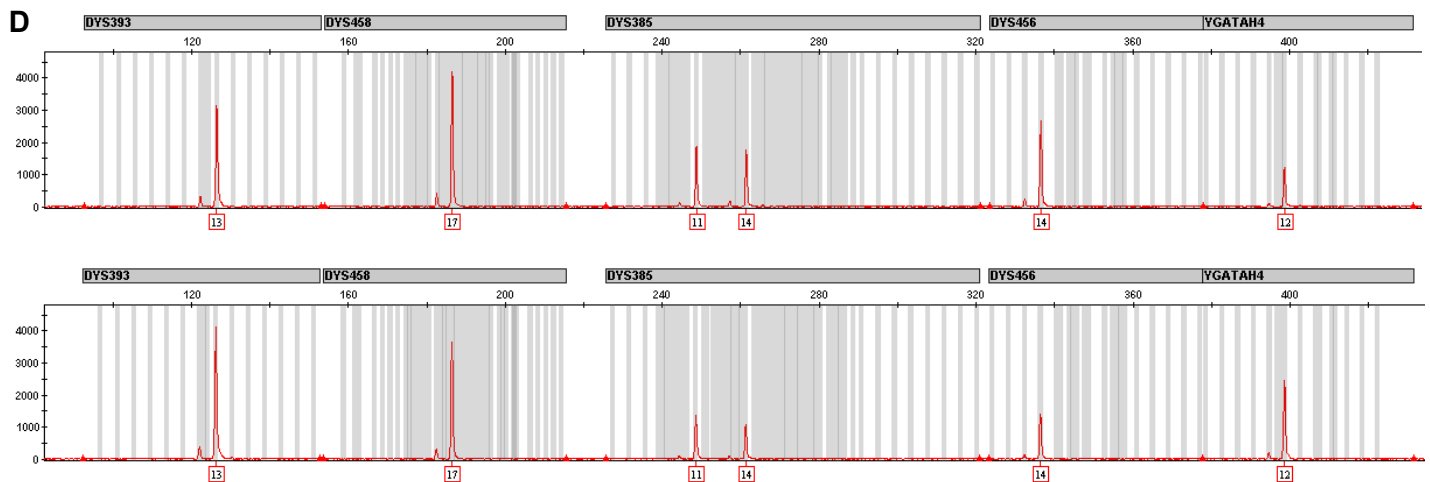


Figure 12- Electropherograms from samples: father on top, and son at the bottom, for each set of markers coloured in blue (A), green (B), black (C), and red (D). The black box highlights the absence of amplification for the DYS448, DYS549 and DYS392 markers in the son's sample.

4.1.2. Population genetic analysis

Two hundred unrelated males from the Portuguese population were selected for population genetic analysis. The corresponding 200 haplotypes showed to be unique (Supplementary Table 1). The set of haplotypes was compared with others from the literature also from the Portuguese population (Salazar, 2013) (N=250), for the same 23 Y-STRs. Pairwise F_{ST} comparisons for the 23 Y-STRs between the two studies were performed, the results listed in Table 5. The results show (nearly) no significant differences between the data tested ($F_{ST} = 0.0025$; $p\text{-value} = 0.0469 \pm 0.0023$), for a significance level of 0.05.

Table 5- Pairwise genetic distances (F_{ST}) (lower diagonal) between the 2 groups: This work (N=200) and Salazar (2013) (N=250), and corresponding non-differentiation P-values (upper diagonal), obtained after 10100 permutations.

	Salazar (2013)	This study
Salazar (2013)	*	0.0469 ± 0.0023
This study	0.0025	*

Since no significant differences were found between the samples, the haplotypes were gathered for haplotype analysis. Out of the 450 haplotypes, 433 were unique, with a relative frequency of 0.0022 ± 0.0022 , while 17 were found twice, exhibiting a frequency of 0.0044 ± 0.0031 . The sample exhibited a haplotype diversity of $1.0000 \pm$

0.0001. A matching probability (MP) of 2.39×10^{-3} , and a discriminatory capacity (DC) equal to 0.9622. This indicates a higher discriminatory power for forensic purposes using the PowerPlex Y23 kit in the Portuguese population.

The allele frequencies for each Y-STR locus were calculated for the complete set of haplotypes (Table 6). A total of 189 different alleles were recorded for the 23 Y-STRs loci, with allele frequencies varying between 0.0022 and 0.6778. The number of alleles for each locus ranged from 4 (DYS391) to 13 (DYS458), while the multi-copy locus DYS385ab exhibited 41 allele combinations. The marker DYS19 exhibited a triplication (alleles 14, 15, 17), while DYS439 showed a duplication (alleles 10, 12). Intermediate alleles were only observed for 2 loci (DYS458 and DYS385ab).

Table 6- Allele frequencies calculated for 23 Y-STRs with data from 450 haplotypes from this work (N=200) and from Salazar (2013) (N=250).

DYS576			DYS19			DYS549			DYS570			DYS439			DYS458			DYS385ab		
Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.
14	6	0.0133	13	55	0.1222	10	12	0.0267	14	2	0.0044	9	2	0.0044	13	1	0.0022	8,14	1	0.0022
15	14	0.0311	14	265	0.5889	11	62	0.1378	15	16	0.0356	10	39	0.0867	14	5	0.0111	9,14	1	0.0022
16	48	0.1067	15	102	0.2267	12	180	0.4000	16	59	0.1311	11	132	0.2933	15	42	0.0933	10,14	3	0.0067
17	120	0.2667	16	20	0.0444	13	154	0.3422	17	157	0.3489	12	211	0.4689	16	109	0.2422	10,15	3	0.0067
18	146	0.3244	17	7	0.0156	14	36	0.0800	18	112	0.2489	13	60	0.1333	17	167	0.3711	11,11	8	0.0178
19	79	0.1756	14,15,17	1	0.0022	15	6	0.0133	19	54	0.1200	14	4	0.0089	17.2	3	0.0067	11,12	9	0.0200
20	25	0.0556							20	14	0.0311	15	1	0.0022	18	78	0.1733	11,13	18	0.0400
21	10	0.0222							21	9	0.0200	10.12	1	0.0022	18.2	8	0.0178	11,14	148	0.3289
22	2	0.0044							22	16	0.0356				19	24	0.0533	11,15	41	0.0911
DYS389I			DYS391			DYS533			DYS635			DYS392			DYS456			DYS385ab		
Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.
11	1	0.0022	9	30	0.0667	8	1	0.0022	19	1	0.0022	11	163	0.3622	12	1	0.0022	12,12	9	0.0200
12	73	0.1622	10	212	0.4711	9	11	0.0244	20	17	0.0378	12	16	0.0356	13	12	0.0267	12,13	5	0.0111
13	294	0.6533	11	202	0.4489	10	21	0.0467	21	86	0.1911	13	240	0.5333	14	40	0.0889	12,14	28	0.0622
14	81	0.1800	12	6	0.0133	11	117	0.2600	22	42	0.0933	14	27	0.0600	15	201	0.4467	12,15	10	0.0222
15	1	0.0022				12	261	0.5800	23	247	0.5489	15	3	0.0067	16	142	0.3156	12,16	3	0.0067
DYS448			DYS481			DYS438			DYS390			DYS643			Y-GATA-H4			DYS385ab		
Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.
17	2	0.0044	19	1	0.0022	9	61	0.1356	20	10	0.0222	8	12	0.0267	10	12	0.0267	12,17	2	0.0044
18	33	0.0733	20	5	0.0111	10	119	0.2644	21	52	0.1156	9	30	0.0667	11	166	0.3689	13,13	5	0.0111
19	223	0.4956	21	40	0.0889	11	22	0.0489	22	115	0.2556	10	244	0.5422	12	240	0.5333	13,14	33	0.0733
20	125	0.2778	22	215	0.4778	12	232	0.5156	23	227	0.5044	11	69	0.1533	13	31	0.0689	13,15	17	0.0378
21	50	0.1111	23	28	0.0622	13	16	0.0356	24	41	0.0911	12	76	0.1689	14	3	0.0067	13,16	12	0.0267
22	16	0.0356	24	28	0.0622				25	9	0.0200	13	15	0.0333	15	1	0.0022	13,17	7	0.0156
23	1	0.0022	25	23	0.0511							14	3	0.0067	16	1	0.0022	13,18	4	0.0089
DYS389II.I			26	19	0.0422				DYS437			DYS393			Y-GATA-H4			DYS385ab		
Allele	N	Freq. rel.	27	15	0.0333	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.	Allele	N	Freq. rel.
14	4	0.0089	28	10	0.0222	13	1	0.0022	21	10	0.0222	12	82	0.1822	10	12	0.0267	14,14	15	0.0333
15	30	0.0667	29	3	0.0067	14	147	0.3267	22	52	0.1156	13	305	0.6778	11	166	0.3689	14,15	10	0.0222
16	284	0.6311				15	252	0.5600	23	115	0.2556	14	54	0.1200	12	240	0.5333	14,16	6	0.0133
17	105	0.2333				16	49	0.1089	24	227	0.5044	15	8	0.0178	13	31	0.0689	14,17	2	0.0044
18	25	0.0556				17	1	0.0022	25	41	0.0911	16	1	0.0022	14	3	0.0067	14,18	5	0.0111
19	2	0.0044							26	5	0.0111				15	1	0.0022	14,19	1	0.0022
															16	1	0.0022	15,15	9	0.0200
															17	1	0.0022	15,16	6	0.0133
															18	1	0.0022	15,17	4	0.0089
															19	1	0.0022	15,18	2	0.0044
															20	1	0.0022	15,19	1	0.0022
															21	1	0.0022	16,16	6	0.0133
															22	1	0.0022	16,17	5	0.0111
															23	1	0.0022	16,18	4	0.0089
															24	1	0.0022	16,19	1	0.0022
															25	1	0.0022	16,21	1	0.0022
															26	1	0.0022	17,17	1	0.0022
																		17,19	1	0.0022

Gene diversities (GD) were calculated for each locus, as shown in Table 7. Those were higher for the multi-copy locus DYS385ab (GD=0.8671), while for single copies the values ranged from 0.4938 (DYS393) to 0.7823 (DYS570). This indicates a high polymorphism in the Portuguese population for the 23 Y-STRs. The results are similar to those found for populations with European ancestry (Purps *et al.*, 2014).

Table 7-Genetic diversity values obtained for 23 Y-STR loci, with data from 450 haplotypes, collected in this work (N=200) and from Salazar (2013) (N=250). The values are sorted in a descending order.

Marker	Gene Diversity
DYS385ab	0.8671
DYS570	0.7823
DYS576	0.7784
DYS458	0.7627
DYS481	0.7145
DYS549	0.6982
DYS456	0.6844
DYS439	0.6702
DYS448	0.6597
DYS390	0.6594
DYS643	0.6491
DYS438	0.6437
DYS635	0.6417
DYS533	0.5881
DYS19	0.5860
DYS392	0.5807
Y-GATA-H4	0.5753
DYS391	0.5732
DYS437	0.5691
DYS389II.I	0.5408
DYS389I	0.5156
DYS393	0.4938

4.1.3. Mutations analysis

Considering the 4,644 allele transfers analysed for 23 Y-STRs from the 202 father-son duos, 15 mutations were observed. The haplotypic configurations of the duos that exhibit mutations are summarized in Table 8. Mutations were observed for 11 loci, and all were single step. In 10 out of the 15 cases, mutations are compatible with the loss of one repeat unit, while in 5 are compatible with the gain of one. No duo exhibited more than one mutation.

Table 8- Father's and son's alleles from the duos where a mutation was observed.

Sample	Y-STR marker	Father's allele	Son's allele
N247/N248	DYS576	18	17
N313/N314		20	21
Z7251/Z7248		19	18
N128/N129	DYS389I	12	13
N83/N54	DYS389II	31	30
N229/N230	DYS19	14	15
N1/N2	DYS481	22	23
N258/N259		29	28
Z6565/Z6564	DYS549	13	12
Z6658/Z6660	DYS635	22	21
Z8295/Z8292	DYS390	24	23
Z8387/Z8386	DYS458	18	17
N210/N209	DYS456	16	15
Z6746/Z7648		16	17
N305/N306	YGATAH4	12	11

During Y-STR mutation counting, 2 incompatibilities were observed for a duo (N128/N129), as the father has the alleles 12 and 29, while his son alleles 13 and 30 for markers DYS389I and DYS389II, respectively; however only one mutation was counted for DYS389I. Since a mutation only in DYS389I would result in the observation of an incompatibility DYS389II.

4.1.3.1. Mutations and father's age

From the 202 father-son duos analysed in this study, it was possible to collect information regarding the father's age at the time of the son's birth, except for one duo. The father's age distribution is represented in Figure 13, it shows that most of the analysed fathers (82%) had a son between 20 and 37 years old, with an average of 31.2 (± 7.7) years. Fathers that transmitted a mutation had on average 31.3 (± 6.6) years, while those that transmitted the allele had on average 31.2 (± 7.8) years. This similarity may be due to the high proportion of analysed fathers around this age. While a positive correlation between the mutation rate and the father's age has been observed for other studies (Brinkmann *et al.*, 1998; Ballantyne *et al.*, 2010; Bugoye, Mulima and Misinzo, 2018).

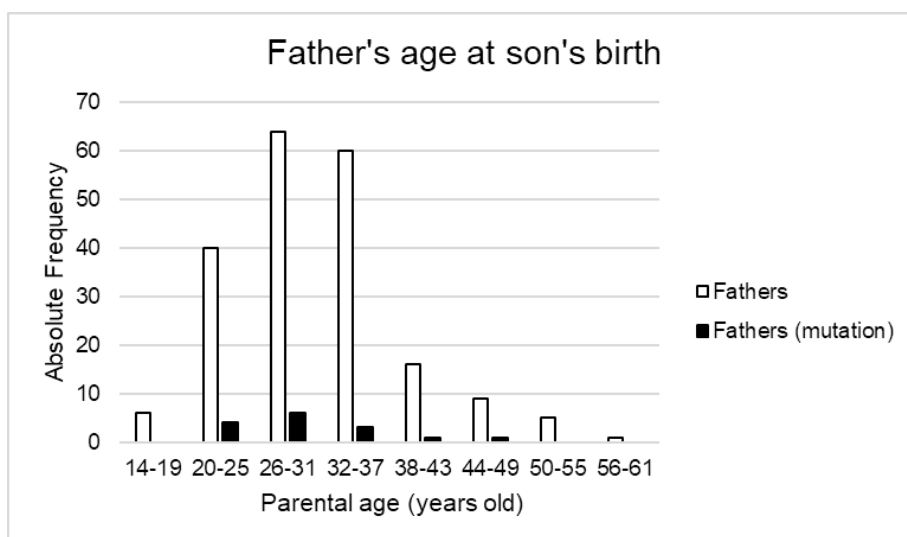


Figure 13- Distribution of the father's age at time of the son's birth, and the father's age that transmitted a mutation.

4.1.3.2. Locus-specific mutation rates

The data gathered in this study was compared with those previously reported in the literature to estimate locus-specific mutation rates (Antão-Sousa, 2017; Yang *et al.*, 2018; Ambrosio *et al.*, 2020; Lin *et al.*, 2020). The results are presented in Supplementary Table 2 (mutation rates) and Supplementary Table 3 (mutation steps), a summary of those is listed in Table 9.

Table 9- Average specific mutation rates obtained for each 23 Y-STR from the PowerPlex Y 23 kit, after pooling the data from this work and the literature (Antão-Sousa, 2017; Yang *et al.*, 2018; Ambrosio *et al.*, 2020; Lin *et al.*, 2020). {+1} - gain of 1 repeat ; >{+1} - gain of more than 1 repeat; {-1} - loss of 1 repeat; >{-1} -loss of more than one repeat.NI- non-integer unit variation

Marker	No. Allele transfers	No. Mutations	Mutation rate X 10 ⁻³	95 % CI X 10 ⁻³	{+1}	{-1}	>{+1}	<{-1}	NI	No. repeat gains	No. repeat losses
DYS576	5857	91	15.5	12.5-19	44	43	1	3	0	45	46
DYS389I	24292	54	2.2	1.7-2.9	23	31	0	0	0	23	31
DYS448	13063	16	1.2	0.7-2	6	9	0	1	0	6	10
DYS389II	24239	89	3.7	2.9-4.5	40	46	3	0	0	43	46
DYS19	25471	50	2.0	1.5-2.6	27	22	0	0	1	27	22
DYS391	25045	51	2.0	1.5-2.7	29	21	1	0	0	30	21
DYS481	2695	14	5.2	2.8-8.7	9	4	0	1	0	9	5
DYS549	1661	5	3.0	1-7	1	4	0	0	0	1	4
DYS533	3505	5	1.4	0.5-3.3	3	2	0	0	0	3	2
DYS438	16181	7	0.4	0.2-0.9	1	1	0	5	0	1	6
DYS437	16161	19	1.2	0.7-1.8	11	8	0	0	0	11	8
DYS570	5903	44	7.5	5.4-10	16	24	1	3	0	17	27
DYS635	13712	40	2.9	2.1-4	14	26	0	0	0	14	26
DYS390	24730	48	1.9	1.4-2.6	22	26	0	0	0	22	26
DYS439	16554	76	4.6	3.6-5.7	38	38	0	0	0	38	38
DYS392	24993	11	0.4	0.2-0.8	7	3	0	1	0	7	4
DYS643	1294	0	0.0	0-2.8	0	0	0	0	0	0	0
DYS393	22986	21	0.9	0.6-1.4	13	8	0	0	0	13	8
DYS458	13095	89	6.8	5.5-8.4	45	41	1	2	0	46	43
DYS385	39378	86	2.2	1.7-2.7	61	20	1	4	0	62	24
DYS456	13646	52	3.8	2.8-5	29	23	0	0	0	29	23
YGATAH4	14224	30	2.1	1.4-3	12	17	0	1	0	12	18
Total	348685	898	2.6	2.4-2.7	451	417	8	21	1	459	438

The DYS576 and DYS570 markers exhibited the highest mutation rates: 1.6×10^{-2} and 7.5×10^{-3} , respectively. Those values are above the mean mutation rate found for the 23 Y-STR markers: 2.6×10^{-3} (95% CI $2.4 \times 10^{-3} - 2.7 \times 10^{-3}$). Markers DYS643, DYS438 and DYS392 exhibited the lowest mutation rates, below average: 0, 4.3×10^{-4} and 4.4×10^{-4} , respectively. The values obtained are similar to the ones reported on the YHRD. High mutation rates were already observed for the DYS576 and DYS570 locus, and those are considered 'rapidly mutating' Y-STRs (RM Y-STR) (Ballantyne, Keerl, Wollstein, Choi, Zuniga, Ralf, Vermeulen, de Knijff, *et al.*, 2012). The analysis of these RM Y-STRs can give higher power of discrimination between samples but can also be challenging for forensic casework due to the recurrent mutations. Higher mutation rates were observed for the markers (DYS576 and DYS570) that belong to the [GAAA] cluster (Table 1), while the ones with lower rates (DYS643 and DYS438) belong to [GAAAA], demonstrating that the mutation rates are affected by the nature of the repetitive motif. Both rapidly mutating STRs exhibit also longer alleles within their allelic range (allelic range from 8 to 25.1 and from 10 to 27, respectively) when compared with DYS643, DYS438 (allelic range 4 to 17 and from 5 to 19, respectively). Therefore, the size of the alleles seems to exert an effect on the mutation rates.

Out of the 348,685 allele transfers, 898 mutations were observed, 459 (51.1%) compatible with the gaining of repeats, 438 (48.8%) with the loss of repeats, while 1 (0.1%) exhibited a non-integer unit variation. Therefore, the data does not indicate an allele contraction or expansion bias as in Yang *et al.*, (2018). Most of the mutations were compatible with one-step ones (96.7%), either gain or lost a repeat unit, those which are more prone to occur by strand slippage as already suggested by the single-step stepwise mutation model (SMM) (Ohta and Kimura, 2008).

The inclusion of data from other studies allowed the increase of the accuracy of the estimates, decreasing the amplitude of the confidence intervals obtained in this study.

4.2. X Chromosome

4.2.1. Genotyping analysis

During X-STR genotyping analysis some microvariants and off-ladder alleles were detected. It was also observed a tri-allelic pattern at DXS10079 (alleles 18, 20 and 21) in one mother's sample (Figure 14), which was confirmed with a second amplification. Her daughter exhibited the alleles 18 and 22 for that locus, while the father the allele 22, suggesting that the daughter inherited only the allele 18 from her mother, while the allele 22 from her father. It is possible to speculate that the mother has an X chromosome with allele 18, while the other contains alleles 20 and 21. Another possibility is the presence of mosaicism, a mixture of 2 cell types, some with alleles 18 and 20, and others with alleles 18 and 21.

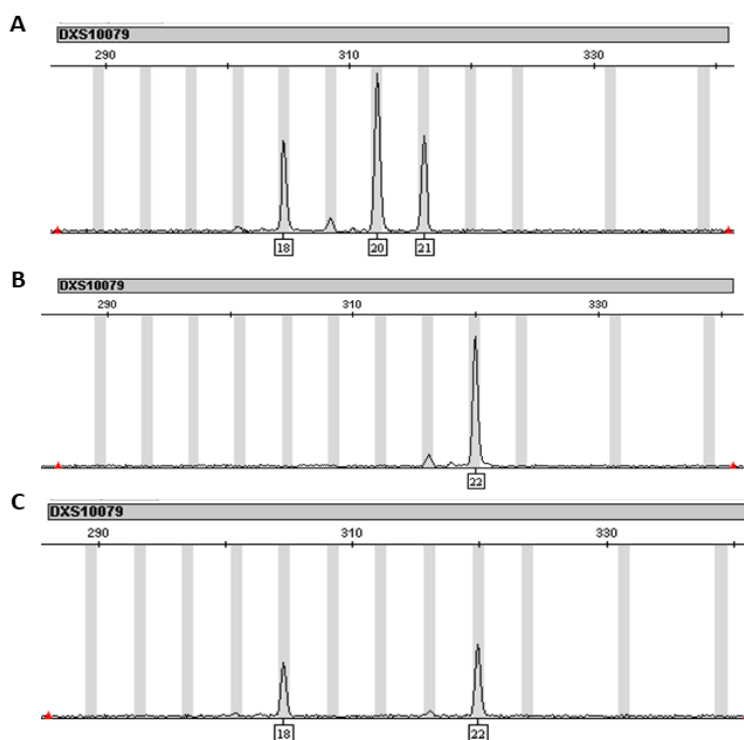


Figure 14- Electropherograms showing the DXS10079 marker from the samples of the mother (A), father (B), and daughter (C), exhibiting the tri-allelic pattern in the mother's sample.

Off-ladder alleles are those that do not align with any of the alleles present in the allelic ladder. Among the analysed samples, some occurred within a locus, while others between loci. In one trio, both father and daughter exhibited an off-ladder allele occurring between DXS10148 and D21S11 (Figure 15). This off-ladder allele was assigned to the

DXS10148 marker, since the father did not show any allele for this marker, while he is heterozygous for D21S11. The daughter inherited this off-ladder allele from her father.

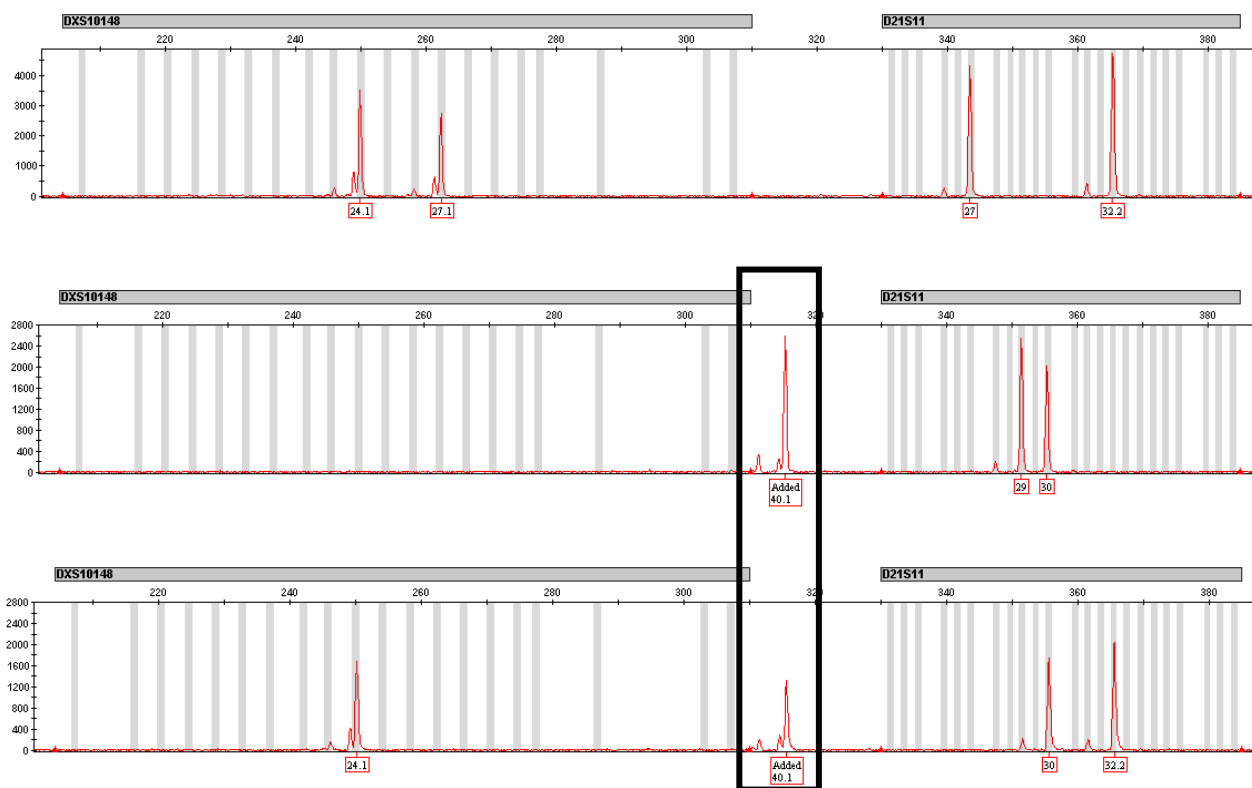


Figure 15- Electropherograms showing the off-ladder between DXS10148 and D21S11, highlighted by the black box, from samples of the mother on top, father at the centre, and daughter at the bottom.

