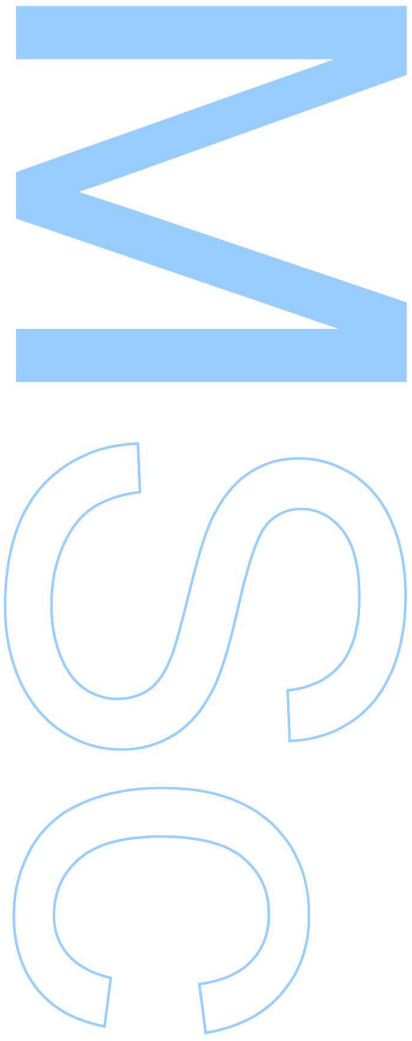
Estimation of mutation rates at Y-STRs

Ana Sofia Antão Sousa
Mestrado em Genética Forense
Departmento de Biologia
2017

Orientador
Doutora Nádia Pinto
Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Portugal
Instituto de Investigação e Inovação em Saúde, I3S, Universidade do Porto, Portugal
Centro de Matemática da Universidade do Porto, Portugal

Coorientador
Professora Doutora Leonor Gusmão
Laboratório de Diagnósticos por DNA (LDD), Universidade do Estado do Rio de Janeiro (UERJ), Brasil



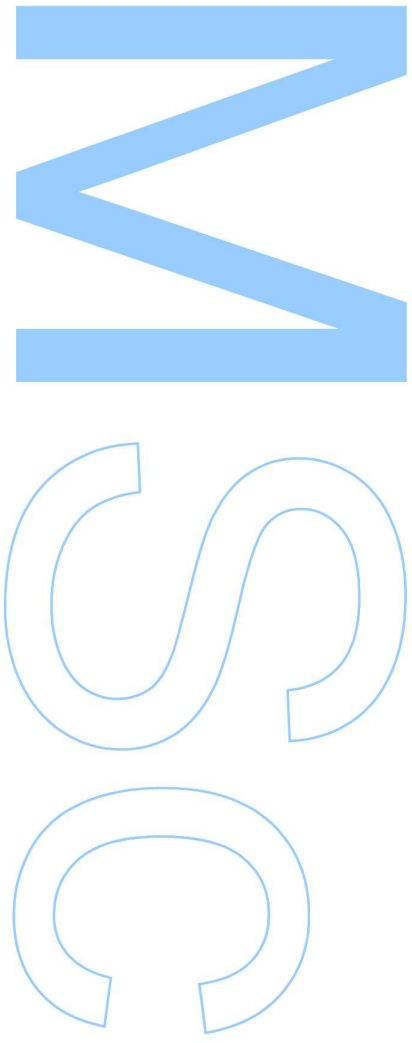
U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

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

O Presidente do Júri,

Porto, _____/_____/_____



INDEX

IMAGE & TABLE INDEX

AGRADECIMENTOS

Obrigada à minha orientadora Doutora Nádia Pinto pelo total apoio, enorme dedicação, amizade e todo o saber que partilhou incansavelmente ao longo deste projeto. Só graças ao seu incrível espírito científico foi possível levar a termo este trabalho.

À minha coorientadora Professora Doutora Leonor Gusmão pela constante disponibilidade para ajudar, nenhuma tarefa foi demasiado grande ou demasiado pequena.

À Doutora Rita Gaio por, com todo o seu saber, ter auxiliado na análise estatística dos dados.

Ao Professor Doutor António Amorim por toda a inspiração científica.

À minha "roomie" Mónica Costa por me ter cedido um espaço em sua casa (e no seu coração) sempre que precisei, por todo o alento e apoio e pelas belas refeições partilhadas. Obrigada também ao Pedro Rodrigues por ter cozinhado para nós, e aos patudos pois, na verdade, o espaço cedido também é deles.

Ao meu colega e amigo Pedro Machado por ter sido um apoio ao longo destes dois anos, por nunca me ter deixado sentir sozinha e por ter aguentado todas as minhas inquietações valentemente.

A todos os meus amigos e família por todo o apoio. À minha mãe Luísa Sousa por ser o meu maior exemplo de resiliência, por nunca ter deixado de acreditar em mim e por todo o seu apoio e amor incondicional. Ao meu pai Luís Sousa, *in memorian*, por me ter deixado a vontade de aprender e a honestidade como maiores heranças. Ao meu irmão Frederico Sousa, por ser a pessoa mais bondosa e incrível, cujo apoio nunca me falhou. Ao meu cão Stitch Sousa por ter estado sempre por perto, literalmente. E, finalmente, ao meu namorado Samuel Prata, por ter caminhado lado a lado comigo ao longo desta jornada, pela sua paciência, compreensão e ajuda e, especialmente, pela sua inabalável capacidade de me fazer rir.

Obrigada!

ABSTRACT

The properties of the Y chromosome make it extremely informative not only for population genetics studies but also for forensic application. Although diverse kind of Y-polymorphisms proved to be valuable in routine forensic casework, short tandem repeats (STRs) have been the most commonly employed particularly due to their high levels of diversity. Y-STR typing is especially useful when DNA from two or more males is mixed, when there is a low amount of male DNA compared to the female DNA in a mixture, or in the so-called deficiency cases in which the alleged father is not available for testing and it is necessary to establish other paternal relationships.

Because it is lacking a homologous chromosome, the Y chromosome does not recombine in most of its extension. For that reason, it is the only chromosome that enables the exact knowledge of which parental allele resulted in which filial one. Since Y-STRs are biologically and analytically like autosomal STRs there is no reason to believe that the knowledge obtained through the study of these polymorphisms cannot be transferred to autosomal (or X-chromosomal) ones.

In this work, we studied mutation rates of Y-STRs by analyzing father-son duos in the framework of three different approaches: the marker approach, the structure approach and the bi-allele approach.

For the traditional approach, an overall mutation rate per marker was computed by proportioning the number of Mendelian incompatibilities between father-son duos. This approach consists of two parts. First, we analyzed unpublished data from a collaborative study by The Spanish and Portuguese - Speaking Working Group of the International Society for Forensic Genetics (GHEP-ISFG), which resulted in the publication of an extended abstract in the Proceeding of the 27th Congress of the ISFG, published by Elsevier, in Forensic Science International: Genetics Supplement Series. Afterwards, gathering the previously mentioned data and data from other published works presenting estimates of mutation rates on Y-STRs, we updated mutation rates for the analyzed Y chromosome markers. This analysis, showed an equilibrium between the number of repeat gains and the number of losses (when analyzed per marker), barring few markers. Also, it is evident that the confidence on the estimation of the mutation rate varies from marker to marker.

For the structure approach, we studied mutation rates by grouping markers with the same repetitive sequences, and thus [GATA], [GAAA], [GAAAA], [GAA] and [ATT] markers were considered and analyzed together. It seems to be an association between the repeat structure and mutation rates, but specially between the repeat structure and the type of mutation (number of mutational steps). However, it seems clear that factors other than the structure of the repetitive motif must be involved in the mutation phenomenon.

For the bi-allele approach, allele and bi-allele mutation rates were computed, allowing the analysis of intra-marker mutation rates. We noted that the number of repeat gains and losses in an intra-marker (or inter-allele) approach is not in equilibrium. Moreover, alleles within the same marker have distinct mutation rates, in some cases confidence intervals do not even intersect.

The bi-allele approach appears to be, from all the three approaches studied in this work, the most satisfactory. Nevertheless, to improve statistical confidence on mutation rate estimates it is peremptory to collect as many complete haplotypic data as possible.

RESUMO

As propriedades do cromossoma Y tornam-no extremamente informativo não só no contexto da genética populacional, mas também em contexto forense. Apesar de diversos polimorfismos do cromossoma Y se terem mostrado úteis na resolução de casos forenses, os *Short Tandem Repeats* (STRs) têm sido os mais utilizados sobretudo devido aos seus elevados níveis de diversidade quando comparados com outros polimorfismos. A tipagem de STRs do cromossoma Y é particularmente vantajosa quando DNA de dois ou mais homens está misturado, quando a quantidade de DNA masculino é baixa em relação à quantidade de DNA feminino, ou em casos em que é necessário o estabelecimento de relações de paternidade em que o alegado pai não está disponível.

Por não ter um cromossoma correspondente homólogo, na maior parte da extensão do cromossoma Y não ocorre recombinação. Por esta razão, é o único cromossoma que permite o conhecimento inequívoco de que alelo parental originou que alelo filial. Assim, já que os STRs do cromossoma Y e os STRs autossomais são biologicamente e analiticamente semelhantes, não há razão para que o conhecimento obtido através do estudo destes polimorfismos não possa ser generalizado para os polimorfismos autossomais.

Neste trabalho, estudamos as taxas de mutação dos STRs do cromossoma Y analisando duos pai-filho no enquadramento de três abordagens: abordagem por marcador, abordagem por estrutura e abordagem bi-alélica.

Na abordagem por marcador, a taxa de mutação foi calculada como a proporção do número de incompatibilidades Mendelianas entre duos pai-filho para cada marcador. Esta abordagem é composta por duas partes. Primeiramente, foram analisados dados não publicados, recolhidos sob o contexto de um trabalho colaborativo pelo Grupo de Línguas Portuguesa e Espanhola da ISFG (GHEP-ISFG), o que resultou na publicação de um *Proceeding* no 27º Congresso da ISFG publicado pela Elsevier na "Forensic Science International: Genetics Supplement Series". Seguidamente, utilizando os dados previamente mencionados e dados recolhidos de artigos referentes a taxas de mutação em STRs do cromossoma Y, calculamos as taxas de mutação por marcador. Esta análise permitiu-nos verificar haver um equilíbrio de ganhos e perdas de repetições (analisando por marcador), com a exceção de alguns marcadores. Também se tornou evidente que a

confiança das estimações de taxas de mutação tem uma grande variação de marcador para marcador.

Na abordagem por estrutura, estudamos taxas de mutação agrupando marcadores consoante a sua sequência repetitiva. Assim, marcadores com as sequências repetitivas [GATA], [GAAA], [GAAAA], [GAA] e [ATT] foram considerados e analisados. Concluímos que parece haver uma associação entre a estrutura do motivo repetitivo e as taxas de mutação e, especialmente, entre o motivo repetitivo e o tipo de mutação (número de passos mutacionais). Contudo, fica claro que outros fatores para além da estrutura do motivo repetitivo deverão estar envolvidos no fenómeno da mutação.

Na abordagem bi-alélica, foram calculadas taxas de mutação alélicas e bi-alélicas para uma análise intra-marcador (ou inter-alelos). O número de ganhos e perdas de repetições intra-marcador não está em equilíbrio. Também, alelos dentro do mesmo marcador têm taxas de mutação distintas e, em alguns casos, os respetivos intervalos de confiança não se intersetam.

A abordagem bi-alélica parece ser, das três estudadas neste trabalho, a mais satisfatória. Contudo, é necessário salientar que para melhorar confiança estatística das taxas de mutação, é essencial a coleção do máximo de informação haplotípica completa quanto possível.

Keywords: Y chromosome; father-son duos; allele approach; bi-allele approach; structure approach

1. INTRODUCTION

1.1. Forensic Genetics

Forensic genetics can be defined rather simply as: “The application of genetics to human and non-human material (in the sense of a science with the purpose of studying inherited characteristics for the analysis of inter- and intra-specific variations in populations) for the resolution of legal conflicts.” (Carracedo, 1998). However, this definition is a traditional and one-sided view as it requires an already established ‘legal conflict’ to be ‘resolved’. Indeed, forensic genetics can be of use: (i) in the investigation phase, even before the conflict between the parties involved has entered the process, as in the case of some paternity tests; (ii) to assist the preparation of the final process as in the case of pleadings and dismissal of cases; (iii) and to serve in prevention of crime as a dissuading factor, since an individual may be dissuaded from committing a crime if he believes he is likely to be caught (Amorim & Budowle, 2016).

The consolidation of the increasingly complex forensic genetics field began over one century ago. In 1900, Karl Landsteiner described the ABO blood grouping system, giving the first step towards forensic haemogenetics (Landsteiner, 1990). When later in 1924, Felix Bernstein demonstrated that the system was transmitted according to rules of Mendelian inheritance, soon it became evident that the ABO system could be applied in solving paternity testing cases and crimes (Bernstein, 1924). These serological tools were limited by the amount of material required to provide discriminating results, by the fact that proteins are prone to degradation on exposure to the environment and the impossibility to analyse body fluids other than blood. During the 1960s and 1970s, developments in molecular biology methods such as RFLP, Sanger sequencing (Sanger *et al.*, 1977) and Southern blotting (Southern, 1975), allowed scientists to examine DNA sequences. In the 1980s the analysis of the first highly polymorphic locus was reported (Wyman & White, 1980).

In 1986, Kary Mullis described the Polymerase Chain Reaction, also known as PCR (Mullis *et al.*, 1986), promoting the development of the sensibility of DNA analyses. The PCR is an enzymatic technique used in molecular biology to amplify a single copy or a few copies of a

segment of DNA generating thousands to millions of copies of a specific DNA sequence, allowing for specific detection and production of large amounts of DNA.

Alec Jeffreys described for the first time, in 1985, “DNA fingerprinting” showing that certain regions of the genome contained repetitive DNA sequences adjacent to each other. He also found out that the number of repeats present could differ from individual to individual. These regions became known as Variable Number of Tandem Repeats (VNTRs) and were the first polymorphisms used in DNA profiling (Jeffreys *et al.*, 1985). By developing a technique to examine the length variation of these DNA repeat sequences, Jeffreys created the ability to perform human identity tests (Butler, 2005, Chapter 1). The technique was called Restriction Fragment Length Polymorphism (RFLP) because it involved the use of a restriction enzyme to cut the regions of DNA surrounding the VNTRs (Butler, 2009, Chapter 1). However, the technique has weaknesses: the alleles are long, which demands a large amount of DNA (preventing the analysis of degraded DNA, for example), comparison between laboratories is difficult, the interpretation is problematic and the analysis is time consuming. Over time, the advances and miniaturization of methodologies made it possible to use other types of genetic data and in the 1990s, the use of VNTRs was replaced by the analysis of Short Tandem Repeats (STRs) (Gill *et al.*, 1994), which became the most commonly used genetic markers for forensic casework.

The fast growth of technology for DNA analysis includes progresses in DNA extraction and quantification methodology, the development of commercial PCR based typing kits and equipment for detecting DNA polymorphisms (Goodwin *et al.*, 2007, Chapter 1).

The comparative analyses of genetic profiles have applications in various contexts, such as (Butler, 2009, Chapter 17):

- Kinship analyses to weight the likelihood of two individuals being related as parent-child or full-siblings, (see, e.g., Green & Mortera, 2017);
- Criminal investigation through the comparison of genetic profiles of, e.g., a sample recovered in a crime scene and a suspect (see, e.g., Pickrahn *et al.*, 2017);
- Identification of unknown remains in historical researches, missing person cases and disaster victim identification through comparison of the collected sample with genetic profiles recovered from personal belongings or from biological relatives (see, e.g.,

Brenner & Weir, 2003);

- Genetic genealogy and ancestry tests, through maternal (mtDNA) or paternal (Y chromosome) lineage determination and also through AIM (Ancestry Informative Marker) analysis (see, e.g., Parson *et al.*, 2008; Romanini *et al.*, 2015; Toscanini *et al.*, 2016);
- Clinical diagnosis of genetic diseases (see, e.g., Seidelmann *et al.*, 2017);
- Clinical investigation, as in the identification of the cell line used so that relevant scientific conclusions can be authenticated (see, e.g., Alonso *et al.*, 2005); or
- Production of genetic profiles for future need (inheritance disputes, missing person, immigration cases, criminal recidivism) (see, e.g., Jeffreys *et al.*, 1985).

1.2. Markers in Forensic Genetics

Genetic markers are commonly characterized as naturally occurring changes in the DNA sequence, where at least two alleles have frequencies greater than 1% in the population (Pereira & Gusmão, 2016). An allele is a variant form of a gene, humans are diploid organisms and have two alleles at each genetic locus, with one allele inherited from each parent.

Genetic variation among individuals is the basis of both pure and applied fields. Any region of the genome can be screened for genetic alterations, whether it is coding or non-coding. However, for population and forensic genetic studies, non-coding markers are preferable as the effects of selection are not directly exerted on them and are thus expected to reflect primarily population level neutral effects, such as drift, expansions, admixture and migration (Wilkinson *et al.*, 2010). The choice for neutral markers in the forensic field also has been recommended to avoid ethical concerns, since they are less prone to disclose information associated with disease or genetic susceptibility (Schneider, 1997).

Depending on the presence or absence of recombination, the different types of available markers can be categorized into two groups: the recombining markers which allow individual identification, the autosomal and X-chromosomal markers; and the non-recombining, mitochondrial DNA and the Y chromosomal DNA (specific zone) that allow the discrimination of maternal and paternal lineages, respectively.

The mode of transmission of the autosomal markers implies that each parent contributes with half of the information to their offspring and, therefore, any pair of individuals related either maternally or paternally as parent-child, unless mutation occurs, will share alleles: the so called identical by descent (or IBD) alleles, as can be seen in **Image 1**.

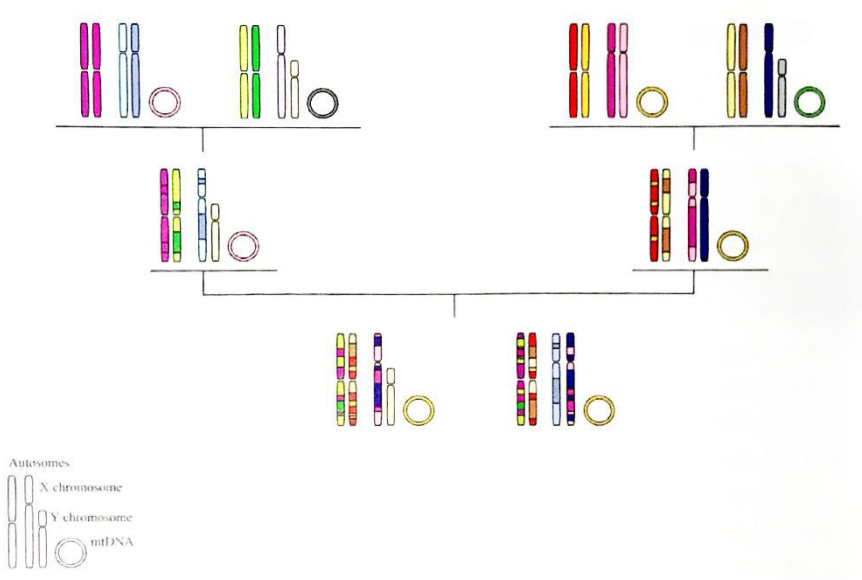


Image 1: Genetic transmission and recombination patterns of the different types of markers.
Source: (Pereira & Gusmão, 2016)

"Mutations can be defined as any permanent, heritable (qualitative or quantitative) change resulting in differences between ancestral and descendant copies of DNA sequences." (Pinto *et al.*, 2014), through which, and for a specific marker, a parent may not share any allele with a child of his/hers. In this case, the identity by descent is broken.

Generally, the kits commercially available consider autosomal markers which provide high power of discrimination, the methods and protocols are well established, and there are guidelines developed on how to properly report data (Bär *et al.*, 1997; Morling *et al.* 2002; Gjertson *et al.*, 2007; Prinz *et al.*, 2007; Gill *et al.*, 2012). Nevertheless, despite autosomal markers being able to provide a reliable answer in a wide framework of kinship analyses, some cases can become complicated when genetic profiles are not complete (due to sample degradation) or the interpretation of the results is not straightforward. In such cases, other types of genetic markers may be recruited. Indeed, combining data from markers with different transmission properties (as autosomes, mtDNA, Y chromosome and X

chromosome) has been successful when addressing such complex cases (see, e.g., Coble *et al.*, 2009).
Uniparental and lineage markers – Y chromosome and mitochondrial DNA – disclose about lineages history. These markers have a low effective population size and present, hence, accentuated genetic differences between human continental groups (see, e.g., Skowronek *et al.*, 2017).

To be useful for forensic genetic purposes, non-coding DNA loci must have some key properties: be highly polymorphic (varying widely amongst individuals within populations), be easy and inexpensive to characterize, give profiles that are simple to interpret and to compare between laboratories, not be under selective pressure and have a low mutation rate (Goodwin *et al.*, 2007, Chapter 2).

Short tandem repeats (STRs) are one of the most abundant types of repeats in the human genome and as previously said they are the primary choice in the field of forensic genetics. They consist of a repeating 2-6 bp motif and span a median of 25 bp. Approximately, 700 000 STR loci exist in the human genome, and in aggregate, they occupy ~1% of its total length (Willems *et al.*, 2016).
The simplest polymorphisms are the Single Nucleotide Polymorphisms (SNPs), single base differences in DNA sequences. SNPs are single nucleotide substitutions present in the genome, they are the most abundant polymorphisms and make up around 85% of human genetic variation (Pereira & Gusmão, 2016).
Another form of polymorphisms is the insertion-deletion markers (indels). An indel can be the insertion or deletion of a segment of DNA ranging from one nucleotide to hundreds of nucleotides, although most have around 3-15 bps. Indel markers can be easily typed and may prove to be particularly useful in ancestry studies (see, e.g., Pereira *et al.*, 2012).
Yet, both SNPs and indels are usually biallelic and therefore are not as polymorphic as STRs. Hence, they have a lower discriminatory power (without significant multiplexing) and do not fit with the ideal properties of DNA polymorphisms for forensic analysis (Butler *et al.*, 2007).
Nevertheless, SNPs and indels have a much lower mutation rate, around an order of magnitude of 10^{-8} (Nachman & Crowell, 2000), compared to STRs, which is around 10^{-3} to

10^{-4} (Ballantyne *et al.*, 2010). This may lead practitioners to think that whenever a mutation occurs it has a higher statistical impact in the result (Phillips *et al.*, 2008). However, Amorim & Pereira (2005) concluded that to match the informative power of STRs, the number of SNP loci needed is much higher; and demonstrated that a SNP battery of loci would be prone, if applied to a routine paternity investigation, to the occurrence of a higher frequency of cases where the statistical evidence is inconclusive. Also, Pinto *et al.* (2013) showed that the use of bi-allelic markers can mislead the investigation when, unknowingly, a close relative of the real father is tested as the alleged one.

All in all, the high degree of polymorphism, the ability to multiplex using PCR, and the ease of typing of STRs, makes their analysis the current method of choice for forensic DNA profiling.

1.3. Y Chromosome in Forensic Genetics

The human Y chromosome is one of the smallest chromosomes of the nuclear genome and contains around 58 million bps, representing 2% of the human male genome. This chromosome is male specific and is transmitted across the paternal lineage. Indeed, the 23rd pair of chromosomes of one female is constituted by two X-chromosomes, unlike the one of males which has one Y and one X chromosome. This condition, exclusive to males, is named hemizygosity.

Conceptually, the Y chromosome can be divided into two genomic territories: one corresponding to a X–Y homology domain involved in meiotic pairing, the pseudoautosomal regions PAR1 and PAR2; the other lacking a homologous chromosome partner, male-specific region – MSY **(Image 2)** (Navarro-Costa, 2012).

The Y chromosome does not recombine in most its extension, albeit the pseudoautosomal regions, which are responsible for the correct pairing of the X and Y chromosomes. These regions, that pair and can recombine with the homologous X chromosome regions during male meiosis, are in the extremities of the chromosome (**Image 2**). The sequences of both pseudoautosomal regions diverge significantly in terms of structural features, reflecting not only differences in their evolutionary history, but also functional constrains associated with the genetic crossing-over requirement in PAR1 (Kauppi *et al.*, 2011), which is essential for successful male meiosis in human species (Mohandas *et al.*, 1992).

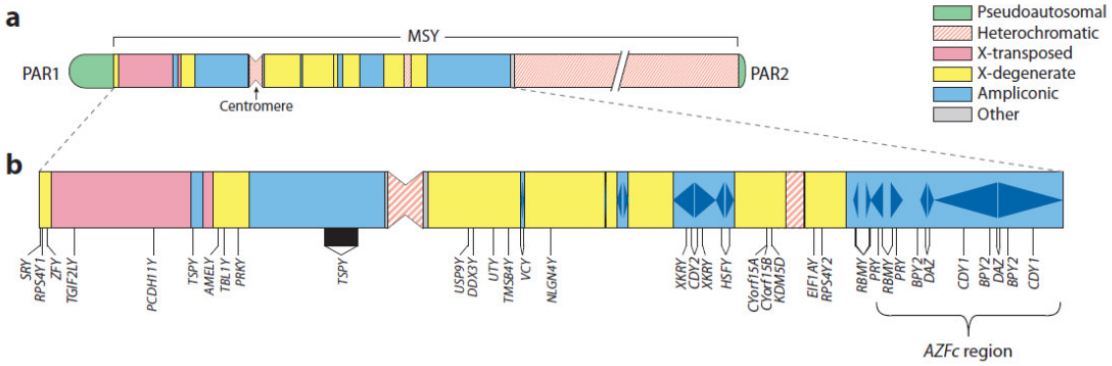


Image 2: Squematic representation of the Y-chromosome, with the male specific region indicated. Palindromes are represented in blue with darker blue triangles. Pseudoautosomal regions 1 (short-arm) and 2 (long-arm) are represented in green.
Source: Hughes & Rozen, 2012

The Male-Specific region (MSY) was once designated as the Non-recombining Region of the Y-chromosome (NRY). However, due to evidence of frequent gene conversion or intrachromosomal recombination, such designation is no longer used (**Image 2**) (Walsh *et al.*, 2004, Chapter 9; Navarro-Costa, 2012; Skaletsky *et al.*, 2003). This region corresponds to 95% of the Y chromosome. It is clonally inherited by the son with no DNA changes apart from those caused by mutation. This region is also rich in palindromes (**Image 2**), regions repeated in reverse along the length of the chromosome, which are prone to intrachromosomal recombination. However, the lack of interchromosomal recombination in the MSY of this chromosome has led to a lower density of functional genes compared to the autosomes. Any deleterious mutation has little to no chance of repair and cannot be removed by recombination (Walsh *et al.*, 2004, Chapter 9).

Structurally, three domains have been identified in the reference MSY: the euchromatic territory spanning approximately 23 Mb, the centromeric region (~1 Mb), and two Yq

heterochromatin blocks, the more distal of which extends for about 40 Mb and exhibits a length polymorphism that ultimately accounts for the significant size variation of the Y in the male population (Repping *et al.*, 2006) (**Image 2**).

1.3.1. Applications

There are several characteristics that qualify the Y chromosome as a special tool for forensic genetics: the small effective population size that tends to create population-specific haplotype distributions on the Y chromosome, the male specificity for most of its length and the absence of recombination, which provides male lineages (Roewer, 2003).

The employment of the Y chromosome for male sex determining purposes in forensic applications started around 40 years ago when luminescence microscopy was applied for detecting Y chromosomes in cells from cadaver material (Radam & Strauch, 1973). However, analyzing only Y-specific DNA for male sex determination is semioptimal, as the absence of the signal can mean either the presence of female material or a negative result due to technical reasons. Hence, since the early 1990s (Akane *et al.*, 1992) the method used for human sex determination has been the amelogenin system which takes advantage of the homologous nature of the human X and Y chromosomes, targeting sites that display sequences with length polymorphisms between the copies.

As previously discussed in general terms, to achieve paternal lineage differentiation for forensic purposes, more polymorphic markers are preferred, that is, those with a higher mutation rate such as Y-STRs that have an average mutation rate about 100 000 times higher that Y-SNPs (Goedbloed *et al.*, 2009), which results in a much higher intra population variability. The beginning of usage of Y-STRs for paternal lineage identification was rather straightforward for forensic biology, as they are biologically and analytically similar to autosomal STRs. Since Y-STRs were introduced to forensic science, they have been used for one main purpose: to identify male lineages with the aim of identifying and excluding suspects (Roewer, 2009; Kayser, 2007).

Y-STR typing is particularly useful when DNA from two or more males is mixed, when there is a low amount of male DNA compared to the female DNA in a mixture (See **Image 3** for a DNA profile analysis of a female-male mixture with autosomal vs. Y-chromosome DNA

markers), or in the so-called deficiency cases in which the alleged father is not available for testing (Pena & Chakraborty, 1994) and is necessary to establish other paternal relationships.

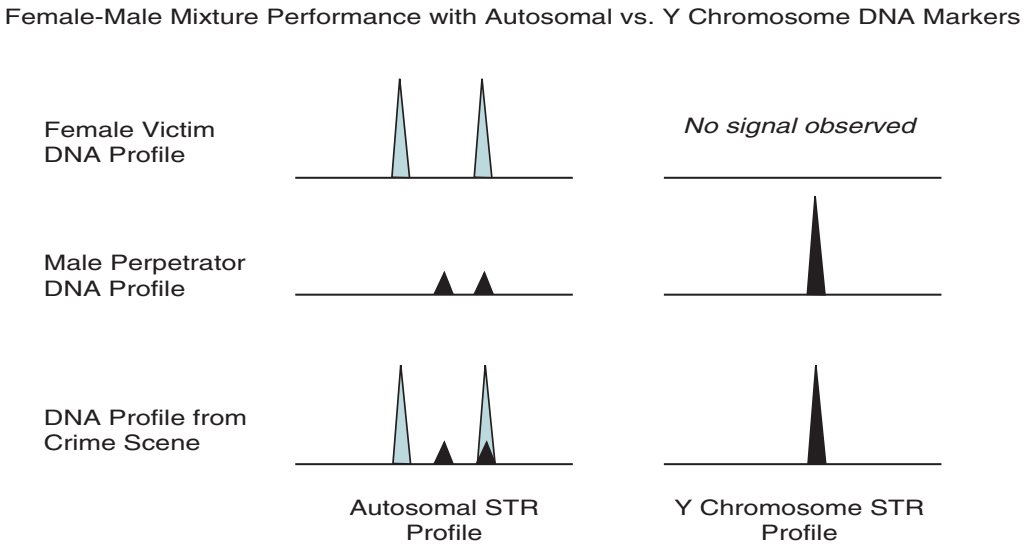


Image 3: Schematic illustrating the types of autosomal or Y-STR profiles that might be observed with sexual assault evidence where mixtures of high amounts of female DNA may mask the STR profile of the perpetrator. Y-STR testing permits isolation of the male component without having to perform a differential lysis.
Source: Butler, 2005, Chapter 9

If legally permitted, Y-STR profiling can also be highly useful in DNA dragnets or mass screenings, in cases where the true perpetrator escapes from voluntarily participation.

However, the characteristics of the Y chromosome that make it so helpful to forensic DNA testing can also be its downfall. The occurrence of mutation is the only source of variation in the Y chromosome amongst male members of the same patrilineal lineage, as practically the whole chromosome is transferred clonally from father to son. Therefore, while exclusions based on the analysis of the Y chromosome DNA testing aid forensic investigations, a match permits only to conclude that the individual could have contributed to the forensic evidence, but so could every other male member from his paternal lineage. Inclusions with Y chromosome testing are not as consequential as autosomal STR matches from a random

match probability perspective (de Knijff, 2003).

Yet, having relatives sharing the same chromosome can be crucial in missing person cases and mass disaster victim identification, since it increases the number of possible reference samples. Paternity tests involving a son and an unavailable putative father can also benefit from Y chromosome markers (Santos *et al.*, 1993).

The lack of recombination in the Y chromosome allows comparison of male individuals belonging to the same paternal lineage and separated by a large period, for this reason Y chromosome testing has become valuable for making inferences on human migration and other population genetics matters as well as in evolutionary studies.

Genealogical investigations and historical research has also been making use of Y chromosome testing as surnames are typically transmitted across the paternal lineage (Jobling, 2001; Sykes & Irven, 2000; Zerjal *et al.*, 2003; Jobling & Tyler-Smith, 2003; Brown, 2002; Iida & Kishi, 2005; Gusmão *et al.*, 1999; Alshamali *et al.*, 2004).

1.3.2. Y Chromosome STR Markers

The use of a common nomenclature for STRs is crucial, in both population and forensic genetics fields, to allow inter-laboratory communication and data comparison. Accordingly, the DNA Commission of the International Society of Forensic Genetics (ISFG) issued recommendations on the use of Y-STRs for forensic analysis (Gusmão *et al.*, 2006).

Forensic practice relies on commercially available kits to perform DNA analysis. Besides the fact that most laboratories do not have the time nor resources to design primers, optimize PCR multiplexes and control primer synthesis quality, the convenience of using kits is also reinforced by the fact that previously tested primer sets and conditions allow improved circumstances to share data between laboratories with reduced chances of silent alleles. Hence, many laboratories were reluctant to start Y-STR typing until kits were available (Butler, 2012, Chapter 13). Another advantage of commercial kits is the availability of common allelic ladders since they allow quality assurance of the results, as well as, compatibility of data introduced into DNA databases. (Butler, 2005, Chapter 9). Although several kits have been released, PowerPlex® Y (comprising 21 loci) and Yfiler® (comprising 17 loci) are the most widely used.

The first STR locus identified on the Y chromosome was DYS19 (Roewer *et al.*, 1992).

Afterwards, dozens of other Y chromosome STRs have been described.

Generally, STRs with four nucleotide motif units are plentiful and more stable than two or three nucleotide repeats hence, they have been favoured when designing the commercially available forensic kits (Pereira & Gusmão, 2016). Nevertheless, Y-STRs with three or five nucleotide motif units are resorted to in commercial kits and thus, they will be also studied in this work (see for example DYS388 and DYS438, **table 1**).

According to the structure of the repetitive sequence, STRs can be classified into simple, compound or complex (Bär *et al.*, 1997). Systems including a homogeneous repeat region where the sequence is uninterrupted are called simple (see, e.g., DYS392), whereas systems including two or more different repeat motifs adjacent to each other are called compound systems (see, e.g., DYS437) and systems including several blocks of repeat motifs varying in length and sequence are called complex systems.

When the Y-STR has a complex sequence, the size of the fragment (sequence) might not correspond to the number of repetitions of the variable motif (see DYS19, **table 1**).

Table 1: Example of the Y-STRs included in the Yfiler®, their sequence and variable motif.

Marker	Sequence	Variable Motif
DYS19	[TAGA]3TAGG[TAGA]7-15	[TAGA]7-15
DYS385	[GAAA]7-28	[GAAA]7-28
DYS389I	[TCTG]3[TCTA]6-13	[TCTA]6-13
DYS389II	[TCTG]4[TCTA]11[TCTG]3[TCTA]8 or [TCTG]5[TCTA]10[TCTG]3[TCTA]8	
DYS390	[TCTG]8[TCTA]5ACTA[TCTA]2[TCTG]1[TCTA]4 [TCTG]8[TCTA]9-14[TCTG]1[TCTA]4	
DYS391	[TCTA]6-14	[TCTA]6-14
DYS392	[TAT]6-17	[TAT]6-17
DYS393	[AGAT]9-17	[AGAT]9-17
DYS437	[TCTA]7-11[TCTG]2[TCTA]4	[TCTA]7-11
DYS438	[TTTTC]6-14	[TTTTC]6-14
DYS439	[GATA]9-14	[GATA]9-14
DYS448	[AGAGAT]11-13N42[AGAGAT]8-9	[AGAGAT]11-13, [AGAGAT]8-9
DYS456	[AGAT]12-18	[AGAT]12-18
DYS458	[GAAA]13-20	[GAAA]13-20
DYS627	[AAAG]11-27	[AAAG]11-27
DYS635	[TCTA]4(TGTA)2[TCTA]2(TGTA)2[TCTA]2(TGTA)0,2[TCTA]n	
GATA H4	[TAGA]8-13 N12 [GATC]2 AA [TAGA]4	[TAGA]8-13

Due to the duplicated, palindromic regions of the Y chromosome, some Y-STR loci occur more than once and, when amplified with a locus-specific set of primers, produce more than

one PCR product, which is the case of DYS385. The entire region around this marker is duplicated and separated by 40,775 bps. However, it is not correct to designate them "DYS385a" and "DYS385b" since it is impossible to assign neither fragment to a defined locus (Gusmão *et al.*, 2006), unless a locus-specific PCR is performed. DYS389I/II also generates two PCR products. However, this marker possesses two primary repeat regions that are flanked on one side by a similar sequence making DYS389I a subset of DYS389II amplicon because the forward PCR primer binds to the flanking region of two different repeat regions that are approximately 120 bps apart (Butler, 2012, Chapter 13) (see **Image 3**).

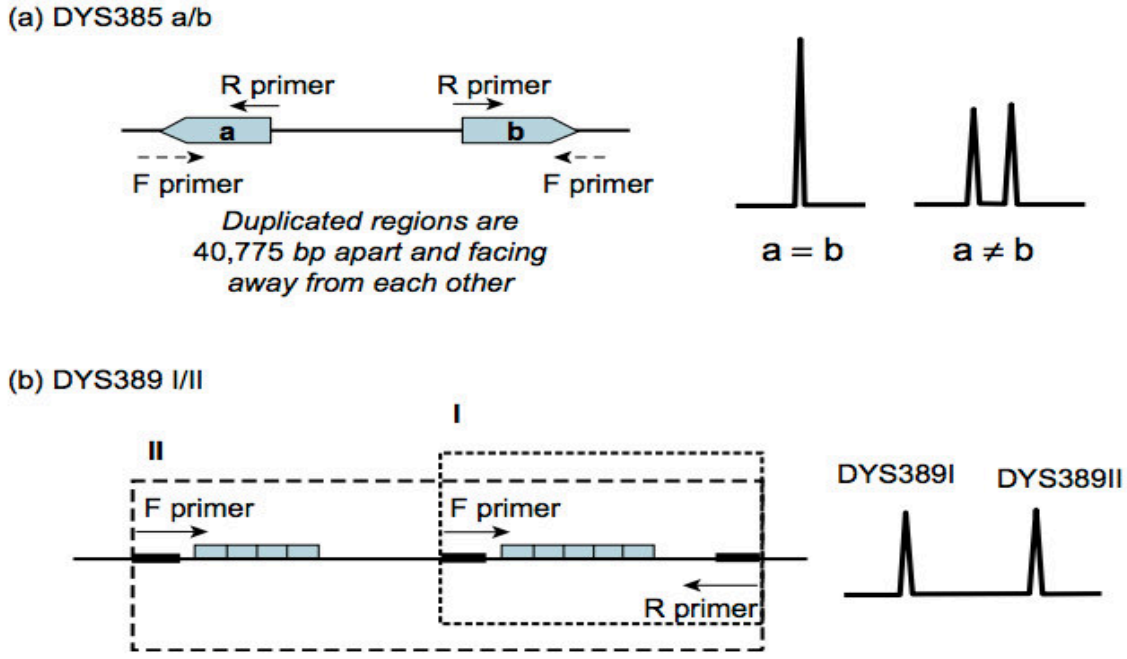


Image 4: Schematic illustration of how multiple PCR primer binding sites give rise to multi-copy PCR products for (a) DYS385a/b and (b) DYS389I/II. Arrows represent forward "F" and reverse "R" primers.

Source: Butler, 2012, Chapter 13

Y chromosome STRs, are highly variable among and between populations, which lead to different allele distributions and frequencies. For instance, the allele with 11 repetitions of the repetitive motif (i.e. allele 11) could be the most frequent allele in population A while in population B the most common is the allele with 13 repetitions (i.e. allele 13). This has consequences in the weight of the evidence, since, for example, the sharing of a rare allele is, of course, statistically more informative than the sharing of a common one.

Massive Parallel Sequencing (MPS) technology, also known as Next Generation Sequencing (NGS), is a technology that permits parallel sequencing analyses of many targeted regions of multiple samples at desirable depth of coverage. Recent studies applying MPS to type Y-STRs (see Zhao *et al.*, 2015; Warshauer *et al.*, 2015; Wendt *et al.*, 2016; Kwon *et al.*, 2016), observed many sequence variants with identical core sequence lengths of the Y-STR loci which had never been previously reported. This entails precaution in the treatment of data obtained by capillary electrophoresis (CE), as alleles that have the same length and that would be referred to by the same number may not have identical sequences, and should not, for that reason, be treated as identical alleles. This implies that when using CE, allele frequencies might be overestimated and the sequence variation underestimated. Notwithstanding, by employing MPS this issue should be easily put to rest, since the full sequence will be available for analysis and comparison, making it possible to identify different alleles, besides them having the same length.

Because male paternal relatives share the same Y chromosome haplotype, the forensic community has begun to shed some attention on rapidly mutating Y-STRs. Since these markers have a higher mutation rate, they are expected to provide a higher probability of distinguishing between closely related individuals, and can add power to the current Y-STR analysis. The most recent commercial kits already include some of these markers (Ballantyne *et al.*, 2012 and Pickrahn *et al.*, 2016).

1.3.3. Y Chromosome Databases

There are several Y chromosome databases that can be classified, essentially, in genetic genealogy databases and forensic databases. Genetic genealogy databases, such as Ysearch (http://www.ysearch.org/), contain Y-STR haplotype information gathered by genetic genealogy companies with different sets of loci to infer male genealogical connections. Hence, the haplotypes in these genealogy databases are associated with individual's surnames (Butler, 2012, Chapter 13). Forensic databases, like YHRD (https://yhrd.org) and US Y-STR (https://www.usystrdatabase.org), contain anonymized collections of haplotypes that can be used to estimate allelic or haplotypic frequencies. These databases are, indeed, crucial for the statistical evaluation of the evidence since different populations may have different allele frequencies.

1.3.3.1. YHRD

The YHRD (Y-Chromosome Haplotype Reference Database) is the largest and most widely used Y-STR database in forensic and population genetic fields. It consists in a freely accessed website which intends to contribute with empirical data to decipher the Y-specific population differentiation and reckon its effects on the frequency estimation process. It was created by Lutz Roewer and colleagues at Humbolt University in Berlin, Germany, and it has been available since 2000 at http://www.yrhd.org/. It is an interactive database that allows the user to search for Y-STR and Y-SNP haplotypes and haplogroups in different layouts and within specified national databases and metapopulations. As of September 12th 2017, the database gathered 149,141 haplotypes for the Power Plex® kit and 136,443 for the YFiler™. Indeed, diagnostic and research laboratories worldwide have joined in a collaborative effort to collect population data, creating a large reference database. All laboratories involved in this project must take part in a mandatory quality control exercise (Willuweit & Roewer, 2000; Willuweit & Roewer, 2015).

1.3.3.2. US Y-STR Database

A population-specific Y-STR Database for United States of America (US Y-STR) was launched in December 2007 to allow haplotype frequency estimates on five different U.S. groups for the 11 loci recommended by the Scientific Working Group on DNA Analysis Methods (SWGDAM). To ensure that no duplicates are inserted in the database, whenever samples possess the same Y-STR profiles, the US Y-STR also requires autosomal STR profiles. As of September 12th 2017, the database has contained 35,660 haplotypes. The main data contributor is the Federal Bureau of Investigation (FBI) (30.2% of the total data contributed), followed by Promega (16.6%) and Applied Biosystems (16.2%). Among other contributors are various state police departments and laboratories.

1.3.2. Mutation

As the Y chromosome is carried along the paternal lineages accumulating diversity only by mutational processes, the chromosome diversity in the general population is reduced, which

alongside a smaller effective population size contributes to a more rapid drift (Walsh *et al.*, 2004, Chapter 9).

It is broadly assumed that random mutations create STRs (Levinson & Gutman, 1987; Schlötterer, 2000) and that they gain or lose repeat units due to a process of DNA-replication and slippage, a mutation mechanism specific to tandemly repeated sequences (Schlötterer, 2000; Ellegren, 2000).

The discovery of different Y alleles among a father-son duo is evidence of mutation. The search for mutations in STR loci involves examining numerous parent/child (father-son duos in the case of Y-markers) allele transfers because the mutation rate is rather low in most STRs (Butler, 2005, Chapter 6).

To estimate mutation rates in the Y chromosome two different methods have been used: deep-rooting pedigrees (see, for example, Heyer *et al.*, 1997, Bonne-Tamir *et al.*, 2003) and male germ-line transmissions from confirmed father-son pairs (see, e.g., Bianchi *et al.*, 1998, Kayser *et al.*, 2000, Dupuy *et al.*, 2001, Dupuy *et al.*, 2004, Kurihara *et al.*, 2004, Gusmão *et al.* 2005, Budowle *et al.*, 2005, Decker *et al.*, 2008, Ge *et al.*, 2009).

The pedigree approach, despite not being necessary to run as many samples, in cases where differences are seen it is difficult to properly identify the generation where the mutation occurred (see Bonné-Tamir *et al.*, 2003) or potential illegitimacy (see Heyer *et al.*, 1997).

When using the father-son pairs approach, the mutation rate is estimated per marker, computing the ratio between the number of cases where Mendelian incompatibilities were observed and the total number of meiotic transfers (see, e.g., Sánchez-Diz *et al.*, 2008, Forster *et al.*, 2015).

Considering the lack of a homologous chromosome, the Y chromosome does not recombine in most of its extension (MSY). For that reason, it is the only chromosome that enables the exact knowledge of which parental allele resulted in which filial one, which turns unambiguous the insight of which allele mutated (or not) in which (Pinto *et al.*, 2014).

Also, since mutation rates for Y-STRs are in the same range as those of autosomal STRs, namely about 1-4 per thousand generational events (Butler, 2005, Chapter 9), and because there is no biological or analytical reason to believe these markers have different mutation mechanisms, studying mutation in this chromosome, would allow insights on autosomal (and X-chromosomal) mechanisms.

These mutation-related subjects will be the focus of this work, where a bi-allele (depending

on the origin, parental, and destination, filial allele) and a structure based (depending on the structure of the repetitive motif) analysis will be performed, aiming to study the relation between these factors and the mutation rates on Y-STRs.

2. AIMS

Our aim was to study mutation rates of Y-STRs, by analyzing father-son duos. For this, we considered multiple factors as the number of gain and loss of repeats, the marker structure, or the frequency of the paternal alleles, in the framework of three different approaches:

2.1. Marker Approach

In this framework, we considered the standard approach analyzing the proportion of inconsistencies between father-son pairs per marker. Our specific aims were:

- To analyze unpublished data resulting from a collaborative study by The Spanish and Portuguese - Speaking Working Group of the International Society for Forensic Genetics (GHEP-ISFG);
- To gather and update information on markers' overall mutation rate, considering data from other published works.

2.2. Structure Approach

In this framework, we studied mutation rates grouping markers with the same repetitive sequences and thus [GATA] markers, [GAAA] markers, [GAAAA] markers, [GAA] markers and [ATT] markers were considered. Our specific aim was:

- To analyze the structure and the repetitive sequence of the markers, evaluating the possibility of presenting estimates for mutation rates depending on the type of the repetitive sequence and on the length of the paternal allele.

2.3. Bi-allele Approach

In this framework, we considered the approach presented by Pinto *et al.* (2014). Our specific aim was:

- To analyze haplotypic information from father-son duos and present estimates for mutation rates depending on the marker, parental and filial alleles.

3. APPROACHES

In this work, data were analyzed under three different frameworks: 3.1. Marker Approach, 3.2. Structure Approach, and 3.3. Bi-allele Approach.

3.1. Marker Approach

The marker approach is the standard methodology in mutation rate estimations and consists on calculating mutation rates for each marker by the simple method of proportion, computing the ratio between the number of Mendelian incompatibilities found and the number of analyzed meiotic transfers (see, for example, Bianchi *et al.*, 1998, Kayser *et al.*, 2000, Dupuy *et al.*, 2001, Dupuy *et al.*, 2004).

3.1.1. Mutation Rates and Segregation Data on 16 Y-STRs: An Update to Previous GHEP-ISFG Studies

Under coordination of Professor Leonor Gusmão, co-supervisor of this thesis, a collaborative study, to improve the estimation of mutation rates of the Y-STRs included in the YFiler® kit, was proposed in the XIII meeting of the GHEP-ISFG Working Group: see http://www.gep-isfg.org/pt/comisses-trabalho/cromossomas-sexuais/exercicio-colaborativo-cromossoma-Y-2009.html for more details (in Portuguese). In such collaborative exercise, eleven laboratories from Brazil, Portugal, Argentina, Colombia, Turkey and Spain have participated, and data from 27,170 meiotic transfers was collected and analyzed. A global analysis considering previous works of the same group (Gusmão *et al.*, 2005; Sánchez-Diz *et al.*, 2008) was also computed.

This work resulted in the presentation of one poster to the 27th Congress of the International Society for Forensic Genetics, Seoul, September 2017, as well as in the publication of an extended abstract in the Proceeding of the conference, published by Elsevier, in Forensic Science International: Genetics Supplement Series.

The results of this section are then described in such article, which is presented next.