Off-ladder alleles within the DXS10146 were observed. Several alleles typed as 39.2 exhibited slight differences in their electrophoretic mobility (of less than one bp variation) compared with the allelic ladder (e.g., Figure 16). These shifts could result from 39.2 isoalleles, variants that exhibit the same size but a different internal sequence. Another hypothesis is the presence of 39.1 alleles, some have shown less than 1bp variation from allele 39.2 (when comparing their electrophoretic mobility), as reported in the Investigator® Argus X-12 QS Handbook, Qiagen, 2015. Hence, it was not possible to discriminate alleles 39.2 from 39.1 or possible isoalleles, as they need to be sequenced. This represents a limitation of the kit as these observations are prone to introduce typing errors, affecting the reproducibility of the results between laboratories.

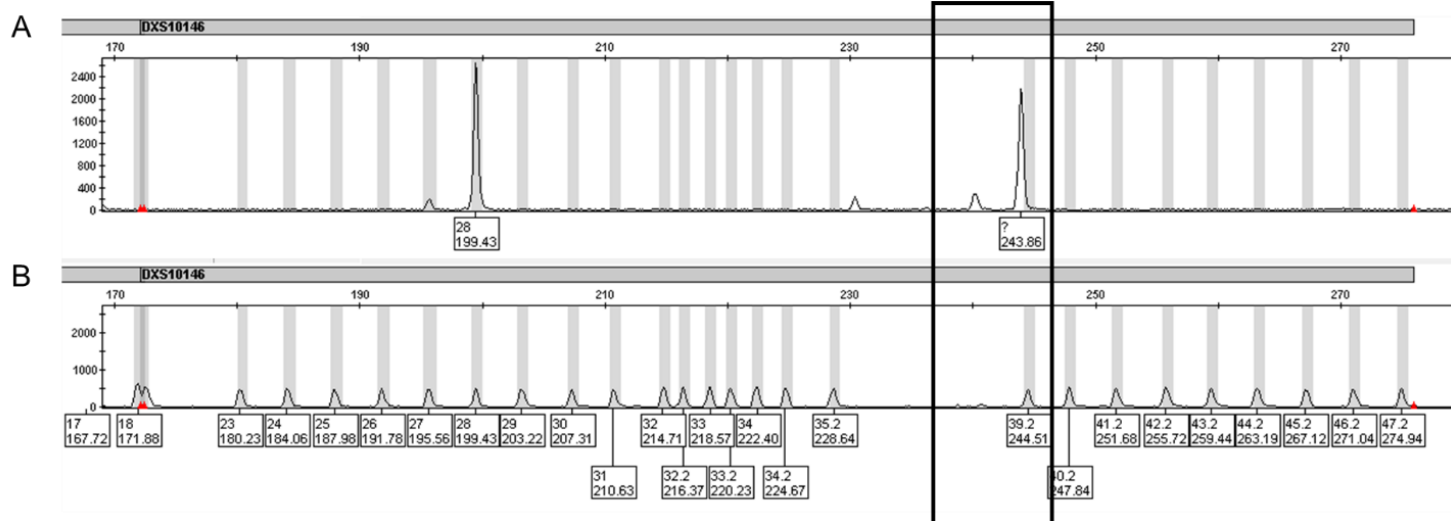


Figure 16- Electropherograms showing the off-ladder allele in DXS10146 marker, highlighted by the black box, from a person's sample (A), and from the Allelic ladder Argus X-12 QS (B).

The absence of amplification of the locus DXS10101 for a father's sample and the analysis of the complete trio illustrated in Figure 17, indicated the presence of a silent (or "null") allele. In a first analysis, the electropherograms indicated that both mother and daughter are homozygous for allele 30.2 at DXS10101. The daughter is homozygous if she inherited from both parents the allele 30.2, however, the evidence does not support the transmission from father to daughter of this allele, which may be explained by the presence of a paternal silent allele, that failed to amplify in both father and daughter.

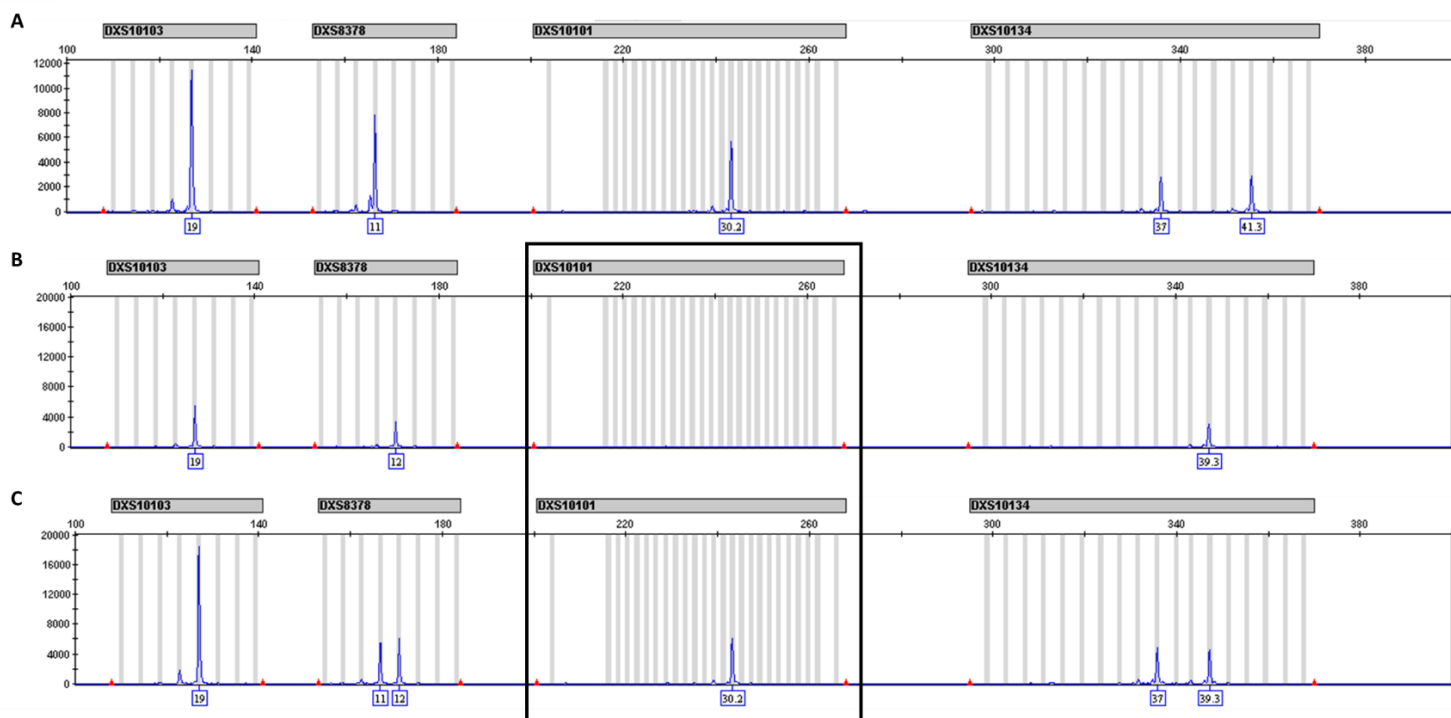


Figure 17- Electropherograms showing the markers coloured in blue for the samples of the mother (A), father(B), and daughter (C), showing the absence of amplification for the marker DXS10101 on the father's sample, and of one allele in the daughter's sample, highlighted by the black box.

However, this should be confirmed by sequencing. Silent alleles at this locus were already observed in Elakkary *et al.*, (2014).

4.2.2. Population genetic analysis

A total of 129 unrelated daughters were selected for population analysis, for which the gametic phase was inferred by the comparison with the father's haplotype and mother's genotype. It was thus possible to determine 258 unrelated haplotypes for each linkage group, and calculate their frequencies, the results for linkage groups 1, 2, 3 and 4 are presented in Supplementary Tables 4, 5, 6 and 7, respectively. This set of unrelated haplotypes were compared with others collected from the Portuguese population in Pinto *et al.*, (2020) (N=422) and Cainé, Costa and Pinheiro, (2013) (N=150), giving a total of 830 haplotypes. Pairwise F_{st} comparisons for 3 X-STR markers inside each linkage group (LG1 to LG4) were performed for the 3 studies referred, the results are listed in Table 10.

Table 10- Pairwise genetic distances (F_{st}) (lower diagonal) between the 3 groups (This work, Pinto *et al.*, (2020), and Cainé, Costa and Pinheiro, (2013)) and corresponding non-differentiation P-values (upper diagonal), for the 3 X-STRs included in linkage group 1 (A), linkage group 2 (B), linkage group 3 (C) and linkage group 4 (D) obtained after 1023 permutations.

A	LG1			
		Cainé. Costa and Pinheiro (2013)	This study	Pinto <i>et al.</i> . (2020)
	Cainé. Costa and Pinheiro (2013)	*	0.4063	0.9648
	This study	0.0000	*	0.3906
	Pinto <i>et al.</i> . (2020)	-0.0022	0.0001	*
B	LG2			
		Cainé. Costa and Pinheiro (2013)	This study	Pinto <i>et al.</i> . (2020)
	Cainé. Costa and Pinheiro (2013)	*	0.2441	0.3467
	This study	0.0011	*	0.8154
	Pinto <i>et al.</i> . (2020)	0.0005	-0.0011	*
C	LG3			
		Cainé. Costa and Pinheiro (2013)	This study	Pinto <i>et al.</i> . (2020)
	Cainé. Costa and Pinheiro (2013)	*	0.1953	0.7979
	This study	0.0014	*	0.4209
	Pinto <i>et al.</i> . (2020)	-0.0017	0.0000	*
D	LG4			
		Cainé. Costa and Pinheiro (2013)	This study	Pinto <i>et al.</i> . (2020)
	Cainé. Costa and Pinheiro (2013)	*	0.6543	0.9639
	This study	-0.0008	*	0.5518
	Pinto <i>et al.</i> . (2020)	-0.0025	-0.0003	*

No significant differences were found between the haplotype samples from the different studies for each linkage group ($F_{st} \leq 0.0014$; $p\text{-value} \geq 0.1953$), for a significance level of 0.0166 ($\alpha=0.05/3$) after applying the Bonferroni correction. Hence, the samples were grouped to test Linkage disequilibrium and Hardy-Weinberg equilibrium.

4.2.3. Linkage disequilibrium

Pairwise linkage disequilibrium (LD) analyses were performed between markers within linkage groups 1, 2, 3, and 4 for a total of 830 haplotypes from the 3 groups. The results are summarised in Table 11, and indicate the presence of associations between DXS8378 and DXS10135 within LG1, between DXS10074 and DXS10079 within LG2, and among the markers within LG 3 and 4, based on a significance level of 0.00417 ($\alpha=0.05/12$) after applying the Bonferroni correction. These associations highlight that these markers cannot be considered as independently transmitted.

Table 11- Pairwise linkage disequilibrium test performed in a sample containing 830 haplotypes from the Portuguese population (from This work, Pinto et al., (2020), and Cainé, Costa and Pinheiro, (2013)) between 3 X-STR inside each linkage groups from 1 to 4. In red are the significant p-values for a significance level of 0.00417 ($\alpha=0.05/12$) after applying Bonferroni correction, for 100172 permutations.

		LG1			LG2			LG3			LG4		
		DXS10148	DXS10135	DXS8378	DXS7132	DXS10079	DXS10074	DXS10103	HPRTB	DXS10101	DXS10146	DXS10134	DXS7423
LG1	DXS10148	*											
	DXS10135	0.3961	*										
	DXS8378	0.1144	0.0000	*									
LG2	DXS7132				*								
	DXS10079				0.0558	*							
	DXS10074				0.1984	0.0000	*						
LG3	DXS10103							*					
	HPRTB							0.0000	*				
	DXS10101							0.0000	0.0000	*			
LG4	DXS10146										*		
	DXS10134										0.0000	*	
	DXS7423										0.0000	0.0000	*

4.2.4. Hardy-Weinberg equilibrium

The Hardy-Weinberg equilibrium was tested for a sample containing only the genotypes from this study ($N=129$), and Pinto *et al.*, (2020) ($N=211$), since the genotypes from Cainé, Costa and Pinheiro, (2013) were not available. The results are shown in Table 12. No statistically significant deviations were found between the observed and expected heterozygotes ($p\text{-value} \geq 0.0209$), considering a significance level of 0.00417 ($\alpha=0.05/12$) after applying the Bonferroni correction. This indicates that there is no evidence of disequilibrium in the 12 X-STRs tested in the samples from the Portuguese population.

Table 12-Results on the observed heterozygosity (H_{obs}), and expected heterozygosity (H_{exp}), and p-value of the exact test of Hardy-Weinberg equilibrium for 12 X-STRs, for 1001000 permutations.

Marker	H_{obs}	H_{exp}	P-value
DXS10148	0.8794	0.8991	0.1262
DXS10135	0.9324	0.9414	0.2629
DXS8378	0.6677	0.6998	0.2006
DXS7132	0.7559	0.7474	0.4730
DXS10079	0.8000	0.8260	0.0209
DXS10074	0.8559	0.8438	0.8061
DXS10103	0.7529	0.7413	0.7470
HPRTB	0.7412	0.7485	0.0703
DXS10101	0.9088	0.9070	0.5056
DXS10146	0.8941	0.8963	0.0752
DXS10134	0.8794	0.8767	0.6236
DXS7423	0.7029	0.7073	0.7584

4.2.5. Genetic diversity and forensic parameters

A total of 830 unrelated haplotypes per linkage group were gathered to form a Portuguese database for 12 X-STRs since the samples from the 3 groups did not exhibit significant genetic variation between them, while some markers exhibited associations within linkage groups. High haplotype diversities were obtained for each linkage group (≥ 0.9996) (Table 13), since a total of 453 different haplotypes were observed for linkage group 1, 248 for linkage group 2, 228 for linkage group 3, and 358 for linkage group 4. Forensic parameters applied to each linkage group (LG) were determined, the results are listed in Table 13.

Table 13- Forensic parameters calculated for the each of the 4 linkage groups, from a sample containing 830 haplotypes from the Portuguese population (from This work (N=258), Pinto et al., (2020) (N=422), and Cainé, Costa and Pinheiro, (2013) (N=150), namely the power of discrimination (PD) for males and females, the mean exclusion chance (MEC) for father-daughter duos and father-mother-daughter trios and paternal half-sisters (PHS). LG1- linkage group 1, LG2- linkage group 2, LG3 linkage group 3, LG4-linkage group 4.

Parameters	LG1	LG2	LG3	LG4
Haplotype Diversity	0.9999	0.9997	0.9996	0.9999
PD male	0.9967	0.9911	0.9891	0.9941
PD female	1.0000	0.9998	0.9998	0.9999
MEC Duos	0.9935	0.9824	0.9783	0.9882
MEC Trios	0.9967	0.9911	0.9889	0.9940
MEC PHS	0.9871	0.9652	0.9573	0.9767

The results show that linkage group 1 is more informative for forensic purposes relative to the others, since it exhibited the highest values for PD for males and females, as well as for the MEC for father-daughter duos, father-mother-daughter trios and paternal half-sisters.

Further allele frequencies, genetic diversity and forensic parameters were calculated for each X-STR using more data. It was used 976 gene copies (for each locus): 258 gene copies from this study (129 females); 422 gene copies from Pinto *et al.*, (2020) (211 females), and 296 gene copies from Cainé, Costa and Pinheiro, (2013) (150 males and 73 females). The allele frequencies are listed in Table 14 and Table 15. A total of 225 different alleles were observed, with a median of 14 alleles per locus and a range of 7 alleles (DSX8378 and DXS10103) to 37 alleles (DXS10146). Out of the 12 loci, only 6 loci exhibit intermediate alleles. The loci DXS10079 has a duplicated allele (10,12).

Table 14- Allele frequencies of the 3 X-STR markers within the linkage groups 1 and 2, from a sample containing 976 gene copies from the Portuguese population (from this work (N=258), Pinto et al., (2020) (N=422), and Cainé, Costa and Pinheiro, (2013)(N=296)).

Linkage Group 1			Linkage Group 2		
DSX10148			DSX8378		
Allele	N	Freq. rel.	Allele	N	Freq. rel.
13.3	2	0.0020	8	1	0.0010
14	1	0.0010	9	22	0.0225
17	5	0.0051	10	309	0.3166
18	137	0.1404	11	348	0.3566
19	20	0.0205	12	267	0.2736
20	10	0.0102	13	27	0.0277
21	4	0.0041	14	2	0.0020
21.1	2	0.0020			
22	6	0.0061			
22.1	24	0.0246			
23	44	0.0451			
23.1	65	0.0666			
24	19	0.0195			
24.1	136	0.1393			
24.2	1	0.0010			
25	2	0.0020			
25.1	150	0.1537			
26.1	132	0.1352			
27	1	0.0010			
27.1	111	0.1137			
28.1	49	0.0502			
29.1	21	0.0215			
29.2	1	0.0010			
30	2	0.0020			
30.1	21	0.0215			
31	1	0.0010			
31.1	1	0.0010			
36.1	2	0.0020			
38.1	2	0.0020			
39.1	2	0.0020			
40.1	1	0.0010			
43.1	1	0.0010			
DSX10135			DSX7132		
Allele	N	Freq. rel.	Allele	N	Freq. rel.
15	3	0.0031	11	16	0.0164
16	6	0.0061	12	118	0.1209
17	28	0.0287	12	1	0.0010
17.1	5	0.0051	13	264	0.2705
18	51	0.0523	14	366	0.3750
18.1	2	0.0020	15	171	0.1752
19	63	0.0645	16	37	0.0379
19.1	15	0.0154	17	2	0.0020
20	58	0.0594	18	1	0.0010
20.1	12	0.0123			
21	78	0.0799			
21.1	12	0.0123			
22	62	0.0635			
22.1	12	0.0123			
23	89	0.0912			
23.1	5	0.0051			
24	79	0.0809			
24.1	3	0.0031			
25	88	0.0902			
25.1	1	0.0010			
26	66	0.0676			
27	62	0.0635			
28	52	0.0533			
28.1	1	0.0010			
29	41	0.0420			
30	34	0.0348			
31	23	0.0236			
32	10	0.0102			
33	6	0.0061			
34	4	0.0041			
35	3	0.0031			
36	1	0.0010			
36.1	1	0.0010			
DSX10079			DSX10074		
Allele	N	Freq. rel.	Allele	N	Freq. rel.
14	10	0.0102	5	1	0.0010
15	32	0.0328	7	36	0.0369
16	25	0.0256	8	142	0.1455
17	60	0.0615	9	11	0.0113
18	138	0.1414	10	2	0.0020
19	264	0.2705	11	2	0.0020
20	235	0.2408	12	1	0.0010
21	135	0.1383	13	14	0.0143
22	66	0.0676	14	27	0.0277
23	8	0.0082	15	90	0.0922
24	2	0.0020	16	181	0.1855
19,20	1	0.0010	17	237	0.2428
			18	167	0.1711
			19	57	0.0584
			20	7	0.0072
			21	1	0.0010

Table 15- Allele frequencies for the 3 X-STR markers within the linkage groups 3 and 4, from a sample containing 976 gene copies from the Portuguese population (from this work (N=258), Pinto et al., (2020) (N=422), and Cainé, Costa and Pinheiro, (2013) (N=296)). *During genotyping some called alleles 39.2 from the DXS10146 showed differences in their mobility (<1 bp) comparatively to the allele 39.2 from the allelic ladder.

Linkage Group 3			Linkage Group 4		
DXS10103			HPRTB		
Allele	N	Freq. rel.	Allele	N	Freq. rel.
15	7	0.0072	8	1	0.0010
16	115	0.1178	9	5	0.0051
17	114	0.1168	10	9	0.0092
18	162	0.1660	11	150	0.1537
19	437	0.4477	12	338	0.3463
20	123	0.1260	13	283	0.2900
21	18	0.0184	14	148	0.1516
			15	32	0.0328
			16	10	0.0102
DXS10101			DXS10146		
Allele	N	Freq. rel.	Allele	N	Freq. rel.
24.2	3	0.0031	23	2	0.0020
25.2	2	0.0020	24	10	0.0102
26.2	13	0.0133	25	55	0.0564
27	4	0.0041	26	71	0.0727
27.2	55	0.0564	27	114	0.1168
28	37	0.0379	27.2	1	0.0010
28.1	13	0.0133	28	168	0.1721
28.2	86	0.0881	29	172	0.1762
29	50	0.0512	29.1	1	0.0010
29.2	137	0.1404	29.2	1	0.0010
30	53	0.0543	30	99	0.1014
30.2	135	0.1383	30.1	1	0.0010
31	103	0.1055	31	29	0.0297
31.2	90	0.0922	32	1	0.0010
32	81	0.0830	32.2	2	0.0020
32.2	30	0.0307	33	3	0.0031
33	47	0.0482	33.2	2	0.0020
33.2	12	0.0123	34	1	0.0010
34	18	0.0184	34.2	3	0.0031
35	4	0.0041	36.2	1	0.0010
36	2	0.0020	37.2	1	0.0010
null	1	0.0010	38.2	6	0.0061
			39.2*	41	0.0420
			40	9	0.0092
			40.2	33	0.0338
			41.2	33	0.0338
			42	3	0.0031
			42.1	1	0.0010
			42.2	18	0.0184
			43.2	36	0.0369
			44	8	0.0082
			44.2	25	0.0256
			45.2	15	0.0154
			46.2	6	0.0061
			47.2	2	0.0020
			48.2	1	0.0010
			52.2	1	0.0010
DXS10134			DXS7423		
Allele	N	Freq. rel.	Allele	N	Freq. rel.
26	1	0.0010	11	1	0.0010
28	1	0.0010	12	2	0.0020
29.1	1	0.0010	13	64	0.0656
30	9	0.0092	14	316	0.3238
31	3	0.0031	15	362	0.3709
32	12	0.0123	16	198	0.2029
33	46	0.0471	17	28	0.0287
33.2	1	0.0010	18	1	0.0010
34	107	0.1096	25	1	0.0010
35	181	0.1855	26	2	0.0020
35.2	3	0.0031	27	1	0.0010
35.3	2	0.0020			
36	190	0.1947			
36.2	1	0.0010			
37	158	0.1619			
37.2	5	0.0051			
37.3	6	0.0061			
38	65	0.0666			
38.3	26	0.0266			
39	23	0.0236			
39.2	4	0.0041			
39.3	32	0.0328			
40	2	0.0020			
40.3	26	0.0266			
41.3	34	0.0348			
42.1	1	0.0010			
42.2	3	0.0031			
42.3	19	0.0195			
43.3	11	0.0113			
44.3	3	0.0031			

The results from the analysis of genetic diversity and forensic parameters for each X-STRs are summarized in Table 16. Genetic diversities or expected heterozygosity (H_{exp}) ranged from 0.6972 to 0.9403, since those are above 0.5 it suggests a high polymorphism in the Portuguese population for the 12 X-STRs. The power of discrimination (PD) for males ranged from 0.6965 to 0.9394, while for females from 0.8476 to 0.9930. The mean exclusion chance (MEC) for duos father/daughter ranged from 0.4906 and 0.8830, while for trios father/mother/daughter from 0.6362 and 0.9360. The probability of exclusion when paternal half-sisters are available for testing (MEC PHS) was also calculated, exhibiting values between 0.2651 and 0.7810. The intervals described for the lowest and highest values belonged to the markers DXS8378 and DXS10135, which correlates with their low and high gene diversities, respectively.

Table 16- Expected heterozygosity and forensic parameters calculated for the 12 X-STR markers, namely the power of discrimination (PD) for males and females, the mean exclusion chance (MEC) for duos, trios and paternal-half-sisters (PHS).

Forensic Parameters	Linkage Group 1			Linkage Group 2			Linkage Group 3			Linkage Group 4		
	DSX10148	DSX10135	DSX8378	DSX7132	DSX10079	DSX10074	DSX10103	HPRTB	DSX10101	DSX10146	DSX10134	DSX7423
H_{exp}	0.8954	0.9403	0.6972	0.7399	0.8203	0.8426	0.7289	0.7489	0.9135	0.8999	0.8786	0.7120
PD male	0.8945	0.9394	0.6965	0.7392	0.8195	0.8418	0.7282	0.7481	0.9125	0.8990	0.8777	0.7113
PD female	0.9796	0.9930	0.8476	0.8902	0.9443	0.9560	0.8938	0.8956	0.9858	0.9818	0.9736	0.8649
MEC Duos	0.8026	0.8830	0.4906	0.5581	0.6788	0.7134	0.5559	0.5692	0.8344	0.8120	0.7752	0.5162
MEC Trios	0.8852	0.9360	0.6362	0.6974	0.7964	0.8228	0.6959	0.7072	0.9060	0.8910	0.8662	0.6596
MEC PHS	0.6491	0.7810	0.2651	0.3327	0.4738	0.5184	0.3340	0.3439	0.6994	0.6643	0.6097	0.2915

4.2.6. Mutations analysis

In this work, 270 meioses from 135 trios were analysed to improve the estimation of the mutation rate of 12 X-STRs. Out of the 3,240 allele transfers, 15 mutations were observed in 7 loci (DXS10103, DXS10074, DXS7132, DXS10135, DXS7423, DXS10146, DXS10079). No trio exhibited more than one mutation. The observed incompatibilities were all compatible with single-step mutations, 13 exhibited a paternal most parsimonious origin; while for 2 both paternal and maternal single-step mutations could explain the genotypic configuration, as in the example: mother (13,13), father (13) and daughter (12,13), the allele 12 mutated from allele 13 however either mother or father could have transmitted the mutation. The genotypic configurations for the trios where an incompatibility was found are represented in Table 17.

Table 17- Genotypic configurations of each trio that showed an incompatibility, and the most parsimonious parental origin.

Sample	X-STR markers	Mother genotype		Father genotype	Daughter genotype		Most parsimonious parental origin
N277/276/279	DXS10103	16	19	18	16	19	Paternal
N349/348/350	DXS10074	15	17	16	17	17	Paternal
N299/300/298	DXS7132	13	14	15	14	14	Paternal
Z8286/8285/8287		13	14	16	14	15	Paternal
Z7902/7964/7903		13	13	13	12	13	Indeterminate
N182/181/441	DXS10135	20	27	22	21	27	Paternal
Z7865/7863/7866		20	31	34	31	35	Paternal
Z7051/7077/7052	DXS7423	15	15	15	15	16	Indeterminate
N161/160/162	DXS10146	29	43.2	31	29	30	Paternal
N191/192/190		28	28	26	27	28	Paternal
N207/208/206		23	29	33.2	23	32.2	Paternal
Z7316/7320/7317		28	30	32	28	31	Paternal
N390/389/391	DXS10079	19	20	18	17	20	Paternal
Z9044/9043/9045		19	19	20	19	21	Paternal
Z7749/7748/7750		16	19	20	16	19	Paternal

4.2.6.1. Mutations and parents' age

The distribution of the parental ages at the time of the daughter's birth, as well as for those that transmitted mutations are illustrated in Figure 18.

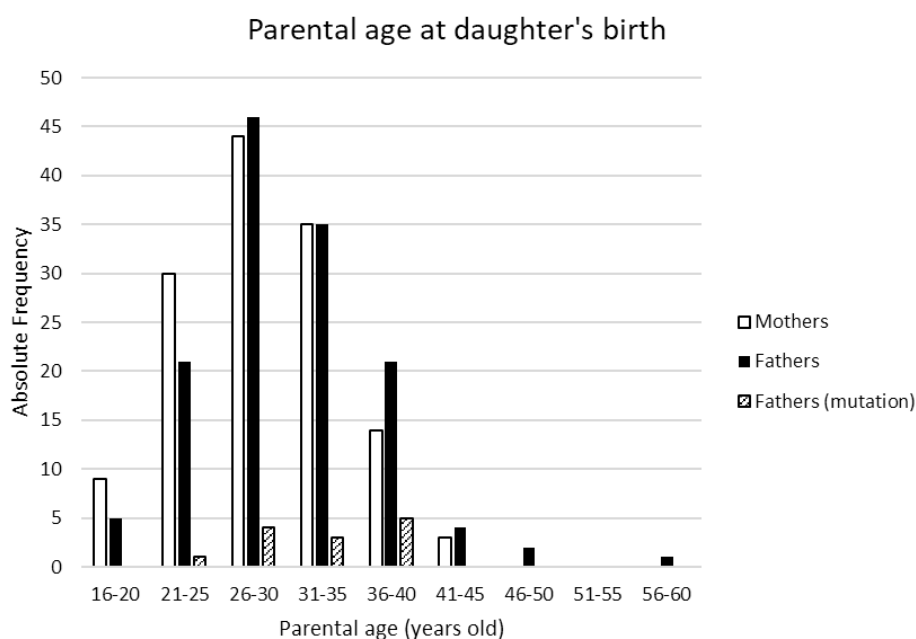


Figure 18- Distribution of the parental age at daughter's birth, for mothers, fathers, and father's that transmitted a mutation. Only mutations for which a most parsimonious parental origin was possible to determine were considered.

From the graph above we can see that most of the parents have between 21 and 35 years at the time of their daughter's birth. Fathers tend to be older than mothers with an average of 30.9 (± 6.5), and 29.0 (± 5.7) years, respectively. The fathers that transmitted mutations were older than those that did not transmit it, with an average of 32.9 (± 5.4) and 30.7 (± 6.6) years, respectively. Mutations with a maternal origin were not observed in this study.

4.2.6.2. Locus-specific mutation rates

Mutation data from this work on 135 mother-father-daughter trios was analysed together with data from other studies where trios were analysed for this set of markers (Athanasiadou *et al.*, 2003; Turrina and De Leo, 2004; Tabbada *et al.*, 2005; Turrina, Atzei and De Leo, 2007; Becker *et al.*, 2008; Tariq *et al.*, 2008; Liu *et al.*, 2008, 2012; Pico *et al.*, 2008; Tetzlaff, Wegener and Lindner, 2012; Nishi *et al.*, 2013; Diegoli *et al.*, 2014; Chen *et al.*, 2020; Pinto *et al.*, 2020). The gathered data was used to estimate specific mutation rates for each X-STR, for paternal and maternal meiosis, and their 95% confidence intervals, the overall results obtained are listed in Supplementary Table 8 (overall mutation rates) and in Supplementary Table 9 (mutation rates for maternal and paternal meiosis); a summary of those results is listed in Table 18. The mutation rate per locus ranged from 7.9×10^{-4} (at DXS7423) to 9.9×10^{-3} (at DXS10135), with an average of 3.6×10^{-3} (95% CI $3.1\text{--}4.2 \times 10^{-3}$) across the 12 X-STR loci. With a mutation rate near 5.4 times higher for fathers than for mothers. The amplitude of the confidence intervals (CI) was reduced by joining more data to the information gathered in this study, representing a higher sample size.

Table 18- Average specific mutation rates obtained for each 12 X-STR from the Investigator Argus X-12 QS kit, after pooling the data from this work (N=135 trios) and the literature (Athanasiadou *et al.*, 2003; Turrina and De Leo, 2004; Tabbada *et al.*, 2005; Turrina, Atzei and De Leo, 2007; Becker *et al.*, 2008; Tariq *et al.*, 2008; Liu *et al.*, 2008, 2012; Pico *et al.*, 2008; Tetzlaff, Wegener and Lindner, 2012; Nishi *et al.*, 2013; Diegoli *et al.*, 2014; Chen *et al.*, 2020; Pinto *et al.*, 2020).

Marker	Total				Most Parsimonious Parental Origin								
					Paternal				Maternal				Indeterminate
	No. Allele transfers	No. Mutations	Mutation Rate x 10 ⁻³	95 % CI x 10 ⁻³	No. Allele transfers	No. Mutations	Mutation Rate x 10 ⁻³	95 % CI x 10 ⁻³	No. Allele transfers	No. Mutations	Mutation Rate x 10 ⁻³	95 % CI x 10 ⁻³	No. mutations
HPRTB	5392	7	1.3	0.5-2.7	2696	3	1.1	0.2-3.2	2696	2	0.7	0.1-2.7	2
DXS7132	6512	32	4.9	3.4-6.9	3256	24	7.4	4.7-10.9	3256	3	0.9	0.2-2.7	5
DXS8378	5122	5	1.0	0.3-2.3	2561	2	0.8	0.1-2.8	2561	0	0.0	0-1.4	3
DXS7423	5006	4	0.8	0.2-2	2503	2	0.8	0.1-2.9	2503	1	0.4	0-2.2	1
DXS10135	3952	39	9.9	6.8-13.2	1976	31	15.7	10.7-22.2	1976	7	3.5	1.4-7.3	1
DXS10074	4612	20	4.3	2.7-6.7	2306	18	7.8	4.6-12.3	2306	0	0.0	0-1.6	2
DXS10101	3952	11	2.8	1.4-5	1976	8	4.0	1.7-8	1976	1	0.5	0-2.8	2
DXS10134	4108	17	4.1	2.4-6.6	2054	11	5.4	2.7-9.6	2054	5	2.4	0.8-5.7	1
DXS10148	3494	14	4.0	2.2-6.7	1747	11	6.3	3.1-11.2	1747	3	1.7	0.4-5	0
DXS10079	4114	25	6.1	3.9-9	2057	19	9.2	5.6-14.4	2057	2	1.0	0.1-3.5	4
DXS10103	3494	7	2.0	0.8-4.1	1747	6	3.4	1.3-7.5	1747	0	0.0	0-2.1	1
DXS10146	3494	13	3.7	2-6.4	1747	10	5.7	2.7-10.5	1747	2	1.1	0.1-4.1	1
Total	53252	194	3.6	3.1-4.2	26626	145	5.4	4.6-6.4	26626	26	1.0	0.6-1.4	23

The type of mutation observed from the data above, for each most parsimonious origin, either maternal, paternal or indeterminate, is listed in Supplementary Tables 10, 11 and 12. A summary of those results can be consulted in Table 19.

Table 19- Summary of the number of mutations counted for each parsimonious parental origin and the type of mutations observed in this study (N=135 trios) and the literature (Athanasiadou *et al.*, 2003; Turrina and De Leo, 2004; Tabbada *et al.*, 2005; Turrina, Atzei and De Leo, 2007; Becker *et al.*, 2008; Tariq *et al.*, 2008; Liu *et al.*, 2008, 2012; Pico *et al.*, 2008; Tetzlaff, Wegener and Lindner, 2012; Nishi *et al.*, 2013; Diegoli *et al.*, 2014; Chen *et al.*, 2020; Pinto *et al.*, 2020). Type of mutation: {+1} represents the gain of one repeat unit, >{+1}- gain of more than one repeat, {-1}- loss of a repeat, <{-1}- the loss of more than one repeat, {+1}V{-1} mutation explained either by the gain or loss of one repeat, NI- non integer repeat unit variation.

Most parsimonious parental origin	Type of mutation						
	<{-1}	{-1}	{+1}	>{+1}	{-1}V{+1}	NI	Total
Paternal	2	82	58	2	0	1	145
Maternal	0	8	15	1	1	1	26
Indeterminate	0	5	12	0	6	0	23
Total	2	95	85	3	7	2	194

Out of the 53252 allele transfers, 194 mutations were observed, 187 of which are compatible with one-step mutations, either gain or lost one repeat unit, 5 are only explainable if considering multi-step mutations, while 2 correspond to a non-integer repeat unit variation. In 7 one-step mutations, it is not possible to determine if they gain or lost a repeat unit, since both directions are compatible with the observations. One-step mutations were considered every time they explained the observed incompatibilities; which may underestimate the number of multi-step mutations (Antão-Sousa, Conde-Sousa, *et al.*, 2019; Pinto *et al.*, 2020).

The most parsimonious parental origin of the mutations was 5.6 times more attributed to the paternal side (N=145), than the maternal one (N=26), probably due to the high number of divisions for the gamete formation in men than in women (Hardy, 2010) or due to the higher probability of maternal mutations that result in hidden mutations when compared with paternal ones as shown by preliminary results of Antão-Sousa *et al.*, (2019). For 23 mutations the parental origin was indeterminate.

5. Conclusions

The present study gathered new haplotypic and segregation data on 23 Y-STRs and 12 X-STRs from the Portuguese population (N=202 duos, and N=135 trios, respectively). These sets of markers are widely used in forensic genetics, therefore the haplotype and genetic diversity, as well as the mutation rates estimated in this work for both types of markers, will increase the statistical weight of the interpretation of the genetic results, specifically for the Portuguese population.

The genotyping analysis of the Y-STRs revealed intermediate alleles, duplicated and triplicated loci. Surprisingly, a *de novo* deletion of 3 adjacent markers was also observed; however, this finding should be confirmed by sequencing. These patterns are observed likely due to the frequent intrachromosomal rearrangements of the Y chromosome during meiosis. The genotype analysis of the X-STR markers revealed a tri-allelic pattern, an apparent silent allele, as well as several microvariants. The presence of several X-STR microvariants that were not included in the allelic ladder and the possible presence of isoalleles represents a major limitation of the 12 X-STR kit used, since errors are more prone to occur, affecting the reproducibility of the results between laboratories.

Intra studies population analysis revealed no significant differences in pairwise F_{ST} genetic distance comparisons between the new data and those we were able to gather from the literature for both Y-STRs and X-STRs. X-STRs linkage disequilibrium analysis revealed associations between markers within the same linkage group (for 1, 2, 3 and 4). While Hardy-Weinberg equilibrium tests did not reveal significant deviations from the observed and expected heterozygosity.

The Portuguese population exhibited an overall high genetic diversity for the sets of 23 Y-STRs and 12 X-STRs, having been the highest values observed for DXS10135 on the X chromosome, DYS385ab (multi-copy) and DYS570 on the Y chromosome. The high genetic diversity contributes to the high haplotype diversity and discriminatory ability observed for the tested markers on the analysed population. This indicates the statistical power of the use of the 23 Y-STR panel for forensic purposes in cases of non-identity, as exclusion is highly likely, however for kinship analysis even close relatives from the same male lineage can exhibit incompatibilities due to recurrent mutations (Purps *et al.*, 2014). The panel of 12 X-STRs also shows a powerful application for forensic casework, such as complex kinship cases, as estimated forensic parameters exhibited a high mean exclusion chance in paternity cases for father-daughter duos, father-mother-daughter trios and paternal half-sisters.

Taken together, the above findings indicate that X-STRs haplotypes for each linkage group from the different studies can be joined to form a Portuguese database, and the 200 haplotypes for the 23 Y-STRs can be deposited in the Portuguese database, at YHRD.

The new results from mutation analysis of this study were gathered with other data previously published to obtain more accurate marker specific mutation rate estimates for the 12 X-STRs, as well as their paternal and maternal mutation rates, and also for the 23 Y-STRs. The first was estimated considering the segregation of X-chromosome in mother-father-daughter trios. Trios were selected instead of duos to prevent a conservative mutation rate analysis since, in the last, hidden mutations (not detected) are more probable to occur (Antão-Sousa, Conde-Sousa, *et al.*, 2019). Fathers showed higher mutation rates than mothers, which can be due to the high number of cell divisions in gametogenesis in men than in women (Hardy, 2010), another explanation may be the higher probability of maternal mutations that result in hidden mutations when compared with paternal ones as shown by preliminary results of Antão-Sousa *et al.*, (2019). For Y-STRs mutation analysis, the haplotypic transmission in father-son duos was observed, where mutation origin (from the father) and mutation result (from the son) are unambiguous, thus ideal to study mutations.

The majority of the mutations were one-step, thus either compatible with the gain or loss of a repeat unit. The highest mutation rate estimates were obtained for DXS10135 for the X transmission analysis, and DYS576 and DYS570 for the Y chromosomal one. The Y-STR markers that belong to the [GAAA] cluster and with longer alleles within their allelic range exhibited higher mutation rates. Both DYS576 and DYS570 are considered 'rapidly mutating' STRs (RM-STRs), providing a higher discriminatory power to the 23 Y-STR Powerplex kit. However, those can be troublesome for kinship analysis as more incompatibilities are prone to be observed, highlighting the importance of obtaining their mutation rates estimates.

A limitation of this study is that when STRs are amplified and analysed by CE a mutation could have occurred in the flanking region instead of the repeats without our knowledge since we are only observing size changes (Dauber *et al.*, 2012). For example, an amplicon with a 4 bp deletion in the flanking region has the same size as one with a loss of a tetrameric repeat unit by CE without sequencing (Dauber *et al.*, 2012). Other mutations without changing the length of the amplicon would also not be detected (Hill 2007), such as variants that may exhibit shifts of less than one bp as for DXS10146.

Mutations are rare, consequently, to obtain better estimates more duos and trios should be analysed for Y-STRs and X-STRs, respectively. Notwithstanding the results obtained in this work contribute to the knowledge on the genetic polymorphisms (allele and haplotype frequencies) and mutation rates estimates observed for both Y-STRs and X-STRs for the Portuguese population. Those will be available to the scientific community, contributing to the advancement of the state of the art not only of the genetic characterization of the Portuguese population, but also of STR mutation modelling, enhancing the interpretation and statistical weight of the genetic proof in kinship analysis and forensic genetics analysis.

6. References

- Alves, C. *et al.* (2007) 'Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFISTR Yfiler) in Portugal', *Forensic Science International*, 171, pp. 250–255. doi: 10.1016/j.forsciint.2006.10.022.
- Ambrosio, I. B. *et al.* (2020) 'Mutational data and population profiling of 23 Y-STRs in three Brazilian populations', *Forensic Science International: Genetics*, 48, p. 102348. doi: 10.1016/j.fsigen.2020.102348.
- Amorim, A. (2009) *A espécie das origens*. 2nd edn. Gradiva.
- Amorim, A., Fernandes, T. and Taveira, N. (2019) 'Mitochondrial DNA in human identification: a review.', *PeerJ*, 7. doi: 10.7717/peerj.7314.
- Amorim, A. and Pinto, N. (2019) 'Estimates of mutation rates from incompatibilities are misleading - guidelines for publication and retrieval of mutation data urgently needed', *Forensic Science International: Genetics Supplement Series*, 7(1), pp. 612–613. doi: 10.1016/j.fsigss.2019.10.110.
- Amster, G. *et al.* (2020) 'Changes in life history and population size can explain the relative neutral diversity levels on X and autosomes in extant human populations', *Proceedings of the National Academy of Sciences*, 117(33), pp. 20063–20069. doi: 10.1073/pnas.1915664117.
- Andersen, M. M., Eriksen, P. S. and Morling, N. (2013) 'The discrete Laplace exponential family and estimation of Y-STR haplotype frequencies', *Journal of Theoretical Biology*, 329, pp. 39–51. doi: 10.1016/J.JTBI.2013.03.009.
- Antão-Sousa, S. (2017) *Estimation of mutation rates at Y-STRs*.
- Antão-Sousa, S. *et al.* (2017) 'Mutation rates and segregation data on 16 Y-STRs: An update to previous GHEP-ISFG studies', *Forensic Science International: Genetics Supplement Series*, 6(October), pp. e601–e602. doi: 10.1016/j.fsigss.2017.10.008.
- Antão-Sousa, S., Amorim, A., *et al.* (2019) 'Mutation in Y STRs: Repeat motif gains vs. losses', *Forensic Science International: Genetics Supplement Series*, 7(1), pp. 240–242. doi: 10.1016/j.fsigss.2019.09.092.
- Antão-Sousa, S., Conde-Sousa, E., *et al.* (2019) 'Underestimation and misclassification of mutations at X chromosome STRs depend on population's allelic profile', *Forensic Science International: Genetics Supplement Series*, 7(1), pp. 718–720. doi: 10.1016/j.fsigss.2019.10.150.
- Athanasiadou, D. *et al.* (2003) 'Development of a quadruplex PCR system for the genetic analysis of X-chromosomal STR loci', *International Congress Series*, 1239(C), pp. 311–314. doi: 10.1016/S0531-5131(02)00479-X.
- Bachtrog, D. and Charlesworth, B. (2001) 'Towards a complete sequence of the human Y chromosome', *Genome biology*, 2(5). doi: 10.1186/gb-2001-2-5-reviews1016.
- Bai, R. *et al.* (2016) 'Developmental Validation of a novel 5 dye Y-STR System comprising the 27 YfilerPlus loci', *Scientific Reports*, 6(1), p. 29557. doi: 10.1038/srep29557.
- Ballantyne, K. N. *et al.* (2010) 'Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications', *The American Journal of Human Genetics*, 87(3), pp. 341–353. doi: 10.1016/j.ajhg.2010.08.006.
- Ballantyne, K. N., Keerl, V., Wollstein, A., Choi, Y., Zuniga, S. B., Ralf, A., Vermeulen, M., De Knijff, P., *et al.* (2012) 'A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal

lineages', *Forensic Science International: Genetics*, 6(2), pp. 208–218. doi: 10.1016/J.FSIGEN.2011.04.017.

Ballantyne, K. N., Keerl, V., Wollstein, A., Choi, Y., Zuniga, S. B., Ralf, A., Vermeulen, M., de Knijff, P., *et al.* (2012) 'A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages', *Forensic Science International: Genetics*, 6(2), pp. 208–218. doi: 10.1016/j.fsigen.2011.04.017.

Barbujani, G. *et al.* (1997) 'An apportionment of human DNA diversity', *Proceedings of the National Academy of Sciences*, 94(9), pp. 4516–4519. doi: 10.1073/pnas.94.9.4516.

Batini, C. and Jobling, M. A. (2017) 'Detecting past male-mediated expansions using the Y chromosome', *Human Genetics*, 136(5), pp. 547–557. doi: 10.1007/s00439-017-1781-z.

Becker, D. *et al.* (2008) 'Population genetic evaluation of eight X-chromosomal short tandem repeat loci using Mentype Argus X-8 PCR amplification kit', *Forensic Science International: Genetics*, 2(1), pp. 69–74. doi: 10.1016/J.FSIGEN.2007.08.013.

Beleza, S. *et al.* (2006) 'Micro-phylogeographic and demographic history of Portuguese male lineages', *Annals of Human Genetics*, 70(2), pp. 181–194. doi: 10.1111/j.1529-8817.2005.00221.x.

Beltramo, J., Pena, M. A. and 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, pp. e141–e143. doi: 10.1016/j.figss.2015.09.057.

Bento, A. M. *et al.* (2009) 'Distribution of Y-chromosomal haplotypes in the Central Portuguese population using 17-STRs', *Forensic Science International: Genetics*, 4(1), pp. 35–36. doi: 10.1016/j.fsigen.2009.02.006.

Bodner, M. *et al.* (2016) 'Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER)', *Forensic Science International: Genetics*, 24, pp. 97–102. doi: 10.1016/j.fsigen.2016.06.008.

Brenner, C. H. (2010) 'Fundamental problem of forensic mathematics—The evidential value of a rare haplotype', *Forensic Science International: Genetics*, 4(5), pp. 281–291. doi: 10.1016/j.fsigen.2009.10.013.

Brinkmann, B. *et al.* (1998) 'Mutation rate in human microsatellites: Influence of the structure and length of the tandem repeat', *American Journal of Human Genetics*, 62(6), pp. 1408–1415. doi: 10.1086/301869.

Butler, J. M. *et al.* (2005) 'Chromosomal Duplications Along the Y-Chromosome and Their Potential Impact on Y-STR Interpretation', *Journal of Forensic Sciences*, 50(4), pp. 1–7. doi: 10.1520/jfs2004481.

Butler, J. M. (2010) *Fundamentals of Forensic DNA Typing*. Elsevier. doi: 10.1016/C2009-0-01945-X.

Butler, J. M. (2012a) *Advanced Topics in Forensic DNA typing: Methodology*. Elsevier. doi: <https://doi.org/10.1016/C2011-0-04189-3>.

Butler, J. M. (2012b) *Advanced Topics in Forensic DNA Typing: Methodology*. Elsevier. doi: 10.1016/C2011-0-04189-3.