ARTICLE IN PRESS

Forensic Science International: Genetics Supplement Series xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

Research paper

Mutation rates and segregation data on 16 Y-STRs: An update to previous GHEP-ISFG studies

S. Antão-Sousa[a,b,c], P. Sánchez-Diz[d], M. Abovich[e], J.C. Alvarez[f], E.F. Carvalho[g], C.M.D. Silva[h], P. Domingues[g], M.J. Farfán[i], A. Gutierrez[j], L. Pontes[k], M.J. Porto[l], Y. Posada[m], T. Restrepo[m], R. Rodenbusch[h], O.A. Santapá[e,n], S. Schumacher[h], D. Suárez[j], C.V. Silva[o], C. Vullo[p], N. Pinto[b,c,q], L. Gusmão[b,c,g,*]

[a] Faculty of Sciences of the University of Porto, Porto, Portugal
[b] Institute of Pathology and Molecular Immunology from University of Porto (IPATIMUP), Portugal
[c] Instituto de Investigação e Inovação em Saúde, I3S, Universidade do Porto, Portugal
[d] Forensic Genetics Unit, Institute of Forensic Science, University of Santiago de Compostela, Galicia, Spain
[e] Banco Nacional de Datos Genéticos, Buenos Aires, Argentina
[f] Lab. Identificación Genética, Dpto. de Medicina Legal, Toxicología y Antropología Física, Universidad de Granada, Spain
[g] DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Brazil
[h] Laboratório de Investigação de Paternidade, Centro de Desenvolvimento Científico e Tecnológico (CDCT), Fundação Estadual de Produção e Pesquisa em Saúde (FEPPS), Porto Alegre, RS, Brazil
[i] Instituto Nacional de Toxicología y Ciencias Forenses (INTCF), Madrid, Spain
[j] Laboratorio de Biología Molecular, Fundación Arthur Stanley Gillow, Bogotá, Colombia
[k] Instituto Nacional de Medicina Legal e Ciências Forenses, Delegação do Norte, Porto, Portugal
[l] Instituto Nacional de Medicina Legal e Ciências Forenses, Delegação do Centro, Coimbra, Portugal
[m] IdentiGEN – Genetic Identification Laboratory and Research Group of Genetic Identification, Institute of Biology, School of Natural and Exact Sciences (FCEN), University of Antioquia, Medellin, Colombia
[n] Laboratorio Central, Pesquisa Neonatal, Hospital Carlos G. Durand, Buenos Aires, Argentina
[o] Instituto Nacional de Medicina Legal e Ciências Forenses, Delegação do Sul, Lisboa, Portugal
[p] DNA Forensic Laboratory, Argentinean Forensic Anthropology Team (EAAF), Córdoba, Argentina
[q] Center of Mathematics of the University of Porto, Portugal

ARTICLE INFO

Keywords:
Y chromosome
Y-STRs
Mutation rates
Mutation rates of Y-STRs

ABSTRACT

The increasing relevance of human Y-STRs in forensic science demands reliable estimates of their mutation rates. Therefore, a collaborative study was carried out by the Spanish and Portuguese working group of the International Society for Forensic Genetics (GHEP-ISFG) in the interest of extending the data on Y-chromosomal short tandem repeat (Y-STR) mutation rates. Sixteen Y-STRs were considered in the analyses: DYS456, DYS389I, DYS389II, DYS390, DYS458, DYS19, DYS385, DYS391, DYS392, DYS393, DYS439, DYS635, DYS437, DYS438, DYS448, GATA H4. Among the sample of 1598 father-son duos analyzed, 46 mutations were observed, 45 of which were a single-step change and 1 was a double-step change. A total of 28 repeat losses were observed against 18 gains, with a ratio of 1:1.5. Eleven duos showing double alleles at the Y-STR loci DYS19, DYS391, DYS439 and DYS448 without allelic discrepancy between the father-son duo were also observed. This new data was added to the previous studies from the GHEP-ISFG working group, totalizing 63 496 allele transmissions (varying between 2298 and 7347 per locus). The average mutation rate across all 16 Y-STRs loci was 0.00187 (95% CI 0.00155–0.00224). The average mutation rates per marker varied between 0.00057 (95% CI 0.00007–0.00206) at DYS438 and 0.00606 (95% CI 0.00375–0.00925) at DYS439.

1. Introduction

Y chromosome short tandem repeats (Y-STRs) are particularly useful in the analysis of DNA mixtures where male DNA is diluted in female DNA. To improve the statistical power of Y chromosome analyses, it is crucial to expand the quantity/quality of available data, namely in what concerns the study of mutation rates. This work was developed in the scope of a collaborative study of the Spanish and

* Corresponding author at: DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Brazil.
E-mail address: leonorbgusmao@gmail.com (L. Gusmão).

http://dx.doi.org/10.1016/j.fsigss.2017.10.008
Received 11 October 2017; Accepted 12 October 2017

Please cite this article as: Antão-Sousa, S., Forensic Science International: Genetics Supplement Series (2017), http://dx.doi.org/10.1016/j.fsigss.2017.10.008

Table 1
Mutation rates per locus from the latest exercise and from the assemblage with previous works [1,2].

Marker	Latest exercise		Total[c]					
Y-STR	No. of mutations	Allele transfers	No. of mutations	Meiotic transmissions	Mutation rate	CI (95%)	Repeat gains	Repeat losses
DYS19	4	1599[a]	11	5111	0.00215	0.00107–0.00385	8	3
DYS385	0	3196	6	7347	0.00082	0.00030–0.00178	6	0
DYS389 I	5	1598	8	4092	0.00196	0.00084–0.00385	3	5
DYS389 II	3	1598	6	4080	0.00147	0.00054–0.00320	4	2
DYS390	2	1598	7	5115	0.00137	0.00055–0.00282	2	5
DYS391	3	1599[a]	15	5115	0.00293	0.00164–0.00483	9	6
DYS392	1	1598	4	5102	0.00078	0.00021–0.00201	3	1
DYS393	1	1598	4	3868	0.00103	0.00028–0.00265	2	2
DYS437	2	1598	5	3446	0.00145	0.00047–0.00338	2	3
DYS438	1	1596[b]	2	3509	0.00057	0.00007–0.00206	0	2
DYS439	7	1599[a]	21	3466	0.00606	0.00375–0.00925	14	7
DYS448	1	1601[a,b]	2	2302	0.00087	0.00011–0.00313	1	1
DYS456	3	1598	5	2298	0.00218	0.00071–0.00507	2	3
DYS458	10	1598	12	2298	0.00522	0.00270–0.00910	5	7
DYS635	1	1598	6	3173	0.00189	0.00069–0.00411	4	2
GATA H4	2	1598	5	3174	0.00158	0.00051–0.00367	2	3
TOTAL	46	27170	119	63496	0.00187	0.00155–0.00224	67	52

[a] Presence of double alleles that were considered as two different alleles.
[b] Presence of null alleles.
[c] Also including data from Gusmão et al. [1] and Sánchez-Diz et al. [2].

Portuguese Working Group of the International Society for Forensic Genetics (GHEP-ISFG) which aimed to collect new mutation data on 16 Y-STR loci, commonly employed in forensic casework, by compiling haplotype information from confirmed father-son duos. Eleven laboratories participated. Data from previous works was obtained under the same purpose and methodology [1,2], and what we now present is an update of such results.

2. Material and methods

The sample comprised 1598 father-son duos from Brazil, Portugal, Argentina, Spain, Turkey and Colombia. All individuals gave informed consent prior to inclusion in the study and the biological kinships were previously confirmed through autosomal STRs.

All samples were genotyped for the AmpFLSTR®YFiler® PCR Amplification kit (Applied Biosystems, Foster City, CA) markers, following manufacturer's instructions.

Mutation rates were estimated per marker computing the ratio between the number of mendelian incompatibilities observed and the number of allele transfers.

Clopper-Pearson confidence intervals (CI) for mutation rates (overall and per marker) were estimated, assuming a level of confidence of 5%.

3. Results and discussion

3.1. Latest exercise

Null alleles were found at markers DYS438, DYS439 and DYS448 and duplications were detected in markers DYS19, DYS439 and DYS448. Among the 27170 allele transfers analyzed, 46 mutations were observed, 45 of which were single-step and 1 was double-step (at DYS458). Double mutations within the same father-son duo were not found. The average age of the fathers involved in mutation events was 30.3 years, similar to the one of the complete set of analyzed fathers (30.4).

3.2. Update of data

Marker mutation rates were estimated for the 16 Y-STRs, considering also the results previously reported, resulting from GHEP collaborative exercises [1,2] – see Table 1.

Among a total of 63,496 meiotic transmissions, 119 mutations were observed, 67 of which were repeat gains and 52 were losses. The obtained locus-specific mutations rates varied between 0.00057 (95% CI 0.00007–0.00206) at DYS438 and 0.00606 (95% Ci 0.00375–0.00925) at DYS439, with an average mutation rate of 0.00187 (95% CI 0.00155–0.00224).

In conclusion, by increasing the total number of father-son duos analyzed, our work allowed to improve the accuracy of the estimates of the mutation rates for the 16 Y-STRs included in the Yfiler kit. Indeed, the average range of the confidence intervals decreased from 0.0055 in [1] and 0.0086 in [2], to 0.0031 when the data from the three studies was analyzed together. The results obtained support previous findings showing that average mutation rates at Y-STRs can vary up to a factor of ten across these loci (YHRD; https://yhrd.org/pages/resources/mutation_rates).

Conflict of interests

All authors declare no conflict of interests.

Acknowledgment

FCT – Fundação para a Ciência e a Tecnologia, Portugal, financed this work through the post-doctoral grant SFRH/BPD/97414/2013.

References

[1] L. Gusmão, P. Sánchez-Diz, F. Calafell, et al., Mutation rates at Y chromosome specific microsatellites, Hum. Mutat. 26 (2005) 520–528.
[2] P. Sánchez-Diz, C. Alves, E. Carvalho, et al., Population and segregation data on 17 Y-STRs: results of a GEP ISFG collaborative study, Int. J. Legal Med. 122 (2008) 529–533.

3.1.2. Update of Marker Mutation Rates

To achieve a statistically significant estimate in Y-STR mutation rates analyses, it is essential to gather as many allele transfers between father-son duos as possible. Assuming the traditional approach here considered, the information needed to compute a marker mutation rate is the number of Mendelian incompatibilities and the number of total allele transfers analyzed for each marker. Hence, for this approach, no haplotypic information was needed and only published data was considered.

3.1.2.1. Material and Methods

Data was retrieved from the work presented in 3.1.1. and from other published works presenting estimates of mutation rates on Y-STRs. Detailed information on analyzed data can be found in **Table 2**.

Table 2: Information gathered to estimate Y-STR mutation rates (analysis per marker).

Authors, date	Study	Father-son duos	Population of Origin
Wang *et al.*, 2016	Genetic polymorphisms and mutation rates of 27 Y-chromosomal STRs in a Han population from Guangdong Province, Southern China. *Forensic Science International: Genetics*, *21*, 5-9.	1,033	China
Turrina *et al.*, 2006	Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay. *International journal of legal medicine, 120*(1), 56-59.	50	Italy
Oh *et al.*, 2015	Haplotype and mutation analysis for newly suggested Y-STRs in Korean father–son pairs. *Forensic Science International: Genetics*, *15*, 64-68.	363	Korea
Tsai *et al.*, 2002	Haplotype frequencies of nine Y-chromosome STR loci in the Taiwanese Han population. *International journal of legal medicine*, *116*(3), 179-183.	109	Taiwan
Berger *et al.*, 2005	Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay. *International journal of legal medicine*, *119*(4), 241-246.	70	Austria
Ballantyne *et al.*, 2014	Toward Male Individualization with Rapidly Mutating Y-Chromosomal Short Tandem Repeats. *Human mutation*, *35*(8), 1021-1032.	2,378	44 countries from Africa, America, Asia, Europe and Oceania
Antão-Sousa *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, in press*	1,598***	Brazil, Portugal, Argentina, Spain, Turkey and Colombia

Gusmão *et al.*, 2005	Mutation rates at Y chromosome specific microsatellites. *Human mutation*, *26*(6), 520-528.	3,026	Argentina, Brazil, Colombia, Portugal, Spain
Sánchez-Diz *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), 529-533.	701	Argentina, Brazil, Portugal
Robino *et al.*, 2015	Development of an Italian RM Y-STR haplotype database: results of the 2013 GEFI collaborative exercise. *Forensic Science International: Genetics*, *15*, 56-63.	409	Italy
Lessig *et al.*, 1998	Y chromosome polymorphisms and haplotypes in West Saxony (Germany). *International journal of legal medicine*, *111*(4), 215-218.	41	Germany
Kayser *et al.*, 2000	Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *The American Journal of Human Genetics*, *66*(5), 1580-1588.	996	Germany and Poland
Dupuy *et al.*, 2001	Y-chromosome variation in a Norwegian population sample. *Forensic science international*, *117*(3), 163-173.	150	Norway
Ballard *et al.*, 2005	A study of mutation rates and the characterization of intermediate, null and duplicated alleles for 13 Y chromosome STRs. *Forensic science international*, *155*(1), 65-70.	245	Britain and Ireland
Budowle *et al.*, 2005	Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America. *Forensic science international*, *150*(1), 1-15.	692	North America (mixed group)
Hohoff *et al.*, 2007	Y-chromosomal microsatellite mutation rates in a population sample from northwestern Germany. *International journal of legal medicine*, *121*(5), 359-363.	1,027	Germany
Kurihara *et al.*, 2004	Mutations in 14 Y-STR loci among Japanese father-son haplotypes. *International journal of legal medicine*, *118*(3), 125-131.	147	Japan
Padilla-Gutiérrez *et al.*, 2008	Population data and mutation rate of nine Y-STRs in a mestizo Mexican population from Guadalajara, Jalisco, Mexico. *Legal Medicine*, *10*(6), 319-320.	189	Mexico
Goedbloed *et al.*, 2009	Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR® Yfiler® PCR amplification kit. *International journal of legal medicine*, *123*(6), 471.	1,757	Germany and Poland
Decker *et al.*, 2008	Analysis of mutations in father–son pairs with 17 Y-STR loci. *Forensic Science International: Genetics*, *2*(3), e31-e35.	389	North America (mixed)
Ge *et al.*, 2009	Mutation rates at Y chromosome short tandem repeats in Texas populations. *Forensic Science International: Genetics*, *3*(3), 179-184.	2,918	North America (mixed)
Pontes *et al.*, 2007	Allele frequencies and population data for 17 Y-STR loci (AmpFℓSTR® Y-filer™) in a Northern Portuguese population sample. *Forensic science international*, *170*(1), 62-67.	45	Portugal

Domingues *et al.*, 2007	Sub-Saharan Africa descendents in Rio de Janeiro (Brazil): population and mutational data for 12 Y-STR loci. *International journal of legal medicine*, *121*(3), 238-241.	135	Sub-Saharan Africa
Dupuy *et al.*, 2004	Y-chromosomal microsatellite mutation rates: Differences in mutation rate between and within loci. *Human mutation*, *23*(2), 117-124.	1,766	Norway
de Souza Góes *et al.*, 2005	Population and mutation analysis of 17 Y-STR loci from Rio de Janeiro (Brazil). *International journal of legal medicine*, *119*(2), 70-76.	119	Brazil
Pestoni *et al.*, 1999	Genetic data on three complex STRs (ACTBP2, D21S11 and HUMFIBRA/FGA) in the Galician population (NW Spain). *International journal of legal medicine*, *112*(5), 337-339.	35	Spain
TOTAL		20,388	

*** - New data only

Data was collected and organized in **Table 3**. Loci mutation rates were calculated and Clopper-Pearson confidence intervals (CI) for mutation rates per marker were estimated, assuming a level of confidence of 0.05.

Table 3: Overall mutation rates per marker attained after gathering the information described in **Table 2**.

			Mutational steps																	
Marker	No. of Mutations	No. of Meiosis	+1	+2	+3	+4	+5	+6	-1	-2	-3	-4	-5	-6	NI	Gains	Losses	Locus mutation rate	Confidence interval (0.95)	Confidence interval amplitude
DYS19	47	23,638	26	0	0	0	0	0	20	0	0	0	0	0	1	26	20	0.00199	0.00146-0.00264	0.00118
DYS385	71	35,724	52	1	0	0	0	0	14	3	1	0	0	0	0	53	18	0.00199	0.00162-0.00259	0.00096
DYS388	1	2,025	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0.00049	0.00001-0.00275	0.00274
DYS389I	49	22,465	21	0	0	0	0	0	28	0	0	0	0	0	0	21	28	0.00218	0.00161-0.00288	0.00127
DYS389II	77	22,411	35	2	1	0	0	0	39	0	0	0	0	0	0	38	39	0.00344	0.00271-0.00429	0.00158
DYS390	40	23,543	18	0	0	0	0	0	22	0	0	0	0	0	0	18	22	0.00170	0.00121-0.00231	0.00110
DYS391	50	23,218	28	1	0	0	0	0	21	0	0	0	0	0	0	29	21	0.00215	0.00160-0.00284	0.00124
DYS392	9	23,168	6	0	0	0	0	0	2	0	0	1	0	0	0	6	3	0.00039	0.00018-0.00074	0.00056
DYS393	19	21,799	12	0	0	0	0	0	7	0	0	0	0	0	0	12	7	0.00087	0.00052-0.00136	0.00084
DYS413	1	4,999	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0.00020	0.00001-0.00111	0.00110
DYS435	0	147	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00000	0.00000-0.02478	0.02478
DYS437	18	14,334	11	0	0	0	0	0	7	0	0	0	0	0	0	11	7	0.00126	0.00074-0.00198	0.00124
DYS438	7	14,354	1	0	0	0	0	0	1	3	0	2	0	0	0	1	6	0.00049	0.00020-0.00100	0.00080
DYS439	65	14,723	30	0	0	0	0	0	35	0	0	0	0	0	0	30	35	0.00441	0.00341-0.00562	0.00221
DYS448	12	11,236	4	0	0	0	0	0	7	1	0	0	0	0	0	4	8	0.00107	0.00055-0.00185	0.00131
DYS449	53	4,447	29	0	0	0	0	0	24	0	0	0	0	0	0	29	24	0.01192	0.00894-0.01556	0.00662
DYS456	43	11,241	25	0	0	0	0	0	18	0	0	0	0	0	0	25	18	0.00383	0.00277-0.00515	0.00238
DYS458	76	11,268	38	0	1	0	0	0	35	2	0	0	0	0	0	39	37	0.00674	0.00532-0.00843	0.00311
DYS460	13	2,757	5	0	0	0	0	0	8	0	0	0	0	0	0	5	8	0.00472	0.00251-0.00805	0.00554

DYS461	0	992	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00000	0.00000-0.00371	0.00371
DYS481	8	1,446	5	0	0	0	0	0	2	1	0	0	0	0	0	5	3	0.00553	0.00239-0.01087	0.00848
DYS518	75	4,405	35	2	3	1	0	0	30	2	0	1	0	0	1	41	33	0.01703	0.01342-0.02130	0.00788
DYS526_A	10	3,119	4	1	1	0	0	0	3	1	0	0	0	0	0	6	4	0.00321	0.00154-0.00589	0.00435
DYS526_B	43	3,173	23	1	0	0	0	0	18	1	0	0	0	0	0	24	19	0.01355	0.00982-0.01821	0.00839
DYS533	5	2,256	3	0	0	0	0	0	2	0	0	0	0	0	0	3	2	0.00222	0.00072-0.00516	0.00444
DYS547	55	3,159	22	0	0	0	0	0	33	0	0	0	0	0	0	22	33	0.01741	0.01314-0.02260	0.00946
DYS549	1	413	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0.00242	0.00006-0.01342	0.01336
DYS570	39	4,654	14	1	0	0	0	0	21	3	0	0	0	0	0	15	24	0.00838	0.00597-0.01144	0.00547
DYS576	71	4,609	37	0	0	0	0	0	31	1	2	0	0	0	0	37	34	0.01540	0.01205-0.01939	0.00734
DYS612	54	3,188	29	4	0	0	0	0	19	2	0	0	0	0	0	33	21	0.01694	0.01275-0.02204	0.00929
DYS626	34	3,138	11	0	0	0	0	0	21	1	0	0	0	1	0	11	23	0.01083	0.00751-0.01511	0.00760
DYS627	65	4,572	28	1	1	0	0	0	34	1	0	0	0	0	0	30	35	0.01422	0.01099-0.01809	0.00710
DYS635	33	11,885	11	0	0	0	0	0	22	0	0	0	0	0	0	11	22	0.00278	0.00191-0.00390	0.00199
DYS643	0	686	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00000	0.00000-0.00536	0.00536
GATA A10	5	1,065	2	0	0	0	0	0	3	0	0	0	0	0	0	2	3	0.00469	0.00153-0.01092	0.00939
GATA C4	0	119	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00000	0.00000-0.03052	0.03052
GATA H4	26	12,398	10	0	0	0	0	0	15	1	0	0	0	0	0	10	16	0.00210	0.00137-0.00307	0.00170
YCAIab	0	4,999	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00000	0.00000-0.00074	0.00074
YCAIIab	3	7,055	1	0	0	0	0	0	2	0	0	0	0	0	0	1	2	0.00043	0.00009-0.00124	0.00115
DXYS156-Y	0	1,027	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00000	0.00000-0.00359	0.00359
TOTAL	1,178	365,855	576	14	7	1	0	0	546	24	3	4	0	1	2	598	578	0.00322	0.00304-0.00341	0.00037

NI – non integer mutations (i.e. mutation between alleles from different microvariant groups).

3.1.1.2. Results and Discussion

Among 365,855 allele transfers, 1,178 mutations were observed, 1,111 of which were single-step, 38 were double-step, 10 were triple-step, 5 were four-step and 1 was a six-step mutation.

We should however remark that marker DYS385 had a different analysis approach due to the inability to assign neither fragment to a defined locus thus, rendering it impossible to determine which allele mutated to which. Despite this, we verified that any other mutational alternatives would involve a greater number of mutations (two or more) than the number assumed. In this work, as in all the others, the simplest explanation: one single mutation, and one step mutation whenever possible, was assumed. However, we must be aware that we can be biasing the results. For example, when a paternal haplotype was 13-15 and the filial haplotype was 14-15, we assumed that the allele 13 mutated to allele 14 and that allele 15 did not suffer a mutation. However, this might not be the case. Indeed, allele 13 may have mutated to 15 and allele 15 to allele 14, and assuming the first, we are underestimating both mutation and multi-step mutation rates. This is, however, the only marker where there is ambiguity in which allele in the father originated which allele in the son.

Two of the 1,178 mutations found were mutations between alleles from different microvariant groups, one from 14 to 14.2 in marker DYS19 and one from 36 to 36.2 in marker DYS518.

The obtained locus-specific mutations rates varied between 0.00000 at DYS435 (0.95 CI 0.00000-0.02478), DYS461 (0.95 CI 0.00000-0.00371), DYS643 (0.95 CI 0.00000-0.00536), GATA C4 (0.95 CI 0.00000-0.03052), YCAIab (0.95 CI 0.00000-0.00074) and DXYS156-Y (0.00000-0.00359), and 0.01741 at DYS547 (0.95 CI 0.01314-0.02260), with an overall average mutation rate of 0.00322 (0.95 CI 0.00304-0.00341). However, it is worth noting that despite markers GATA C4 and DYS435, e.g., having a mutation rate of 0.00000, they also have the highest confidence interval amplitude (0.03052 and 0.02478, respectively). So, even though the markers might present a lower mutation rate than the average and appear to be less mutable, there is very little statistical confidence in this estimation due to the small number of meiotic transfers (the average for these markers is 226). Regrettably, this is the case for many markers. YCAIab, on the contrary, has also a null mutation rate estimate, but a confidence interval amplitude of 0.00074, since the total

of meiotic transfers gathered for this marker were 4999, twenty-two times more than GATA C4 and DYS435 (see **Table 3**). Hence, the confidence in the estimation of marker mutation rates varies highly between all markers analyzed, even when considering markers present in commercial kits, e.g., DYS458 has three times more confidence interval amplitude than DYS19, both included in the YFiler® commercial kit, and thus there is more statistical confidence in the DYS19 marker estimations.

From the 1,178 mutations, 598 were gains and 578 losses of repeats. A Chi-square test (significant level equal to 0.05) was computed to measure the statistical significance of these differences for each marker. As can be seen in **Table A1** in the Appendix, no statistically significant differences between the overall markers' gains and losses were found ($p = 0.55975$). When computing this analysis per marker, differences were statistically significant for marker DYS385 ($p = 0.00003$) and DYS626 ($p = 0.03959$), and marginally non significant for markers DYS438 ($p = 0.05878$) and DYS635 ($p = 0.05551$) (see **Table A1**). Indeed, 27 out of the 34 markers where mutations were found revealed a p-value greater than 0.20, when the number of gains and losses of repeats were compared. To compute this Chi-square goodness of fit test, an Excel mathematical formula was used.

3.2. Structure approach

To explore the hypothesis of the variable motif of a marker having influence on mutation mechanisms and frequencies, a structure approach ensued. The structure approach differs from the previous as it aims to study mutation rates and mechanisms, gathering information from markers with the same variable motif.

3.2.1. Material and Methods

The first step was to describe the sequence and repetitive motif of each marker to then cluster them accordingly (see **Table 4**).

Next, markers were organized into 5 clusters: [GATA], [GAAA], [GAAAA], [GAA] and [ATT] markers, according to their repetitive motif (see **Table 5**).