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

- Cainé, L., Costa, S. and Pinheiro, M. F. (2013) 'Population data of 12 X-STR loci in a North of Portugal sample', *International Journal of Legal Medicine*, 127(1), pp. 63–64. doi: 10.1007/s00414-012-0672-z.
- Carvalho, M. *et al.* (2003) 'Y-chromosome STR haplotypes in two population samples: Azores Islands and Central Portugal', *Forensic Science International*, 134(1), pp. 29–35. doi: 10.1016/S0379-0738(03)00097-5.
- Cavelier, L., Johannisson, A. and Gyllensten, U. (2000) 'Analysis of mtDNA Copy Number and Composition of Single Mitochondrial Particles Using Flow Cytometry and PCR', *Experimental Cell Research*, 259(1), pp. 79–85. doi: 10.1006/excr.2000.4949.
- Chakraborty, R. *et al.* (1997) 'Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci', *Proceedings of the National Academy of Sciences of the United States of America*, 94(3), pp. 1041–1046. doi: 10.1073/pnas.94.3.1041.
- Chen, M. *et al.* (2020) 'Genetic polymorphisms and mutation rates of 16 X-STRs in a Han Chinese population of Beijing and application examples in second-degree kinship cases', *International Journal of Legal Medicine*, 134(1), pp. 163–168. doi: 10.1007/s00414-019-02047-8.
- Clopper, C. J. and Pearson, E. S. (1934) 'The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial', *Biometrika*, 26(4), p. 404. doi: 10.2307/2331986.
- Dauber, E.-M. *et al.* (2012) 'Germline mutations of STR-alleles include multi-step mutations as defined by sequencing of repeat and flanking regions', *Forensic Science International: Genetics*, 6(3), pp. 381–386. doi: 10.1016/j.fsigen.2011.07.015.
- Desmarais, D. *et al.* (1998) 'Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA).', *Journal of forensic sciences*, 43(5), pp. 1046–1049.
- Diederiche, M. *et al.* (2005) 'A case of double alleles at three Y-STR loci: forensic implications', *International Journal of Legal Medicine*, 119(4), pp. 223–225. doi: 10.1007/s00414-005-0526-z.
- Diegoli, T. M. *et al.* (2014) 'Mutation rates of 15 X chromosomal short tandem repeat markers', *International Journal of Legal Medicine*, 128(4), pp. 579–587. doi: 10.1007/s00414-014-1016-y.
- Diegoli, T. M. (2015) 'Forensic typing of short tandem repeat markers on the X and Y chromosomes', *Forensic Science International: Genetics*, 18, pp. 140–151. doi: 10.1016/j.fsigen.2015.03.013.
- Diegoli, T. M. and Coble, M. D. (2015) 'Characterization of X Chromosomal Short Tandem Repeat Markers for Forensic Use', p. 136. Available at: <https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=271654>.
- Elakkary, S. *et al.* (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), pp. 26–30. doi: 10.1016/j.fsigen.2014.02.007.
- Ellegren, H. (2004) 'Microsatellites: Simple sequences with complex evolution', *Nature Reviews Genetics*, 5(6), pp. 435–445. doi: 10.1038/nrg1348.
- Excoffier, L. and Lischer, H. E. L. (2010) 'Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows', *Molecular Ecology Resources*, 10(3), pp. 564–567. doi: 10.1111/j.1755-0998.2010.02847.x.
- Fan, H. and Chu, J. Y. (2007) 'A Brief Review of Short Tandem Repeat Mutation', *Genomics, Proteomics and Bioinformatics*, 5(1), pp. 7–14. doi: 10.1016/S1672-0229(07)60009-6.

- Fernandes, A. T. *et al.* (2001) 'Y-chromosome STR haplotypes in the Madeira archipelago population', *Forensic Science International*, 122(2–3), pp. 178–180. doi: 10.1016/S0379-0738(01)00495-9.
- Fisher, R. A. (1951) 'Standard calculations for evaluating a blood-group system', *Heredity*, 5(1), pp. 95–102. doi: 10.1038/hdy.1951.5.
- Foster, E. A. *et al.* (1999) 'Reply: The Thomas Jefferson paternity case', *Nature*, 397(6714), pp. 32–32. doi: 10.1038/16181.
- Gao, H. *et al.* (2019) 'Application of CPI cutoff value based on parentage testing of duos and trios typed by four autosomal kits', *PLOS ONE*. Edited by A. Rogobete, 14(11), p. e0225174. doi: 10.1371/journal.pone.0225174.
- Gomes, C. *et al.* (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), pp. 917–921. doi: 10.1007/s00414-012-0768-5.
- Gomes, I. *et al.* (2016) 'New sequence variants detected at DXS10148, DXS10074 and DXS10134 loci', *Forensic Science International: Genetics*, 20, pp. 112–116. doi: 10.1016/j.fsigen.2015.10.005.
- Gomes, I. *et al.* (2020) 'Twenty Years Later: A Comprehensive Review of the X Chromosome Use in Forensic Genetics', *Frontiers in Genetics*, 11(September), pp. 1–17. doi: 10.3389/fgene.2020.00926.
- Goodwin, W., Linacre, A. and Hadi, S. (2010) *An introduction to forensic genetics*. 2nd edn. Wiley-Blackwell.
- Gusmão, L. *et al.* (2005) 'Mutation rates at Y chromosome specific microsatellites', *Human Mutation*, 26(6), pp. 520–528. doi: 10.1002/humu.20254.
- Gusmão, L. *et al.* (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), pp. 227–234. doi: 10.1007/s00414-008-0309-4.
- Hallast, P. *et al.* (2021) 'A Southeast Asian origin for present-day non-African human Y chromosomes', *Human Genetics*, 140(2), pp. 299–307. doi: 10.1007/s00439-020-02204-9.
- Hardy, K. (2010) 'Embryology', in *Basic Science in Obstetrics and Gynaecology*. Elsevier, pp. 25–47. doi: 10.1016/B978-0-443-10281-3.00007-5.
- Hares, D. R. (2015) 'Selection and implementation of expanded CODIS core loci in the United States', *Forensic Science International: Genetics*, 17, pp. 33–34. doi: 10.1016/j.fsigen.2015.03.006.
- Huang, Q.-Y. *et al.* (2002) 'Mutation Patterns at Dinucleotide Microsatellite Loci in Humans', *American Journal of Human Genetics*, 70(3), p. 625. doi: 10.1086/338997.
- Ivanov, P. L. *et al.* (1996) 'Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II', *Nature Genetics* 1996 12:4, 12(4), pp. 417–420. doi: 10.1038/ng0496-417.
- Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985) 'Individual-specific "fingerprints" of human DNA', *Nature*, 316(6023), pp. 76–79. doi: 10.1038/316076a0.
- Jobling, M. *et al.* (2013) *Human Evolutionary Genetics*.
- Jobling, M. A. (2008) 'Copy number variation on the human Y chromosome', *Cytogenetic and Genome Research*, 123(1–4), pp. 253–262. doi: 10.1159/000184715.
- Jobling, M. A. and Tyler-Smith, C. (2017) 'Human Y-chromosome variation in the

genome-sequencing era', *Nature Reviews Genetics*, 18(8), pp. 485–497. doi: 10.1038/nrg.2017.36.

Kayser, M. and Sajantila, A. (2001) 'Mutations at Y-STR loci: Implications for paternity testing and forensic analysis', *Forensic Science International*, 118(2–3), pp. 116–121. doi: 10.1016/S0379-0738(00)00480-1.

Kimura, M. (1968) 'Evolutionary Rate at the Molecular Level', *Nature*, 217(5129), pp. 624–626. doi: 10.1038/217624a0.

Kishida, T. et al. (1997) 'Duplex PCR of the Y-27H39 and HPRT loci with reference to Japanese population data on the HPRT locus.', *Nihon hoigaku zasshi = The Japanese journal of legal medicine*, 51(2), pp. 67–69.

Kling, D. et al. (2014) 'A general model for likelihood computations of genetic marker data accounting for linkage, linkage disequilibrium, and mutations', *International Journal of Legal Medicine* 2014 129:5, 129(5), pp. 943–954. doi: 10.1007/S00414-014-1117-7.

Kling, D., Dell'Amico, B. and Tillmar, A. O. (2015) 'FamLinkX – implementation of a general model for likelihood computations for X-chromosomal marker data', *Forensic Science International: Genetics*, 17, pp. 1–7. doi: 10.1016/j.fsigen.2015.02.007.

Kornberg, A. et al. (1964) 'ENZYMATIC SYNTHESIS OF DEOXYRIBONUCLEIC ACID, XVI. OLIGONUCLEOTIDES AS TEMPLATES AND THE MECHANISM OF THEIR REPLICATION', *Proceedings of the National Academy of Sciences*, 51(2), pp. 315–323. doi: 10.1073/pnas.51.2.315.

Kowalczyk, M. et al. (2018) 'Molecular markers used in forensic genetics', *Medicine, Science and the Law*, 58(4), pp. 201–209. doi: 10.1177/0025802418803852.

Lahn, B. T. and Page, D. C. (1999) 'Four Evolutionary Strata on the Human X Chromosome', *Science*, 286(5441), pp. 964–967. doi: 10.1126/science.286.5441.964.

Lee, J. S. et al. (1999) 'Relative stabilities of dinucleotide and tetranucleotide repeats in cultured mammalian cells', *Human Molecular Genetics*, 8(13), pp. 2567–2572. doi: 10.1093/hmg/8.13.2567.

Levinson, G. and Gutman, G. A. (1987) 'Slipped-strand mispairing: a major mechanism for DNA sequence evolution.', *Molecular Biology and Evolution*, 4(3), pp. 203–221. doi: 10.1093/oxfordjournals.molbev.a040442.

Lewontin, R. C. (1972) 'The Apportionment of Human Diversity', in *Evolutionary Biology*. New York, NY: Springer US, pp. 381–398. doi: 10.1007/978-1-4684-9063-3_14.

Lewontin, R. and Hartl, D. (1991) 'Population genetics in forensic DNA typing', *Science*, 254(5039), pp. 1745–1750. doi: 10.1126/science.1845040.

Lin, H. et al. (2020) 'Analyzing genetic polymorphism and mutation of 44 Y-STRs in a Chinese Han population of Southern China', *Legal Medicine*, 42(November 2019), p. 101643. doi: 10.1016/j.legalmed.2019.101643.

Liu, Q.-L. et al. (2012) 'Development of multiplex PCR system with 15 X-STR loci and genetic analysis in three nationality populations from China', *ELECTROPHORESIS*, 33(8), pp. 1299–1305. doi: 10.1002/elps.201100558.

Liu, Q. et al. (2008) 'Development of a five ChX STRs loci typing system', *International Journal of Legal Medicine*, 122(3), pp. 261–265. doi: 10.1007/s00414-008-0223-9.

Mršić, G. et al. (2018) 'Expanded Croatian 12 X-STR loci database with an overview of anomalous profiles', *Forensic Science International: Genetics*, 34(September 2017), pp. 249–256. doi: 10.1016/j.fsigen.2018.03.004.

Mulero, J. J. et al. (2006) 'Development and Validation of the AmpF ℓ STR $\text{\textregistered}$ Yfiler TM

PCR Amplification Kit: A Male Specific, Single Amplification 17 Y-STR Multiplex System*, *Journal of Forensic Sciences*, 51(1), pp. 64–75. doi: <https://doi.org/10.1111/j.1556-4029.2005.00016.x>.

Nachman, M. W. and Crowell, S. L. (2000) 'Estimate of the mutation rate per nucleotide in humans.', *Genetics*, 156(1), pp. 297–304. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10978293>.

Nadir, E. *et al.* (1996) 'Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications.', *Proceedings of the National Academy of Sciences of the United States of America*, 93(13), p. 6470. doi: 10.1073/PNAS.93.13.6470.

Nei, M. (1978) 'Estimation of average heterozygosity and genetic distance from a small number of individuals', *Genetics*, 89(3).

Nishi, T. *et al.* (2013) 'Application of a novel multiplex polymerase chain reaction system for 12 X-chromosomal short tandem repeats to a Japanese population study', *Legal Medicine*, 15(1), pp. 43–46. doi: 10.1016/J.LEGALMED.2012.07.001.

Ohno, S. (1967) *Sex Chromosomes and Sex Linked Genes*. Berlin: Springer Verlag.

Ohno, Y., Sebetan, I. M. and Akaishi, S. (1982) 'A simple method for calculating the probability of excluding paternity with any number of codominant alleles', *Forensic Science International*, 19(1), pp. 93–98. doi: 10.1016/0379-0738(82)90155-4.

Ohta, T. and Kimura, M. (2008) 'A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population', *Genetics Research*, 89(5–6), pp. 367–370. doi: 10.1017/S0016672308009531.

Page, D. C. *et al.* (1984) 'Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution', *Nature*, 311(5982), pp. 119–123. doi: 10.1038/311119a0.

Pakendorf, B. and Stoneking, M. (2005) 'Mitochondrial DNA and human evolution', *Annual Review of Genomics and Human Genetics*, 6, pp. 165–183. doi: 10.1146/annurev.genom.6.080604.162249.

Pereira, L. and Ribeiro, F. M. (2009) *O património genético Português*. 2nd edn. Gradiva.

Pereira, R. *et al.* (2007) 'Genetic diversity of 10 X chromosome STRs in northern Portugal', *International Journal of Legal Medicine*, 121(3), pp. 192–197. doi: 10.1007/s00414-006-0144-4.

Pereira, R. *et al.* (2009) 'A new multiplex for human identification using insertion/deletion polymorphisms', *Electrophoresis*, 30(21), pp. 3682–3690. doi: 10.1002/elps.200900274.

Pereira, R. *et al.* (2012) 'Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing', *PLoS ONE*, 7(1). doi: 10.1371/journal.pone.0029684.

Pereira, V. *et al.* (2017) 'Evaluation of the Precision ID Ancestry Panel for crime case work: A SNP typing assay developed for typing of 165 ancestral informative markers', *Forensic Science International: Genetics*, 28, pp. 138–145. doi: 10.1016/j.fsigen.2017.02.013.

Pereira, V. and Gusmão, L. (2016) 'CHAPTER 9: Types of Genomes, Sequences and Genetic Markers (Repeats, SNPs, Indels, Haplotypes)', in *Handbook of Forensic Genetics- Biodiversity and Heredity in Civil and Criminal Investigation*, pp. 163–191. doi: doi.org/10.1142/9781786340788_0009.

- Pico, A. *et al.* (2008) 'Genetic profile characterization and segregation analysis of 10 X-STRs in a sample from Santander, Colombia', *International Journal of Legal Medicine* 2007 122:4, 122(4), pp. 347–351. doi: 10.1007/S00414-007-0215-1.
- Pinto, N. *et al.* (2013) 'Paternity exclusion power: Comparative behaviour of autosomal and X-chromosomal markers in standard and deficient cases with inbreeding', *Forensic Science International: Genetics*, 7(2), pp. 290–295. doi: 10.1016/j.fsigen.2012.12.002.
- Pinto, N. *et al.* (2020) 'Paternal and maternal mutations in X-STRs: A GHEP-ISFG collaborative study', *Forensic Science International: Genetics*, 46(June 2019), p. 102258. doi: 10.1016/j.fsigen.2020.102258.
- Pinto, N., Gusmão, L. and 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), pp. 27–32. doi: 10.1016/j.fsigen.2010.01.011.
- Pinto, N., Gusmão, L. and Amorim, A. (2014) 'Mutation and mutation rates at Y chromosome specific Short Tandem Repeat Polymorphisms (STRs): A reappraisal', *Forensic Science International: Genetics*, 9(1), pp. 20–24. doi: 10.1016/j.fsigen.2013.10.008.
- Piraino, S. W. and Furney, S. J. (2016) 'Beyond the exome: the role of non-coding somatic mutations in cancer', *Annals of Oncology*, 27(2), pp. 240–248. doi: 10.1093/ANNONC/MDV561.
- Pontes, M. L. *et al.* (2007) 'Allele frequencies and population data for 17 Y-STR loci (AmpFISTR Y-filer) in a Northern Portuguese population sample', *Forensic Science International*, 170, pp. 62–67. doi: 10.1016/j.forsciint.2006.04.008.
- Pontes, M.L. *et al.* (2007) 'Allele frequencies and population data for 17 Y-STR loci (AmpFISTR® Y-filer™) in a Northern Portuguese population sample', *Forensic Science International*, 170(1), pp. 62–67. doi: 10.1016/j.forsciint.2006.04.008.
- Purps, J. *et al.* (2014) 'A global analysis of Y-chromosomal haplotype diversity for 23 STR loci', *Forensic Science International: Genetics*, 12, pp. 12–23. doi: 10.1016/j.fsigen.2014.04.008.
- Ralf, A. *et al.* (2020) 'Identification and characterization of novel rapidly mutating Y-chromosomal short tandem repeat markers', *Human Mutation*, 41(9), pp. 1680–1696. doi: 10.1002/humu.24068.
- Roewer, L. *et al.* (2020) 'DNA commission of the International Society of Forensic Genetics (ISFG): Recommendations on the interpretation of Y-STR results in forensic analysis', *Forensic Science International: Genetics*, 48. doi: 10.1016/j.fsigen.2020.102308.
- Ross, M. T. *et al.* (2005) 'The DNA sequence of the human X chromosome', *Nature*, 434(7031), pp. 325–337. doi: 10.1038/nature03440.
- Salazar, R. A. P. (2013) *Extending Y-STR Loci in Portugal for Forensic and Population Studies: The PowerPlex® Y23 Experience.*
- Sánchez-Diz, P. *et al.* (2008) 'Population and segregation data on 17 Y-STRs: results of a GEP-ISFG collaborative study', *International Journal of Legal Medicine*, 122(6), pp. 529–533. doi: 10.1007/s00414-008-0265-z.
- Santos, F. R. *et al.* (1996) 'Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism DYS19', *Human Genetics*, 97(3), pp. 309–313. doi: 10.1007/BF02185760.
- Schaffner, S. F. (2004) 'The X chromosome in population genetics', *Nature Reviews*

Genetics, 5(1), pp. 43–51. doi: 10.1038/nrg1247.

Schwarz, G. (1978) 'Estimating the Dimension of a Model', *The Annals of Statistics*, 6(2), pp. 461–464. doi: 10.1214/aos/1176344136.

Sender, R., Fuchs, S. and Milo, R. (2016) 'Revised Estimates for the Number of Human and Bacteria Cells in the Body', *PLOS Biology*, 14(8), p. e1002533. doi: 10.1371/journal.pbio.1002533.

Skaletsky, H. *et al.* (2003) 'The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes', *Nature*, 423(6942), pp. 825–837. doi: 10.1038/nature01722.

Soares-Vieira, J. A. *et al.* (2008) 'Population and mutation analysis of Y-STR loci in a sample from the city of São Paulo (Brazil) Short Communication', *Genetics and Molecular Biology*, 31, pp. 651–656. Available at: www.sbg.org.br (Accessed: 19 September 2021).

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

Szibor, R. *et al.* (2009) 'Nomenclature discrepancies in the HPRTB short tandem repeat', *International Journal of Legal Medicine*, 123(2), pp. 185–186. doi: 10.1007/s00414-008-0314-7.

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

Taanman, J.-W. (1999) 'The mitochondrial genome: structure, transcription, translation and replication', *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1410(2), pp. 103–123. doi: 10.1016/S0005-2728(98)00161-3.

Tabbada, K. A. *et al.* (2005) 'Development of a pentaplex X-chromosomal short tandem repeat typing system and population genetic studies', *Forensic Science International*, 154(2–3), pp. 173–180. doi: 10.1016/j.forsciint.2004.10.005.

Tariq, M. A. *et al.* (2008) 'Allele frequency distribution of 13 X-chromosomal STR loci in Pakistani population', *International Journal of Legal Medicine*, 122(6), pp. 525–528. doi: 10.1007/s00414-008-0263-1.

Tetzlaff, S., Wegener, R. and Lindner, I. (2012) 'Population genetic investigation of eight X-chromosomal short tandem repeat loci from a northeast German sample', *Forensic Science International: Genetics*, 6(6), pp. e155–e156. doi: 10.1016/J.FSigen.2012.03.007.

Thompson, J. M. *et al.* (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), pp. 240–250. doi: 10.1016/J.FSigen.2012.10.013.

Tillmar, A. O. *et al.* (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, pp. 269–275. doi: 10.1016/j.fsigen.2017.05.005.

Tomas, C., Pereira, V. and Morling, N. (2012) 'Analysis of 12 X-STRs in greenlanders, danes and somalis using argus X-12', *International Journal of Legal Medicine*, 126(1), pp. 121–128. doi: 10.1007/s00414-011-0609-y.

Triggs, C. M., Buckleton, J. S. and Walsh, S. J. (eds) (2004) *Forensic DNA Evidence Interpretation*. CRC Press. doi: 10.1201/9781420037920.

Turrina, S., Atzei, R. and De Leo, D. (2007) 'Polymorphism of four X-chromosomal

STRs: DXS7423, DXS7424, DXS8378 and DXS6809 in a North Italian population sample and their use in kinship testing', *Forensic Science International*, 168(2–3), pp. 241–243. doi: 10.1016/J.FORSCIINT.2006.02.044.

Turrina, S. and De Leo, D. (2004) 'Population genetic comparisons of three X-chromosomal STRs (DXS7132, DXS7133 and GATA172D05) in North and South Italy', *International Congress Series*, 1261(C), pp. 302–304. doi: 10.1016/S0531-5131(03)01626-1.

Urquhart, A. *et al.* (1994) 'Variation in Short Tandem Repeat sequences- a survey of twelve microsatellite loci for use as forensic identification markers', *International Journal of Legal Medicine*, 107(1), pp. 13–20. doi: 10.1007/BF01247268.

Vieira-Silva, C. *et al.* (2009) 'Y-STR mutational rates determination in South Portugal Caucasian population', *Forensic Science International: Genetics Supplement Series*, 2(1), pp. 60–61. doi: 10.1016/j.fsigss.2009.08.159.

Walsh, P. S., Metzger, D. A. and Higuchi, R. (1991) 'Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material.', *BioTechniques*, 10(4), pp. 506–13. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1867860>.

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

Willuweit, S. and Roewer, L. (2015) 'The new y chromosome haplotype reference database', *Forensic Science International: Genetics*, 15, pp. 43–48. doi: 10.1016/j.fsigen.2014.11.024.

Wright, S. (1949) 'THE GENETICAL STRUCTURE OF POPULATIONS', *Annals of Eugenics*, 15(1), pp. 323–354. doi: 10.1111/j.1469-1809.1949.tb02451.x.

Yang, Y. *et al.* (2018) 'Haplotypic polymorphisms and mutation rate estimates of 22 Y-chromosome STRs in the Northern Chinese Han father – son pairs', *Scientific Reports*, 8(April), pp. 6–11. doi: 10.1038/s41598-018-25362-3.

Zhao, Z. *et al.* (2015) 'Mutation rate estimation for 15 autosomal STR loci in a large population from Mainland China', *Meta Gene*, 5, pp. 150–156. doi: 10.1016/j.mgene.2015.07.006.

7. Appendix

Supplementary Table 1- Y-haplotypes of the 200 unrelated males analysed. ID- Haplotype identification number. N-Absolute frequency of the haplotype

ID	N	DYS76	DYS389I	DYS448	DYS389II	DYS19	DYS391	DYS481	DYS549	DYS533	DYS438	DYS437	DYS570	DYS635	DYS390	DYS439	DYS392	DYS643	DYS393	DYS458	DYS385ab	DYS456	YGATAH4
h1	1	16	15	19	32	14	10	23	14	12	9	14	17	21	23	11	13	10	12	18	14,14	16	11
h2	1	17	12	19	28	14	11	24	12	12	12	15	18	23	24	11	13	11	13	17	11,15	17	12
h3	1	16	14	19	30	14	10	23	11	10	9	16	18	21	23	11	11	10	12	15	14,18	16	10
h4	1	18	13	19	30	15	11	22	13	12	12	15	16	23	24	11	13	10	13	17	11,14	15	12
h5	1	16	12	19	28	15	10	23	13	13	9	16	17	21	24	11	11	9	13	16	13,17	13	11
h6	1	19	13	19	29	14	11	22	13	12	12	14	17	23	23	11	13	10	13	18	11,14	16	13
h7	1	17	14	19	30	14	11	22	13	12	12	14	18	23	24	11	13	11	13	17	11,13	16	11
h8	1	16	12	20	28	14	10	25	13	11	10	16	19	22	22	11	11	12	13	16	13,14	14	11
h9	1	19	12	20	29	15	11	24	12	11	9	16	18	21	22	11	11	8	12	17	13,16	16	11
h10	1	16	12	19	27	14	10	22	14	12	12	15	17	23	24	12	13	11	14	17	10,15	17	12
h11	1	20	13	20	29	14	11	21	13	12	11	15	17	23	24	11	13	10	12	16	11,14	15	13
h12	1	17	13	19	29	14	11	22	13	12	12	14	16	23	25	12	13	9	13	18	11,12	15	12
h13	1	16	13	19	29	14	11	25	13	12	12	14	20	23	24	11	13	9	13	17	11,14	15	11
h14	1	17	12	22	28	15	11	21	13	10	10	16	18	21	21	13	11	12	15	16	14,15	15	11
h15	1	18	14	19	31	14	11	22	12	11	12	15	17	25	24	13	13	10	14	17	11,14	15	12
h16	1	18	13	19	30	14	11	22	12	11	12	17	16	23	24	12	13	10	13	17	11,14	14	12
h17	1	17	13	20	30	16	10	23	12	12	11	14	18	23	24	10	11	10	13	17	11,14	15	13
h18	1	17	14	20	30	13	10	29	11	11	10	14	23	21	24	10	11	12	13	17	13,14	16	12
h19	1	17	13	19	30	15	10	22	14	13	12	14	17	25	24	12	13	10	13	17	11,15	15	13
h20	1	16	12	20	29	14	10	25	13	11	10	16	18	22	22	11	11	12	13	16	14,15	14	11
h21	1	17	12	20	28	14	10	25	13	11	10	16	16	24	22	11	11	12	13	15	13,15	14	11
h22	1	17	13	18	29	14	11	22	13	11	12	14	18	23	24	12	13	10	13	17	11,14	15	11

h23	1	20	14	19	30	14	10	22	13	12	12	15	16	23	24	12	13	10	13	19	11,14	16	12
h24	1	17	13	19	28	14	11	22	12	12	12	15	19	23	24	11	14	10	14	16	11,15	15	12
h25	1	19	14	19	31	14	11	23	14	12	13	15	19	24	24	12	13	10	14	17	11,11	16	12
h26	1	17	13	18	29	14	11	22	12	11	12	15	16	23	24	12	13	10	13	18	11,14	16	12
h27	1	18	13	21	28	14	9	21	11	10	9	15	16	23	22	11	11	10	12	16	13,17	15	11
h28	1	18	14	20	30	13	9	26	11	11	10	14	22	21	24	10	11	12	13	18	13,14	15	12
h29	1	19	13	20	31	14	10	24	11	10	10	14	18	20	22	12	12	9	12	18,2	14,19	15	11
h30	1	19	13	21	28	16	11	22	11	12	10	15	22	22	22	12	11	12	13	17	11,12	13	12
h31	1	18	13	20	28	16	11	22	10	12	10	15	19	21	22	13	11	13	13	17	11,12	14	12
h32	1	17	14	20	29	14	9	22	12	12	10	15	20	25	25	13	13	10	12	16	13,14	17	11
h33	1	19	13	19	29	14	11	22	12	12	12	15	17	23	24	11	13	10	13	16	12,14	16	12
h34	1	18	13	19	29	14	10	22	12	12	12	15	18	24	24	13	13	11	13	18	11,15	15	12
h35	1	18	14	19	30	14	11	22	13	12	12	15	17	23	24	12	12	10	13	17	11,14	16	12
h36	1	16	12	21	28	16	10	22	14	9	10	16	19	21	22	12	11	11	14	16	14,15	17	12
h37	1	17	13	19	30	15	11	23	12	12	12	15	18	23	23	13	13	10	13	17	11,14	16	12
h38	1	17	13	18	29	14	10	22	14	13	12	14	18	23	24	12	13	10	13	16	10,14	16	11
h39	1	18	13	19	30	14	11	22	13	12	12	14	17	24	24	12	13	11	13	17	11,15	15	12
h40	1	16	13	19	29	14	11	23	11	12	12	15	17	23	24	13	13	10	13	18	11,15	15	12
h41	1	19	13	19	30	14	10	22	12	12	12	15	17	24	24	12	13	10	13	20	11,14	15	12
h42	1	17	14	19	30	14	11	23	13	13	12	15	17	23	24	12	13	10	12	16	11,15	16	12
h43	1	20	13	19	29	14	11	22	14	13	12	15	17	23	24	12	14	10	13	16	11,15	15	11
h44	1	15	13	21	29	14	10	26	12	11	10	16	18	21	23	11	11	12	13	16	13,14	14	11
h45	1	18	13	19	30	14	11	23	14	12	12	14	18	23	24	12	13	10	13	15	11,14	15	11
h46	1	15	13	21	29	14	10	22	12	11	9	15	17	24	22	11	11	8	12	16	13,15	15	10
h47	1	18	13	20	29	15	12	22	12	12	12	15	17	23	24	12	13	10	13	18	11,15	16	11
h48	1	18	13	19	29	14	11	22	14	12	12	15	18	23	23	13	13	10	13	15	11,14	15	11
h49	1	19	14	19	29	14	11	23	12	12	12	15	17	24	24	12	14	10	12	20	12,2,14	15	13
h50	1	21	13	19	29	14	11	22	13	12	12	15	17	23	24	14	13	10	13	17	11,14	16	12
h51	1	18	13	19	28	14	10	22	12	12	12	15	19	23	25	11	13	10	13	16	11,11	15	12
h52	1	17	13	19	30	13	10	23	13	12	10	14	24	21	24	12	11	11	13	16	15,18	17	11

h53	1	17	13	19	29	13	11	22	12	12	12	15	17	23	24	13	13	10	13	16	11,14	15	12
h54	1	16	14	19	31	14	11	22	13	12	12	15	19	23	24	11	13	10	13	17	11,14	15	12
h55	1	19	13	18	30	14	11	22	12	12	12	14	17	25	24	12	13	10	13	17	11,11	15	12
h56	1	18	14	19	30	14	11	22	13	12	12	14	18	23	24	12	13	11	13	17	11,13	15	11
h57	1	20	13	21	29	15	10	24	12	12	9	15	16	22	25	11	11	11	12	20	13,13	15	11
h58	1	18	12	19	27	14	10	22	13	12	12	15	17	23	24	12	13	11	14	17	10,15	17	12
h59	1	19	13	20	30	13	10	24	10	13	11	14	18	22	22	12	16	12	13	16	15,18	16	10
h60	1	19	13	19	29	14	10	23	13	13	12	15	16	23	24	11	13	10	13	17	11,14	17	12
h61	1	17	13	19	30	15	10	23	13	12	9	15	18	21	25	11	11	9	12	18	14,16	14	11
h62	1	20	12	19	28	14	11	22	12	12	12	15	16	23	24	12	13	10	13	20	11,14	17	12
h63	1	18	13	19	29	14	11	22	12	12	12	15	18	23	25	13	13	10	14	18	11,13	17	13
h64	1	17	13	19	29	14	11	23	13	12	12	15	17	23	24	13	15	11	13	17	11,14	15	12
h65	1	18	12	19	27	14	10	22	14	12	12	15	17	23	24	12	13	11	14	17	10,15	17	12
h66	1	17	13	19	29	13	10	21	14	11	9	15	18	20	23	12	13	11	14	16	16,16	17	12
h67	1	18	13	20	31	15	11	23	12	12	11	14	19	23	25	10	11	10	14	15	11,14	15	12
h68	1	18	13	19	28	14	10	23	13	12	12	15	17	23	24	12	13	10	13	17	12,14	16	12
h69	1	17	14	19	30	13	10	28	13	13	9	15	18	22	23	12	13	12	13	17	15,16	14	11
h70	1	17	13	20	30	14	11	22	12	12	12	15	17	23	24	12	13	11	13	19	11,14	16	12
h71	1	18	13	20	31	14	10	25	12	11	10	14	20	21	23	12	11	9	12	17,2	13,18	16	11
h72	1	18	13	19	29	15	10	22	12	12	12	15	17	23	24	13	13	10	13	16	11,14	16	12
h73	1	18	13	18	30	14	11	23	14	12	12	15	16	23	24	12	13	10	12	16	11,15	16	12
h74	1	21	13	19	29	14	11	22	12	12	12	15	15	23	26	12	13	11	12	16	11,14	15	12
h75	1	19	14	17	29	15	10	22	13	12	12	15	17	23	24	11	13	10	13	16	11,14	15	12
h76	1	19	13	19	29	14	11	22	12	13	10	15	17	25	24	11	13	10	13	17	9,14	15	13
h77	1	19	13	20	29	13	11	22	13	12	12	15	16	25	24	12	14	10	13	17	11,14	15	12
h78	1	17	13	19	30	14	11	23	13	12	12	14	16	23	23	12	13	10	13	16	14,15	16	11
h79	1	18	13	19	29	14	11	22	13	12	12	15	16	23	23	12	13	10	13	16	11,14	17	12
h80	1	17	14	21	30	14	10	22	12	11	9	15	17	21	24	11	11	10	12	15	14,17	15	11
h81	1	16	12	19	28	14	10	26	13	12	10	16	19	21	24	11	11	11	12	16	13,14	15	11
h82	1	21	13	19	29	14	11	22	11	12	12	15	15	23	23	12	13	10	13	17	11,14	16	12

h83	1	16	12	20	28	14	10	24	12	11	9	15	16	20	23	12	11	10	12	16	13,16	15	11
h84	1	18	13	19	30	14	11	22	14	12	12	15	17	23	25	12	13	11	15	19	11,14	16	12
h85	1	18	13	19	29	14	10	23	13	12	11	15	17	23	23	12	13	11	13	18	11,12	15	11
h86	1	18	13	19	30	14	11	22	13	13	12	14	18	24	24	12	13	11	13	17	11,15	16	12
h87	1	19	13	20	29	13	9	26	11	11	10	14	21	21	24	10	11	11	13	17	13,14	16	12
h88	1	17	13	20	29	13	9	27	11	11	10	14	22	21	24	11	11	12	13	17	13,13	16	11
h89	1	19	13	19	29	14	11	24	12	12	12	15	17	23	23	12	13	10	13	18	11,15	15	12
h90	1	16	12	19	29	16	11	23	13	12	9	16	18	22	24	12	11	9	12	15	13,16	13	11
h91	1	17	13	19	29	14	11	22	13	12	12	15	17	24	24	12	13	10	13	15	11,14	15	12
h92	1	18	14	19	30	14	11	23	12	13	12	15	17	23	23	12	13	10	13	17	11,14	15	12
h93	1	16	12	19	28	15	11	24	15	11	9	16	16	21	24	12	11	9	12	16	14,17	13	11
h94	1	20	13	19	30	14	11	22	13	12	12	16	15	25	23	13	13	10	13	18	11,15	15	0
h95	1	18	13	20	30	15	10	27	12	12	10	14	17	21	23	11	12	12	15	15	15,16	15	11
h96	1	19	13	20	29	14	11	22	12	12	11	15	17	23	24	12	13	10	12	16	11,14	14	13
h97	1	18	13	19	29	14	11	22	12	12	12	14	17	23	24	13	13	10	13	17	11,14	15	11
h98	1	16	14	19	31	14	10	23	13	13	12	15	17	23	24	11	13	10	13	19	11,14	15	12
h99	1	17	14	20	31	15	10	27	13	11	10	14	19	21	23	11	12	12	15	15	15,15	14	10
h100	1	15	12	22	29	15	11	21	11	10	10	16	18	21	22	11	11	12	14	17	13,14	15	11
h101	1	17	13	18	29	14	11	22	13	12	13	15	16	24	25	12	12	10	13	16	12,15	15	12
h102	1	19	13	19	29	14	11	22	12	11	12	16	16	23	24	12	13	10	13	17	11,11	14	12
h103	1	18	14	21	30	14	10	25	13	12	10	14	19	20	23	11	11	10	12	18.2	13,13	14	11
h104	1	19	14	19	30	15	10	22	13	12	12	15	17	23	23	12	13	10	13	17	11,14	16	12
h105	1	17	14	21	31	14	10	23	11	11	9	14	17	23	23	11	11	10	13	18	14,15	16	11
h106	1	17	14	20	30	13	9	27	11	11	10	14	22	21	24	10	11	14	13	18	13,14	15	12
h107	1	15	13	19	30	15	10	26	13	12	9	14	20	22	23	11	15	10	13	15	16,17	15	11
h108	1	15	13	20	30	14	10	26	13	12	10	14	17	20	23	11	11	9	12	19.2	13,19	14	11
h109	1	20	13	20	32	14	10	24	11	10	10	14	18	20	22	12	11	9	12	18.2	14,18	15	11
h110	1	18	13	19	29	15	11	22	12	12	10	15	16	23	25	13	13	10	13	16	11,14	15	12
h111	1	19	13	19	29	15	11	21	13	13	12	15	18	23	22	11	13	10	12	18	10,14	15	12
h112	1	16	12	19	29	16	10	23	13	12	9	16	19	21	26	11	11	9	12	15	13,16	13	11

h113	1	18	13	19	29	14	11	22	12	11	12	15	18	23	23	12	13	10	13	17	12,14	16	11
h114	1	19	14	20	32	17	10	28	11	11	10	15	21	21	25	12	12	13	14	18	12,15	15	10
h115	1	19	13	19	29	14	10	21	12	11	12	16	16	23	24	12	13	10	13	17	11,14	14	12
h116	1	19	13	19	29	14	11	22	12	12	12	15	18	23	23	13	13	11	13	19	11,14	15	12
h117	1	18	13	20	30	14	10	23	13	11	9	15	15	21	23	12	11	11	12	17	13,16	15	10
h118	1	18	13	20	29	14	11	22	13	13	12	14	17	23	24	11	13	10	13	19	11,0	16	12
h119	1	18	13	19	29	14	10	22	13	12	12	15	18	24	24	13	13	10	13	17	11,15	15	12
h120	1	17	13	19	29	15	11	22	13	12	10	15	17	24	24	13	13	11	13	17	11,14	16	12
h121	1	18	13	19	30	14	11	22	14	13	12	14	18	24	24	12	13	11	13	17	11,15	16	12
h122	1	20	13	18	29	14	10	25	12	13	12	15	17	23	24	10, 12	13	10	13	16	12,12	15	12
h123	1	18	13	18	29	15	10	22	14	12	12	15	19	23	24	11	13	10	13	17	12,15	16	11
h124	1	18	13	20	29	14	11	22	13	12	12	15	20	23	23	12	13	10	13	16	11,15	17	12
h125	1	16	13	21	29	16	9	23	12	12	9	14	19	25	23	12	11	10	12	14	13,13	16	12
h126	1	20	13	19	29	14	10	23	12	12	13	16	16	23	24	12	13	10	13	17	11,14	15	12
h127	1	18	12	19	26	15	11	20	14	11	12	15	17	24	24	10	13	10	13	16	12,14	16	12
h128	1	17	14	19	30	14	11	22	13	12	12	15	17	23	24	12	13	10	13	19	11,14	15	12
h129	1	19	14	20	30	13	10	26	10	11	10	14	23	21	24	10	11	12	13	17	13,14	17	12
h130	1	19	12	19	28	14	11	22	12	11	12	15	17	24	24	12	13	11	13	17	11,15	16	12
h131	1	19	13	19	29	14	11	22	12	12	12	15	18	23	24	13	13	10	13	15	11,14	15	11
h132	1	19	14	19	28	14	11	22	12	12	12	15	17	24	24	12	14	10	12	20	12,2,14	15	12
h133	1	16	12	20	28	15	10	24	12	12	10	16	19	22	22	11	11	12	13	15	13,14	15	11
h134	1	17	13	19	30	14	11	21	12	11	12	14	20	23	23	13	13	10	13	17	12,14	15	13
h135	1	19	13	22	29	15	10	23	15	11	9	15	15	21	23	12	11	11	12	15	13,18	15	11
h136	1	18	13	19	29	14	11	21	11	11	12	15	17	23	23	12	13	10	12	17	11,13	16	12
h137	1	19	13	20	30	15	10	22	12	12	12	15	18	23	24	12	13	10	13	17	11,15	15	13
h138	1	18	13	22	29	15	10	22	15	11	9	15	16	21	23	11	11	11	12	15	13,18	15	11
h139	1	17	13	20	29	13	9	27	11	11	10	14	25	22	24	10	11	12	13	18	13,14	17	12
h140	1	19	12	19	28	14	11	22	13	12	12	15	16	23	24	11	13	10	13	16	11,14	16	12
h141	1	18	13	21	28	17	10	23	10	12	10	15	18	21	23	11	11	12	13	17	12,12	14	12
h142	1	19	13	19	29	14	11	21	13	12	12	15	16	23	23	12	13	10	14	16	11,14	15	12

h143	1	17	13	21	29	14, 15, 17	10	22	11	10	10	16	18	21	22	10	11	11	14	19	13, 15	14	11
h144	1	17	12	19	28	14	11	22	13	12	12	15	16	23	25	13	13	12	13	16	11, 14	15	12
h145	1	17	13	19	29	14	11	22	13	12	12	15	18	23	24	11	14	10	14	16	11, 15	15	12
h146	1	19	14	19	30	14	11	22	13	12	12	15	18	23	24	12	13	10	13	17	12, 14	15	11
h147	1	17	14	19	30	14	11	22	12	11	13	15	18	23	24	11	13	10	13	17	11, 15	15	12
h148	1	19	13	19	30	14	11	22	13	12	12	15	17	23	25	12	13	10	13	17	11, 14	16	12
h149	1	17	13	20	30	14	10	25	12	10	10	14	18	21	23	12	11	9	13	18, 2	15, 17	15	11
h150	1	18	13	20	30	15	10	22	12	12	12	15	18	23	23	12	13	10	13	16	11, 15	15	12
h151	1	18	13	19	29	14	11	22	12	11	13	15	16	23	24	12	13	10	13	16	11, 13	16	12
h152	1	17	13	20	30	13	10	22	12	12	10	14	18	24	24	11	11	0	13	15	16, 18	15	12
h153	1	18	13	21	28	16	10	23	10	12	10	14	18	21	23	12	11	13	13	17	12, 12	14	12
h154	1	19	13	19	29	14	11	23	14	11	12	15	18	23	23	12	13	10	13	17	11, 14	17	11
h155	1	18	13	20	31	15	10	26	13	8	10	14	20	22	24	11	11	10	12	15	13, 18	15	10
h156	1	18	13	19	29	14	11	21	13	12	12	15	17	23	23	12	13	10	13	16	11, 14	17	11
h157	1	18	13	19	29	14	11	22	13	11	12	15	18	23	23	12	13	10	13	17	11, 14	16	11
h158	1	17	12	19	28	14	12	22	13	12	12	15	17	23	24	12	13	10	13	16	11, 15	17	11
h159	1	19	13	19	29	14	11	22	15	12	12	15	17	23	22	13	13	10	13	17	11, 14	15	12
h160	1	17	13	19	29	14	11	22	13	12	12	15	16	21	25	12	12	10	13	17	11, 11	16	12
h161	1	15	12	21	29	15	10	21	12	9	10	16	18	20	23	12	11	11	14	17	14, 14	15	12
h162	1	14	12	21	29	15	10	21	12	9	10	16	18	20	22	11	11	11	14	16	14, 14	15	13
h163	1	16	13	20	29	15	10	24	12	12	9	14	16	21	23	12	13	10	13	16	14, 18	15	11
h164	1	22	12	19	28	14	10	22	11	12	12	15	17	23	23	11	12	10	14	19	11, 14	16	12
h165	1	18	13	21	28	17	10	23	10	12	10	15	19	22	23	11	11	12	13	16	11, 12	14	12
h166	1	18	14	20	31	15	10	25	11	11	10	15	17	21	23	12	12	12	14	19	16, 16	13	12
h167	1	18	13	19	29	13	11	22	13	12	13	15	17	23	24	12	13	10	13	16	11, 14	16	12
h168	1	16	12	19	28	15	10	24	12	11	9	14	18	21	24	11	14	10	13	16	14, 16	15	11
h169	1	17	13	20	30	14	10	22	12	12	12	15	17	23	24	12	13	11	13	20	12, 14	16	12
h170	1	18	14	20	30	13	9	28	12	10	10	14	25	21	23	11	11	12	13	17	13, 14	16	12
h171	1	18	12	19	28	14	11	22	13	12	12	14	18	24	23	15	13	10	13	16	12, 14	15	12
h172	1	18	13	19	29	14	11	23	13	12	12	14	16	23	23	10	14	10	13	17	11, 14	16	12

h173	1	17	13	21	29	15	9	22	11	12	9	14	16	21	23	12	11	10	12	14	13,16	15	11
h174	1	18	13	19	29	13	11	24	12	12	12	14	18	23	23	11	13	10	13	17	12,14	17	11
h175	1	18	13	20	29	14	10	25	13	11	11	14	17	23	24	12	14	10	13	17	13,14	16	11
h176	1	17	13	20	29	13	9	26	11	11	10	14	22	21	25	10	12	12	13	18	14,14	16	12
h177	1	17	12	21	28	16	10	21	14	10	10	16	18	21	22	11	11	12	14	17	13,14	15	11
h178	1	17	14	19	29	14	11	23	13	12	12	15	18	23	24	12	14	10	13	17	10,14	16	12
h179	1	16	14	20	31	13	9	23	13	11	9	14	19	22	24	11	11	12	14	15	15,17	14	12
h180	1	19	13	19	29	14	11	22	12	14	12	15	18	23	24	12	13	11	13	18	11,14	16	12
h181	1	18	13	21	28	16	11	22	10	12	10	15	18	21	23	11	11	11	13	19	12,12	14	12
h182	1	19	14	20	30	13	9	25	11	11	10	14	22	21	23	10	11	10	13	18	12,15	15	12
h183	1	17	13	19	29	13	11	21	13	11	9	15	18	20	23	12	13	12	14	16	15,16	17	12
h184	1	19	13	20	30	14	10	22	12	11	10	14	19	24	25	11	11	12	14	17	16,17	15	11
h185	1	19	13	20	29	14	11	22	12	12	9	15	15	24	23	11	11	10	12	17	13,15	17	11
h186	1	21	13	20	29	14	11	23	14	12	12	15	18	24	25	13	13	10	13	17	11,14	15	11
h187	1	19	13	20	29	14	11	22	13	13	12	15	18	23	24	10	13	10	13	17	11,14	16	11
h188	1	19	14	19	30	14	11	22	12	12	12	15	17	23	24	12	13	10	13	16	11,14	15	12
h189	1	18	13	19	30	15	10	24	12	12	12	15	17	23	24	12	13	11	13	18	11,14	16	11
h190	1	18	12	19	27	14	11	22	14	13	12	15	16	23	26	11	13	10	14	17	11,14	15	12
h191	1	17	14	19	30	14	10	22	12	12	12	15	16	23	24	13	13	10	13	17	11,14	16	11
h192	1	19	13	19	29	14	11	22	13	12	13	14	17	23	23	12	13	10	12	18	12,15	16	12
h193	1	19	14	20	30	13	9	27	11	11	10	14	22	22	24	10	11	12	13	18	13,14	15	12
h194	1	18	13	20	29	14	10	22	12	12	12	15	17	23	24	11	13	10	13	17	11,14	15	13
h195	1	18	13	20	31	14	10	24	11	10	10	14	18	20	22	12	11	9	12	18,2	14,18	15	11
h196	1	17	14	21	30	16	10	22	10	12	10	15	17	22	23	12	11	13	13	17	12,12	15	12
h197	1	18	13	19	29	14	11	21	11	11	12	15	17	24	23	12	13	10	12	17	11,14	15	12
h198	1	18	12	19	26	15	11	20	14	11	12	15	17	24	24	11	13	10	13	16	12,14	16	12
h199	1	19	13	19	30	14	11	22	12	12	12	15	17	24	24	11	13	10	13	19	11,14	15	12
h200	1	17	14	19	30	14	10	22	13	12	12	15	18	23	24	11	13	10	14	16	11,13	15	12

Supplementary Table 2- Y-STRs specific mutation rate estimates based on data from this study (N=202 father-son duos), and the literature. (*)-collected data from other studies.