Due to their complex structure, DYS389II, DYS526_B, DYS547 and DYS635 markers were excluded from the analysis, since they harbor different repeat structures varying in number, which does not allow to discern the number of repeats in each motif, based on the length of the allele.
Markers DYS435, DYS448 and DYS526_A were not included in the analysis because their repetitive motif did not match any others.
Clopper-Pearson confidence intervals (CI) were estimated, assuming a level of confidence of 0.05.

Table 4: Structure of the considered Y-STR markers.

Marker	Sequence	Repetitive Motif	Source
DYS19	[TAGA]3TAGG[TAGA]7-15	[TAGA]7-15	http://www.cstl.nist.gov/biotech/strbase/str_y19.htm
DYS385	[GAAA]7-28	[GAAA]7-28	http://www.cstl.nist.gov/biotech/strbase/str_y385.htm
DYS388	[ATT]10-16	[ATT]10-16	http://www.cstl.nist.gov/strbase/str_y388.htm
DYS389I	[TCTG]3[TCTA]6-13	[TCTA]6-13	http://www.cstl.nist.gov/biotech/strbase/str_y389.htm
DYS389II	[TCTG]v[TCTA]t[TCTG]3[TCTA]8		http://www.cstl.nist.gov/biotech/strbase/str_y389.htm
DYS390	$[TCTG]_n$ $[TCTA]_m[TCTG]_p[TCTA]_q$		http://www.cstl.nist.gov/biotech/strbase/str_y390.htm
DYS391	[TCTA]6-14	[TCTA]6-14	http://www.cstl.nist.gov/biotech/strbase/str_y391.htm
DYS392	[TAT]6-17	[TAT]6-17	http://www.cstl.nist.gov/biotech/strbase/str_y392.htm
DYS393	[AGAT]9-17	[AGAT]9-17	http://www.cstl.nist.gov/biotech/strbase/str_y393.htm
DYS435	[TGGA]n	[TGGA]9-13	http://lobstr.teamerlich.org/ystr-codis.html
DYS437	[TCTA]7-11[TCTG]2[TCTA]4	[TCTA]7-11	http://www.cstl.nist.gov/biotech/strbase/str_y437.htm
DYS438	[TTTTC]6-14	[TTTTC]6-14	http://www.cstl.nist.gov/biotech/strbase/str_y438.htm
DYS439	[GATA]9-14	[GATA]9-14	http://www.cstl.nist.gov/biotech/strbase/str_y439.htm
DYS448	[AGAGAT]11-13N42[AGAGAT]8-9	[AGAGAT]11-13, [AGAGAT]8-9	http://www.cstl.nist.gov/biotech/strbase/str_y448.htm
DYS449	[TTTC]n N50 [TTTC]p	[TTTC]22-40	D'Amato *et al.*, 2010
DYS456	[AGAT]12-18	[AGAT]12-18	http://www.cstl.nist.gov/biotech/strbase/str_y456.htm
DYS458	[GAAA]13-20	[GAAA]13-20	http://www.cstl.nist.gov/biotech/strbase/str_y458.htm
DYS460	[ATAG]7-13	[ATAG]7-13	http://www.cstl.nist.gov/biotech/strbase/str_y460.htm
DYS461	[TAGA]9-13	[TAGA]9-13	http://lobstr.teamerlich.org/ystr-codis.html
DYS481	[CTT]17-32	[CTT]17-32	D'Amato *et al.*, 2010
DYS518	(AAAG)15 (GGAG)1 (AAAG)4 N6 (AAAG)13	[AAAG]32-49	D'Amato *et al.*, 2010
DYS526_A	[CCTT]n	[CCTT]n	Butler, 2014
DYS526_B	[CTTT]o[CCTT]pN113[CCTT]n		Butler, 2014
DYS533	[ATCT]7-15	[ATCT]7-15	http://lobstr.teamerlich.org/ystr-codis.html
DYS547	[CCTT]n T(CTTC)oN56(TTTC)pN10(CCTT)4(TCTC)1(TTTC)q		Butler, 2014
DYS549	[GATA]10-15	[GATA]10-15	http://www.cstl.nist.gov/div831/strbase/training/ISHI2014-ChrYinfo.pdf
DYS570	[TTTC]10-25	[TTTC]10-25	http://www.cstl.nist.gov/div831/strbase/training/ISHI2014-ChrYinfo.pdf
DYS576	[AAAG]11-23	[AAAG]11-23	http://www.cstl.nist.gov/div831/strbase/training/ISHI2014-ChrYinfo.pdf

DYS612	(CCT)5 (CTT)1 (TCT)4(CTT)1(TCT)15-29	[TCT]15-29	D'Amato *et al.*, 2010
DYS626	(AAAG)15-22 (AGAA)2 (AGAG)1 (GAAG)3 (AAAG)3	[AAAG]15-22	D'Amato *et al.*, 2010
DYS627	[AAAG]11-27	[AAAG]11-27	http://www.cstl.nist.gov/div831/strbase/training/ISHI2014-ChrYinfo.pdf
DYS635	[TCTA]4(TGTA)2[TCTA]2(TGTA)2[TCTA]2(TGTA)0,2[TCTA]n		http://www.cstl.nist.gov/strbase/str_y635.htm
DYS643	[CTTTT]6-17	[CTTTT]6-17	http://www.cstl.nist.gov/div831/strbase/training/ISHI2014-ChrYinfo.pdf
GATA A10	[TCCA]2[TAGA]13-18	[TAGA]13-18	http://www.cstl.nist.gov/strbase/str_ya10.htm
GATA H4	[TAGA]8-13 N12 [GATC]2 AA [TAGA]4	[TAGA]8-13	http://www.cstl.nist.gov/biotech/strbase/str_yh4.htm

Table 5: Composition of the clusters.

Clusters	Markers
GATA	DYS19, DYS389I, DYS390, DYS391, DYS393, DYS437, DYS439, DYS456, DYS460, DYS461, DYS533, DYS549, GATA H4 and GATA A10
GAAA	DYS385, DYS449, DYS458, DYS518, DYS570, DYS576, DYS626 and DYS627
GAAAA	DYS438 and DYS643
GAA	DYS481 and DYS612
ATT	DYS388 and DYS392

3.2.2. Results and Discussion

Considering the markers clustered into categories defined by the structure of the repetitive motif, mutations rates and the respective confidence intervals were computed (significance level equal to 0.05), see **Table 6**.

Comparing the two tetra-nucleotide clusters, the [GAAA] cluster has a mutation rate 3.1 times higher than [GATA], having the respective confidence intervals no intersection. It is also remarkable the fact that the [GAAA] multi-step mutation rate is 45 times higher than the one in the [GATA] cluster, again with no intersection of the confidence intervals.

The cluster with the highest mutation rate is the [GAA] cluster: 0.01451 (0.95 CI 0.01114 – 0.01856). This is also the case for the single-step mutation rate: 0.01287 (0.95 CI 0.00971-0.01672) and multi-step mutations rate: 0.00164 (0.95 CI 0.00066-0.00337). Comparing this cluster to the other tri-nucleotide cluster, [ATT], there are major differences in mutation rates and respective confidence intervals. Additionally, we noticed that of the two markers included in the [GAA] cluster, the DYS612 marker has a mutation rate 2.30 times higher than DYS481, being responsible for the high mutation rate of the cluster (see **Table 4** and **Table 6**). Indeed, computing the Chi-squared test for these four markers (within and between clusters – see **Table 7**) we can see that it is expected that the differences of the mutations on markers DYS481 and DYS612 are not due to the sampling size ($p = 0.0333$), which is in concordance to the fact that confidence intervals shown in **Table 3** do not intersect.

Table 6: Types of mutation, their mutation rates and confidence intervals (0.95) for each cluster.

	Repetitive Motif				
	GATA	**GAAA**	**GAAAA**	**ATT**	**GAA**
Markers	14	8	2	2	2
Allele Transfers	174,598	70,548	14,556	25,275	4,273
Mutations	390	488	7	11	62
Mutations ≠ 1 step	2	32	5	2	7
Mutation Rate	0.00223	0.00692	0.00048	0.00044	0.01451
Confidence Interval	0.00202-0.00247	0.00632-0.00756	0.00019-0.00099	0.00022-0.00078	0.01114-0.01856
Mutation rate = 1 step	0.00222	0.00646	0.00014	0.00036	0.01287
Confidence Interval	0.00200-0.00245	0.00587-0.00707	0.00002-0.00050	0.00016-0.00068	0.00971-0.01672
Mutation rate ≠ 1 step	0.00001	0.00045	0.00034	0.00008	0.00164
Confidence Interval	0.00000-0.00004	0.00031-0.00064	0.00011-0.00080	0.00001-0.00029	0.00066-0.00337

Table 7: Chi-square test for markers belonging to clusters [GAAAA] and [GAA].

		[GAAAA]		[GAA]	
		DYS438	DYS643	DYS481	DYS612
[GAAAA]	DYS438	-	-	-	-
	DYS643	0.8224	-	-	-
[GAA]	DYS481	< 0.0001	0.8020	-	-
	DYS612	< 0.0001	0.3441	0.0333	-

For this reason, we decided to analyse the flanking region of the markers theorizing that perhaps if the preceding flanking region of the repetitive motif mostly consisted of the same nucleotides as the repetitive motif, then that could lead to more replication slippage (see **Image 5**). Indeed, we observed that the marker DYS612 has a preceding flanking region constituted by the same nucleotides as the repetitive motif, unlike any of the other markers.

DYS481 [GAA]

>chrY:8558308+8558428 121bp GAATGTGGCTAACGCTGTTC
TCACCAGAAGGTTGCAAGAC
GAATGTGGCTAACGCTGTTCagcatgctg[**cttcttcttcttcttcttctt**
Cttcttcttcttcttcttcttcttcttcttcttcttcttcttctt]ttttg
aGTCTTGCAACCTTCTGGTGA

DYS612 [GAA]

>chrY:13640701+13640883 183bp GAAGTTTCACACAGGTTCAGAGG
AAAAAGGGAACTGAGGGAAGG
GAAGTTTCACACAGGTTCAGAGGtttgcctcctcctcctcctctttcttc
Ttcttctc[**cttcttcttcttcttcttcttcttcttcttcttcttcttctt**
Cttcttcttcttcttcttcttcttcttctt]ctgtcacttttccaaat
tattttcttttgCCTTCCCTCAGTTCCCTTTTT

DYS643 [GAAAA]

>chrY:15314105+15314249 145bp AAGCCATGCCTGGTTAAACT
TGTAACCAAACACCACCCATT
AAGCCATGCCTGGTTAAACTactgtgc[**cttttcttttcttttcttttctt**
Ttctttttctttttctttttctttttctttt]ctttctttttaaaacttt
ttacttcagtagaattttgggggggAATGGGTGGTGTTTGGTTACA

DYS438 [GAAAA]

>chrY:12825850+12826179 330bp CCAAAATTAGTGGGGAATAGTTG
GATCACCCAGGGTCTGGAGTT
CCAAAATTAGTGGGGAATAGTTGaacggtaaacagtatatttt[**cttttct**
Tttcttttcttttcttttcttttcttttcttttcttttcttttctttt]ct
atttgaaatggagtttcactcttgttgcccaggctgaaatgcaatggtgt
gatctcgactcaccacaacctccacttcccaggttcaagcgattctcctg
catcagcctcccaggtagctgggattataggcgtctgccaccacgcccag
ctaatttttgtgtttttagtagagacagggtttcaccatgttggtgagg
ctggtctcgAACTCCAGACCCTGGGTGATC

Image 5: Flanking regions of markers DYS481, DYS612, DYS643 and DYS438.

Also for markers with [GATA] and [GAAA] repetitive motifs, when a Chi-square test (significance level equal to 0.05) was performed to measure the pairwise differences considering the outcomes: mutation and no mutation, for any pair of markers within and between [GATA] and [GAAA] clusters, statistically significant differences were found both within and between clusters.

For the [GATA] cluster (see **Table 8**) 29 out of 91 of the pairwise comparisons revealed statistically significant differences. DYS393, e.g., has statistically significant differences with 9 out of the 13 remaining alleles of the cluster.

On the other hand, for the markers with repetitive motif with structure [GAAA] (see **Table 10**) 16 out 28 of the pairwise comparisons revealed statistically significant differences. DYS385, e.g., has statistically significant differences with all the other markers from the cluster.

Finally, comparing markers belonging to different clusters: [GATA] and [GAAA] (see **Table 10**), we can verify that 91 out of the 112 pairwise comparisons shown statistically significant differences.
Analyzing **Tables 8** and **Table 9** it seems that four markers from the [GATA] cluster (DYS393, DYS439, DYS456, DYS460) and two markers from the [GAAA] cluster (DYS385, DYS458), accumulate more differences than the others. Indeed, excluding these six markers from the analyses, the differences between markers belonging to different clusters and the similarities of those belonging to the same is clear (see **Table 11**).

Facing these results, we considered only the markers described in **Table 11**, and recalculated the estimates presented in **Table 6** (see **Table 12**). The results still showing that [GAAA] markers have a higher trend to mutate, inclusively by more than one step.

Table 8: Chi-square test for markers belonging to [GATA] cluster.

	DYS19	DYS389I	DYS390	DYS391	DYS393	DYS437	DYS439	DYS456	DYS460	DYS461	DYS533	DYS549	GATA A10	GATA H4
DYS19	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DYS389I	0.7234	-	-	-	-	-	-	-	-	-	-	-	-	-
DYS390	0.5416	0.291	-	-	-	-	-	-	-	-	-	-	-	-
DYS391	0.9291	0.8688	0.4254	-	-	-	-	-	-	-	-	-	-	-
DYS393	0.0038	0.001	0.0265	0.0022	-	-	-	-	-	-	-	-	-	-
DYS437	0.0847	0.0375	0.258	0.061	0.5117	-	-	-	-	-	-	-	-	-
DYS439	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	-	-	-	-	-	-	-	-
DYS456	0.0029	0.0107	0.0003	0.0051	<0.0001	<0.0001	0.4394	-	-	-	-	-	-	-
DYS460	0.0209	0.042	0.006	0.0282	<0.0001	0.0005	0.9424	0.6972	-	-	-	-	-	-
DYS461	0.2853	0.25	0.3492	0.2702	0.6599	0.5272	0.0591	0.0914	0.067	-	-	-	-	-
DYS533	0.8545	0.727	0.8464	0.8003	0.6575	0.966	0.0879	0.1778	0.1371	0.5185	-	-	-	-
DYS549	0.545	0.5914	0.4714	0.5643	0.2183	0.3214	0.9778	0.8861	0.4824	0.1667	0.5106	-	-	-
GATA A10	0.1429	0.2029	0.0754	0.1662	0.0021	0.016	0.9603	0.881	0.9994	0.0868	0.2316	0.5027	-	-
GATA H4	0.9456	0.9419	0.5044	0.9759	0.0065	0.0929	0.001	0.0207	0.0449	0.2729	0.8077	0.5739	0.1911	-

Note: P-values below 0.05 are indicated in red.

Table 9: Chi-square test for markers belonging to the [GAAA] cluster.

	DYS385	DYS449	DYS458	DYS518	DYS570	DYS576	DYS626	DYS627
DYS385	-	-	-	-	-	-	-	-
DYS449	<0.0001	-	-	-	-	-	-	-
DYS458	<0.0001	0.001	-	-	-	-	-	-
DYS518	<0.0001	0.0518	<0.0001	-	-	-	-	-
DYS570	<0.0001	0.257	0.0856	0.0014	-	-	-	-
DYS576	<0.0001	0.3665	<0.0001	0.3265	0.0311	-	-	-
DYS626	<0.0001	0.4925	0.05	0.0108	0.8195	0.1075	-	-
DYS627	<0.0001	0.299	<0.0001	0.4000	0.022	0.9592	0.0833	-

Note: P-values below 0.05 are indicated in red.

Table 10: Chi-square test for markers belonging to the [GATA] cluster vs. markers belonging to the [GAAA] cluster.

	DYS385	DYS449	DYS458	DYS518	DYS570	DYS576	DYS626	DYS627
DYS19	0.955	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DYS389I	0.5799	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DYS390	0.5936	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DYS391	0.7918	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DYS393	0.0033	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DYS437	0.0878	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DYS439	<0.0001	<0.0001	0.0087	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DYS456	0.0007	<0.0001	0.0013	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DYS460	0.0133	0.0022	0.241	<0.0001	0.0306	0.0002	0.0193	<0.0001
DYS461	0.2944	0.0006	0.0138	<0.0001	0.003	0.0002	0.002	<0.0001
DYS533	0.8883	<0.0001	0.0078	<0.0001	0.0007	<0.0001	0.0004	<0.0001
DYS549	0.5287	0.7689	0.7643	0.7641	0.9676	0.9376	0.9057	0.914
GATA A10	0.1207	0.0357	0.4702	0.0019	0.1421	0.0094	0.105	0.0077
GATA H4	0.835	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Note: P-values below 0.05 are indicated in red.

Table 11: Chi-square re-analysis removing the six markers that accumulate more differences.

		[GAAA] markers						[GATA] markers									
		DYS449	DYS518	DYS570	DYS576	DYS626	DYS627	DYS19	DYS389I	DYS390	DYS391	DYS437	DYS461	DYS533	DYS549	GATA A10	GATA H4
[GAAA] markers	DYS449	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	DYS518	0.0518	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	DYS570	0.257	0.0014	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	DYS576	0.3665	0.3265	0.0311	-	-	-	-	-	-	-	-	-	-	-	-	-
	DYS626	0.4925	0.0108	0.8195	0.1075	-	-	-	-	-	-	-	-	-	-	-	-
	DYS627	0.299	0.4000	0.022	0.9592	0.0833	-	-	-	-	-	-	-	-	-	-	-
[GATA] markers	DYS19	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	-	-	-	-	-	-	-	-	-	-
	DYS389I	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.7234	-	-	-	-	-	-	-	-	-
	DYS390	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.5416	0.291	-	-	-	-	-	-	-	-
	DYS391	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.9291	0.8688	0.4254	-	-	-	-	-	-	-
	DYS437	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0847	0.0375	0.258	0.061	-	-	-	-	-	-
	DYS461	0.0006	<0.0001	0.003	0.0002	0.002	<0.0001	0.2853	0.25	0.3492	0.2702	0.5272	-	-	-	-	-
	DYS533	<0.0001	<0.0001	0.0007	<0.0001	0.0004	<0.0001	0.8545	0.727	0.8464	0.8003	0.966	0.5185	-	-	-	-
	DYS549	0.7689	0.7641	0.9676	0.9376	0.9057	0.914	0.545	0.5914	0.4714	0.5643	0.3214	0.1667	0.5106	-	-	-
	GATA A10	0.0357	0.0019	0.1421	0.0094	0.105	0.0077	0.1429	0.2029	0.0754	0.1662	0.016	0.0868	0.2316	0.5027	-	-
	GATA H4	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.9456	0.9419	0.5044	0.9759	0.0929	0.2729	0.8077	0.5739	0.1911	-

Table 12: Re-analysis of the clusters removing the six markers that accumulate more differences.

	Repetitive Motif	
	GATA	**GAAA**
Markers	10	6
Allele Transfers	124207	23804
Mutations	246	334
Mutations ≠ 1	2	24
Mutation rate	0.00198	0.01403
Confidence interval	0.00174-0.00224	0.01258-0.01561
Mutation rate = 1	0.00196	0.01302
Confidence interval	0.00172-0.00222	0.01126-0.01415
Mutation rate ≠ 1	0.00002	0.00101
Confidence interval	0.00000-0.00006	0.00092-0.00190

Summarizing, the results show that the structure of the marker may influence the mutation rate but it seems not to be the only factor since, there are differences between markers from the same cluster and similarities between markers belonging to different clusters. On this regard, we considered other factors such as the location on the chromosome (see **Image 6**), the flanking regions and the expected allele's length, not reaching concluding inferences. For example, two markers of the [GAAA] cluster: DYS570 and DYS576, have the same simple structure of [GAAA] repetitive motif, their adjacent regions are similar, they are located in the same chromosome region, they have different expected allele length (but this is the case for other markers with no statistically significant differences between them as well), and they show statistically significant differences between them in what concerns the event of mutation ($p = 0.0311$).

Nevertheless, the results suggest that the repeat sequence influences the number of steps involved in the mutational event, [GAAA] being more prone to multi-step mutations than [GATA] repeats.

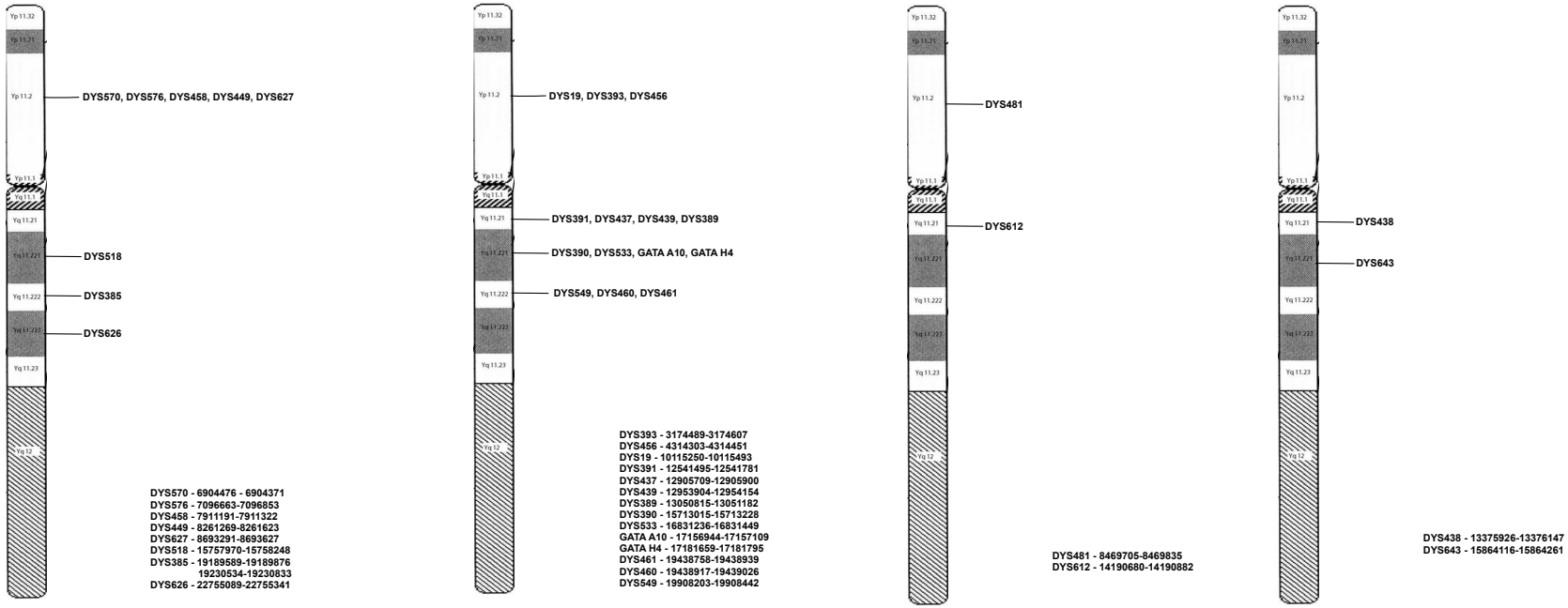


Image 6: Markers' location on the Y-Chromosome: [GAAA] cluster, [GATA] cluster, [GAAAA] cluster and [GAA] cluster.

3.3. Bi-allele Approach

The bi-allele approach differs from the previous approaches because it takes into consideration the paternal allele (initial or original allele) and the filial allele (final or destination allele) involved in the allele transfer, assuming that different alleles of the same marker can have different mutation rates. In this work, we intend to apply the framework presented in Pinto *et al.* (2014) "(…) for forensic casework the relevant parameter for incorporation in a likelihood ratio is biallelic specific, i.e., the mutation rate estimate corresponds to the probability of the specific allelic transition observed (…)".

3.3.1. Material and Methods

For this approach, haplotypic information is essential to evaluate allele transfer and mutation frequencies, since it is essential to acknowledge not only the alleles that mutated but also those that did not, being possible to infer, this way, the allele and bi-allele mutation rates.
We started the collection by posting in the YHRD site (https://yhrd.org/pages/Projects/P1) a request for collaboration. The requested data must include the haplotype information and the non-mutated transmissions per allele, as this information is essential to test STR mutation models (Pinto *et al.*, 2014) (see **Image 7**).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Code	Population	Age	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448
2	Father_1	A	31	16	13	24	29	17	14	1319	12	10	11	22	11	11	15	9	20
3	Son_1	A	28	15	13	23	29	16	15	1316	12	10	11	23	11	11	15	9	21
4	Father_2	A	31	15	13	24	29	16	15	1114	12	11	13	23	13	11	15	12	19
5	Son_2	A	70	15	13	23	29	17	14	1114	13	11	11	23	13	12	14	12	19
6	Father_3	A	18	15	12	22	29	16	15	14	14	10	11	20	11	12	16	10	21
7	Son_3	A	19	17	13	23	29	16	13	1516	14	10	13	20	13	12	15	9	19
8	Father_4	A	24	16	13	24	30	17	14	1114	13	11	12	23	13	12	15	13	19
9	Son_4	A	45	16	13	24	29	18	15	1014	13	10	12	24	13	12	15	12	19
10	Father_5	A	30	17	14	24	30	17	14	1214	13	11	12	23	13	12	15	12	19
11	Son_5	A	36	17	13	24	30	15	13	1719	13	11	11	21	11	11	14	10	20

Image 7: Requested format for Y-haplotypic information, posted in the YHRD site in the project page as a request for collaboration.

Since complete haplotypic information is rarely included in published works, the authors of several works on Y-STRs mutation rates were directly contacted to contribute to the project, by sending the haplotypic data they used in their research. At this point we should remark that although it was our goal to study father's age influence in mutation rates, we could not

gather enough information on that. Detailed information on haplotypic data analyzed can be found in the **Table 13**.

Table 13: Data collected with complete haplotypic information.

Authors, date	Study	Father/son duos	Origin
Wang *et al.*, 2016	Genetic polymorphisms and mutation rates of 27 Y-chromosomal STRs in a Han population from Guangdong Province, Southern China. *Forensic Science International: Genetics*, *21*, 5-9.	981*	China
Turrina *et al.*, 2006	Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay. *International journal of legal medicine*, *120*(1), 56-59.	104**	Italy
Oh, Y. *et al.*, 2015	Haplotype and mutation analysis for newly suggested Y-STRs in Korean father–son pairs. *Forensic Science International: Genetics*, *15*, 64-68.	355*	Korea
Rukhsana Parveen	Not published	110	Pakistan
Tsai, L. *et al.*, 2002	Haplotype frequencies of nine Y-chromosome STR loci in the Taiwanese Han population. *International journal of legal medicine*, *116*(3), 179-183.	101*	Taiwan
Berger, B. *et al.*, 2005	Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay. *International journal of legal medicine*, *119*(4), 241-246.	70	Austria
Ballantyne, K. *et al.*, 2014	Toward Male Individualization with Rapidly Mutating Y-Chromosomal Short Tandem Repeats. *Human mutation*, *35*(8), 1021-1032.	2378	44 countries from Africa, America, Asia, Europe and Oceania
Antão-Sousa, 2017	Mutation rates and segregation data on 16 Y-STRs: an update to previous GHEP-ISFG studies, *Forensic Science International: Genetics Supplement Series, under review*	1598***	Brazil, Portugal, Argentina, Spain, Turkey and Colombia
Gusmão *et al.*, 2005	Mutation rates at Y chromosome specific microsatellites. *Human mutation*, *26*(6), 520-528.	3026	Argentina, Brazil, Colombia, Portugal, Spain
Sánchez-Diz *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), 529-533.	701	Argentina, Brazil, Portugal
Robino *et al.*, 2015	Development of an Italian RM Y-STR haplotype database: results of the 2013 GEFI collaborative exercise. *Forensic Science International: Genetics*, *15*, 56-63.	409	Italy
TOTAL		9833	

*Excluding fathers with more than one son
**Author sent extra data that is not included in the publication
*** New data only

Excepting one sample of 110 duos from Pakistan, and one of 54 from Italy, all the analyzed data is published in international peer reviewed journals.

Nevertheless, because data from Pakistan is not published, FST genetic distances and the corresponding non-differentiation probabilities were calculated on Arlequin 3.5 software.

The genetic distance between the population studied in this work and other population samples from Pakistan did not reveal statistically significant differences.

New haplotypic information was analyzed under a framework of a collaborative exercise of the GHEP-ISFG working group and submitted as a proceeding, as previously mentioned in 3.1.1., for the 17th Conference Volume of the International Society for Forensic Genetics.

Incongruities between published data and supplementary material were found in Domingues, P. M., *et al.* (2007). According to the article one of the two mutations that occurred in DYS385 was from allele 19 to 18 (N106 in Table S1), however N106 genotype for this marker is 11-14.

Data was organized using Microsoft Excel. For each marker, a document was created, comprising a sheet with a mutation matrix showing allele transfers for each studied population and one summarizing the information from all the populations (see **Table 14** as an example). The first column of each table represents the parental allele and the first row the filial alelle. This way, studying allele transfers and mutations is straightforward. For example, for the 3960 allele transfers in allele 14, eight suffered a mutation: one mutated to allele 14.2, one to allele 13 and six mutated to allele 15 – see **Table 14**.

Table 14: Mutation matrix for marker DYS19.

Filial allele / Paternal allele	11	12	13	14	14.1	14.2	14.3	15	16	17	18	19	Total
11	90	0	0	0	0	0	0	0	0	0	0	0	90
12	0	8	1	0	0	0	0	0	0	0	0	0	9
13	0	0	967	1	0	0	0	0	0	0	0	0	968
14	0	0	1	3952	0	1	0	6	0	0	0	0	3960
14.1	0	0	0	0	1	0	0	0	0	0	0	0	1
14.2	0	0	0	0	0	0	0	0	0	0	0	0	0
14.3	0	0	0	0	0	0	1	0	0	0	0	0	1
15	0	0	0	0	0	0	0	2383	4	0	0	0	2387
16	0	0	0	0	0	0	0	1	1142	2	0	0	1145
17	0	0	0	0	0	0	0	0	5	502	1	0	508
18	0	0	0	0	0	0	0	0	0	0	4	0	4
19	0	0	0	0	0	0	0	0	0	0	0	1	1
Total allele tranfers													9074

After organizing all data in mutation matrices as shown in **Table 14** for the 35 markers, the total allele transfers and mutations for each marker were computed resorting to mathematical formulae.

Moreover, the allelic length of GATA H4 marker is around 28 because the primers amplify two loci: H4.1 and H4.2, the latter being non-polymorphic in humans. For this study only the polymorphic H4.1 locus is relevant and the allelic range of some data had to be adjusted.

In case of DYS385, where two loci are simultaneously amplified with one pair of primers, the number of allele transfers was doubled.

To estimate mutation rates (allele and bi-allele), the total number of Mendelian incompatibilities found was divided by the total number of allele transfers. Clopper-Pearson confidence intervals (CI) for mutation rates were estimated, assuming a level of confidence of 0.05.

Tables analogous to **Table 15** were built for each marker, including estimates for allele and bi-allele mutation rates and their respective 0.95 confidence intervals. These tables are presented in the **Appendix**.

Noting that some markers have a higher number of observations than others, we evaluated the possibility of modal alleles of specific (more studied) markers having more observations than some (less studied) markers (**Table 16)**.

On the other hand, by analyzing the mutation matrices we noted that alleles tend to mutate to the most frequent adjacent allele (belonging to the same microvariant group). Thus, we evaluated the difference on the occurrence of gains and losses, in a per allele perspective, classifying the alleles in three disjunctive categories: alleles shorter than the modal allele, the modal allele, and alleles longer than the modal allele (**Table 17**).

Finally, and since confidence intervals for allele and bi-allele estimates are wide, we gathered the information concerning markers where the repetitive motif is [GATA] expecting to gain statistical power. We used the [GATA] cluster because it is the one with more observations.

Table 15: Allele and bi-allele mutation rates for marker DYS19.

Marker	No. of mutations	No. of meiosis	Original allele	Fillial allele	No. of observations	Frequency of the paternal allele	Allele mutation rate	Confidence interval (0.95)	Bi-allele mutation rate	Confidence interval (0.95)
DYS19	23	9,074	11	11	90	0.00992	0.00000	0.00000-0.04016		
			12	12	8	0.00099	0.11111	0.00281-0.48250		
				13	1				0.11111	0.00281-0.48250
			13	13	967	0.10668	0.00103	0.00003-0.00574		
				14	1				0.00103	0.00003-0.00574
			14	13	1	0.43641	0.00202	0.00087-0.00398	0.00025	0.00001-0.00141
				14	3,952					
				14.2	1				0.00025	0.00001-0.00141
				15	6				0.00152	0.00056-0.00329
			14.1	14.1	1	0.00011	0.00000	0.00000-0.97500		
			14.3	14.3	1	0.00011	0.00000	0.00000-0.97500		
			15	15	2,383	0.26306	0.00168	0.00046-0.00428		
				16	4				0.00168	0.00046-0.00428
			16	15	1	0.12618	0.00262	0.00054-0.00764	0.00087	0.00002-0.00486
				16	1,142					
				17	2				0.00175	0.00021-0.00630
			17	16	5	0.05598	0.01181	0.00435-0.02553	0.00984	0.00320-0.02282
				17	502					
				18	1				0.00197	0.00005-0.01092
			18	18	4	0.00044	0.00000	0.00000-0.60236		
			19	19	1	0.00011	0.00000	0.00000-0.97500		

Table 16: Comparison of quantity of data without and with haplotypic information.

	Published Data (**Table 2**)		Haplotypic Data (**Table 13**)				
Marker	No. of mutations	Allele transfers	No. of mutations	Allele transfers	Modal Allele*	Frequency of the modal allele	Number of markers with less published observations than the modal allele (haplotypic data)
DYS19	49	23,728	23	9,074	14	3,960	16
DYS385	73	35,378	29	15,370	14	3,755	16
DYS388	2	2,017	1	101			
DYS389I	51	22,555	22	8,084	13	4,444	21
DYS389II	84	22,501	40	8,047	29	2,928	11
DYS390	42	23,633	19	9,015	24	3,865	16
DYS391	50	23,308	24	9,108	10	5,321	23
DYS392	9	23,258	7	9,092	13	3,593	16
DYS393	21	21,889	11	7,861	13	5,087	23
DYS413	1	4,999	0	0			
DYS435	0	147	0	147			
DYS437	18	14,431	10	7,335	14	3,557	16
DYS438	7	14,452	4	7,398	10	2,663	11
DYS439	68	14,821	33	7,355	12	3,089	11
DYS448	12	11,334	3	6,040	19	2,286	10
DYS449	52	4,032	52	4,032			
DYS456	44	11,339	22	6,036	15	2,763	11
DYS458	81	11,366	45	6,043	17	1,728	8
DYS460	11	2,342	10	2,002			
DYS461	0	992	0	873			
DYS481	8	1,085	8	1,085			
DYS518	74	3,990	74	3,990			
DYS526A	10	3,119	10	3,119			
DYS526B	43	3,173	43	3,173			
DYS533	3	1,895	3	1,895			
DYS547	55	3,159	55	3,159			
DYS549	1	104	1	104			
DYS570	43	4,293	43	4,293	18	1,107	8
DYS576	65	4,194	65	4,194	18	1,367	8
DYS612	54	3,188	54	3,188			
DYS626	34	3,138	34	3,138			

DYS627	66	4,157	66	4,157			
DYS635	35	11,997	20	6,916	23	2,931	11
DYS643	0	104	0	104			
GATA A10	5	1,065	4	874			
GATA C4	0	119	0	0			
GATA H4	27	12,496	13	7,128	12	3,753	16
YCAIab	0	4,999	0	0			
YCAIIab	3	7,055	0	0			
DXYS15Y	0	1,027	0	0			
TOTAL	1201	362,879	848	173,530			

*Of markers with frequency greater than the median (4194)

Table 17: Repeat gains and losses considering alleles classified in three disjunctive categories: alleles shorter than the modal allele, in the modal allele, and in alleles longer than the modal allele.

Marker	Modal allele	Allele < than modal		Modal allele		Allele > than modal		N
		Gains	Losses	Gains	Losses	Gains	Losses	
DYS19	14	2	0	1	6	7	6	9,074
DYS385	14	8	2	4	1	7	5	15,370
DYS388	12	1	0	0	0	0	0	101
DYS389I	13	4	1	5	1	1	10	8,084
DYS389II	29	3	0	11	3	8	15	8,047
DYS390	24	0	1	6	4	0	8	9,105
DYS391	10	1	0	5	0	7	11	9,107
DYS392	13	0	1	3	0	2	0	9,092
DYS393	13	2	0	3	2	1	3	7,861
DYS435	11	0	0	0	0	0	0	147
DYS437	14	0	0	2	0	3	5	7,335
DYS438	10	0	0	2	0	0	2	7,398
DYS439	12	7	0	5	1	6	14	7,355
DYS448	19	0	0	1	1	0	1	6,040
DYS449	31	7	8	4	3	14	16	4,032
DYS456	15	1	0	6	1	5	9	6,036
DYS458	17	7	5	7	8	11	7	6,043
DYS460	11	1	3	1	3	0	2	2,002
DYS461	12	0	0	0	0	0	0	873
DYS481	23	1	0	1	0	5	1	1,085
DYS518	38	10	3	7	4	25	25	3,990
DYS526_A	14	3	1	2	1	1	2	3,119
DYS526_B	37	8	6	4	2	12	11	3,173
DYS533	11	0	0	0	0	1	2	1,085
DYS547	48	13	7	4	4	5	22	3,159
DYS549	12	0	0	0	0	0	1	104
DYS570	18	8	4	4	6	6	15	4,293
DYS576	18	13	1	12	10	10	19	4,194
DYS612	36	16	5	3	3	14	13	3,188
DYS626	30	3	3	1	1	7	19	3,138
DYS627	21	14	7	4	8	9	24	4,157
DYS635	23	3	8	1	3	2	3	6,916
DYS643	10	0	0	0	0	0	0	104
GATA A10	15	1	1	0	2	0	0	874
GATA H4	12	3	0	1	7	0	2	7,128
	TOTAL	140	67	110	85	169	273	172,809

For the statistical analysis, as the goal of this approach is to compare only the motif repeat extension, the non-polymorphic portions of DYS19, DYS389I, DYS437, GATA H4, GATA A10, DYS518 and DYS626 markers were removed. For example, considering marker DYS19 (see **Table 4**) the monomorphic region corresponding to four (tetranucleotide) repeats was subtracted to each allele, that is, instead of an allelic range of 11 to 19, the range was shifted to 7 to 15. Grounded in the same reasoning, to marker DYS626 nine repeats were subtracted, to DYS389I three, to DYS437 six, to DYS518 eleven and to GATA A10 two. Also, any non-consensus alleles, e.g. 14.1, 14.2 or 14.3 in marker DYS19, were not considered in the analysis since the repeat motifs are not complete.

Here we would like to note that, marker DYS390 was removed from the analysis since the NIST STRBase (http://strbase.nist.gov/str_y390.htm) reports the occurrence of a different sequence for allele 21, preventing its use in this approach.

Likewise, if the variable motif is interrupted, it becomes impossible, without sequencing, to distinguish between alleles with the same length but different sequence. For example, in marker DYS449: "[TTTC]10 N50 [TTTC]14" and "[TTTC]13 N50 [TTTC]11" have the same length and thus the same designation: allele 24, but the sequence is different and mutation-wise they might behave differently. So, markers DYS449 and DYS518 were not included in the analysis.

The information from the cluster [GATA]: number of meiosis, number and type of mutations, was then compiled into a table in a Microsoft Excel document (see, e.g., **Table 18**).

Table 18: Example of data organization for the statistical analysis.

Size	Marker	No. of meiosis (NM)	Allele Frequency (AF)	Relative frequency (RF=AF/NM)	No. of mutations (M)	Mutation/Allelic Frequency (M/AF)	Type of mutation
8	DYS19	9017	9	0.000998	1	=0.111111	+1
8	DYS19	9017	9	0.000998	0	=0	0
8	DYS19	9017	9	0..000998	0	=0	-1

3.3.2. Results and Discussion

Table 15 and the other bi-allele tables presented in the **Appendix** show that alleles of the same marker do not have the same mutation rates and, for some, the mutation rate's confidence intervals do not even intersect. For example, in marker DYS389I, allele 13 mutates to allele 12 with a mutation rate of 0.00023 (0.95 CI 0.00001-0.00125) and allele

14 mutates to allele 13 with a mutation rate of 0.00512 (0.95 CI 0.00221-0.01006), the lower bound of the confidence interval for the latter mutational event is almost twice the upper bound of the first mutational event. In this case, the longer allele has a higher mutation rate than the shorter allele. In marker DYS612, allele 32 mutates to allele 33 with a mutation rate of 0.07692 (0.95 CI 0.01615-0.20870) and allele 36 mutates to allele 35 and 37 with a mutation rate of 0.00333 (0.95 CI 0.00069-0.00970), the lower bound of the first mutational event is higher than the upper bound of the second mutational event. In this case, the shorter allele has a higher mutation rate than the longer allele, which is, somehow, contrary to the commonly accepted idea that longer alleles mutate more than shorter alleles.

In the complete set of allele transfers, when the paternal allele mutated to only one of the adjacent alleles (see, e.g., allele 12, in **Table 15**), the allele and the bi-allele mutation rates and their respective confidence intervals coincide. Yet, when this is not the case, as it is in alleles 14, 16 and 17 (see **Table 15**), it is worth noting that the allele mutates more frequently to one of the adjacent alleles than the other. Indeed, in an intra-marker analysis, the mutation phenomenon is not always symmetrical, differing from allele to allele. In allele 14, e.g., a mutation to allele 15 is more frequent than to allele 13. In allele 17, it was 5 times more frequent the mutation to allele 16 than to allele 18 (see **Table 15**).

Indeed, analyzing repeat gains and losses in alleles other than the modal allele, we observe a mutational tendency towards the modal allele, i.e., when mutation occurred in an allele shorter than the modal allele it was 2.1 times more likely to observe a repeat gain (see **Table 17**). From the 35 analyzed markers, this is not the case for 5 and applying the Chi-square test we verified that none of these exceptions were statistically significant (level of significance equal to 0.05). On the other hand, when mutation occurred in an allele longer than the modal it was 1.6 times more likely to observe one repeat loss (see **Table 17**) than one gain. From the 35 analyzed markers, this is not the case for 8 and applying the Chi-square test we verified that none of these exceptions were statistically significant (level of significance equal to 0.05). Whereas when mutation occurred in the modal allele it was 1.3 times more likely to observe a repeat gain (see **Table 17**). From the 35 analyzed markers, this is not the case for 6, except for 6 of the 35 markers studied, and applying the Chi-square test we verified that only one of these

exceptions was marginally statistically significant (level of significance equal to 0.05). Hence, even though, in the analysis per marker it seems that there is a balance between repeat gains and losses (see **Table 3**), in the analysis per allele it is evident a difference in mutation direction depending on the length of the allele.
Concerning marker and allele mutation rates it should be noticed that these values can be significantly different. For example, despite the marker DYS19 having an overall mutation rate of 0.00253 (0.00161-0.00380), the allele (and bi-allele) mutation rates are not, for most alleles, close to this number (see **Table 15**). Therefore, instead of using the marker overall average mutation rate for kinship analyses, it would be more adequate to use allele's specific mutation rates. For this, it is essential to start making available complete haplotypic data. Note that we gathered only 9,833 complete haplotypes from the 20,388 father-son duos used in the publications described in **Table 2**.

Nevertheless, and considering only the information given by the complete haplotypes, in some cases we have better estimates for some specific alleles than for less studied markers. See, e.g., **Table 16** where we show that for 18 of the 40 markers analyzed the number of the observations of the modal allele (resorting just to the haplotypic information we gathered) is greater than the observations of some markers considering all the published data (number of markers with less observations than the modal allele varying between 8 and 23, 20% and 57.5%, respectively). For example, 23 of the 40 markers analyzed (see **Table 16**) have less allele transfers published than the frequency of the modal allele of marker DYS391 (considering only the haplotypic data).

For the statistical analysis considering the structure of the repetitive motif, as mentioned before, only the [GATA] cluster had a sufficiently reasonable sample size to allow for the application of statistical parametric models.
The effect of the allele size on the number of mutations was evaluated by a Poisson hurdle regression model, as the empirical distribution for the number of mutations was seen to be zero-inflated (as zero represents the no-mutation events). In contrast to zero-inflated models, hurdle models treat zero-counts (i.e. number of no mutations) and non-zero (i.e. number of mutations) outcomes as two separate categories, rather than treating the zero-count outcomes as a mixture of structural and sampling zeros. A random effect accounting for the inter-marker variability was also considered.

Firstly, we fit a truncated Poisson distribution to the number of mutations, using the frequency of the alleles as an offset, a random effect by "marker", and the category "size 11" as an explanatory variable (see **Table 19**).

Table 19: Results from the statistical model analyzing the effect of the allele size on the number of mutations.

Variables		Fixed-effects				Random-effects
	Coeficient	Exp(Coef)	Std error	p-value	0.95 CI for Exp(Coef)	Std dev.
Intercept	-6.761	0.001	0.271	<0.001	-----	0.425
Size >=11	1.076	2.933	0.254	<0.001	(1.8, 4.8)	-----
< 11	Reference	-----	-----	-----	-----	-----

The longer alleles are positively (and significantly) associated with the mutation phenomenon. More precisely, alleles with a size larger than or equal to 11 are expected to mutate 2.9 (0.95 CI: 1.8-4.8) times more frequently than the alleles with less than 11 repeats.

Then, we fit a mixed-effects logistic regression model to the binary part of the data (zero vs. non-zero mutations, i.e., no-mutation vs. mutation), consisting of an intercept-level random effects only, and with the allele frequency as an explanatory variable (see **Table 20**).

Table 20: Results from the statistical model analyzing the effect of the allele frequency on the existence of mutations.

Variables		Fixed-effects				Random-effects
	Coeficient	Exp(Coef)	Std error	p-value	0.95 CI for Exp(Coef)	Std dev.
Intercept	-1.883	-----	0.381	<0.001	-----	0.0010
Freq. allele	0.007	1.007	0.002	<0.001	(1.003, 1.011)	-----
	Reference	-----	-----	-----	-----	-----

As expected, the existence or nonexistence of mutation is positively (and significantly) associated with the allele frequency. That is, the more frequent an allele is, the higher the odd for a mutation to occur. More precisely, the model estimated an increase of 0.7% (0.95 CI: 0.3%, 11%) in the odds for a mutation for each 1-unit increase in the frequency of the alleles.

We have also attempted to run a model with the allele size as another predictor at the same time (besides the allele frequency); yet, the computational algorithm did not converge due to limitations in sample size for most of the frequencies in the range.

4. CONCLUSION

- A traditional approach, computing an overall mutation rate per marker by proportioning the number of Mendelian incompatibilities between father-son duos, showed that there seems to be in equilibrium between the number of repeat gains and the number of repeat losses (when analyzed per marker), barring few markers. Through this analysis, it is evident that the confidence on the estimation of the mutation rate varies from marker to marker. Indeed, some markers have been extensively studied and used (mainly if they are included in commercial kits) while others, being more recent to the field (or less used) have accordingly fewer observations and lower statistical confidence associated.
- When clustering markers according to the structure of the repetitive motif we noticed that markers with the same repeating structure behave similarly, but differently from those of other clusters, despite some exceptions within clusters having been identified. Nevertheless, and supported by statistical analyses of significance, there seems to be an association between the type of structure and mutation rates, but specially between the repeat structure and the type of mutation (number of mutational steps). Nonetheless, factors other than the structure of the repetitive motif must be involved in the mutation phenomenon.
- When we implemented a bi-allele approach we noted that the number of repeat gains and losses in an intra-marker (or inter-allele) approach is not in equilibrium. Indeed, alleles do not mutate uniformly, tending to mutate to longer alleles if they are shorter than the modal allele, and to shorter ones if they are longer. Alleles within the same marker have distinct mutation rates, in some cases confidence intervals do not even intersect. If the complete haplotypic data were more often made available these estimates would be statistically more powerful.
- Next Generation Sequencing will be essential in the study of the mutation phenomenon and the factor interfering with it. For instance, employing NGS, several alleles with different sequences than those described in **Table 4** were already presented by numerous authors (Zhao *et al.,* 2015, Warshauer *et al.,* 2015, Wendt *et al.*, 2016, Kwon *et al.,* 2016). We did not take this possibility, of alleles apparently equal having different sequences, into consideration in this work, since we used published data on STRs. Nevertheless, and despite being aware that this might skew the results, the described discrepant alleles have low frequency in the population. Next Generation Sequencing will allow the

observation of the sequence of the different alleles, despite them having or not the same length.

- The bi-allele approach appears to be, from all the three approaches studied in this work, the most satisfactory. To improve statistical confidence on mutation rate estimates it is peremptory to collect as many complete haplotypic data as possible.
- Notwithstanding, the method of proportion will hardly become an immaculate method to estimate mutation rates, even for the Y chromosome, regardless the number of observations. Indeed, difficultly the estimation of mutation rates for rare alleles will be reliable (and note that with NGS a greater number of variant rarer alleles will be reported). Thus, the development of a mathematical model to infer mutations, other than proportion, is required even when Y transmission is considered.

REFERENCES

Akane, A., Seki, S., Shiono, H., Nakamura, H., Hasegawa, M., Kagawa, M., ... & Nakagome, Y. (1992). Sex determination of forensic samples by dual PCR amplification of an XY homologous gene. *Forensic science international*, *52*(2), 143-148.

Alonso, A., Alves, C., Suárez-Mier, M. P., Albarrán, C., Pereira, L., de Simón, L. F., ... & Amorim, A. (2005). Mitochondrial DNA haplotyping revealed the presence of mixed up benign and neoplastic tissue sections from two individuals on the same prostatic biopsy slide. *Journal of clinical pathology*, *58*(1), 83-86.