Marker	Reference	No. Allele transfers	No. Mutations	Mutation rate X 10 ⁻³	95 % CI X 10 ⁻³
DYS576	Lin et al., (2020)	640	13	20.3	10.9-34.5
	Ambrosio et al., (2020)	406	4	9.9	2.7-25
	Antão-Sousa, (2017)*	4609	71	15.4	12.1-19.4
	This study	202	3	14.9	3.1-42.8
	Total	5857	91	15.5	12.5-19
DYS389I	Lin et al., (2020)	640	1	1.6	0-8.7
	Ambrosio et al., (2020)	407	1	2.5	0.1-13.6
	Yang et al., (2018)	578	2	3.5	0.4-12.4
	Antão-Sousa,(2017)*	22465	49	2.2	1.6-2.9
	This study	202	1	5.0	0.1-27.3
	Total	24292	54	2.2	1.7-2.9
DYS448	Lin et al., (2020)	640	2	3.1	0.4-11.2
	Ambrosio et al., (2020)	408	1	2.5	0.1-13.6
	Yang et al., (2018)	578	1	1.7	0-9.6
	Antão-Sousa,(2017)*	11236	12	1.1	0.6-1.9
	This study	201	0	0.0	0-18.2
	Total	13063	16	1.2	0.7-2
DYS389II	Lin et al., (2020)	640	3	4.7	1-13.6
	Ambrosio et al., (2020)	408	3	7.4	1.5-21.3
	Yang et al., (2018)	578	5	8.7	2.8-20.1
	Antão-Sousa,(2017)*	22411	77	3.4	2.7-4.3
	This study	202	1	5.0	0.1-27.3
	Total	24239	89	3.7	2.9-4.5
DYS19	Lin et al., (2020)	640	1	1.6	0-8.7
	Ambrosio et al., (2020)	411	0	0.0	0-8.9
	Yang et al., (2018)	578	1	1.7	0-9.6
	Antão-Sousa,(2017)*	23638	47	2.0	1.5-2.6
	This study	204	1	4.9	0.1-27
	Total	25471	50	2.0	1.5-2.6
DYS391	Lin et al., (2020)	640	0	0.0	0-5.7
	Ambrosio et al., (2020)	407	0	0.0	0-9
	Yang et al., (2018)	578	1	1.7	0-9.6
	Antão-Sousa,(2017)*	23218	50	2.2	1.6-2.8
	This study	202	0	0.0	0-18.1
	Total	25045	51	2.0	1.5-2.7

DYS481	Lin et al., (2020)	640	1	1.6	0-8.7
	Ambrosio et al., (2020)	407	3	7.4	1.5-21.4
	Antão-Sousa, (2017)*	1446	8	5.5	2.4-10.9
	This study	202	2	9.9	1.2-35.3
	Total	2695	14	5.2	2.8-8.7
DYS549	Lin et al., (2020)	640	0	0.0	0-5.7
	Ambrosio et al., (2020)	407	3	7.4	1.5-21.4
	Antão-Sousa, (2017)*	413	1	2.4	0.1-13.4
	This study	201	1	5.0	0.1-27.4
	Total	1661	5	3.0	1-7
DYS533	Lin et al., (2020)	640	0	0.0	0-5.7
	Ambrosio et al., (2020)	407	0	0.0	0-9
	Antão-Sousa, (2017)*	2256	5	2.2	0.7-5.2
	This study	202	0	0.0	0-18.1
	Total	3505	5	1.4	0.5-3.3
DYS438	Lin et al., (2020)	640	0	0.0	0-5.7
	Ambrosio et al., (2020)	407	0	0.0	0-9
	Yang et al., (2018)	578	0	0.0	0-6.4
	Antão-Sousa, (2017)*	14354	7	0.5	0.2-1
	This study	202	0	0.0	0-18.1
	Total	16181	7	0.4	0.2-0.9
DYS437	Lin et al., (2020)	640	0	0.0	0-5.7
	Ambrosio et al., (2020)	407	1	2.5	0.1-13.6
	Yang et al., (2018)	578	0	0.0	0-6.4
	Antão-Sousa, (2017)*	14334	18	1.3	0.7-2
	This study	202	0	0.0	0-18.1
	Total	16161	19	1.2	0.7-1.8
DYS570	Lin et al., (2020)	640	2	3.1	0.4-11.2
	Ambrosio et al., (2020)	407	3	7.4	1.5-21.4
	Antão-Sousa, (2017)*	4654	39	8.4	6-11.4
	This study	202	0	0.0	0-18.1
	Total	5903	44	7.5	5.4-10
DYS635	Lin et al., (2020)	640	4	6.3	1.7-15.9
	Ambrosio et al., (2020)	407	1	2.5	0.1-13.6
	Yang et al., (2018)	578	1	1.7	0-9.6
	Antão-Sousa, (2017)*	11885	33	2.8	1.9-3.9
	This study	202	1	5.0	0.1-27.3
	Total	13712	40	2.9	2.1-4
DYS390	Ambrosio et al., (2020)	407	4	9.8	2.7-25
	Yang et al., (2018)	578	3	5.2	1.1-15.1

	Antão-Sousa, (2017)*	23543	40	1.7	1.2-2.3
	This study	202	1	5.0	0.1-27.3
	Total	24730	48	1.9	1.4-2.6
DYS439	Lin et al., (2020)	640	3	4.7	1-13.6
	Ambrosio et al., (2020)	410	4	9.8	2.7-24.8
	Yang et al., (2018)	578	4	6.9	1.9-17.6
	Antão-Sousa, (2017)*	14723	65	4.4	3.4-5.6
	This study	203	0	0.0	0-18
	Total	16554	76	4.6	3.6-5.7
DYS392	Lin et al., (2020)	640	1	1.6	0-8.7
	Ambrosio et al., (2020)	406	0	0.0	0-9
	Yang et al., (2018)	578	1	1.7	0-9.6
	Antão-Sousa, (2017)*	23168	9	0.4	0.2-0.7
	This study	201	0	0.0	0-18.2
	Total	24993	11	0.4	0.2-0.8
DYS643	Ambrosio et al., (2020)	407	0	0.0	0-9
	Antão-Sousa, (2017)*	686	0	0.0	0-5.4
	This study	201	0	0.0	0-18.2
	Total	1294	0	0.0	0-2.8
DYS393	Ambrosio et al., (2020)	407	1	2.5	0.1-13.6
	Yang et al., (2018)	578	1	1.7	0-9.6
	Antão-Sousa, (2017)*	21799	19	0.9	0.5-1.4
	This study	202	0	0.0	0-18.1
	Total	22986	21	0.9	0.6-1.4
DYS458	Lin et al., (2020)	640	4	6.3	1.7-15.9
	Ambrosio et al., (2020)	407	3	7.4	1.5-21.4
	Yang et al., (2018)	578	5	8.7	2.8-20.1
	Antão-Sousa, (2017)*	11268	76	6.7	5.3-8.4
	This study	202	1	5.0	0.1-27.3
	Total	13095	89	6.8	5.5-8.4
DYS385 a/b	Lin et al., (2020)	1280	10	7.8	3.8-14.3
	Ambrosio et al., (2020)	814	1	1.2	0-6.8
	Yang et al., (2018)	1156	4	3.5	0.9-8.8
	Antão-Sousa, (2017)*	35724	71	2.0	1.6-2.5
	This study	404	0	0.0	0-9.1
	Total	39378	86	2.2	1.7-2.7
DYS456	Lin et al., (2020)	640	3	4.7	1-13.6
	Ambrosio et al., (2020)	407	3	7.4	1.5-21.4
	Yang et al., (2018)	1156	1	0.9	0-4.8
	Antão-Sousa, (2017)*	11241	43	3.8	2.8-5.1

	This study	202	2	9.9	1.2-35.3
	Total	13646	52	3.8	2.8-5
YGATAH4	Lin et al., (2020)	640	2	3.1	0.4-11.2
	Ambrosio et al., (2020)	407	0	0.0	0-9
	Yang et al., (2018)	578	1	1.7	0-9.6
	Antão-Sousa, (2017)*	12398	26	2.1	1.4-3.1
	This study	201	1	5.0	0.1-27.4
	Total	14224	30	2.1	1.4-3
Total		348685	898	2.6	2.4-2.7

Supplementary Table 3- Y-STR specific mutation steps based on data from this study (N=202 father-son duos), and from the literature. (*)- collected data from other studies. Mutation steps: {+1}- gain of one repeat; {-1}- loss of one repeat; >{+1}- gain of more than one repeat; <{-1}loss of more than one repeat; NI- non-integer repeat variation.

Marker	Reference	No. Mutations	{+1}	{-1}	>{+1}	<{-1}	NI	No. repeat gains	No. repeat losses
DYS576	Lin et al., (2020)	13	4	8	1	0	0	5	8
	Ambrosio et al., (2020)	4	2	2	0	0	0	2	2
	Antão-Sousa, (2017)*	71	37	31	0	3	0	37	34
	This study	3	1	2	0	0	0	1	2
	Total	91	44	43	1	3	0	45	46
DYS389I	Lin et al., (2020)	1	0	1	0	0	0	0	1
	Ambrosio et al., (2020)	1	0	1	0	0	0	0	1
	Yang et al., 2018	2	1	1	0	0	0	1	1
	Antão-Sousa, (2017)*	49	21	28	0	0	0	21	28
	This study	1	1	0	0	0	0	1	0
	Total	54	23	31	0	0	0	23	31
DYS448	Lin et al., (2020)	2	2	0	0	0	0	2	0
	Ambrosio et al., (2020)	1	0	1	0	0	0	0	1
	Yang et al., 2018	1	0	1	0	0	0	0	1
	Antão-Sousa, (2017)*	12	4	7	0	1	0	4	8
	This study	0	0	0	0	0	0	0	0
	Total	16	6	9	0	1	0	6	10
DYS389II	Lin et al., (2020)	3	1	2	0	0	0	1	2
	Ambrosio et al., (2020)	3	1	2	0	0	0	1	2
	Yang et al., 2018	5	3	2	0	0	0	3	2
	Antão-Sousa, (2017)*	77	35	39	3	0	0	38	39
	This study	1	0	1	0	0	0	0	1
	Total	89	40	46	3	0	0	43	46
DYS19	Lin et al., (2020)	1	0	1	0	0	0	0	1
	Ambrosio et al., (2020)	0	0	0	0	0	0	0	0
	Yang et al., 2018	1	0	1	0	0	0	0	1
	Antão-Sousa, (2017)*	47	26	20	0	0	1	26	20
	This study	1	1	0	0	0	0	1	0
	Total	50	27	22	0	0	1	27	22
DYS391	Lin et al., (2020)	0	0	0	0	0	0	0	0
	Ambrosio et al., (2020)	0	0	0	0	0	0	0	0
	Yang et al., 2018	1	1	0	0	0	0	1	0
	Antão-Sousa, (2017)*	50	28	21	1	0	0	29	21
	This study	0	0	0	0	0	0	0	0
	Total	51	29	21	1	0	0	30	21

DYS481	Lin et al., (2020)	1	1	0	0	0	0	1	0
	Ambrosio et al., (2020)	3	2	1	0	0	0	2	1
	Antão-Sousa, (2017)*	8	5	2	0	1	0	5	3
	This study	2	1	1	0	0	0	1	1
	Total	14	9	4	0	1	0	9	5
DYS549	Lin et al., (2020)	0	0	0	0	0	0	0	0
	Ambrosio et al., (2020)	3	1	2	0	0	0	1	2
	Antão-Sousa, (2017)*	1	0	1	0	0	0	0	1
	This study	1	0	1	0	0	0	0	1
	Total	5	1	4	0	0	0	1	4
DYS533	Lin et al., (2020)	0	0	0	0	0	0	0	0
	Ambrosio et al., (2020)	0	0	0	0	0	0	0	0
	Antão-Sousa, (2017)*	5	3	2	0	0	0	3	2
	This study	0	0	0	0	0	0	0	0
	Total	5	3	2	0	0	0	3	2
DYS438	Lin et al., (2020)	0	0	0	0	0	0	0	0
	Ambrosio et al., (2020)	0	0	0	0	0	0	0	0
	Yang et al., 2018	0	0	0	0	0	0	0	0
	Antão-Sousa, (2017)*	7	1	1	0	5	0	1	6
	This study	0	0	0	0	0	0	0	0
	Total	7	1	1	0	5	0	1	6
DYS437	Lin et al., (2020)	0	0	0	0	0	0	0	0
	Ambrosio et al., (2020)	1	0	1	0	0	0	0	1
	Yang et al., 2018	0	0	0	0	0	0	0	0
	Antão-Sousa, (2017)*	18	11	7	0	0	0	11	7
	This study	0	0	0	0	0	0	0	0
	Total	19	11	8	0	0	0	11	8
DYS570	Lin et al., (2020)	2	0	2	0	0	0	0	2
	Ambrosio et al., (2020)	3	2	1	0	0	0	2	1
	Antão-Sousa, (2017)*	39	14	21	1	3	0	15	24
	This study	0	0	0	0	0	0	0	0
	Total	44	16	24	1	3	0	17	27
DYS635	Lin et al., (2020)	4	2	2	0	0	0	2	2
	Ambrosio et al., (2020)	1	0	1	0	0	0	0	1
	Yang et al., 2018	1	1	0	0	0	0	1	0
	Antão-Sousa, (2017)*	33	11	22	0	0	0	11	22
	This study	1	0	1	0	0	0	0	1
	Total	40	14	26	0	0	0	14	26
DYS390	Ambrosio et al., (2020)	4	2	2	0	0	0	2	2
	Yang et al., 2018	3	2	1	0	0	0	2	1
	Antão-Sousa, (2017)*	40	18	22	0	0	0	18	22

	This study	1	0	1	0	0	0	0	1
	Total	48	22	26	0	0	0	22	26
DYS439	Lin et al., (2020)	3	1	2	0	0	0	1	2
	Ambrosio et al., (2020)	4	3	1	0	0	0	3	1
	Yang et al., 2018	4	4	0	0	0	0	4	0
	Antão-Sousa, (2017)*	65	30	35	0	0	0	30	35
	This study	0	0	0	0	0	0	0	0
	Total	76	38	38	0	0	0	38	38
DYS392	Lin et al., (2020)	1	0	1	0	0	0	0	1
	Ambrosio et al., (2020)	0	0	0	0	0	0	0	0
	Yang et al., 2018	1	1	0	0	0	0	1	0
	Antão-Sousa, (2017)*	9	6	2	0	1	0	6	3
	This study	0	0	0	0	0	0	0	0
	Total	11	7	3	0	1	0	7	4
DYS643	Ambrosio et al., (2020)	0	0	0	0	0	0	0	0
	Antão-Sousa, (2017)*	0	0	0	0	0	0	0	0
	This study	0	0	0	0	0	0	0	0
	Total	0	0	0	0	0	0	0	0
DYS393	Ambrosio et al., (2020)	1	0	1	0	0	0	0	1
	Yang et al., 2018	1	1	0	0	0	0	1	0
	Antão-Sousa, (2017)*	19	12	7	0	0	0	12	7
	This study	0	0	0	0	0	0	0	0
	Total	21	13	8	0	0	0	13	8
DYS458	Lin et al., (2020)	4	3	1	0	0	0	3	1
	Ambrosio et al., (2020)	3	1	2	0	0	0	1	2
	Yang et al., 2018	5	3	2	0	0	0	3	2
	Antão-Sousa, (2017)*	76	38	35	1	2	0	39	37
	This study	1	0	1	0	0	0	0	1
	Total	89	45	41	1	2	0	46	43
DYS385 a/b	Lin et al., (2020)	10	6	4	0	0	0	6	4
	Ambrosio et al., (2020)	1	1	0	0	0	0	1	0
	Yang et al., 2018	4	2	2	0	0	0	2	2
	Antão-Sousa, (2017)*	71	52	14	1	4	0	53	18
	This study	0	0	0	0	0	0	0	0
	Total	86	61	20	1	4	0	62	24
DYS456	Lin et al., (2020)	3	1	2	0	0	0	1	2
	Ambrosio et al., (2020)	3	1	2	0	0	0	1	2
	Yang et al., 2018	1	1	0	0	0	0	1	0
	Antão-Sousa, (2017)*	43	25	18	0	0	0	25	18
	This study	2	1	1	0	0	0	1	1

	Total	52	29	23	0	0	0	29	23
YGATA H4	Lin et al., (2020)	2	1	1	0	0	0	1	1
	Ambrosio et al., (2020)	0	0	0	0	0	0	0	0
	Yang et al., 2018	1	1	0	0	0	0	1	0
	Antão-Sousa, (2017)*	26	10	15	0	1	0	10	16
	This study	1	0	1	0	0	0	0	1
	Total	30	12	17	0	1	0	12	18
Total		898	451	417	8	21	1	459	438

Supplementary Table 4- X-STR haplotypes for the linkage group 1 of the analysed unrelated females (N=129). ID-haplotype identification number.

ID	DXS10148	DXS10135	DXS8378	N	Freq.	s.d.	ID	DXS10148	DXS10135	DXS8378	N	Freq.	s.d.
h1	24.1	27	13	1	0.0039	0.0039	h105	22	22	12	1	0.0039	0.0039
h2	27.1	26	12	1	0.0039	0.0039	h106	23	19	12	2	0.0078	0.0055
h3	18	19	10	1	0.0039	0.0039	h107	25.1	23	11	1	0.0039	0.0039
h4	26.1	17	11	1	0.0039	0.0039	h108	26.1	15	12	1	0.0039	0.0039
h5	26.1	19.1	10	1	0.0039	0.0039	h109	29.1	24	12	1	0.0039	0.0039
h6	17	22	11	1	0.0039	0.0039	h110	23	23	10	2	0.0078	0.0055
h7	24.1	24	11	3	0.0116	0.0067	h111	25.1	23	10	1	0.0039	0.0039
h8	24.1	19	12	1	0.0039	0.0039	h112	24.1	29	12	1	0.0039	0.0039
h9	19	24	12	2	0.0078	0.0055	h113	26.1	25	12	1	0.0039	0.0039
h10	25.1	24	11	1	0.0039	0.0039	h114	18	23	10	1	0.0039	0.0039
h11	29.1	24	11	1	0.0039	0.0039	h115	25.1	23	12	1	0.0039	0.0039
h12	19	20.1	10	2	0.0078	0.0055	h116	20	21	10	1	0.0039	0.0039
h13	24.1	22	10	1	0.0039	0.0039	h117	27.1	19	11	1	0.0039	0.0039
h14	23	21.1	11	1	0.0039	0.0039	h118	30.1	20	10	1	0.0039	0.0039
h15	25.1	21	10	2	0.0078	0.0055	h119	27.1	26	10	1	0.0039	0.0039
h16	19	28	10	1	0.0039	0.0039	h120	23.1	18	12	1	0.0039	0.0039
h17	18	24	11	2	0.0078	0.0055	h121	29.1	21	12	1	0.0039	0.0039
h18	24.1	27	11	1	0.0039	0.0039	h122	23.1	21	10	1	0.0039	0.0039
h19	27.1	22	10	3	0.0116	0.0067	h123	25.1	16	12	1	0.0039	0.0039
h20	27.1	22	11	1	0.0039	0.0039	h124	23.1	29	12	1	0.0039	0.0039
h21	23.1	33	11	1	0.0039	0.0039	h125	25.1	24	10	1	0.0039	0.0039
h22	26.1	22.1	10	1	0.0039	0.0039	h126	24.1	21	10	1	0.0039	0.0039
h23	22.1	34	11	1	0.0039	0.0039	h127	24.1	21.1	12	2	0.0078	0.0055
h24	26.1	24	13	1	0.0039	0.0039	h128	13.3	29	12	1	0.0039	0.0039

h25	27.1	21	12	1	0.0039	0.0039	h129	18	24	10	1	0.0039	0.0039
h26	18	19.1	10	3	0.0116	0.0067	h130	25.1	27	12	1	0.0039	0.0039
h27	24.1	19.1	10	1	0.0039	0.0039	h131	23.1	26	13	1	0.0039	0.0039
h28	23.1	23	10	1	0.0039	0.0039	h132	27.1	21	10	2	0.0078	0.0055
h29	40.1	19.1	10	1	0.0039	0.0039	h133	27.1	17	11	1	0.0039	0.0039
h30	24.1	28	12	1	0.0039	0.0039	h134	26.1	21	11	1	0.0039	0.0039
h31	24	22	11	1	0.0039	0.0039	h135	23	24	13	1	0.0039	0.0039
h32	26.1	19	10	6	0.0233	0.0094	h136	23.1	28	11	1	0.0039	0.0039
h33	24	21	10	1	0.0039	0.0039	h137	23.1	32	12	1	0.0039	0.0039
h34	25.1	21	12	2	0.0078	0.0055	h138	24.1	23	10	1	0.0039	0.0039
h35	27.1	24	13	1	0.0039	0.0039	h139	18	27	10	1	0.0039	0.0039
h36	25.1	30	11	2	0.0078	0.0055	h140	23.1	27	11	1	0.0039	0.0039
h37	24	21	12	1	0.0039	0.0039	h141	26.1	28	12	1	0.0039	0.0039
h38	24.1	23.1	10	1	0.0039	0.0039	h142	23	30	10	1	0.0039	0.0039
h39	27.1	23	12	1	0.0039	0.0039	h143	23.1	25	12	1	0.0039	0.0039
h40	25.1	27	11	2	0.0078	0.0055	h144	26.1	20	10	1	0.0039	0.0039
h41	25.1	19	10	1	0.0039	0.0039	h145	18	19	12	2	0.0078	0.0055
h42	26.1	24	10	2	0.0078	0.0055	h146	25.1	20.1	12	1	0.0039	0.0039
h43	28.1	18	11	1	0.0039	0.0039	h147	26.1	19	11	1	0.0039	0.0039
h44	23	24	11	3	0.0116	0.0067	h148	30.1	22	12	1	0.0039	0.0039
h45	27.1	23	11	3	0.0116	0.0067	h149	24	31	11	1	0.0039	0.0039
h46	27.1	24	11	2	0.0078	0.0055	h150	30.1	20.1	10	1	0.0039	0.0039
h47	22.1	21	12	1	0.0039	0.0039	h151	18	17.1	12	1	0.0039	0.0039
h48	25.1	22	12	2	0.0078	0.0055	h152	23	32	10	1	0.0039	0.0039
h49	25.1	19	11	1	0.0039	0.0039	h153	18	21	12	2	0.0078	0.0055
h50	29.1	28	11	1	0.0039	0.0039	h154	28.1	21	11	1	0.0039	0.0039

h51	28.1	18	10	1	0.0039	0.0039	h155	26.1	19	12	1	0.0039	0.0039
h52	24.1	18	12	2	0.0078	0.0055	h156	27.1	25	11	1	0.0039	0.0039
h53	22.1	23	10	1	0.0039	0.0039	h157	24.1	26	9	1	0.0039	0.0039
h54	25.1	26	11	4	0.0155	0.0077	h158	24.1	27	12	1	0.0039	0.0039
h55	26.1	21	12	1	0.0039	0.0039	h159	29.1	22	10	1	0.0039	0.0039
h56	27.1	26	11	1	0.0039	0.0039	h160	18	30	10	2	0.0078	0.0055
h57	19	21.1	11	1	0.0039	0.0039	h161	23	17	12	1	0.0039	0.0039
h58	29.1	27	11	1	0.0039	0.0039	h162	17	25	11	1	0.0039	0.0039
h59	26.1	20.1	10	1	0.0039	0.0039	h163	22	27	12	1	0.0039	0.0039
h60	29.1	24.1	10	1	0.0039	0.0039	h164	18	23	11	2	0.0078	0.0055
h61	27.1	25	10	1	0.0039	0.0039	h165	18	25	12	1	0.0039	0.0039
h62	18	18	12	1	0.0039	0.0039	h166	26.1	22	10	1	0.0039	0.0039
h63	25.1	25	12	1	0.0039	0.0039	h167	27.1	18	11	1	0.0039	0.0039
h64	26.1	22	11	1	0.0039	0.0039	h168	18	20	10	1	0.0039	0.0039
h65	24.1	20	10	2	0.0078	0.0055	h169	22.1	25	14	1	0.0039	0.0039
h66	23	18	12	1	0.0039	0.0039	h170	26.1	25	9	1	0.0039	0.0039
h67	19	26	10	1	0.0039	0.0039	h171	24.1	31	12	1	0.0039	0.0039
h68	28.1	23	10	1	0.0039	0.0039	h172	24.1	35	10	1	0.0039	0.0039
h69	22.1	30	11	1	0.0039	0.0039	h173	27.1	18	10	1	0.0039	0.0039
h70	23.1	24	11	1	0.0039	0.0039	h174	27.1	30	12	1	0.0039	0.0039
h71	24.1	25	12	1	0.0039	0.0039	h175	23.1	31	10	1	0.0039	0.0039
h72	26.1	18	10	2	0.0078	0.0055	h176	29.1	23	12	1	0.0039	0.0039
h73	26.1	31	11	1	0.0039	0.0039	h177	25.1	22	11	1	0.0039	0.0039
h74	18	23	12	1	0.0039	0.0039	h178	30	17.1	11	1	0.0039	0.0039
h75	25.1	28	11	3	0.0116	0.0067	h179	24.1	26	11	1	0.0039	0.0039
h76	23	22	10	1	0.0039	0.0039	h180	27.1	28	11	1	0.0039	0.0039

h77	23.1	19	10	1	0.0039	0.0039	h181	14	26	11	1	0.0039	0.0039
h78	23.1	25	10	1	0.0039	0.0039	h182	27.1	17	13	1	0.0039	0.0039
h79	23	20	10	1	0.0039	0.0039	h183	18	20	12	1	0.0039	0.0039
h80	18	25	10	3	0.0116	0.0067	h184	26.1	21	10	1	0.0039	0.0039
h81	23.1	22	10	1	0.0039	0.0039	h185	23.1	26	11	2	0.0078	0.0055
h82	18	21	10	2	0.0078	0.0055	h186	24	17.1	12	1	0.0039	0.0039
h83	26.1	28	13	1	0.0039	0.0039	h187	25.1	18	13	1	0.0039	0.0039
h84	26.1	17	10	2	0.0078	0.0055	h188	25.1	21	9	1	0.0039	0.0039
h85	24.1	26	12	1	0.0039	0.0039	h189	27.1	30	10	1	0.0039	0.0039
h86	19	24	11	1	0.0039	0.0039	h190	23	24	12	1	0.0039	0.0039
h87	27.1	20	11	2	0.0078	0.0055	h191	29.1	18	12	1	0.0039	0.0039
h88	23.1	20	10	1	0.0039	0.0039	h192	27.1	15	12	1	0.0039	0.0039
h89	24.1	24	12	1	0.0039	0.0039	h193	24.1	22.1	10	1	0.0039	0.0039
h90	26.1	18	11	1	0.0039	0.0039	h194	18	20	11	1	0.0039	0.0039
h91	24.1	29	11	1	0.0039	0.0039	h195	23.1	19.1	11	1	0.0039	0.0039
h92	18	26	12	1	0.0039	0.0039	h196	28.1	20	12	1	0.0039	0.0039
h93	25.1	29	11	2	0.0078	0.0055	h197	18	26	10	1	0.0039	0.0039
h94	29.1	19.1	10	1	0.0039	0.0039	h198	24.1	21	12	1	0.0039	0.0039
h95	28.1	19	10	1	0.0039	0.0039	h199	27.1	19	10	1	0.0039	0.0039
h96	27.1	19	12	3	0.0116	0.0067	h200	28.1	25	12	1	0.0039	0.0039
h97	18	18	10	1	0.0039	0.0039	h201	20	23	11	1	0.0039	0.0039
h98	18	29	11	1	0.0039	0.0039	h202	18	25	11	1	0.0039	0.0039
h99	27.1	36.1	11	1	0.0039	0.0039	h203	22.1	25	12	1	0.0039	0.0039
h100	25.1	17	11	1	0.0039	0.0039	h204	22.1	22	9	1	0.0039	0.0039
h101	23	27	12	1	0.0039	0.0039	h205	18	27	11	1	0.0039	0.0039
h102	24	22	10	1	0.0039	0.0039	h206	23	20	11	1	0.0039	0.0039

h103	29.1	21	9	1	0.0039	0.0039	h207	28.1	17	10	1	0.0039	0.0039
h104	25.1	24	12	1	0.0039	0.0039	h208	28.1	25	11	1	0.0039	0.0039

Supplementary Table 5- X-STR haplotypes for the linkage group 2 of the analysed unrelated females (N=129). ID-haplotype identification number.