Alshamali, F., Alkhayat, A. Q., Budowle, B., & Watson, N. (2004, April). Y chromosome in forensic casework and paternity testing. In *International Congress Series* (Vol. 1261, pp. 353-356). Elsevier.

Amorim, A., & Pereira, L. (2005). Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic science international*, *150*(1), 17-21.

Amorim, A., & Budowle, B. (Eds.). (2016). Definition and Purpose. *Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation* (Vol. 2). World Scientific.

Antão-Sousa, S., Sánchez-Diz, P., Abovich, M., Alvarez, J.C., Carvalho, E.F., Silva, C.M.D., … , Gusmão, L. , *under review*, Mutation rates and segregation data on 16 Y-STRs: an update to previous GHEP-ISFG studies, *Forensic Science International: Genetics Supplement Series, 6*

Ballantyne, K. N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., ... & Decorte, R. (2010). Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *The American Journal of Human Genetics*, *87*(3), 341-353.

Ballantyne, K. N., Keerl, V., Wollstein, A., Choi, Y., Zuniga, S. B., Ralf, A., ... & Kayser, M. (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), 208-218.

Ballantyne, K. N., Ralf, A., Aboukhalid, R., Achakzai, N. M., Anjos, M. J., Ayub, Q., ... & Bobillo, C. (2014). Toward Male Individualization with Rapidly Mutating Y-Chromosomal Short Tandem Repeats. *Human mutation*, *35*(8), 1021-1032.

Ballard, D. J., Phillips, C., Wright, G., Thacker, C. R., Robson, C., Revoir, A. P., & Court, D. S. (2005). A study of mutation rates and the characterisation of intermediate, null and duplicated alleles for 13 Y chromosome STRs. *Forensic science international*, *155*(1), 65-70.

Bär, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., Lincoln, P., ... & Olaisen, B. (1997). DNA recommendations. *International journal of legal medicine*, *110*(4), 175-176.

Berger, B., Lindinger, A., Niederstätter, H., Grubwieser, P., & Parson, W. (2005). Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay. *International journal of legal medicine*, *119*(4), 241-246.

Bernstein, F. (1924). Ergebnisse einer biostatistischen zusam-menfassenden betrachtung über die erblichen blutstrukturen des menschen. *Journal of Molecular Medicine*, *3*(33), 1495-1497.

Bianchi, N. O., Catanesi, C. I., Bailliet, G., Martinez-Marignac, V. L., Bravi, C. M., Vidal-Rioja, L. B., ... & Lopez-Camelo, J. S. (1998). Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. *The American Journal of Human Genetics*, *63*(6), 1862-1871.

Bonné-Tamir, B., Korostishevsky, M., Redd, A. J., Pel-Or, Y., Kaplan, M. E., & Hammer, M. F. (2003). Maternal and paternal lineages of the Samaritan isolate: mutation rates and time to most recent common male ancestor. *Annals of human genetics*, *67*(2), 153-164.

Brenner, C. H., & Weir, B. S. (2003). Issues and strategies in the DNA identification of World Trade Center victims. *Theoretical population biology*, *63*(3), 173-178.

Brown, K. (2002). Tangled roots? Genetics meets genealogy. *Science*, *295*(5560), 1634-1635.

Budowle, B., Adamowicz, M., Aranda, X. G., Barna, C., Chakraborty, R., Cheswick, D., ... & Ladd, C. (2005). Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America. *Forensic science international*, *150*(1), 1-15.

Butler, J. M. (2005). *Forensic DNA typing: biology, technology, and genetics of STR markers*. Academic Press.

Butler, J. M., Coble, M. D., & Vallone, P. M. (2007). STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic science, medicine, and pathology*, *3*(3), 200-205.

Butler, J. M. (2009). *Fundamentals of forensic DNA typing*. Academic Press.

Butler, J. M. (2012). Advanced Topics in Forensic DNA Typing.

Carracedo, Á. (1998). Investigación de la Paternidad. In Gisbert Calabuig (ed), *Medicina legal y toxicología.* Barcelona: Editorial Masson.

Coble, M. D., Loreille, O. M., Wadhams, M. J., Edson, S. M., Maynard, K., Meyer, C. E., ... & Gill, P. (2009). Mystery solved: the identification of the two missing Romanov children using DNA analysis. *PloS one*, *4*(3), e4838.

D'Amato, M. E., Ehrenreich, L., Cloete, K., Benjeddou, M., & Davison, S. (2010). Characterization of the highly discriminatory loci DYS449, DYS481, DYS518, DYS612, DYS626, DYS644 and DYS710. *Forensic science international: genetics*, *4*(2), 104-110.

Decker, A. E., Kline, M. C., Redman, J. W., Reid, T. M., & Butler, J. M. (2008). Analysis of mutations in father–son pairs with 17 Y-STR loci. *Forensic Science International: Genetics*, *2*(3), e31-e35.

Domingues, P. M., Gusmão, L., da Silva, D. A., Amorim, A., Pereira, R. W., & de Carvalho, E. F. (2007). Sub-Saharan Africa descendents in Rio de Janeiro (Brazil): population and mutational data for 12 Y-STR loci. *International journal of legal medicine*, *121*(3), 238-241.

Dupuy, B. M., Andreassen, R., Flønes, A. G., Tomassen, K., Egeland, T., Brion, M., ... & Olaisen, B. (2001). Y-chromosome variation in a Norwegian population sample. *Forensic science international*, *117*(3), 163-173.

Dupuy, B. M., Stenersen, M., Egeland, T., & Olaisen, B. (2004). Y-chromosomal microsatellite mutation rates: Differences in mutation rate between and within loci. *Human mutation*, *23*(2), 117-124.

Ellegren, H. (2000). Microsatellite mutations in the germline:: implications for evolutionary inference. *Trends in genetics*, *16*(12), 551-558.

Forster, P., Hohoff, C., Dunkelmann, B., Schürenkamp, M., Pfeiffer, H., Neuhuber, F., & Brinkmann, B. (2015, March). Elevated germline mutation rate in teenage fathers. In *Proc. R. Soc. B* (Vol. 282, No. 1803, p. 20142898). The Royal Society.

Ge, J., Budowle, B., Aranda, X. G., Planz, J. V., Eisenberg, A. J., & Chakraborty, R. (2009). Mutation rates at Y chromosome short tandem repeats in Texas populations. *Forensic Science International: Genetics*, *3*(3), 179-184.

Gill, P., Kimpton, C., d'Aloja, E., Andersen, J. F., Bar, W., Brinkmann, B., ... & Nellemann, L. (1994). Report of the European DNA profiling group (EDNAP)—towards standardisation of short tandem repeat (STR) loci. *Forensic science international*, *65*(1), 51-59.

Gill, P., Gusmão, L., Haned, H., Mayr, W. R., Morling, N., Parson, W., ... & Weir, B. S. (2012). DNA commission of the International Society of Forensic Genetics: recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. *Forensic Science International: Genetics*, *6*(6), 679-688.

Gjertson, D. W., Brenner, C. H., Baur, M. P., Carracedo, A., Guidet, F., Luque, J. A., ... & Schneider, P. M. (2007). ISFG: recommendations on biostatistics in paternity testing. *Forensic Science International: Genetics*, *1*(3), 223-231.

Goedbloed, M., Vermeulen, M., Fang, R. N., Lembring, M., Wollstein, A., Ballantyne, K., ... & Lessig, R. (2009). Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR® Yfiler® PCR amplification kit. *International journal of legal medicine*, *123*(6), 471.

Goodwin, W., Linacre, A., & Hadi, S. An Introduction to Forensic Genetics. 2007. *West Sussex, England: John Wiley & Sons, Ltd*.

Green, P. J., & Mortera, J. (2017). Paternity testing and other inference about relationships from DNA mixtures. *Forensic Science International: Genetics*, *28*, 128-137.

Gusmão, L., Brion, M., González-Neira, A., Lareu, M., & Carracedo, A. (1999). Y chromosome specific polymorphisms in forensic analysis. *Legal Medicine*, *1*(2), 55-60.

Gusmão, L., Sánchez-Diz, P., Calafell, F., Martin, P., Alonso, C. A., Álvarez-Fernández, F., ... & Builes, J. J. (2005). Mutation rates at Y chromosome specific microsatellites. *Human mutation*, *26*(6), 520-528.

Gusmão, L., Butler, J. M., Carracedo, A., Gill, P., Kayser, M., Mayr, W. R., ... & Schneider, P. M. (2006). DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *International journal of legal medicine*, *120*(4), 191-200.

Hanson, E. K., & Ballantyne, J. (2006). Comprehensive annotated STR physical map of the human Y chromosome: Forensic implications. *Legal medicine*, *8*(2), 110-120.

Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E., & de Knijff, P. (1997). Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Human molecular genetics*, *6*(5), 799-803.

Hohoff, C., Dewa, K., Sibbing, U., Hoppe, K., Forster, P., & Brinkmann, B. (2007). Y-chromosomal microsatellite mutation rates in a population sample from northwestern Germany. *International journal of legal medicine*, *121*(5), 359-363.

Hughes, J. F., & Rozen, S. (2012). Genomics and genetics of human and primate y chromosomes. *Annual review of genomics and human genetics*, *13*, 83-108.

Iida, R., & Kishi, K. (2005). Identification, characterization and forensic application of novel Y-STRs. *Legal medicine*, *7*(4), 255-258.

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Individual-specific 'fingerprints' of human DNA. *Nature*, *316*(6023), 76-79.

Jeffreys, A. J., Brookfield, J. F., & Semeonoff, R. (1985). Positive identification of an immigration test-case using human DNA fingerprints. *Nature*, *317*(6040), 818-819.

Jobling, M. A. (2001). In the name of the father: surnames and genetics. *TRENDS in Genetics*, *17*(6), 353-357.

Jobling, M. A., & Tyler-Smith, C. (2003). The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetics*, *4*(8), 598-612.

Kauppi, L., Barchi, M., Baudat, F., Romanienko, P. J., Keeney, S., & Jasin, M. (2011). Distinct properties of the XY pseudoautosomal region crucial for male meiosis. *Science*, *331*(6019), 916-920.

Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., ... & Szibor, R. (2000). Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *The American Journal of Human Genetics*, *66*(5), 1580-1588.

Kayser, M. (2007). Uni-parental markers in human identity testing including forensic DNA analysis. *Biotechniques*, *43*(6), S16-S21.

de Knijff, P. (2003) *Profiles in DNA*, 7, 3–5.

Kwon, S. Y., Lee, H. Y., Kim, E. H., Lee, E. Y., & Shin, K. J. (2016). Investigation into the sequence structure of 23 Y chromosomal STR loci using massively parallel sequencing. *Forensic Science International: Genetics*, *25*, 132-141.

Kurihara, R., Yamamoto, T., Uchihi, R., Li, S. L., Yoshimoto, T., Ohtaki, H., ... & Katsumata, Y. (2004). Mutations in 14 Y-STR loci among Japanese father-son haplotypes. *International journal of legal medicine*, *118*(3), 125-131.

Landsteiner, K. (1990). *The specificity of serological reactions*. Courier Corporation.

Levinson, G., & Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular biology and evolution*, *4*(3), 203-221.

Lessig, R., & Edelmann, J. (1998). Y chromosome polymorphisms and haplotypes in West Saxony (Germany). *International journal of legal medicine*, *111*(4), 215-218.

Mohandas, T. K., Speed, R. M., Passage, M. B., Yen, P. H., Chandley, A. C., & Shapiro, L. J. (1992). Role of the pseudoautosomal region in sex-chromosome pairing during male meiosis: meiotic studies in a man with a deletion of distal Xp. *American journal of human genetics*, *51*(3), 526.

Morling, N., Bastisch, I., Gill, P., & Schneider, P. M. (2007). Interpretation of DNA mixtures—European consensus on principles. *Forensic Science International: Genetics*, *1*(3), 291-292.

Mullis, K., Faloona, F., Scharf, S., Saiki, R. K., Horn, G. T., & Erlich, H. (1986, January). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In *Cold Spring Harbor symposia on quantitative biology* (Vol. 51, pp. 263-273). Cold Spring Harbor Laboratory Press.

Nachman, M. W., & Crowell, S. L. (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics*, *156*(1), 297-304.

Navarro-Costa, P. (2012). Sex, rebellion and decadence: the scandalous evolutionary history of the human Y chromosome. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1822*(12), 1851-1863.

Oh, Y. N., Lee, H. Y., Lee, E. Y., Kim, E. H., Yang, W. I., & Shin, K. J. (2015). Haplotype and mutation analysis for newly suggested Y-STRs in Korean father–son pairs. *Forensic Science International: Genetics*, *15*, 64-68.

Padilla-Gutiérrez, J. R., Valle, Y., Quintero-Ramos, A., Hernández, G., Rodarte, K., Olivares, N., & Rivas, F. (2008). Population data and mutation rate of nine Y-STRs in a mestizo Mexican population from Guadalajara, Jalisco, Mexico. *Legal Medicine*, *10*(6), 319-320.

Parson, W., Fendt, L., Ballard, D., Børsting, C., Brinkmann, B., Carracedo, Á., ... & Dupuy, B. M. (2008). Identification of West Eurasian mitochondrial haplogroups by mtDNA SNP screening: Results of the 2006–2007 EDNAP collaborative exercise. *Forensic Science International: Genetics*, *2*(1), 61-68.

Pena, S. D., & Chakraborty, R. (1994). Paternity testing in the DNA era. *Trends in Genetics*, *10*(6), 204-209.

Pereira, R., Phillips, C., Pinto, N., Santos, C., dos Santos, S. E. B., Amorim, A., ... & Gusmão, L. (2012). Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PloS one*, *7*(1), e29684.

Pereira, V., & Gusmão, L. (2016). Types of genomes, sequences and genetic markers. In *Handbook of Forensic Genetics* (pp. 163-191). World Scientific Publishing Co Pte Ltd.

Pestoni, C., Lareu, M. V., López-Gómez, J., & Carracedo, A. (1999). Genetic data on three complex STRs (ACTBP2, D21S11 and HUMFIBRA/FGA) in the Galician population (NW Spain). *International journal of legal medicine*, *112*(5), 337-339.

Phillips, C., Fondevila, M., García-Magariños, M., Rodriguez, A., Salas, A., Carracedo, A., & Lareu, M. V. (2008). Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Science International: Genetics*, *2*(3), 198-204.

Pickrahn, I., Müller, E., Zahrer, W., Dunkelmann, B., Cemper-Kiesslich, J., Kreindl, G., & Neuhuber, F. (2016). Yfiler® Plus amplification kit validation and calculation of forensic parameters for two Austrian populations. *Forensic Science International: Genetics*, *21*, 90-94.

Pickrahn, I., Kreindl, G., Müller, E., Dunkelmann, B., Zahrer, W., Cemper-Kiesslich, J., & Neuhuber, F. (2017). Contamination incidents in the pre-analytical phase of forensic DNA analysis in Austria–statistics of 17 years. *Forensic Science International: Genetics*.

Pinto, N., Magalhães, M., Conde-Sousa, E., Gomes, C., Pereira, R., Alves, C., ... & Amorim, A. (2013). Assessing paternities with inconclusive STR results: the suitability of bi-allelic markers. *Forensic Science International: Genetics*, *7*(1), 16-21.

Pinto, N., Gusmão, L., & Amorim, A. (2014). Mutation and mutation rates at Y chromosome specific Short Tandem Repeat Polymorphisms (STRs): A reappraisal. *Forensic Science International: Genetics*, *9*, 20-24.

Pontes, M. L., Cainé, L., Abrantes, D., Lima, G., & Pinheiro, M. F. (2007). Allele frequencies and population data for 17 Y-STR loci (AmpFℓSTR® Y-filer™) in a Northern Portuguese population sample. *Forensic science international*, *170*(1), 62-67.

Prinz, M., Carracedo, A., Mayr, W. R., Morling, N., Parsons, T. J., Sajantila, A., ... & Schneider, P. M. (2007). DNA Commission of the International Society for Forensic

Genetics (ISFG): recommendations regarding the role of forensic genetics for disaster victim identification (DVI). *Forensic Science International: Genetics*, *1*(1), 3-12.

Radam, G., and H. Strauch. 1973. Lumineszenzmikroskopischer Nachweis des Y Chromosoms in Knochenmarkszellen – Eine neue Methode zur Geschlechtserkennung an Leichenmaterial. *Kriminalistik und Forensische Wissenscha en* 6: 149–151.

Repping, S., van Daalen, S. K., Brown, L. G., Korver, C. M., Lange, J., Marszalek, J. D., ... & Rozen, S. (2006). High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nature genetics*, *38*(4), 463.

Robino, C., Ralf, A., Pasino, S., De Marchi, M. R., Ballantyne, K. N., Barbaro, A., ... & Fabbri, M. (2015). Development of an Italian RM Y-STR haplotype database: results of the 2013 GEFI collaborative exercise. *Forensic Science International: Genetics*, *15*, 56-63.

Roewer, L., Amemann, J., Spurr, N. K., Grzeschik, K. H., & Epplen, J. T. (1992). Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Human genetics*, *89*(4), 389-394.

Roewer, L. (2003, January). The use of the Y chromosome in forensic genetics—current practices and future perspectives. In *International Congress Series* (Vol. 1239, pp. 279-280). Elsevier.

Roewer, L. (2009). Y chromosome STR typing in crime casework. *Forensic science, medicine, and pathology*, *5*(2), 77-84.

Romanini, C., Romero, M., Puerto, M. S., Catelli, L., Phillips, C., Pereira, R., ... & Vullo, C. (2015). Ancestry informative markers: inference of ancestry in aged bone samples using an autosomal AIM-Indel multiplex. *Forensic Science International: Genetics*, *16*, 58-63.

Sánchez-Diz, P., Alves, C., Carvalho, E., Carvalho, M., Espinheira, R., García, O., ... & Silva, C. (2008). Population and segregation data on 17 Y-STRs: results of a GEP-ISFG collaborative study. *International journal of legal medicine*, *122*(6), 529-533.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, *74*(12), 5463-5467.)

Santos, F. R., Epplen, J. T., & Pena, S. D. J. (1993). Testing deficiency paternity cases with a Y-linked tetranucleotide repeat polymorphism. *EXPERIENTIA-BASEL-SUPPLEMENTUM-*, *67*, 261-261.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, *109* (6), 365-371.

Schneider, P. M. (1997). Basic issues in forensic DNA typing. *Forensic science international*, *88*(1), 17-22.

Seidelmann, S. B., Smith, E., Subrahmanyan, L., Dykas, D., Ziki, M. D. A., Azari, B., ... & Jacoby, D. (2017). Application of whole exome sequencing in the clinical diagnosis and management of inherited cardiovascular diseases in adults. *Circulation: Cardiovascular Genetics*, *10*(1), e001573.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., & Cordum, H. S. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, *423*(6942), 825.

Skowronek, M. F., Velazquez, T., Mut, P., Figueiro, G., Sans, M., Bertoni, B., & Sapiro, R. (2017). Associations between male infertility and ancestry in South Americans: a case control study. *BMC medical genetics*, *18*(1), 78.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 98, 503–517.

de Souza Góes, A. C., de Carvalho, E. F., Gomes, I., da Silva, D. A., Gil, É. H. F., Amorim, A., & Gusmão, L. (2005). Population and mutation analysis of 17 Y-STR loci from Rio de Janeiro (Brazil). *International journal of legal medicine*, *119*(2), 70-76.

Sykes, B., & Irven, C. (2000). Surnames and the Y chromosome. *The American Journal of Human Genetics*, *66*(4), 1417-1419.

Toscanini, U., Brisighelli, F., Moreno, F., Pantoja-Astudillo, J. A., Morales, E. A., Bustos, P., ... & Salas, A. (2016). Analysis of Y-chromosome STRs in Chile confirms an extensive introgression of European male lineages in urban populations. *Forensic Science International: Genetics*, *21*, 76-80.

Tsai, L. C., Yuen, T. Y., Hsieh, H. M., Lin, M., Tzeng, C. H., Huang, N. E., ... & Lee, J. I. (2002). Haplotype frequencies of nine Y-chromosome STR loci in the Taiwanese Han population. *International journal of legal medicine*, *116*(3), 179-183.

Turrina, S., Atzei, R., & De Leo, D. (2006). Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay. *International journal of legal medicine*, *120*(1), 56-59.

Walsh, S. J., Triggs, C. M., & Buckleton, J. S. (2004). Forensic DNA evidence interpretation.

Wang, Y., Zhang, Y. J., Zhang, C. C., Li, R., Yang, Y., Ou, X. L., ... & Sun, H. Y. (2016). Genetic polymorphisms and mutation rates of 27 Y-chromosomal STRs in a Han population from Guangdong Province, Southern China. *Forensic Science International: Genetics*, *21*, 5-9.

Warshauer, D. H., Churchill, J. D., Novroski, N., King, J. L., & Budowle, B. (2015). Novel Y-chromosome short tandem repeat variants detected through the use of massively parallel sequencing. *Genomics, proteomics & bioinformatics*, *13*(4), 250-257.

Wendt, F. R., Churchill, J. D., Novroski, N. M., King, J. L., Ng, J., Oldt, R. F., ... & Budowle, B. (2016). Genetic analysis of the yavapai native americans from west-Central arizona using the illumina MiSeq FGx™ forensic genomics system. *Forensic Science International: Genetics*, *24*, 18-23.

Wilkinson, R. D., Steiper, M. E., Soligo, C., Martin, R. D., Yang, Z., & Tavaré, S. (2010). Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Systematic Biology*, *60*(1), 16-31.

Willems, T., Gymrek, M., Poznik, G. D., Tyler-Smith, C., Erlich, Y., & 1000 Genomes Project Chromosome Y Group. (2016). Population-scale sequencing data enable precise estimates of Y-STR mutation rates. *The American Journal of Human Genetics*, *98*(5), 919-933.

Willuweit S, Roewer L. Y Chromosome Haplotype Reference Database (YHRD). http://www.yhrd.org/. 2000.

Willuweit, S. & Roewer, L. (2015). The new Y chromosome haplotype reference database. *Forensic Science International: Genetics*, *15*, 43-48.

Wyman, A.R. & White, R. (1980) A highly polymorphic locus in human DNA. *Proceedings of the National Academy of Sciences of the United States of America* 77, 6754–6758.

Zerjal, T., Xue, Y., Bertorelle, G., Wells, R. S., Bao, W., Zhu, S., ... & Li, P. (2003). The genetic legacy of the Mongols. *The American Journal of Human Genetics*, *72*(3), 717-721.

Zhao, L., Wang, F., Wang, H., Li, Y., Alexander, S., Wang, K., ... & Earle, P. (2015). Next-generation sequencing-based molecular diagnosis of 82 retinitis pigmentosa probands from Northern Ireland. *Human genetics*, *134*(2), 217-230.

APPENDIX

Table A 1: Chi-square test for each marker repeat gains and losses.

Marker	Gains	Losses	p-value
DXYS156-Y	0	0	
DYS19	26	20	0.37634
DYS385	53	18	0.00003
DYS388	0	1	0.31731
DYS389I	21	28	0.31731
DYS389II	38	39	0.90927
DYS390	18	22	0.52709
DYS391	29	21	0.25790
DYS392	6	3	0.31731
DYS393	12	7	0.25135
DYS413	0	1	0.31731
DYS435	0	0	
DYS437	11	7	0.34578
DYS438	1	6	0.05878
DYS439	30	35	0.53514
DYS448	4	8	0.24821
DYS449	29	24	0.49221
DYS456	25	18	0.28575
DYS458	39	37	0.81855
DYS460	5	8	0.40538
DYS461	0	0	
DYS481	5	3	0.47950
DYS518	41	33	0.35238
DYS526_A	6	4	0.52709
DYS526_B	24	19	0.44577
DYS533	3	2	0.65472
DYS547	22	33	0.13801
DYS549	0	1	0.31731
DYS570	15	24	0.14954
DYS576	37	34	0.72181
DYS612	33	21	0.10247
DYS626	11	23	0.03959
DYS627	30	35	0.53514
DYS635	11	22	0.05551
DYS643	0	0	
GATA A10	2	3	0.65472
GATA C4	0	0	
GATA H4	10	16	0.23932
YCAIab	0	0	
YCAIIab	1	2	0.56370
TOTAL	598	578	0.55975

Table A 2: Chi-square tests for the [GAAA] and for the [GATA] cluster.

	DYS 385	DYS 449	DYS 458	DYS 518	DYS 570	DYS 576	DYS 626	DYS 627	DYS 19	DYS 389I	DYS 390	DYS 391	DYS 393	DYS 437	DYS 439	DYS 456	DYS 460	DYS 461	DYS 533	DYS 549	GATA A10	GATA H4
DYS 385	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DYS 449	<0.0001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DYS 458	<0.0001	0.001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DYS 518	<0.0001	0.0518	<0.0001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DYS 570	<0.0001	0.257	0.0856	0.0014	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DYS 576	<0.0001	0.3665	<0.0001	0.3265	0.0311	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DYS 626	<0.0001	0.4925	0.05	0.0108	0.8195	0.1075	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DYS 627	<0.0001	0.299	<0.0001	0.4000	0.022	0.9592	0.0833	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DYS 19	0.955	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DYS 389I	0.5799	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.7234	-	-	-	-	-	-	-	-	-	-	-	-	-
DYS 390	0.5936	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.5416	0.291	-	-	-	-	-	-	-	-	-	-	-	-
DYS 391	0.7918	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.9291	0.8688	0.4254	-	-	-	-	-	-	-	-	-	-	-
DYS 393	0.0033	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0038	0.001	0.0265	0.0022	-	-	-	-	-	-	-	-	-	-
DYS 437	0.0878	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0847	0.0375	0.258	0.061	0.5117	-	-	-	-	-	-	-	-	-
DYS 439	<0.0001	<0.0001	0.0087	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.00010	<0.0001	<0.0001	<0.0001	<0.0001	-	-	-	-	-	-	-	-
DYS 456	0.0007	<0.0001	0.0013	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0029	0.0107	0.0003	0.0051	<0.0001	<0.0001	0.4394	-	-	-	-	-	-	-
DYS 460	0.0133	0.0022	0.241	<0.0001	0.0306	0.0002	0.0193	<0.0001	0.0209	0.042	0.006	0.0282	<0.0001	0.0005	0.9424	0.6972	-	-	-	-	-	-
DYS 461	0.2944	0.0006	0.0138	<0.0001	0.003	0.0002	0.002	<0.0001	0.2853	0.25	0.3492	0.2702	0.6599	0.5272	0.0591	0.0914	0.067	-	-	-	-	-
DYS 533	0.8883	<0.0001	0.0078	<0.0001	0.0007	<0.0001	0.0004	<0.0001	0.8545	0.727	0.8464	0.8003	0.6575	0.966	0.0879	0.1778	0.1371	0.5185	-	-	-	-

DYS549	0.5287	0.7689	0.7643	0.7641	0.9676	0.9376	0.9057	0.914	0.545	0.5914	0.4714	0.5643	0.2183	0.3214	0.9778	0.8861	0.4824	0.1667	0.5106	-	-	-
GATA A10	0.1207	0.0357	0.4702	0.0019	0.1421	0.0094	0.105	0.0077	0.1429	0.2029	0.0754	0.1662	0.0021	0.016	0.9603	0.881	0.9994	0.0868	0.2316	0.5027	-	-
GATA H4	0.835	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.9456	0.9419	0.5044	0.9759	0.0065	0.0929	0.001	0.0207	0.0449	0.2729	0.8077	0.5739	0.1911	-

Table A 3: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS19.

Marker	Nº of mutations	Nº of meiosis	Original allele	Fillial allele	Nº of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS19	23	9,074	11	11	90	0.00992	0.00000	0.00000-0.04016		
			12	12	8	0.00099	0.11111	0.00281-0.48250		
				13	1				0.11111	0.00281-0.48250
			13	13	967	0.10668	0.00103	0.00003-0.00574		
				14	1				0.00103	0.00003-0.00574
			14	13	1	0.43641	0.00202	0.00087-0.00398	0.00025	0.00001-0.00141
				14	3,952					
				14.2	1				0.00025	0.00001-0.00141
				15	6				0.00152	0.00056-0.00329
			14.1	14.1	1	0.00011	0.00000	0.00000-0.97500		
			14.3	14.3	1	0.00011	0.00000	0.00000-0.97500		
			15	15	2,383	0.26306	0.00168	0.00046-0.00428		
				16	4				0.00168	0.00046-0.00428
			16	15	1	0.12618	0.00262	0.00054-0.00764	0.00087	0.00002-0.00486
				16	1,142					
				17	2				0.00175	0.00021-0.00630
			17	16	5	0.05598	0.01181	0.00435-0.02553	0.00984	0.00320-0.02282
				17	502					
				18	1				0.00197	0.00005-0.01092
			18	18	4	0.00044	0.00000	0.00000-0.60236		
			19	19	1	0.00011	0.00000	0.00000-0.97500		

Table A 4: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS385.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS385	29	15,370	8	8	3	0.00020	0.00000	0.00000-0.70760		
			9	9	42	0.00273	0.00000	0.00000-0.08408		
			10	10	484	0.03149	0.00000	0.00000-0.00759		
			11	10	1	0.19473	0.00067	0.00000-0.00123	0.00033	0.00001-0.00186
				11	2,991					
				12	1				0.00033	0.00001-0.00186
			12	12	1,086	0.07066	0.00000	0.00000-0.00339		
			12.2	12.2	1	0.00007	0.00000	0.00000-0.97500		
			13	13	2,077	0.13578	0.00479	0.00230-0.00879		
				14	10				0.00479	0.00230-0.00879
			13.2	13.2	1	0.00007	0.00000	0.00000-0.97500		
			14	13	1	0.24463	0.00133	0.00000-0.00098	0.00027	0.00001-0.00148
				14	3,755					
				15	4				0.00106	0.00029-0.00272
			15	15	1,514	0.09870	0.00198	0.00041-0.00577		
				16	3				0.00198	0.00041-0.00577
			15.3	15.3	3	0.00020	0.00000	0.00000-0.70760		
			16	16	956	0.06226	0.00104	0.00003-0.00581		
				17	1				0.00104	0.00003-0.00581
			17	17	814	0.05303	0.00123	0.00003-0.00682		
				18	1				0.00123	0.00003-0.00682
			17.2	17.2	1	0.00007	0.00000	0.00000-0.97500		

			17.3	17.3	2	0.00013	0.00000	0.00000-0.84189		
			18	17	2	0.05237	0.00497	0.00136-0.01267	0.00248	0.00030-0.00895
				18	801					
				19	2				0.00248	0.00030-0.00895
			19	18	2	0.03006	0.00433	0.00052-0.01555	0.00433	0.00052-0.01555
				19	460					
			20	19	1	0.01379	0.00472	0.00012-0.02600	0.00472	0.00012-0.02600
				20	211					
			21	21	92	0.00599	0.00000	0.00000-0.03930		
			22	22	39	0.00254	0.00000	0.00000-0.09739		
			23	23	6	0.00039	0.00000	0.00000-0.45926		
			25	25	1	0.00007	0.00000	0.00000-0.97500		
			27	27	1	0.00007	0.00000	0.00000-0.97500		

Table A 5: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS388.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS388	1	101	10	10	17	0.16832	0.05556	0.00141-0.27294		
				12	1				0.05556	0.00141-0.27294
			11	11	1	0.00990	0.00000	0.00000-0.97500		
			12	12	73	0.72277	0.00000	0.00000-0.04928		
			13	13	7	0.06931	0.00000	0.00000-0.40962		
			14	14	2	0.01980	0.00000	0.00000-0.84189		

Table A 6: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS389I.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS389I	22	8,084	8	8	1	0.00012	0.00000	0.00000-0.97500		
			9	9	71	0.00878	0.00000	0.00000-0.05063		
			10	10	55	0.00680	0.00000	0.00000-0.06487		
			11	11	72	0.00891	0.00000	0.00000-0.04994		
			12	11	1	0.22514	0.00275	0.00089-0.00640	0.00055	0.00001-0.00306
				12	1,815					
				13	4				0.00220	0.00060-0.00562
			13	12	1	0.54973	0.00135	0.00050-0.00294	0.00023	0.00001-0.00125
				13	4,438					
				14	5				0.00113	0.00037-0.00262
			14	13	8	0.19334	0.00576	0.00264-0.01090	0.00512	0.00221-0.01006
				14	1,554					
				15	1				0.00064	0.00002-0.00356
			15	14	2	0.00544	0.04545	0.00555-0.15473	0.04545	0.00555-0.15473
				15	42					
			16	16	8	0.00099	0.00000	0.00000-0.36942		
			17	17	3	0.00037	0.00000	0.00000-0.70760		
			18	18	3	0.00037	0.00000	0.00000-0.70760		

Table A 7: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS389II.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS389II	40	8,047	23	23	2	0.00025	0.00000	0.00000-0.84189		
			24	24	16	0.00199	0.00000	0.00000-0.20591		
			25	25	47	0.00584	0.00000	0.00000-0.07549		
			26	26	61	0.00783	0.03175	0.00387-0.11002		
				27	2				0.03175	0.00387-0.03399
			27	27	194	0.02423	0.00513	0.00013-0.02824		
				28	1				0.00513	0.00013-0.00755
			28	28	1,260	0.15658		0.00000-0.00292		
			29	28	3	0.36386	0.00478	0.00262-0.00801	0.00102	0.00021-0.00094
				29	2,914					
				30	11				0.00376	0.00188-0.00255
			30	29	6	0.28942	0.00601	0.00329-0.01007	0.00258	0.00095-0.00196
				30	2,315					
				31	7				0.00301	0.00121-0.00221
				32	1				0.00043	0.00001-0.00063
			31	30	5	0.11644	0.00534	0.00173-0.01241	0.00534	0.00173-0.00424
				31	932					
			32	31	3	0.02970	0.01255	0.00260-0.03624	0.01255	0.00260-0.01155
				32	236					
			33	32	1	0.00360	0.03448	0.00087-0.17764	0.03448	0.00087-0.05094
				33	28					
			34	34	2	0.00025	0.00000	0.00000-0.84189		

Table A 8: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS390.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS390	19	9,105	19	19	1	0.00011	0.00000	0.00000-0.97500		
			20	20	5	0.00055	0.00000	0.00000-0.52182		
			21	21	290	0.03185	0.00000	0.00000-0.01264		
			22	22	898	0.09863	0.00000	0.00000-0.00410		
			23	22	1	0.25744	0.00043	0.00001-0.00237	0.00043	0.00001-0.00237
				23	2,343					
			24	23	4	0.42449	0.00259	0.00124-0.00475	0.00103	0.00028-0.00265
				24	3,855					
				25	6				0.00155	0.00057-0.00338
			25	24	4	0.05931	0.00741	0.00202-0.01886	0.00741	0.00202-0.01886
				25	1,536					
			26	25	4	0.01702	0.02581	0.00708-0.06475	0.02581	0.00708-0.06475
				26	151					
			27	27	7	0.00077	0.00000	0.00000-0.40962		

Table A 9: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS391.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS391	24	9,107	5	5	1	0.00011	0.00000	0.00000-0.97500		
			6	6	6	0.00066	0.00000	0.00000-0.45926		
			8	8	9	0.00099	0.00000	0.00000-0.33627		
			9	9	421	0.04634				
				10	1		0.00237	0.00006-0.01313	0.00237	0.00006-0.01313
			10	10	5,316	0.58428				
				11	5		0.00094	0.00031-0.00219		0.00031-0.00219
			11	10	9				0.00280	0.00128-0.00530
				11	3,203	0.35346	0.00497	0.00284-0.00806		
				12	7				0.00217	0.00087-0.00448
			12	11	2					
				12	114	0.01274	0.01724	0.00209-0.06089	0.01724	0.00209-0.06089
			13	12						
				13	13	0.00143	0.00000	0.00000-0.24705		
			14	14	1	0.00011	0.00000	0.00000-0.97500		

Table A 10: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS392.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS392	7	9,092	7	7	2	0.00022	0.00000	0.00000-0.84189		
			9	9	3	0.00033	0.00000	0.00000-0.70760		
			10	10	62	0.00682	0.00000	0.00000-0.05776		
			10.2	10.2	1	0.00011	0.00000	0.00000-0.97500		
			11	10	1	0.39122	0.00028	0.00001-0.00157		
				11	3,556				0.00028	0.00001-0.00157
			11.1	11.1	1	0.00011	0.00000	0.00000-0.97500		
			12	12	670	0.07369	0.00000	0.00000-0.00549		
			13	13	3,590	0.39518	0.00083	0.00017-0.00244	0.00083	0.00017-0.00244
				14	3					
			14	10	1	0.11912	0.00277	0.00057-0.00807	0.00092	0.00002-0.00513
				14	1,080					
				15	2				0.00185	0.00022-0.00665
			15	15	102	0.01122	0.00000	0.00000-0.03552		
			16	16	17	0.00187	0.00000	0.00000-0.19506		
			17	17	1	0.00011	0.00000	0.00000-0.97500		

Table A 11: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS393.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS393	11	7,861	9	9	1	0.00013	0.00000	0.00000-0.97500		
			10	10	12	0.00153	0.00000	0.00000-0.26465		
			11	11	43	0.00547	0.00000	0.00000-0.08221		
			12	12	1,572	0.20023	0.00127	0.00015-0.00458	0.00127	0.00015-0.00458
				13	2					
			13	12	2	0.64712	0.00098	0.00032-0.00229	0.00039	0.00005-0.00142
				13	5,082					
				14	3				0.00059	0.00012-0.00172
			14	13	1	0.12505	0.00203	0.00025-0.00733	0.00102	0.00003-0.00565
				14	981					
				15	1				0.00102	0.00003-0.00565
			15	14	1	0.01870	0.00680	0.00017-0.03732	0.00680	0.00017-0.03732
				15	146					
			16	15	1	0.00178	0.07143	0.00181-0.33868	0.07143	0.00181-0.33868
				16	13					

Table A 12: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS435.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS435	0	147	10	10	1	0.00680	0.00000	0.00000-0.97500		
			11	11	128	0.87075	0.00000	0.00000-0.02841		
			12	12	17	0.11565	0.00000	0.00000-0.19506		
			13	13	1	0.00680	0.00000	0.00000-0.97500		

Table A 13: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS437.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS437	10	7,335	13	13	22	0.002999318	0.00000			
			13.1	13.1	2	0.000272665	0.00000			
			14	14	3,555	0.484935242	0.00056	0.00007-0.00203		
				15	2				0.00056	0.00007-0.00203
			15	14	4	0.407089298	0.00201	0.00074-0.00437	0.00134	0.00037-0.00343
				15	2,980					
				16	2				0.00067	0.00008-0.00242
			16	16	747	0.101976823	0.00134	0.00003-0.00743		
				17	1				0.00134	0.00003-0.00743
			17	16	1	0.002726653	0.05000	0.00127-0.24873	0.05000	0.00127-0.24873
				17	19					

Table A 14: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS438.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS438	4	7,398	7	7	4	0.00054	0.00000	0.00000-0.60236		
			8	8	22	0.00297	0.00000	0.00000-0.15437		
			9	9	542	0.07326	0.00000	0.00000-0.00678		
			9.2	9.2	2	0.00027	0.00000	0.00000-0.84189		
			10	6	1	0.35996	0.00075	0.00009-0.00271	0.000375516	0.00001-0.00209
				9	1				0.000375516	0.00001-0.00209
				10	2,661					
			11	11	1,556	0.21033	0.00000	0.00000-0.00237		
			11.2	11.2	25	0.00338	0.00000	0.00000-0.13719		
			12	10	2			0.00010-0.00312	0.000864304	0.00010-0.00312
				12	2,312	0.31279	0.00086	0.00000-1.00000		
			13	13	260	0.03514	0.00000	0.00000-0.01409		
			14	14	10	0.00135	0.00000	0.00000-0.30850		

Table A 15: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS439.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS439	33	7,355	8	8	4	0.00054	0.00000	0.00000-0.60236		
			9	9	20	0.00272	0.00000	0.00000-0.16843		
			10	10	824	0.11230	0.00242	0.00029-0.00872		
				11	2				0.00242	0.00029-0.00872
			11	11	2,352	0.32046	0.00212	0.00069-0.00494		
				12	5				0.00212	0.00069-0.00494
			11.1	11.1	1	0.00014	0.00000	0.00000-0.97500		
			12	11	1	0.41999	0.00194	0.00071-0.00422	0.00032	0.00001-0.00180
				12	3,083					
				13	5				0.00162	0.00053-0.00377
			13	12	9	0.12386	0.01647	0.00924-0.02701	0.00988	0.00453-0.01867
				13	896					
				14	6				0.00659	0.00242-0.01428
			14	13	5	0.01958	0.03472	0.01137-0.07917	0.03472	0.01137-0.07917
				14	139					
			15	15	3	0.00041	0.00000	0.00000-0.70760		

Table A 16: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS448.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS448	3	6,033	15	15	2	0.00033	0.00000	0.00000-0.84189		
			16	16	5	0.00083	0.00000	0.00000-0.52182		
			17	17	49	0.00812	0.00000	0.00000-0.07252		
			18	18	730	0.12100	0.00000	0.00000-0.00504		
			18.2	18.2	4	0.00066	0.00000	0.00000-0.60236		
			19	18	1	0.37892	0.00087	0.00011-0.00316	0.00044	0.00001-0.00243
				19	2,284					
				20	1				0.00044	0.00001-0.00243
			19.2	19.2	2	0.00033	0.00000	0.00000-0.84189		
			20	19	1	0.36698	0.00045	0.00001-0.00251	0.00045	0.00001-0.00251
				20	2,213					
			21	21	630	0.10443	0.00000	0.00000-0.00584		
			22	22	97	0.01608	0.00000	0.00000-0.03732		
			23	23	17	0.00282	0.00000	0.00000-0.19506		
			24	24	3	0.00050	0.00000	0.00000-0.70760		
			25	25	1	0.00017	0.00000	0.00000-0.97500		

Table A 17: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS449.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS449	52	4,032	24	24	2	0.00050	0.00000	0.00000-0.84189		
			25	25	17	0.00422	0.00000	0.00000-0.19506		
			26	26	89	0.02207	0.00000	0.00000-0.04060		
			27	27	121	0.03051	0.01626	0.00198-0.05750		
				28	2				0.01626	0.00198-0.05750
			27.2	27.2	2	0.00050	0.00000	0.00000-0.84189		
			28	28	289	0.07168	0.00000	0.00000-0.01268		
			28.2	28.2	1	0.00025	0.00000	0.00000-0.97500		
			29	28	4	0.13269	0.01495	0.00648-0.02925	0.00748	0.00204-0.01903
				29	527					
				30	4				0.00748	0.00204-0.01903
			29.2	29.2	2	0.00050	0.00000	0.00000-0.84189		
			29.3	29.3	1	0.00025	0.00000	0.00000-0.97500		
			30	29	4	0.16295	0.00761	0.00248-0.01767	0.00609	0.00166-0.01551
				30	652					
				31	1				0.00152	0.00004-0.00845
			30.2	30.2	1	0.00025	0.00000	0.00000-0.97500		
			31	30	3	0.19420	0.00894	0.00360-0.01833	0.00383	0.00079-0.01116
				31	776					
				32	4				0.00511	0.00139-0.01303
			31.2	31.2	2	0.00050	0.00000	0.00000-0.84189		
			32	31	2	0.16815	0.01327	0.00609-0.02505	0.00295	0.00036-0.01061

				32	669					
				33	7				0.01032	0.00416-0.02116
			32.2	32.2	2	0.00050	0.00000	0.00000-0.84189		
			33	32	3	0.11359	0.01528	0.00617-0.03124	0.00655	0.00135-0.01902
				33	451					
				34	4				0.00873	0.00238-0.02221
			34	33	6	0.06225	0.02390	0.00882-0.05130	0.02390	0.00882-0.05130
				34	245					
			35	34	3	0.02307	0.06452	0.02404-0.13515	0.03226	0.00670-0.09139
				35	87					
				36	3				0.03226	0.00670-0.09139
			36	35	1	0.00942	0.02632	0.00067-0.13810	0.02632	0.00067-0.13810
				36	37					
			37	37	5	0.00124	0.00000	0.00000-0.52182		
			38	37	1	0.00050	0.50000	0.01258-0.98742	0.50000	0.01258-0.98742
				38	1					
			40	40	1	0.00025	0.00000	0.00000-0.97500		

Table A 18: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS456.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS456	22	6,036	11	11	3	0.00050	0.00000	0.00000-0.70760		
			12	12	20	0.00331	0.00000	0.00000-0.16843		
			13	13	156	0.02584	0.00000	0.00000-0.02337		
			14	14	669	0.11100	0.00149	0.00004-0.00829		
				15	1				0.00149	0.00004-0.00829
			15	14	1	0.45775	0.00253	0.00102-0.00521	0.00036	0.00001-0.00201
				15	2,756					
				16	6				0.00217	0.00080-0.00472
			16	15	4	0.27833	0.00417	0.00168-0.00857	0.00238	0.00065-0.00608
				16	1,673					
				17	3				0.00179	0.00037-0.00521
			17	16	3	0.10520	0.00787	0.00256-0.01828	0.00472	0.00098-0.01374
				17	630					
				18	2				0.00315	0.00038-0.01133
			18	17	2	0.01624	0.02041	0.00248-0.07178	0.02041	0.00248-0.07178
				18	96					
			19	19	10	0.00166	0.00000	0.00000-0.30850		
			20	20	1	0.00017	0.00000	0.00000-0.97500		

A 19: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS458.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS458	45	6,043	12	12	6	0.00099	0.00000	0.00000-0.45926		
			13	13	93	0.01539	0.00000	0.00000-0.03889		
			14	13	1	0.03161	0.00524	0.00013-0.02882	0.00524	0.00013-0.02882
				14	190					
			14.1	14.1	5	0.00083	0.00000	0.00000-0.52182		
			15	14	2	0.16052	0.00515	0.00168-0.01199	0.00206	0.00025-0.00743
				15	965					
				16	3				0.00309	0.00064-0.00901
			15.1	15.1	1	0.00017	0.00000	0.00000-0.97500		
			15.2	15.2	2	0.00033	0.00000	0.00000-0.84189		
			16	15	2	0.23945	0.00415	0.00152-0.00900	0.00138	0.00017-0.00498
				16	1,441					
				17	3				0.00207	0.00043-0.00605
				19	1				0.00069	0.00002-0.00384
			16.2	16.2	12	0.00199	0.00000	0.00000-0.26465		
			17	15	1	0.28595	0.00868	0.00487-0.01428	0.00058	0.00001-0.00322
				16	7				0.00405	0.00163-0.00833
				17	1,713					
				18	7				0.00405	0.00163-0.00833
			17.2	17.2	24	0.00397	0.00000	0.00000-0.14247		

			18	17	5	0.16796	0.00985	0.00473-0.01804	0.00493	0.00160-0.01146
				18	1,005					
				19	5				0.00493	0.00160-0.01146
			18.2	18.2	26	0.00430	0.00000	0.00000-0.13227		
			19	18	1	0.06106	0.01897	0.00766-0.03869	0.00271	0.00007-0.01501
				19	362					
				20	6				0.01626	0.00599-0.03505
			19.2	19.2	11	0.00182	0.00000	0.00000-0.28491		
			20	19	1	0.01837	0.00901	0.00023-0.04917	0.00901	0.00023-0.04917
				20	110					
			20.2	20.2	6	0.00099	0.00000	0.00000-0.45926		
			21	21	15	0.00248	0.00000	0.00000-0.21802		
			21.1	21.1	1	0.00017	0.00000	0.00000-0.97500		
			21.2	21.2	1	0.00017	0.00000	0.00000-0.97500		
			22	22	9	0.00149	0.00000	0.00000-0.33627		

Table A 20: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS460.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS460	10	2,002	7	7	2	0.00100	0.00000	0.00000-0.84189		
			8	8	2	0.00100	0.00000	0.00000-0.84189		
			9	9	364	0.18232	0.00274	0.00007-0.01517		
				10	1				0.00274	0.00007-0.01517
			10	9	1	0.39011	0.00384	0.00079-0.01118	0.00128	0.00003-0.00711
				10	778					
				11	2				0.00256	0.00031-0.00922
			11	10	3	0.39510	0.00506	0.00138-0.01290	0.00379	0.00078-0.01104
				11	787					
				12	1				0.00126	0.00003-0.00702
			12	11	2	0.02847	0.03509	0.00428-0.12107	0.03509	0.00428-0.12107
				12	55					
			13	13	4	0.00200	0.00000	0.00000-0.60236		

Table A 21: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS461.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS461	0	873	9	9	1	0.00115	0.00000	0.00000-0.97500		
			10	10	10	0.01145	0.00000	0.00000-0.30850		
			11	11	171	0.19588	0.00000	0.00000-0.02134		
			12	12	541	0.61970	0.00000	0.00000-0.00680		
			13	13	137	0.15693	0.00000	0.00000-0.02657		
			14	14	12	0.01375	0.00000	0.00000-0.26465		
			15	15	1	0.00115	0.00000	0.00000-0.97500		

Table A 22: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS481.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS481	8	1,085	18	18	1	0.00092	0.00000	0.00000-0.97500		
			19	19	3	0.00276	0.00000	0.00000-0.70760		
			20	20	3	0.00276	0.00000	0.00000-0.70760		
			21	21	44	0.04055	0.00000	0.00000-0.08042		
			22	22	166	0.15392	0.00599	0.00015-0.03291		
				23	1				0.00599	0.00015-0.03291
			23	23	333	0.30783	0.00299	0.00008-0.01657		
				24	1				0.00299	0.00008-0.01657
			24	24	193	0.17972	0.01026	0.00124-0.03656		
				25	2				0.01026	0.00124-0.03656
			25	23	1	0.15023	0.00613	0.00016-0.03371	0.00613	0.00016-0.03371
				25	162					
			26	26	87	0.08203	0.02247	0.00273-0.07883		
				27	2				0.02247	0.00273-0.07883
			27	27	47	0.04332	0.00000	0.00000-0.07549		
			28	28	27	0.02581	0.03571	0.00090-0.18348		
				29	1				0.03571	0.00090-0.18348
			29	29	8	0.00737	0.00000	0.00000-0.36942		
			30	30	2	0.00184	0.00000	0.00000-0.84189		
			31	31	1	0.00092	0.00000	0.00000-0.97500		