ID	DXS7132	DXS10079	DXS10074	N	Freq.	s.d.	ID	DXS7132	DXS10079	DXS10074	N	Freq.	s.d.
1	12	19	15	1	0.0039	0.0039	75	15	18	17	1	0.0039	0.0039
2	15	17	8	1	0.0039	0.0039	76	14	17	19	1	0.0039	0.0039
3	12	17	8	2	0.0078	0.0055	77	15	18	16	1	0.0039	0.0039
4	14	20	16	7	0.0271	0.0101	78	12	20	8	2	0.0078	0.0055
5	13	21	15	2	0.0078	0.0055	79	14	18	15	1	0.0039	0.0039
6	14	22	15	1	0.0039	0.0039	80	16	21	16	1	0.0039	0.0039
7	14	21	16	3	0.0116	0.0067	81	12	22	14	1	0.0039	0.0039
8	11	19	17	1	0.0039	0.0039	82	15	19	10	1	0.0039	0.0039
9	14	22	8	1	0.0039	0.0039	83	14	20	7	3	0.0116	0.0067
10	15	21	17	1	0.0039	0.0039	84	14	21	15	2	0.0078	0.0055
11	14	20	13	2	0.0078	0.0055	85	15	22	8	2	0.0078	0.0055
12	14	21	14	1	0.0039	0.0039	86	15	17	15	1	0.0039	0.0039
13	13	15	16	2	0.0078	0.0055	87	11	20	14	1	0.0039	0.0039
14	13	18	17	1	0.0039	0.0039	88	12	19	19	1	0.0039	0.0039
15	16	21	17	1	0.0039	0.0039	89	14	20	17	2	0.0078	0.0055
16	12	20	16	4	0.0155	0.0077	90	16	22	17	1	0.0039	0.0039
17	13	20	18	4	0.0155	0.0077	91	13	18	19	1	0.0039	0.0039
18	14	23	13	1	0.0039	0.0039	92	14	21	18	2	0.0078	0.0055
19	12	19	16	2	0.0078	0.0055	93	15	21	8	1	0.0039	0.0039
20	13	18	8	6	0.0233	0.0094	94	12	22	19	2	0.0078	0.0055
21	12	19	18	4	0.0155	0.0077	95	16	19	17	1	0.0039	0.0039
22	15	20	8	2	0.0078	0.0055	96	14	18	17	3	0.0116	0.0067
23	13	18	18	2	0.0078	0.0055	97	14	20	15	1	0.0039	0.0039
24	14	17	8	3	0.0116	0.0067	98	13	21	17	4	0.0155	0.0077
25	15	20	18	2	0.0078	0.0055	99	14	19	15	4	0.0155	0.0077
26	14	20	18	3	0.0116	0.0067	100	15	17	14	1	0.0039	0.0039
27	12	14	16	2	0.0078	0.0055	101	11	16	17	1	0.0039	0.0039
28	11	20	18	1	0.0039	0.0039	102	11	17	8	1	0.0039	0.0039

29	14	22	18	2	0.0078	0.0055	103	13	15	17	1	0.0039	0.0039
30	13	19	8	4	0.0155	0.0077	104	14	18	18	1	0.0039	0.0039
31	15	19	8	3	0.0116	0.0067	105	14	21	13	1	0.0039	0.0039
32	12	18	19	2	0.0078	0.0055	106	13	18	16	2	0.0078	0.0055
33	14	19	8	2	0.0078	0.0055	107	13	20	19	1	0.0039	0.0039
34	13	20	7	3	0.0116	0.0067	108	14	18	16	4	0.0155	0.0077
35	14	19	16	4	0.0155	0.0077	109	13	17	8	1	0.0039	0.0039
36	15	22	18	2	0.0078	0.0055	110	13	21	16	2	0.0078	0.0055
37	14	21	17	2	0.0078	0.0055	111	14	17	18	1	0.0039	0.0039
38	15	18	19	1	0.0039	0.0039	112	12	20	20	1	0.0039	0.0039
39	13	20	17	3	0.0116	0.0067	113	12	18	17	1	0.0039	0.0039
40	15	15	16	1	0.0039	0.0039	114	12	18	16	1	0.0039	0.0039
41	13	19	16	8	0.0310	0.0108	115	15	22	20	1	0.0039	0.0039
42	15	15	18	1	0.0039	0.0039	116	14	21	19	1	0.0039	0.0039
43	14	15	19	2	0.0078	0.0055	117	13	22	18	2	0.0078	0.0055
44	13	19	18	3	0.0116	0.0067	118	12	18	15	1	0.0039	0.0039
45	15	19	18	1	0.0039	0.0039	119	14	21	8	1	0.0039	0.0039
46	14	16	16	1	0.0039	0.0039	120	12	18	8	1	0.0039	0.0039
47	13	22	8	1	0.0039	0.0039	121	12	18	7	1	0.0039	0.0039
48	14	17	16	1	0.0039	0.0039	122	14	14	19	1	0.0039	0.0039
49	14	24	17	1	0.0039	0.0039	123	16	18	18	1	0.0039	0.0039
50	13	19	15	3	0.0116	0.0067	124	14	18	8	1	0.0039	0.0039
51	14	19	13	1	0.0039	0.0039	125	14	14	18	1	0.0039	0.0039
52	12	20	17	2	0.0078	0.0055	126	18	16	18	1	0.0039	0.0039
53	13	20	16	2	0.0078	0.0055	127	15	16	17	1	0.0039	0.0039
54	15	17	17	1	0.0039	0.0039	128	14	19	19	1	0.0039	0.0039
55	14	19	18	5	0.0194	0.0086	129	14	19	9	1	0.0039	0.0039
56	13	17	17	1	0.0039	0.0039	130	14	15	18	1	0.0039	0.0039
57	14	19	17	9	0.0349	0.0114	131	13	19	19	1	0.0039	0.0039
58	14	15	13	1	0.0039	0.0039	132	12	20	13	1	0.0039	0.0039
59	15	20	17	3	0.0116	0.0067	133	13	23	16	1	0.0039	0.0039

60	14	20	8	4	0.0155	0.0077	134	13	18	7	1	0.0039	0.0039
61	11	20	17	1	0.0039	0.0039	135	14	21	7	1	0.0039	0.0039
62	16	15	17	1	0.0039	0.0039	136	12	21	14	1	0.0039	0.0039
63	14	16	8	1	0.0039	0.0039	137	13	22	17	1	0.0039	0.0039
64	13	15	18	3	0.0116	0.0067	138	13	23	17	1	0.0039	0.0039
65	13	19	14	1	0.0039	0.0039	139	16	17	17	1	0.0039	0.0039
66	15	18	15	1	0.0039	0.0039	140	15	19	16	1	0.0039	0.0039
67	13	19	7	3	0.0116	0.0067	141	13	18	20	1	0.0039	0.0039
68	12.3	18	17	1	0.0039	0.0039	142	12	22	17	1	0.0039	0.0039
69	15	21	16	1	0.0039	0.0039	143	14	21	9	1	0.0039	0.0039
70	15	19	15	1	0.0039	0.0039	144	15	19	17	1	0.0039	0.0039
71	13	19	17	3	0.0116	0.0067	145	14	17	14	1	0.0039	0.0039
72	15	19	14	1	0.0039	0.0039	146	15	22	17	1	0.0039	0.0039
73	13	22	15	1	0.0039	0.0039	147	16	21	15	1	0.0039	0.0039
74	15	19	19	1	0.0039	0.0039							

Supplementary Table 6- X-STR haplotypes for the linkage group 3 of the analysed unrelated females (N=129). ID-Haplotype identification number.

ID	DXS10103	HPRTB	DXS10101	N	Freq.	s.d.	ID	DXS10103	HPRTB	DXS10101	N	Freq.	s.d.
1	18	12	27.2	3	0.0116	0.0067	65	17	12	31	3	0.0116	0.0067
2	16	13	30.2	2	0.0078	0.0055	66	19	13	33	1	0.0039	0.0039
3	18	12	24.2	1	0.0039	0.0039	67	19	12	28	5	0.0194	0.0086
4	20	13	32	1	0.0039	0.0039	68	18	12	32.2	1	0.0039	0.0039
5	17	15	30	2	0.0078	0.0055	69	16	14	28	1	0.0039	0.0039
6	20	11	29.2	4	0.0155	0.0077	70	20	12	31.2	1	0.0039	0.0039
7	19	12	32.2	3	0.0116	0.0067	71	19	12	26.2	1	0.0039	0.0039
8	19	13	30.2	10	0.0388	0.012	72	18	13	26.2	1	0.0039	0.0039
9	19	12	null	1	0.0039	0.0039	73	17	12	30	1	0.0039	0.0039
10	19	13	32.2	1	0.0039	0.0039	74	20	11	31	2	0.0078	0.0055
11	19	12	31.2	9	0.0349	0.0114	75	19	13	29	2	0.0078	0.0055
12	20	14	31.2	3	0.0116	0.0067	76	15	14	32	2	0.0078	0.0055
13	20	12	27.2	5	0.0194	0.0086	77	19	13	29.2	4	0.0155	0.0077
14	19	12	28.2	5	0.0194	0.0086	78	19	13	33.2	1	0.0039	0.0039
15	16	12	28.2	3	0.0116	0.0067	79	16	11	31	1	0.0039	0.0039
16	18	14	30	2	0.0078	0.0055	80	17	13	32	4	0.0155	0.0077
17	18	14	33	1	0.0039	0.0039	81	19	14	31	2	0.0078	0.0055
18	18	14	32	2	0.0078	0.0055	82	20	13	29.2	2	0.0078	0.0055
19	17	16	31	2	0.0078	0.0055	83	17	13	33	2	0.0078	0.0055
20	18	13	30	2	0.0078	0.0055	84	19	11	29.2	6	0.0233	0.0094
21	18	11	28.2	3	0.0116	0.0067	85	20	14	29	1	0.0039	0.0039
22	20	12	29	1	0.0039	0.0039	86	19	14	33.2	1	0.0039	0.0039
23	20	14	31	1	0.0039	0.0039	87	16	13	29.2	3	0.0116	0.0067
24	19	14	30.2	1	0.0039	0.0039	88	19	11	26.2	1	0.0039	0.0039
25	19	15	30.2	2	0.0078	0.0055	89	21	12	32.2	1	0.0039	0.0039
26	18	9	31	1	0.0039	0.0039	90	18	11	27.2	2	0.0078	0.0055
27	19	12	29	4	0.0155	0.0077	91	18	11	25.2	1	0.0039	0.0039
28	19	13	31.2	3	0.0116	0.0067	92	19	14	29.2	3	0.0116	0.0067

29	17	13	31	4	0.0155	0.0077	93	20	13	28.2	2	0.0078	0.0055
30	19	14	32.2	1	0.0039	0.0039	94	16	12	32	1	0.0039	0.0039
31	20	12	26.2	1	0.0039	0.0039	95	19	11	27.2	2	0.0078	0.0055
32	19	11	31.2	5	0.0194	0.0086	96	19	11	32.2	1	0.0039	0.0039
33	17	13	29.2	1	0.0039	0.0039	97	16	12	27.2	1	0.0039	0.0039
34	18	13	31.2	1	0.0039	0.0039	98	18	11	31.2	1	0.0039	0.0039
35	19	11	30.2	7	0.0271	0.0101	99	16	14	31	2	0.0078	0.0055
36	16	11	28.2	1	0.0039	0.0039	100	19	12	33.2	1	0.0039	0.0039
37	20	13	30.2	6	0.0233	0.0094	101	19	13	28	1	0.0039	0.0039
38	17	13	30	1	0.0039	0.0039	102	20	12	29.2	4	0.0155	0.0077
39	18	13	27.2	1	0.0039	0.0039	103	16	13	28.2	1	0.0039	0.0039
40	17	15	31	3	0.0116	0.0067	104	18	12	29.2	2	0.0078	0.0055
41	19	14	27.2	1	0.0039	0.0039	105	18	11	26.2	1	0.0039	0.0039
42	16	12	25.2	1	0.0039	0.0039	106	16	13	31.2	1	0.0039	0.0039
43	16	14	33	3	0.0116	0.0067	107	18	14	35	1	0.0039	0.0039
44	17	16	32	1	0.0039	0.0039	108	18	13	34	1	0.0039	0.0039
45	19	14	32	2	0.0078	0.0055	109	18	12	31.2	1	0.0039	0.0039
46	19	12	27	1	0.0039	0.0039	110	19	12	27.2	1	0.0039	0.0039
47	19	12	32	3	0.0116	0.0067	111	16	13	32	1	0.0039	0.0039
48	19	11	28.2	2	0.0078	0.0055	112	16	10	28.2	1	0.0039	0.0039
49	21	13	30.2	2	0.0078	0.0055	113	20	16	29.2	1	0.0039	0.0039
50	18	11	29.2	2	0.0078	0.0055	114	19	14	31.2	2	0.0078	0.0055
51	18	9	32.2	1	0.0039	0.0039	115	15	14	33	1	0.0039	0.0039
52	19	12	29.2	7	0.0271	0.0101	116	20	12	24.2	1	0.0039	0.0039
53	19	12	30.2	5	0.0194	0.0086	117	16	13	31	2	0.0078	0.0055
54	19	13	31	1	0.0039	0.0039	118	18	12	28	1	0.0039	0.0039
55	20	14	33.2	1	0.0039	0.0039	119	16	14	28.2	1	0.0039	0.0039
56	17	12	26.2	1	0.0039	0.0039	120	20	12	30.2	1	0.0039	0.0039
57	17	12	32	2	0.0078	0.0055	121	18	12	31	1	0.0039	0.0039
58	18	11	34	1	0.0039	0.0039	122	19	10	28.2	1	0.0039	0.0039
59	15	13	33	1	0.0039	0.0039	123	20	12	32.2	1	0.0039	0.0039

60	16	12	29.2	2	0.0078	0.0055	124	20	11	30.2	2	0.0078	0.0055
61	19	13	30	2	0.0078	0.0055	125	20	14	30.2	1	0.0039	0.0039
62	20	13	26.2	1	0.0039	0.0039	126	16	13	33	1	0.0039	0.0039
63	19	12	31	2	0.0078	0.0055	127	17	15	32	1	0.0039	0.0039
64	18	13	29.2	1	0.0039	0.0039	128	18	12	30.2	1	0.0039	0.0039

Supplementary Table 7- X-STR haplotypes for the linkage group 4 of the analysed unrelated females (N=129). ID- Haplotype identification number.

ID	DXS10146	DXS10134	DXS7423	N	Freq.	s.d.	ID	DXS10146	DXS10134	DXS7423	N	Freq.	s.d.
1	28	37	13	1	0.0039	0.0039	87	26	38.3	15	2	0.0078	0.0055
2	31	35	14	1	0.0039	0.0039	88	46.2	40.3	13	1	0.0039	0.0039
3	28	37	14	3	0.0116	0.0067	89	45.2	39	14	1	0.0039	0.0039
4	28	36	14	9	0.0349	0.0114	90	42.2	41.3	16	1	0.0039	0.0039
5	31	42.3	15	1	0.0039	0.0039	91	41.2	35	16	1	0.0039	0.0039
6	29	35	15	5	0.0194	0.0086	92	26	35	15	2	0.0078	0.0055
7	28	35	15	4	0.0155	0.0077	93	44.2	37	14	3	0.0116	0.0067
8	39.2	42.3	16	2	0.0078	0.0055	94	29	35	17	1	0.0039	0.0039
9	39.2	39.3	16	1	0.0039	0.0039	95	27	38.0	15	2	0.0078	0.0055
10	30	37	14	2	0.0078	0.0055	96	29	33	16	2	0.0078	0.0055
11	27	34	15	2	0.0078	0.0055	97	40.2	40.3	16	1	0.0039	0.0039
12	27	37	14	7	0.0271	0.0101	98	38.2	41.3	15	1	0.0039	0.0039
13	27	35	15	1	0.0039	0.0039	99	31	32	13	1	0.0039	0.0039
14	28	37	15	4	0.0155	0.0077	100	40.2	36	15	2	0.0078	0.0055
15	29	36	13	1	0.0039	0.0039	101	41.2	35	15	1	0.0039	0.0039
16	43.2	35	14	1	0.0039	0.0039	102	34.2	31	16	1	0.0039	0.0039
17	25	36	15	3	0.0116	0.0067	103	28	35	14	1	0.0039	0.0039
18	41.2	36	15	1	0.0039	0.0039	104	29	35.2	13	1	0.0039	0.0039
19	30	33	15	2	0.0078	0.0055	105	26	36	14	2	0.0078	0.0055
20	41.2	43.3	16	1	0.0039	0.0039	106	26	35	14	2	0.0078	0.0055
21	27	37	16	1	0.0039	0.0039	107	39.2	35	16	1	0.0039	0.0039
22	25	37	13	1	0.0039	0.0039	108	25	36	14	1	0.0039	0.0039
23	30	35	15	2	0.0078	0.0055	109	29	34	16	2	0.0078	0.0055
24	42.2	44.3	15	1	0.0039	0.0039	110	43.2	41.3	16	1	0.0039	0.0039
25	26	39.3	16	1	0.0039	0.0039	111	29	36	15	2	0.0078	0.0055
26	27	36	14	2	0.0078	0.0055	112	47.2	33	14	1	0.0039	0.0039
27	28	34	16	2	0.0078	0.0055	113	42.2	37	14	1	0.0039	0.0039
28	43.2	36	15	1	0.0039	0.0039	114	27	39	16	1	0.0039	0.0039

29	40.2	35.2	15	1	0.0039	0.0039	115	39.2	41.3	16	2	0.0078	0.0055
30	28	38.0	15	2	0.0078	0.0055	116	29	37.2	15	1	0.0039	0.0039
31	27	38.0	14	2	0.0078	0.0055	117	30	33	16	2	0.0078	0.0055
32	27	35	14	4	0.0155	0.0077	118	25	39.3	15	1	0.0039	0.0039
33	29	34	14	3	0.0116	0.0067	119	26	37	15	2	0.0078	0.0055
34	44.2	41.3	14	1	0.0039	0.0039	120	26	39.3	15	1	0.0039	0.0039
35	26	42.1	15	1	0.0039	0.0039	121	30	34	16	2	0.0078	0.0055
36	26	38.0	14	1	0.0039	0.0039	122	30	43.3	14	1	0.0039	0.0039
37	28	38.0	14	1	0.0039	0.0039	123	28	33	16	1	0.0039	0.0039
38	28	34	14	2	0.0078	0.0055	124	28	43.3	16	1	0.0039	0.0039
39	39.2	42.3	17	1	0.0039	0.0039	125	26	36	16	2	0.0078	0.0055
40	31	39.3	15	1	0.0039	0.0039	126	25	38.0	14	1	0.0039	0.0039
41	30	39.3	14	1	0.0039	0.0039	127	27	36	15	4	0.0155	0.0077
42	28	36	15	2	0.0078	0.0055	128	43.2	37	15	1	0.0039	0.0039
43	29	40.3	16	1	0.0039	0.0039	129	39.2	38.3	15	1	0.0039	0.0039
44	37.2	39	15	1	0.0039	0.0039	130	39.2	41.3	15	1	0.0039	0.0039
45	40.2	42.3	16	1	0.0039	0.0039	131	27	34	14	2	0.0078	0.0055
46	29	33	14	2	0.0078	0.0055	132	42.2	36	15	2	0.0078	0.0055
47	41.2	39	14	2	0.0078	0.0055	133	45.2	37	15	1	0.0039	0.0039
48	29	37	15	4	0.0155	0.0077	134	44.2	35	16	1	0.0039	0.0039
49	39.2	34	13	1	0.0039	0.0039	135	42.2	35	14	1	0.0039	0.0039
50	26	37.3	14	1	0.0039	0.0039	136	40.2	42.3	15	1	0.0039	0.0039
51	29	39.3	15	1	0.0039	0.0039	137	43.2	35	16	1	0.0039	0.0039
52	28	32	16	2	0.0078	0.0055	138	33	30	15	1	0.0039	0.0039
53	31	35	16	2	0.0078	0.0055	139	24	37	14	1	0.0039	0.0039
54	43.2	38.3	16	1	0.0039	0.0039	140	39.2	42.3	14	1	0.0039	0.0039
55	31	35	15	2	0.0078	0.0055	141	26	37	16	1	0.0039	0.0039
56	45.2	33	14	1	0.0039	0.0039	142	29	35	14	1	0.0039	0.0039
57	30	34	14	1	0.0039	0.0039	143	40.2	38.0	16	1	0.0039	0.0039
58	29	38.0	16	1	0.0039	0.0039	144	26	39	14	1	0.0039	0.0039
59	27	35	16	1	0.0039	0.0039	145	41.2	36.2	14	1	0.0039	0.0039

60	42.2	35	16	1	0.0039	0.0039	146	45.2	35	16	1	0.0039	0.0039
61	30	36	15	4	0.0155	0.0077	147	28	35	16	1	0.0039	0.0039
62	41.2	37.2	14	1	0.0039	0.0039	148	44.2	38.3	14	1	0.0039	0.0039
63	26	34	13	1	0.0039	0.0039	149	27	39	14	1	0.0039	0.0039
64	40.2	39.3	16	1	0.0039	0.0039	150	30	35	16	1	0.0039	0.0039
65	27	37	15	1	0.0039	0.0039	151	41.2	37	13	1	0.0039	0.0039
66	23	36	14	1	0.0039	0.0039	152	24	29.1	14	1	0.0039	0.0039
67	32.2	34	17	1	0.0039	0.0039	153	30	34	15	2	0.0078	0.0055
68	30	36	14	4	0.0155	0.0077	154	39.2	39.3	14	1	0.0039	0.0039
69	43.2	35	15	1	0.0039	0.0039	155	31	37	15	1	0.0039	0.0039
70	28	38.0	16	1	0.0039	0.0039	156	40.2	35	15	1	0.0039	0.0039
71	28	33	15	2	0.0078	0.0055	157	26	36	15	1	0.0039	0.0039
72	29	36	16	2	0.0078	0.0055	158	45.2	36	14	1	0.0039	0.0039
73	30	43.3	15	1	0.0039	0.0039	159	25	35	13	1	0.0039	0.0039
74	29	37	14	2	0.0078	0.0055	160	26	35	16	1	0.0039	0.0039
75	29.1	34	14	1	0.0039	0.0039	161	43.2	37	14	1	0.0039	0.0039
76	43.2	37	16	1	0.0039	0.0039	162	26	32	14	1	0.0039	0.0039
77	30	37	15	2	0.0078	0.0055	163	29	35	16	2	0.0078	0.0055
78	32	36	14	1	0.0039	0.0039	164	28	34	15	1	0.0039	0.0039
79	29	37	17	1	0.0039	0.0039	165	29	36	14	1	0.0039	0.0039
80	27	38.0	16	2	0.0078	0.0055	166	31	30	14	1	0.0039	0.0039
81	25	35	15	1	0.0039	0.0039	167	31	36	15	1	0.0039	0.0039
82	28	36	16	2	0.0078	0.0055	168	39.2	37	14	1	0.0039	0.0039
83	40.2	36	13	1	0.0039	0.0039	169	26	41.3	16	1	0.0039	0.0039
84	44.2	37	16	1	0.0039	0.0039	170	29	43.3	15	1	0.0039	0.0039
85	29	39.3	14	1	0.0039	0.0039	171	41.2	40	14	1	0.0039	0.0039
86	24	35	15	1	0.0039	0.0039							

Supplementary Table 8- X-STRs specific mutation rate estimates based on data collected from this study (N=135 mother-father-daughter trios), and from the literature.