Table A 23: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS518.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS518	74	3,990	32	32	4	0.00100	0.00000	0.00000-0.60236		
			33	33	21	0.00551	0.04545	0.00115-0.22844		
				34	1				0.04545	0.00115-0.22844
			33.1	33.1	1	0.00025	0.00000	0.00000-0.97500		
			34	34	58	0.01454	0.00000	0.00000-0.06162		
			34.1	34.1	1	0.00025	0.00000	0.00000-0.97500		
			35	35	158	0.04010	0.01250	0.00152-0.04442		
				36	1				0.00625	0.00016-0.03433
				38	1				0.00625	0.00016-0.03433
			36	35	1	0.06842	0.00733	0.00089-0.02621	0.00366	0.00009-0.02024
				36	271					
				36.2	1				0.00366	0.00009-0.02024
			36.2	36.2	3	0.00075	0.00000	0.00000-0.70760		
			36.3	36.3	1	0.00025	0.00000	0.00000-0.97500		
			37	36	2	0.12782	0.01569	0.00680-0.03067	0.00392	0.00048-0.01409
				37	502					
				38	5				0.00980	0.00319-0.02273
				40	1				0.00196	0.00005-0.01088
			37.2	37.2	8	0.00201	0.00000	0.00000-0.36942		
			37.3	37.3	1	0.00025	0.00000	0.00000-0.97500		
			38	37	4	0.19799	0.01392	0.00697-0.02478	0.00506	0.00138-0.01291
				38	779					

				39	5				0.00633	0.00206-0.01471
				41	1				0.00127	0.00003-0.00703
				42	1				0.00127	0.00003-0.00703
			38.2	38.2	3	0.00075	0.00000	0.00000-0.70760		
			39	38	8	0.17368	0.01732	0.00898-0.03005	0.01154	0.00500-0.02262
				39	681					
				40	4				0.00577	0.00157-0.01471
			39.2	39.2	2	0.00050	0.00000	0.00000-0.84189		
			40	38	1	0.14862	0.01686	0.00812-0.03079	0.00169	0.00004-0.00936
				39	5				0.00843	0.00274-0.01957
				40	583					
				41	3				0.00506	0.00104-0.01471
				42	1				0.00169	0.00004-0.00936
			40.2	40.2	1	0.00025	0.00000	0.00000-0.97500		
			41	37	1	0.10627	0.02594	0.01302-0.04595	0.00236	0.00006-0.01307
				40	4				0.00943	0.00258-0.02398
				41	413					
				42	6				0.01415	0.00521-0.03054
			42	41	2	0.06115	0.03689	0.01700-0.06886	0.00820	0.00099-0.02929
				42	235					
				43	7				0.02869	0.01161-0.05821
			43	42	2	0.03133	0.04000	0.01311-0.09088	0.01600	0.00194-0.05660
				43	120					
				44	3				0.02400	0.00498-0.06854

			44	43	1	0.01128	0.04444	0.00543-0.15149	0.02222	0.00056-0.11770
				44	43					
				46	1				0.02222	0.00056-0.11770
			45	43	1	0.00551	0.04545	0.00115-0.22844	0.04545	0.00115-0.22844
				45	21					
			46	46	5	0.00125	0.00000	0.00000-0.52182		
			47	47	1	0.00025	0.00000	0.00000-0.97500		

Table A 24: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS526_A.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS526_A	10	3,119	10	10	1	0.00032	0.00000	0.00000-0.97500		
			11	11	97	0.03110	0.00000	0.00000-0.03732		
			12	12	302	0.09715	0.00330	0.00008-0.01825		
				15	1				0.00330	0.00008-0.01825
			13	12	1	0.15678	0.00613	0.00127-0.01782	0.00204	0.00005-0.01134
				13	486					
				14	2				0.00409	0.00050-0.01470
			14	12	1	0.32062	0.00300	0.00062-0.00874	0.00100	0.00003-0.00556
				14	997					
				15	1				0.00100	0.00003-0.00556
				16	1				0.00100	0.00003-0.00556
			15	14	2	0.29080	0.00221	0.00027-0.00794	0.00221	0.00027-0.00794
				15	905					
			16	16	272	0.08753	0.00366	0.00009-0.02024		
				17	1				0.00366	0.00009-0.02024
			17	17	44	0.01411	0.00000	0.00000-0.08042		
			18	18	3	0.00096	0.00000	0.00000-0.70760		
			19	19	1	0.00032	0.00000	0.00000-0.97500		
			20	20	1	0.00032	0.00000	0.00000-0.97500		

Table A 25: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS526_B.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS526_B	43	3,173	30	30	5	0.00158	0.00000	0.00000-0.52182		
			31	31	101	0.03183	0.00000	0.00000-0.03586		
			32	32	145	0.04570	0.00000	0.00000-0.02512		
			33	32	1	0.03908	0.00806	0.00020-0.04411	0.00806	0.00020-0.04411
				33	123					
			34	34	241	0.07658	0.00823	0.00100-0.02941		
				35	2				0.00823	0.00100-0.02941
			35	34	2	0.10432	0.00906	0.00187-0.02626	0.00604	0.00073-0.02166
				35	328					
				36	1				0.00302	0.00008-0.01672
			35.2	35.2	1	0.00032	0.00000	0.00000-0.97500		
			36	35	3	0.16987	0.01484	0.00643-0.02903	0.00557	0.00115-0.01618
				36	531					
				37	5				0.00928	0.00302-0.02151
			37	35	1	0.17933	0.01054	0.00388-0.02281	0.00176	0.00004-0.00975
				36	1				0.00176	0.00004-0.00975
				37	563					
				38	4				0.00703	0.00192-0.01790
			37.1	37.1	1	0.00032	0.00000	0.00000-0.97500		
			38	37	3	0.16451	0.01533	0.00664-0.02997	0.00575	0.00119-0.01670
				38	514					
				39	4				0.00766	0.00209-0.01950

				40	1				0.00192	0.00005-0.01063
			39	38	4	0.12606	0.02000	0.00867-0.03903	0.01000	0.00273-0.02540
				39	392					
				40	4				0.01000	0.00273-0.02540
			40	39	2	0.04916	0.03205	0.01049-0.07321	0.01282	0.00156-0.04554
				40	151					
				41	3				0.01923	0.00398-0.05517
			41	40	2	0.00788	0.08000	0.00984-0.26031	0.08000	0.00984-0.26031
				41	23					
			42	42	10	0.00315	0.00000	0.00000-0.30850		
			43	43	1	0.00032	0.00000	0.00000-0.97500		

Table A 26: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS533.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS533	3	1,085	8	8	2	0.00184	0.00000	0.00000-0.84189		
			9	9	15	0.01382	0.00000	0.00000-0.21802		
			10	10	146	0.13456	0.00000	0.00000-0.02495		
			11	11	584	0.53825	0.00000	0.00000-0.00630		
			12	11	1	0.26452	0.00697	0.00085-0.02495	0.00348	0.00009-0.01926
				12	285					
				13	1				0.00348	0.00009-0.01926
			13	12	1	0.03594	0.02564	0.00065-0.13476	0.02564	0.00065-0.13476
				13	38					
			14	14	12	0.01106	0.00000	0.00000-0.26465		

Table A 27: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS547.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS547	55	3,159	40	40	1	0.00313	0.00000	0.00000-0.97500		
			41	41	5	0.01567	0.00000	0.00000-0.52182		
			42	41	2	0.02821	0.22222	0.02814-0.60009	0.22222	0.02814-0.60009
				42	7					
			43	43	48	0.15047	0.00000	0.00000-0.07397		
			44	44	87	0.27586	0.01136	0.00029-0.06169		
				45	1				0.01136	0.00029-0.06169
			45	45	193	0.60815	0.00515	0.00013-0.02838		
				46	1				0.00515	0.00013-0.02838
			46	45	1	1.01881	0.01846	0.00680-0.03975	0.00308	0.00008-0.01702
				46	319					
				47	5				0.01538	0.00501-0.03554
			46.2	46.2	1	0.00313	0.00000	0.00000-0.97500		
			47	46	4	1.57367	0.01793	0.00823-0.03376	0.00797	0.00218-0.02028
				47	493					
				48	5				0.00996	0.00324-0.02309
			47.2	47.2	11	0.03762	0.08333			
				48.2	1			0.00211-0.38480	0.08333	0.00211-0.38480
			48	47	4	2.37931	0.01054	0.00456-0.02066	0.00527	0.00144-0.01344
				48	751					
				49	4				0.00527	0.00144-0.01344
			48.2	48.2	98	0.30721	0.00000	0.00000-0.03694		

			49	48	6	1.84326	0.01361	0.00589-0.02663	0.01020	0.00375-0.02208
				49	580					
				50	2				0.00340	0.00041-0.01223
			49.2	49.2	18	0.05643	0.00000	0.00000-0.18530		
			49.3	49.3	1	0.00313	0.00000	0.00000-0.97500		
			50	49	10	1.12539	0.03343	0.01739-0.05766	0.02786	0.01344-0.05063
				50	347					
				51	2				0.00557	0.00068-0.01998
			50.2	50.2	4	0.01254	0.00000	0.00000-0.60236		
			51	50	3	0.30721	0.04082	0.01123-0.10122	0.03061	0.00636-0.08686
				51	94					
				52	1				0.01020	0.00026-0.05554
			52	51	2	0.11285	0.05556	0.00680-0.18664	0.05556	0.00680-0.18664
				52	34					
			53	52	1	0.03135	0.10000	0.00253-0.44502	0.10000	0.00253-0.44502
				53	9					
			54	54	2	0.00627	0.00000	0.00000-0.84189		
			55	55	1	0.00313	0.00000	0.00000-0.97500		

Table A 28: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS549.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS549	1	104	11	11	13	0.12500	0.00000	0.00000-0.24705		
			12	12	47	0.45192	0.00000	0.00000-0.07549		
			13	12	1	0.38462	0.02500	0.00063-0.13159	0.02500	0.00063-0.13159
				13	39					
			14	14	4	0.03846	0.00000	0.00000-0.60236		

Table A 29: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS570.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS570	43	4,293	13	13	2	0.00047	0.00000	0.00000-0.84189		
			13.3	13.3	1	0.00023	0.00000	0.00000-0.97500		
			14	14	21	0.00489	0.00000	0.00000-0.16110		
			15	15	67	0.01584	0.01471	0.00037-0.07923		
				16	1				0.01471	0.00037-0.07923
			16	16	390	0.09108	0.00256	0.00006-0.01417		
				17	1				0.00256	0.00006-0.01417
			17	14	1	0.25553	0.00912	0.00438-0.01670	0.00091	0.00002-0.00507
				15	1				0.00091	0.00002-0.00507
				16	2				0.00182	0.00022-0.00657
				17	1087					
				18	5				0.00456	0.00148-0.01060
				19	1				0.00091	0.00002-0.00507

			18	17	6	0.25786	0.00903	0.00434-0.01655	0.00542	0.00199-0.01176
				18	1097					
				19	4				0.00361	0.00099-0.00923
			19	18	5	0.22455	0.00830	0.00359-0.01629	0.00519	0.00169-0.01206
				19	956					
				20	3				0.00311	0.00064-0.00907
			19.3	19.3	4	0.00093	0.00000	0.00000-0.60236		
			20	18	2	0.10529	0.01770	0.00767-0.03458	0.00442	0.00054-0.01589
				19	3				0.00664	0.00137-0.01927
				20	444					
				21	3				0.00664	0.00137-0.01927
			20.2	20.2	2	0.00047	0.00000	0.00000-0.84189		
			20.3	20.3	2	0.00047	0.00000	0.00000-0.84189		
			21	20	4	0.02958	0.03150	0.00865-0.07868	0.03150	0.00865-0.07868
				21	123					
			22	21	1	0.00885	0.02632	0.00067-0.13810	0.02632	0.00067-0.13810
				22	37					
			23	23	14	0.00326	0.00000	0.00000-0.23164		
			24	24	3	0.00070	0.00000	0.00000-0.70760		

Table A 30: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS576.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS576	65	4,194	12	12	1	0.00024	0.00000	0.00000-0.97500		
			12.1	12.1	1	0.00024	0.00000	0.00000-0.97500		
			13	13	10	0.00238	0.00000	0.00000-0.30850		
			14	14	34	0.00811	0.00000	0.00000-0.10282		
			15	15	164	0.03934	0.00606	0.00015-0.03330		
				16	1				0.00606	0.00015-0.03330
			16	16	599	0.14378	0.00663	0.00181-0.01690		
				17	4				0.00663	0.00181-0.01690
			17	16	1	0.24154	0.00888	0.00407-0.01680	0.00099	0.00002-0.00549
				17	1004					
				18	8				0.00790	0.00342-0.01550
			18	17	10	0.32594	0.01609	0.01011-0.02427	0.00732	0.00351-0.01341
				18	1,345					
				19	12				0.00878	0.00454-0.01528
			18.2	18.2	1	0.00024	0.00000	0.00000-0.97500		
			19	16	2	0.16857	0.02829	0.01736-0.04335	0.00283	0.00034-0.01018
				18	11				0.01556	0.00779-0.02767
				19	687					
				20	7				0.00990	0.00399-0.02029
			19.1	19.1	1	0.00024	0.00000	0.00000-0.97500		
			20	18	1	0.05651	0.02110	0.00688-0.04854	0.00422	0.00011-0.02328

				19	1				0.00422	0.00011-0.02328
				20	232					
				21	3				0.01266	0.00262-0.03654
			20.2	20.2	3	0.00072	0.00000	0.00000-0.70760		
			21	20		0.01097	0.08696	0.02420-0.20792	0.08696	0.02420-0.20792
				21	42					
			22	22	5	0.00119	0.00000	0.00000-0.52182		

Table A 31: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS612.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS612	54	3,188	25	25	1	0.00031	0.00000	0.00000-0.97500		
			27	27	1	0.00031	0.00000	0.00000-0.97500		
			28	28	1	0.00031	0.00000	0.00000-0.97500		
			30	30	5	0.00157	0.00000	0.00000-0.52182		
			31	31	9	0.00314	0.10000	0.00253-0.44502		
				32	1				0.10000	0.00253-0.44502
			32	32	36	0.01223	0.07692	0.01615-0.20870		
				33	3				0.07692	0.01615-0.20870
			33	33	81	0.02604	0.02410	0.00293-0.08435		
				34	1				0.01205	0.00030-0.06531
				35	1				0.01205	0.00030-0.06531
			34	32	1	0.07967	0.00787	0.00096-0.02815	0.00394	0.00010-0.02174
				34	252					
				35	1				0.00394	0.00010-0.02174
			35	34	4	0.14806	0.02754	0.01474-0.04664	0.00847	0.00231-0.02156
				35	459					
				36	8				0.01695	0.00735-0.03312
				37	1				0.00212	0.00005-0.01175
			36	35	3	0.28262	0.00666	0.00245-0.01444	0.00333	0.00069-0.00970
				36	895					
				37	3				0.00333	0.00069-0.00970

			37	36	5	0.21455	0.01754	0.00910-0.03044	0.00731	0.00238-0.01698
				37	672					
				38	7				0.01023	0.00412-0.02097
			38	36	1	0.13049	0.02644	0.01327-0.04682	0.00240	0.00006-0.01332
				37	5				0.01202	0.00391-0.02782
				38	405					
				39	4				0.00962	0.00263-0.02444
				40	1				0.00240	0.00006-0.01332
			39	38	2	0.06932	0.01357	0.00281-0.03916	0.00905	0.00110-0.03231
				39	218					
				40	1				0.00452	0.00011-0.02495
			40	40	42	0.01317	0.00000	0.00000-0.08408		
			40.1	40.1	33	0.01035	0.00000	0.00000-0.10576		
			41	41	16	0.00502	0.00000	0.00000-0.20591		
			41.1	41.1	6	0.00188	0.00000	0.00000-0.45926		
			42	42	1	0.00063	0.50000	0.01258-0.98742		
				44	1				0.50000	0.01258-0.98742
			42.1	42.1	1	0.00031	0.00000	0.00000-0.97500		

Table A 32: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS626.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS626	34	3,138	22	22	1	0.00032	0.00000	0.00000-0.97500		
			23	23	2	0.00064	0.00000	0.00000-0.84189		
			24	24	18	0.00574	0.00000	0.00000-0.18530		
			25	25	106	0.03378	0.00000	0.00000-0.03420		
			26	26	175	0.05577	0.00000	0.00000-0.02086		
			27	25	1	0.04589	0.01389	0.00169-0.04927	0.00694	0.00018-0.03808
				26	1				0.00694	0.00018-0.03808
				27	142					
			28	28	183	0.05895	0.01081	0.00131-0.03850		
				29	2				0.01081	0.00131-0.03850
			29	28	1	0.18101	0.00352	0.00043-0.01266	0.00176	0.00004-0.00977
				29	566					
				30	1				0.00176	0.00004-0.00977
			29.1	29.1	1	0.00032	0.00000	0.00000-0.97500		
			30	29	1	0.18961	0.00336	0.00041-0.01209	0.00168	0.00004-0.00933
				30	593					
				31	1				0.00168	0.00004-0.00933
			30.2	30.2	1	0.00032	0.00000	0.00000-0.97500		
			31	30	5	0.18547	0.01203	0.00485-0.02462	0.00859	0.00280-0.01993
				31	575					
				32	2				0.00344	0.00042-0.01236

			32	31	4	0.13576	0.01643	0.00663-0.03356	0.00939	0.00256-0.02387
				32	419					
				33	3				0.00704	0.00145-0.02044
			33	27	1	0.07393	0.03448	0.01500-0.06681	0.00431	0.00011-0.02378
				32	6				0.02586	0.00955-0.05544
				33	224					
				34	1				0.00431	0.00011-0.02378
			33.2	33.2	1	0.00032	0.00000	0.00000-0.97500		
			34	33	1	0.02581			0.01235	0.00031-0.06688
				34	79		0.02469	0.00300-0.08636		
				35	1				0.01235	0.00031-0.06688
			35	34	1	0.00510	0.06250	0.00158-0.30232	0.06250	0.00158-0.30232
				35	15					
			36	35	1	0.00064	0.50000	0.01258-0.98742	0.50000	0.01258-0.98742
				36	1					
			39	39	1	0.00032	0.00000	0.00000-0.97500		
			45.2	45.2	1	0.00032	0.00000	0.00000-0.97500		

Table A 33: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS627.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS627	66	4,157	13	13	1	0.00024	0.00000	0.00000-0.97500		
			14	14	3	0.00072	0.00000	0.00000-0.70760		
			15	15	25	0.00601	0.00000	0.00000-0.13719		
			16	16	274	0.06591	0.00000	0.00000-0.01337		
			16.2	16.2	2	0.00048	0.00000	0.00000-0.84189		
			17	17	379	0.09117	0.00000	0.00000-0.00969		
			17.1	17.1	1	0.00024	0.00000	0.00000-0.97500		
			17.2	17.2	4	0.00096	0.00000	0.00000-0.60236		
			18	17	1	0.07289	0.01320	0.00361-0.03345	0.00330	0.00008-0.01825
				18	299					
				19	2				0.00660	0.00080-0.02364
				21	1				0.00330	0.00008-0.01825
			18.2	18.2	1	0.00024	0.00000	0.00000-0.97500		
			19	17	1	0.11210	0.01073	0.00349-0.02486	0.00215	0.00005-0.01190
				18	2				0.00429	0.00052-0.01542
				19	461					
				20	2				0.00429	0.00052-0.01542
			19.2	19.2	2	0.00048	0.00000	0.00000-0.84189		
			20	18	1	0.16719	0.01727	0.00895-0.02997	0.00144	0.00004-0.00799
				19	2				0.00288	0.00035-0.01036
				20	683					
				21	8				0.01151	0.00498-0.02255

				22	1				0.00144	0.00004-0.00799
			20.2	20.2	1	0.00024	0.00000	0.00000-0.97500		
			20.3	20.3	1	0.00024	0.00000	0.00000-0.97500		
			21	20	8	0.18884	0.01529	0.00792-0.02655	0.01019	0.00441-0.01998
				21	773					
				22	4				0.00510	0.00139-0.01299
			21.2	21.2	1	0.00024	0.00000	0.00000-0.97500		
			22	20	1	0.17344	0.02080	0.01169-0.03408	0.00139	0.00004-0.00770
				21	11				0.01526	0.00764-0.02713
				22	706					
				23	3				0.00416	0.00086-0.01211
			22.2	22.2	2	0.00048	0.00000	0.00000-0.84189		
			23	22	7	0.08275	0.02907	0.01403-0.05281	0.02035	0.00822-0.04148
				23	334					
				24	3				0.00872	0.00180-0.02527
			23.3	23.3	1	0.00024	0.00000	0.00000-0.97500		
			24	23	3	0.02550	0.04717	0.01549-0.10665	0.02830	0.00587-0.08049
				24	101					
				25	2				0.01887	0.00229-0.06650
			25	24	1	0.00770	0.06250	0.00766-0.20807	0.03125	0.00079-0.16217
				25	30					
				26	1				0.03125	0.00079-0.16217
			26	25	1	0.00144	0.16667	0.00421-0.64123	0.16667	0.00421-0.64123
				26	5					

			27	27	1	0.00024	0.00000	0.00000-0.97500		

Table A 34: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS635.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS635	20	6,916	17	17	2	0.00029	0.00000	0.00000-0.84189		
			18	18	1	0.00014	0.00000	0.00000-0.97500		
			19	19	150	0.02169	0.00000	0.00000-0.02429		
			20	20	622	0.08994	0.00000	0.00000-0.00591		
			21	20	1	0.22932	0.00252	0.00069-0.00644	0.00063	0.00002-0.00351
				21	1,582					
				22	3				0.00189	0.00039-0.00552
			22	21	7	0.12175	0.00831	0.00335-0.01705	0.00831	0.00335-0.01705
				22	835					
			23	22	3	0.42380	0.00136	0.00037-0.00349	0.00102	0.00021-0.00299
				23	2,927					
				24	1				0.00034	0.00001-0.00190
			24	23	2	0.09268	0.00312	0.00038-0.01123	0.00312	0.00038-0.01123
				24	639					
			25	24	1	0.01822	0.02381	0.00494-0.06800	0.00794	0.00020-0.04343
				25	123					
					2				0.01587	0.00193-0.05616
			26	26	14	0.00202	0.00000	0.00000-0.23164		
			27	27	1	0.00014	0.00000	0.00000-0.97500		

Table A 35: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker DYS643.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
DYS643	0	104	9	9	10	0.09615	0.00000	0.00000-0.30850		
			10	10	46	0.44231	0.00000	0.00000-0.07706		
			11	11	18	0.17308	0.00000	0.00000-0.18530		
			12	12	29	0.27885	0.00000	0.00000-0.11944		
			13	13	1	0.00962	0.00000	0.00000-0.97500		

Table A 36: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker GATA H4.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
GATA H4	13	7,128	8	8	1	0.00014	0.00000	0.00000-0.97500		
			9	9	31	0.00435	0.00000	0.00000-0.11219		
			10	10	314	0.04405	0.00000	0.00000-0.01168		
			11	11	2,423	0.34035	0.00124	0.00026-0.00361		
				12	3				0.00124	0.00026-0.00361
			12	11	7	0.52652	0.00213	0.00092-0.00420	0.00187	0.00075-0.00384
				12	3,745					
				13	1				0.00027	0.00001-0.00148
			13	12	2	0.07870	0.00357	0.00043-0.01282	0.00357	0.00043-0.01282
				13	559					
			14	14	40	0.00561	0.00000	0.00000-0.08810		
			15	15	2	0.00028	0.00000	0.00000-0.84189		

Table A 37: Mutation and bi-allele mutation rates and the confidence interval (95%) for marker GATA A10.

Marker	No. of mutations	No. of meiosis	Original allele	Filial allele	No. of observations	Frequency of the original allele	Allele mutation rate	Confidence interval (95%)	Bi-allele mutation rate	Confidence interval (95%)
GATA A 10	4	874	13	13	33	0.03776	0.00000	0.00000-0.10576		
			14	13	1	0.32380	0.00707	0.00086-0.02529	0.00353	0.00009-0.01953
				14	281					
				15	1				0.00353	0.00009-0.01953
			15	14	2	0.49542	0.00462	0.00056-0.01658	0.00462	0.00056-0.01658
				15	431					
			16	16	110	0.12586	0.00000	0.00000-0.03298		
			17	17	14	0.01602	0.00000	0.00000-0.23164		
			18	18	1	0.00114	0.00000	0.00000-0.97500		