	Reference	Total			
		No. Allele transfers	No. Mutations	Mutation Rate (x 10 ⁻³)	95 % CI (x 10 ⁻³)
HPRTB	Athanasiadou et al., (2003)	66	0	0.0	0-54.4
	Turrina et al., (2004)	80	0	0.0	0-45.1
	Turrina et al. (2007)	80	0	0.0	0-45.1
	Tariq et al., (2008)	100	0	0.0	0-36.2
	Pico et al., (2008)	102	0	0.0	0-35.5
	Becker et al., (2008)	248	0	0.0	0-14.8
	Tetzlaff et al. (2012)	210	1	4.8	0.1-26.2
	Diegoli et al., (2014)	834	1	1.2	0-6.7
	Chen et al., (2020)	116	0	0.0	0-31.3
	Tabbada et al., (2005)	62	0	0.0	0-57.8
	Pinto et al., (2020)	3224	5	1.6	0.5-3.6
	This study	270	0	0.0	0-13.6
	Total	5392	7	1.3	0.5-2.7
DXS7132	Turrina et al. (2004)	80	0	0.0	0-45.1
	Turrina et al. (2007)	80	0	0.0	0-45.1
	Tariq et al., (2008)	100	0	0.0	0-36.2
	Pico et al., (2008)	102	1	9.8	0.2-53.4
	Liu et al., (2008)	588	3	5.1	1.1-14.8
	Becker et al., (2008)	248	1	4.0	0.1-22.3
	Tetzlaff et al. (2012)	210	1	4.8	0.1-26.2
	Liu et al., (2012)	620	3	4.8	1-14.1
	Nishi et al., (2013)	40	0	0.0	0-88.1
	Diegoli et al., (2014)	834	1	1.2	0-6.7
	Chen et al., (2020)	116	0	0.0	0-31.3
	Pinto et al., (2020)	3224	19	5.9	3.6-9.2
	This study	270	3	11.1	2.3-32.1
	Total	6512	32	4.9	3.4-6.9
DXS8378	Turrina et al. (2007)	80	0	0.0	0-45.1
	Tariq et al., (2008)	100	0	0.0	0-36.2
	Becker et al., (2008)	248	0	0.0	0-14.8
	Tetzlaff et al. (2012)	210	1	4.8	0.1-26.2
	Nishi et al., (2013)	40	0	0.0	0-88.1
	Diegoli et al., (2014)	834	2	2.4	0.3-8.6
	Chen et al., (2020)	116	0	0.0	0-31.3
	Pinto et al., (2020)	3224	2	0.6	0.1-2.2
	This study	270	0	0.0	0-13.6

	Total	5122	5	1.0	0.3-2.3
DXS7423	Turrina et al. (2007)	80	0	0.0	0-45.1
	Tariq et al., (2008)	100	0	0.0	0-36.2
	Becker et al., (2008)	248	0	0.0	0-14.8
	Tetzlaff et al. (2012)	210	0	0.0	0-17.4
	Nishi et al., (2013)	40	0	0.0	0-88.1
	Diegoli et al., (2014)	834	0	0.0	0-4.4
	Pinto et al., (2020)	3224	3	0.9	0.2-2.7
	This study	270	1	3.7	0.1-20.5
	Total	5006	4	0.8	0.2-2
DXS10135	Becker et al., (2008)	248	0	0.0	0-14.8
	Tetzlaff et al. (2012)	210	2	4.8	0.1-26.2
	Pinto et al. (2020)	3224	35	10.9	7.6-15.1
	This study	270	2	7.4	0.9-26.5
	Total	3952	39	9.9	6.8-13.2
DXS10074	Becker et al., (2008)	248	0	0.0	0-14.8
	Tetzlaff et al. (2012)	210	1	4.8	0.1-26.2
	Liu et al., (2012)	620	1	1.6	0-9
	Nishi et al., (2013)	40	0	0.0	0-88.1
	Pinto et al., (2020)	3224	17	5.3	3.1-8.4
	This study	270	1	3.7	0.1-20.5
	Total	4612	20	4.3	2.7-6.7
DXS10101	Becker et al., (2008)	248	0	0.0	0-14.8
	Tetzlaff et al. (2012)	210	1	4.8	0.1-26.2
	Pinto et al., (2020)	3224	10	3.1	1.5-5.7
	This study	270	0	0.0	0-13.6
	Total	3952	11	2.8	1.4-5
DXS10134	Becker et al., (2008)	248	0	0.0	0-14.8
	Tetzlaff et al. (2012)	210	1	4.8	0.1-26.2
	Nishi et al., (2013)	40	0	0.0	0-88.1
	Chen et al., (2020)	116	0	0.0	0-31.3
	Pinto et al., (2020)	3224	16	5.0	2.8-8
	This study	270	0	0.0	0-13.6
	Total	4108	17	4.1	2.4-6.6
DXS10148	Pinto et al., (2020)	3224	14	4.3	2.4-7.3
	This study	270	0	0.0	0-13.6
	Total	3494	14	4.0	2.2-6.7
DXS10079	Liu et al., (2012)	620	2	3.2	0.4-11.6
	Pinto et al., (2020)	3224	20	6.2	3.8-9.6
	This study	270	3	11.1	2.3-32.1
	Total	4114	25	6.1	3.9-9

DXS10103	Pinto et al., (2020)	3224	6	1.9	0.7-4
	This study	270	1	3.7	0.1-20.5
	Total	3494	7	2.0	0.8-4.1
DXS10146	Pinto et al., (2020)	3224	9	2.8	1.3-5.3
	This study	270	4	14.8	4.1-37.5
	Total	3494	13	3.7	2-6.4
Total		53252	194	3.6	3.1-4.2

Supplementary Table 9- X-STRs specific mutation rates for male and female meiosis using data from this study (N=135 mother-father-daughter trios), and from the literature. Note: the number of allele transfers is the for each paternal and maternal meiosis.

Markers	Reference	Most Parsimonious Parental Origin							
		No. Allele transfers	Paternal			Maternal			Indeterminate
			No. Mutations	Mutation Rate (x 10 ⁻³)	95 % CI (x 10 ⁻³)	No. Mutations	Mutation Rate (x 10 ⁻³)	95 % CI (x 10 ⁻³)	No. Mutations
HPRTB	Athanasiadou et al., (2003)	33	0	0.0	0-105.8	0	0.0	0-105.8	0
	Turrina et al., (2004)	40	0	0.0	0-88.1	0	0.0	0-88.1	0
	Turrina et al. (2007)	40	0	0.0	0-88.1	0	0.0	0-88.1	0
	Tariq et al., (2008)	50	0	0.0	0-71.1	0	0.0	0-71.1	0
	Pico et al., (2008)	51	0	0.0	0-69.8	0	0.0	0-69.8	0
	Becker et al., (2008)	124	0	0.0	0-29.3	0	0.0	0-29.3	0
	Tetzlaff et al. (2012)	105	0	0.0	0-34.5	1	9.5	0.2-51.9	0
	Diegoli et al., (2014)	417	1	2.4	0.1-13.3	0	0.0	0-8.8	0
	Chen et al., (2020)	58	0	0.0	0-61.6	0	0.0	0-61.6	0
	Tabbada et al., (2005)	31	0	0.0	0-112.2	0	0.0	0-112.2	0
	Pinto et al., (2020)	1612	2	1.2	0.2-4.5	1	0.6	0-3.5	2
	This study	135	0	0.0	0-27	0	0.0	0-27	0
	Total	2696	3	1.1	0.2-3.2	2	0.7	0.1-2.7	2
DXS7132	Turrina et al. (2004)	40	0	0.0	0-88.1	0	0.0	0-88.1	0
	Turrina et al. (2007)	40	0	0.0	0-88.1	0	0.0	0-88.1	0
	Tariq et al., (2008)	50	0	0.0	0-71.1	0	0.0	0-71.1	0
	Pico et al., (2008)	51	0	0.0	0-69.8	0	0.0	0-69.8	1
	Liu et al., (2008)	294	3	10.2	2.1-29.5	0	0.0	0-12.5	0
	Becker et al., (2008)	124	1	8.1	0.2-44.1	0	0.0	0-29.3	0
	Tetzlaff et al. (2012)	105	1	9.5	0.2-51.9	0	0.0	0-34.5	0
	Liu et al., (2012)	310	3	9.7	2-28	0	0.0	0-11.8	0
	Nishi et al., (2013)	20	0	0.0	0-168.4	0	0.0	0-168.4	0
	Diegoli et al., (2014)	417	1	2.4	0.1-13.3	0	0.0	0-8.8	0
	Chen et al., (2020)	58	0	0.0	0-61.6	0	0.0	0-61.6	0
	Pinto et al., (2020)	1612	13	8.1	4.3-13.8	3	1.9	0.4-5.4	3
	This study	135	2	14.8	1.8-52.5	0	0.0	0-27	1
	Total	3256	24	7.4	4.7-10.9	3	0.9	0.2-2.7	5
DXS8378	Turrina et al. (2007)	40	0	0.0	0-88.1	0	0.0	0-88.1	0
	Tariq et al., (2008)	50	0	0.0	0-71.1	0	0.0	0-71.1	0
	Becker et al., (2008)	124	0	0.0	0-29.3	0	0.0	0-29.3	0
	Tetzlaff et al. (2012)	105	1	9.5	0.2-51.9	0	0.0	0-34.5	0

	Nishi et al., (2013)	20	0	0.0	0-168.4	0	0.0	0-168.4	0
	Diegoli et al., (2014)	417	1	2.4	0.1-13.3	0	0.0	0-8.8	1
	Chen et al., (2020)	58	0	0.0	0-61.6	0	0.0	0-61.6	0
	Pinto et al., (2020)	1612	0	0.0	0-2.3	0	0.0	0-2.3	2
	This study	135	0	0.0	0-27	0	0.0	0-27	0
	Total	2561	2	0.8	0.1-2.8	0	0.0	0-1.4	3
DXS7423	Turrina et al. (2007)	40	0	0.0	0-88.1	0	0.0	0-88.1	0
	Tariq et al., (2008)	50	0	0.0	0-71.1	0	0.0	0-71.1	0
	Becker et al., (2008)	124	0	0.0	0-29.3	0	0.0	0-29.3	0
	Tetzlaff et al. (2012)	105	0	0.0	0-34.5	0	0.0	0-34.5	0
	Nishi et al., (2013)	20	0	0.0	0-168.4	0	0.0	0-168.4	0
	Diegoli et al., (2014)	417	0	0.0	0-8.8	0	0.0	0-8.8	0
	Pinto et al., (2020)	1612	2	1.2	0.2-4.5	1	0.6	0-3.5	0
	This study	135	0	0.0	0-27	0	0.0	0-27	1
	Total	2503	2	0.8	0.1-2.9	1	0.4	0-2.2	1
DXS10135	Becker et al., (2008)	124	0	0.0	0-29.3	0	0.0	0-29.3	0
	Tetzlaff et al. (2012)	105	2	9.5	0.2-51.9	0	0.0	0-34.5	0
	Pinto et al. (2020)	1612	27	16.7	11.1-24.3	7	4.3	1.7-8.9	1
	This study	135	2	14.8	1.8-52.5	0	0.0	0-27	0
	Total	1976	31	15.7	10.7-22.2	7	3.5	1.4-7.3	1
DXS10074	Becker et al., (2008)	124	0	0.0	0-29.3	0	0.0	0-29.3	0
	Tetzlaff et al. (2012)	105	1	9.5	0.2-51.9	0	0.0	0-34.5	0
	Liu et al., (2012)	310	1	3.2	0.1-17.8	0	0.0	0-11.8	0
	Nishi et al., (2013)	20	0	0.0	0-168.4	0	0.0	0-168.4	0
	Pinto et al., (2020)	1612	15	9.3	5.2-15.3	0	0.0	0-2.3	2
	This study	135	1	7.4	0.2-40.6	0	0.0	0-27	0
	Total	2306	18	7.8	4.6-12.3	0	0.0	0-1.6	2
DXS10101	Becker et al., (2008)	124	0	0.0	0-29.3	0	0.0	0-29.3	0
	Tetzlaff et al. (2012)	105	0	0.0	0-34.5	0	0.0	0-34.5	1
	Pinto et al., (2020)	1612	8	5.0	2.1-9.8	1	0.6	0-3.5	1
	This study	135	0	0.0	0-27	0	0.0	0-27	0
	Total	1976	8	4.0	1.7-8	1	0.5	0-2.8	2
DXS10134	Becker et al., (2008)	124	0	0.0	0-29.3	0	0.0	0-29.3	0
	Tetzlaff et al. (2012)	105	1	9.5	0.2-51.9	0	0.0	0-34.5	0
	Nishi et al., (2013)	20	0	0.0	0-168.4	0	0.0	0-168.4	0
	Chen et al., (2020)	58	0	0.0	0-61.6	0	0.0	0-61.6	0
	Pinto et al., (2020)	1612	10	6.2	3-11.4	5	3.1	1-7.2	1
	This study	135	0	0.0	0-27	0	0.0	0-27	0

	Total	2054	11	5.4	2.7-9.6	5	2.4	0.8-5.7	1
DXS10148	Pinto et al., (2020)	1612	11	6.8	3.4-12.2	3	1.9	0.4-5.4	0
	This study	135	0	0.0	0-27	0	0.0	0-27	0
	Total	1747	11	6.3	3.1-11.2	3	1.7	0.4-5	0
DXS10079	Liu et al., (2012)	310	1	3.2	0.1-17.8	1	3.2	0.1-17.8	0
	Pinto et al., (2020)	1612	15	9.3	5.2-15.3	1	0.6	0-3.5	4
	This study	135	3	22.2	4.6-63.6	0	0.0	0-27	0
	Total	2057	19	9.2	5.6-14.4	2	1.0	0.1-3.5	4
DXS10103	Pinto et al., (2020)	1612	5	3.1	1-7.2	0	0.0	0-2.3	1
	This study	135	1	7.4	0.2-40.6	0	0.0	0-27	0
	Total	1747	6	3.4	1.3-7.5	0	0.0	0-2.1	1
DXS10146	Pinto et al., (2020)	1612	6	3.7	1.4-8.1	2	1.2	0.2-4.5	1
	This study	135	4	29.6	8.1-74.1	0	0.0	0-27	0
	Total	1747	10	5.7	2.7-10.5	2	1.1	0.1-4.1	1
Total		26626	145	5.4	4.6-6.4	26	1.0	0.6-1.4	23

Supplementary Table 10- X-STR overall mutation steps based on data from this study (N=135 mother-father-daughter), and the literature. Mutation steps: {+1}- gain of one repeat; {-1}- loss of one repeat; [-1]V{+1}- gain or loss of one repeat; >{+1}-gain of more than one repeat; <{-1}loss of more than one repeat; NI- non-integer repeat variation.

Marker	Reference	Total							
		No. allele transfers	No. mutations	{+1}	{-1}	{-1}V{+1}	>{+1}	<{-1}	NI
HPRTB	Tetzlaff, Wegener and Lindner, (2012)	210	1	1	0	0	0	0	0
	Diegoli et al., (2014)	834	1	1	0	0	0	0	0
	Pinto et al., (2020)	2276	5	3	1	1	0	0	0
	Total	3320	7	5	1	1	0	0	0
DXS7132	Pico et al., (2008)	102	1	0	0	1	0	0	0
	Liu et al., (2008)	588	3	1	2	0	0	0	0
	Becker et al., (2008)	248	1	0	1	0	0	0	0
	Tetzlaff, Wegener and Lindner, (2012)	210	1	0	1	0	0	0	0
	Liu et al., (2012)	620	3	1	2	0	0	0	0
	Diegoli et al., (2014)	834	1	1	0	0	0	0	0
	Pinto et al., (2020)	2994	19	12	5	1	1	0	0
	This study	270	3	0	3	0	0	0	0
	Total	5866	32	15	14	2	1	0	0
DXS8378	Diegoli et al., (2014)	834	2	0	1	1	0	0	0
	Tetzlaff, Wegener and Lindner, (2012)	210	1	1	0	0	0	0	0
	Pinto et al., (2020)	1852	2	1	0	1	0	0	0
	Total	2896	5	2	1	2	0	0	0
DXS7423	Pinto et al., (2020)	1704	3	1	2	0	0	0	0
	This study	270	1	1	0	0	0	0	0
	Total	1974	4	2	2	0	0	0	0
DXS10135	Tetzlaff, Wegener and Lindner, (2012)	210	2	0	1	0	0	0	1
	Pinto et al., (2020)	2994	35	15	20	0	0	0	0
	This study	270	2	1	1	0	0	0	0
	Total	3474	39	16	22	0	0	0	1
DXS10074	Tetzlaff, Wegener and Lindner, (2012)	210	1	0	1	0	0	0	0
	Liu et al., (2012)	620	1	0	1	0	0	0	0
	Pinto et al., (2020)	2994	17	5	11	0	1	0	0
	This study	270	1	1	0	0	0	0	0
	Total	4094	20	6	13	0	1	0	0
DXS10101	Tetzlaff, Wegener and Lindner, (2012)	210	1	1	0	0	0	0	0
	Pinto et al., (2020)	2424	10	4	5	0	0	1	0

	Total	2634	11	5	5	0	0	1	0
DXS10134	Tetzlaff, Wegener and Lindner, (2012)	210	1	0	1	0	0	0	0
	Pinto et al., (2020)	2848	16	9	5	0	1	0	1
	Total	3058	17	9	6	0	1	0	1
DXS10148	Pinto et al., (2020)	2990	14	6	7	0	0	1	0
	Total	2990	14	6	7	0	0	1	0
DXS10079	Liu et al., (2012)	620	2	0	2	0	0	0	0
	Pinto et al., (2020)	3202	20	9	10	1	0	0	0
	This study	270	3	1	2	0	0	0	0
	Total	4092	25	10	14	1	0	0	0
DXS10103	Pinto et al., (2020)	1850	6	3	3	0	0	0	0
	This study	270	1	1	0	0	0	0	0
	Total	2120	7	4	3	0	0	0	0
DXS10146	Pinto et al., (2020)	2064	9	4	4	1	0	0	0
	This study	270	4	1	3	0	0	0	0
	Total	2334	13	5	7	1	0	0	0
Total		38852	194	85	95	7	3	2	2

Supplementary Table 11- X-STR mutation steps counted for a paternal and maternal most parsimonious origin, based on data from this study (N=135 mother-father-daughter), and the literature. Mutation steps: {+1}- gain of one repeat; {-1}- loss of one repeat; [-1]V{+1}- gain or loss of one repeat; >{+1}-gain of more than one repeat; <{-1}-loss of more than one repeat; NI- non-integer repeat variation.

Marker	Reference	Most Parsimonious Parental origin															
		Paternal								Maternal							
		No. allele transfers	No. mutations	{+1}	{-1}	{-1}V{+1}	>{+1}	<{-1}	NI	No. allele transfers	No. mutations	{+1}	{-1}	{-1}V{+1}	>{+1}	<{-1}	NI
HPRTB	Tetzlaff, Wegener and Lindner, (2012)	105	0	0	0	0	0	0	0	105	1	1	0	0	0	0	0
	Diegoli et al., (2014)	417	1	1	0	0	0	0	0	417	0	0	0	0	0	0	0
	Pinto 2020	1138	2	2	0	0	0	0	0	1138	1	1	0	0	0	0	0
	Total	1660	3	3	0	0	0	0	0	1660	2	2	0	0	0	0	0
DXS7132	Pico et al., (2008)	51	0	0	0	0	0	0	0	51	0	0	0	0	0	0	0
	Liu et al., (2008)	294	3	1	2	0	0	0	0	294	0	0	0	0	0	0	0
	Becker et al., (2008)	124	1	0	1	0	0	0	0	124	0	0	0	0	0	0	0
	Tetzlaff, Wegener and Lindner, (2012)	105	1	0	1	0	0	0	0	105	0	0	0	0	0	0	0
	Liu et al., (2012)	310	3	1	2	0	0	0	0	310	0	0	0	0	0	0	0
	Diegoli et al., (2014)	417	1	1	0	0	0	0	0	417	0	0	0	0	0	0	0
	Pinto et al., (2020)	1497	13	8	5	0	0	0	0	1497	3	1	0	1	1	0	0
	This study	135	2	0	2	0	0	0	0	135	0	0	0	0	0	0	0
	Total	2933	24	11	13	0	0	0	0	2933	3	1	0	1	1	0	0
DXS8378	Diegoli et al., (2014)	417	1	0	1	0	0	0	0	417	0	0	0	0	0	0	0
	Tetzlaff, Wegener and Lindner, (2012)	105	1	1	0	0	0	0	0	105	0	0	0	0	0	0	0
	Pinto et al., (2020)	926	0	0	0	0	0	0	0	926	0	0	0	0	0	0	0
	Total	1448	2	1	1	0	0	0	0	1448	0	0	0	0	0	0	0

DXS7423	Pinto et al., (2020)	852	2	0	2	0	0	0	0	852	1	1	0	0	0	0	0
	This study	135	0	0	0	0	0	0	0	135	0	0	0	0	0	0	0
	Total	987	2	0	2	0	0	0	0	987	1	1	0	0	0	0	0
DXS10135	Tetzlaff, Wegener and Lindner, (2012)	105	2	0	1	0	0	0	1	105	0	0	0	0	0	0	0
	Pinto et al., (2020)	1497	27	12	15	0	0	0	0	1497	7	2	5	0	0	0	0
	This study	135	2	1	1	0	0	0	0	135	0	0	0	0	0	0	0
	Total	1737	31	13	17	0	0	0	1	1737	7	2	5	0	0	0	0
DXS10074	Tetzlaff, Wegener and Lindner, (2012)	105	1	0	1	0	0	0	0	105	0	0	0	0	0	0	0
	Liu et al., (2012)	310	1	0	1	0	0	0	0	310	0	0	0	0	0	0	0
	Pinto et al., (2020)	1497	15	4	10	0	1	0	0	1497	0	0	0	0	0	0	0
	This study	135	1	1	0	0	0	0	0	135	0	0	0	0	0	0	0
	Total	2047	18	5	12	0	1	0	0	2047	0	0	0	0	0	0	0
DXS10101	Tetzlaff, Wegener and Lindner, (2012)	105	0	0	0	0	0	0	0	105	0	0	0	0	0	0	0
	Pinto et al., (2020)	1212	8	2	5	0	0	1	0	1212	1	1	0	0	0	0	0
	Total	1317	8	2	5	0	0	1	0	1317	1	1	0	0	0	0	0
DXS10134	Tetzlaff, Wegener and Lindner, (2012)	105	1	0	1	0	0	0	0	105	0	0	0	0	0	0	0
	Pinto et al., (2020)	1424	10	4	5	0	1	0	0	1424	5	4	0	0	0	0	1
	Total	1529	11	4	6	0	1	0	0	1529	5	4	0	0	0	0	1
DXS10148	Pinto et al., (2020)	1495	11	4	6	0	0	1	0	1495	3	2	1	0	0	0	0
	Total	1495	11	4	6	0	0	1	0	1495	3	2	1	0	0	0	0

DXS10079	Liu et al., (2012)	310	1	0	1	0	0	0	0	310	1	0	1	0	0	0	0
	Pinto et al., (2020)	1601	15	7	8	0	0	0	0	1601	1	1	0	0	0	0	0
	This study	135	3	1	2	0	0	0	0	135	0	0	0	0	0	0	0
	Total	2046	19	8	11	0	0	0	0	2046	2	1	1	0	0	0	0
DXS10103	Pinto et al., (2020)	925	5	2	3	0	0	0	0	925	0	0	0	0	0	0	0
	This study	135	1	1	0	0	0	0	0	135	0	0	0	0	0	0	0
	Total	1060	6	3	3	0	0	0	0	1060	0	0	0	0	0	0	0
DXS10146	Pinto et al., (2020)	1032	6	3	3	0	0	0	0	1032	2	1	1	0	0	0	0
	This study	135	4	1	3	0	0	0	0	135	0	0	0	0	0	0	0
	Total	1167	10	4	6	0	0	0	0	1167	2	1	1	0	0	0	0
Total		19426	145	58	82	0	2	2	1	19426	26	15	8	1	1	0	1

Supplementary Table 12- X-STR mutation steps counted for an indeterminate most parsimonious origin, based on data from this study (N=135 mother-father-daughter), and the literature. Mutation steps: {+1}- gain of one repeat; {-1}- loss of one repeat; [-1]V{+1}- gain or loss of one repeat.

Marker	Reference	Most Parsimonious Parental origin			
		Indeterminate			
		No. mutations	{+1}	{-1}	{-1} V {+1}
HPR1B	Tetzlaff, Wegener and Lindner, (2012)	0	0	0	0
	Diegoli et al., (2014)	0	0	0	0
	Pinto 2020	2	0	1	1
	Total	2	0	1	1
DXS7132	Pico et al., (2008)	1	0	0	1
	Liu et al., (2008)	0	0	0	0
	Becker et al., (2008)	0	0	0	0
	Tetzlaff, Wegener and Lindner, (2012)	0	0	0	0
	Liu et al., (2012)	0	0	0	0
	Diegoli et al., (2014)	0	0	0	0
	Pinto et al., (2020)	3	3	0	0
	This study	1	0	1	0
	Total	5	3	1	1
DXS8378	Diegoli et al., (2014)	1	0	0	1
	Tetzlaff, Wegener and Lindner, (2012)	0	0	0	0
	Pinto et al., (2020)	2	1	0	1
	Total	3	1	0	2
DXS7423	Pinto et al., (2020)	0	0	0	0
	This study	1	1	0	0
	Total	1	1	0	0
DXS10135	Tetzlaff, Wegener and Lindner, (2012)	0	0	0	0
	Pinto et al., (2020)	1	1	0	0
	This study	0	0	0	0
	Total	1	1	0	0
DXS10074	Tetzlaff, Wegener and Lindner, (2012)	0	0	0	0
	Liu et al., (2012)	0	0	0	0
	Pinto et al., (2020)	2	1	1	0
	This study	0	0	0	0
	Total	2	1	1	0
DXS10101	Tetzlaff, Wegener and Lindner, (2012)	1	1	0	0
	Pinto et al., (2020)	1	1	0	0
	Total	2	2	0	0
DXS10134	Tetzlaff, Wegener and Lindner, (2012)	0	0	0	0

	Pinto et al., (2020)	1	1	0	0
	Total	1	1	0	0
DXS10148	Pinto et al., (2020)	0	0	0	0
	Total	0	0	0	0
DXS10079	Liu et al., (2012)	0	0	0	0
	Pinto et al., (2020)	4	1	2	1
	This study	0	0	0	0
	Total	4	1	2	1
DXS10103	Pinto et al., (2020)	1	1	0	0
	This study	0	0	0	0
	Total	1	1	0	0
DXS10146	Pinto et al., (2020)	1	0	0	1
	This study	0	0	0	0
	Total	1	0	0	1
Total		23	12	5	6