Forensic Science International: Genetics 34 (2018) 1–10

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Determining Y-STR mutation rates in deep-routing genealogies: Identification of haplogroup differences

Sofie Claerhout[a,*], Michiel Vandenbosch[a], Kelly Nivelle[a], Leen Gruyters[b], Anke Peeters[a], Maarten H.D. Larmuseau[a,b,1], Ronny Decorte[a,c,1]

[a] KU Leuven, Forensic Biomedical Sciences, Department of Imaging & Pathology, Leuven, Belgium
[b] KU Leuven, Laboratory of Socioecology and Social Evolution, Department of Biology, Leuven, Belgium
[c] UZ Leuven, Laboratory of Forensic Genetics and Molecular Archaeology, Leuven, Belgium

ARTICLE INFO

Keywords:
Y-chromosome
Short tandem repeat (STR)
Haplotype
Single nucleotide polymorphism (SNP)
Haplogroup
Mutation rate

ABSTRACT

Knowledge of Y-chromosomal short tandem repeat (Y-STR) mutation rates is essential to determine the most recent common ancestor (MRCA) in familial searching or genealogy research. Up to now, locus-specific mutation rates have been extensively examined especially for commercially available forensic Y-STRs, while haplogroup specific mutation rates have not yet been investigated in detail. Through 450 patrilineally related namesakes distributed over 212 deep-rooting genealogies, the individual mutation rates of 42 Y-STR loci were determined, including 27 forensic Y-STR loci from the Yfiler® Plus kit and 15 additional Y-STR loci (DYS388, DYS426, DYS442, DYS447, DYS454, DYS455, DYS459a/b, DYS549, DYS607, DYS643, DYS724a/b and YCAIIa/b). At least 726 mutations were observed over 148,596 meiosis and individual Y-STR mutation rates varied from 2.83×10^{-4} to 1.86×10^{-2}. The mutation rate was significantly correlated with the average allele size, the complexity of the repeat motif sequence and the age of the father. Significant differences in average Y-STR mutations rates were observed when haplogroup 'I & J' (4.03×10^{-3} mutations/generation) was compared to 'R1b' (5.35×10^{-3} mutations/generation) and to the overall mutation rate (5.03×10^{-3} mutations/generation). A difference in allele size distribution was identified as the only cause for these haplogroup specific mutation rates. The haplogroup specific mutation rates were also present within the commercially available Y-STR kits (Yfiler®, PowerPlex® Y23 System and Yfiler® Plus). This observation has consequences for applications where an average Y-STR mutation rate is used, e.g. tMRCA estimations in familial searching and genealogy research.

1. Introduction

The Y-chromosome acts as a unique tool for forensic investigations since it is inherited through the patrilineal line in a relatively conserved manner. Around 95% of the Y-chromosome does not recombine with the X-chromosome, which refers to the 'non-recombining region on the Y-chromosome' (NRY). The only source of genetic variation on the Y-chromosome between men is the occurrence of mutations on the NRY during meiosis. The Y-chromosome therefore provides evidence for a biological paternal kinship and enables different forensic applications such as paternity tests, identification of suspects, familial searching and tracing a biological relationship between two men. On top, it is an interesting tool for forensic genetics as it can be used to identify a male DNA profile in male-female mixture samples. Genotyping for the male specific Y-chromosome is not hindered by any female DNA, facilitates exclusion of other male suspects from complicity, detects potential multiple male contributors and accelerates the identification of the male contributor [1].

The NRY contains slowly mutating bi-allelic Y-chromosomal single nucleotide polymorphisms (Y-SNPs) (ca. 2×10^{-8} mutations/generation) [2], which divide the male human population into 20 evolutionary Y-chromosome haplogroups and deep subhaplogroups. Around 13,000 SNPs have been identified from whole genome sequencing data [3], while only a core set of Y-SNPs is needed for forensic or genealogical applications to identify geographical origins [4]. Fast mutating multi-allelic Y-chromosomal short tandem repeats (Y-STRs) are DNA polymorphisms with a relatively high mutation frequency (ca. 5.91×10^{-3} mutations/generation) [2], which are used to obtain a unique Y-haplotype. This fast mutation rate provides a high Y-STR heterozygosity, theoretically giving every non-paternally related man a

* Corresponding author at: KU Leuven, Forensic Biomedical Sciences, Kapucijnenvoer 33, B-3000 Leuven, Belgium.
E-mail address: sofie.claerhout@kuleuven.be (S. Claerhout).
[1] These authors share senior authorship.

https://doi.org/10.1016/j.fsigen.2018.01.005
Received 15 November 2017; Received in revised form 10 January 2018; Accepted 14 January 2018
Available online 17 January 2018
1872-4973/

different haplotype. A difference in a single locus is sufficient to successfully distinguish two close relatives [5]. This makes it interesting to analyze both Y-STRs with slow mutation rates to increase the chances of finding relatives, and Y-STRs with fast mutation rates to distinguish close relatives. Ballantyne et al. studied 186 Y-STRs through genotyping 2000 father-son couples, making it possible to identify 13 rapidly mutating (RM) Y-STRs with a median mutation rate of 1.97×10^{-2} [2]. With the discovery of RM Y-STRs the average mutation rate increases by a factor ten compared to the conventional Y-STRs, which makes it more likely for mutations to occur in the meiosis separating the patrilineal relatives and allowing forensic DNA analysis to individualize relatives. Known molecular factors influencing mutability rates of Y-STRs are the length (base pairs) of the repeat unit, the average allele size (number of repeats), the complexity of the repetitive sequence and, although not always confirmed, the age of the father at the time of Y-chromosome transmission [2,6–8]. Since most studies on Y-STR mutation rates focus on the same sets of Y-STRs used in forensic genetics (Yfiler™, PowerPlex® Y23 or Yfiler® Plus), it is important to expand our knowledge concerning mutation rates and molecular factors for other less studied Y-STRs.

Besides differences in mutation rates between Y-STRs, Dupuy et al. indicated in 2004 small differences in average Y-STR mutation rates between six (sub)haplogroups defined by five Y-SNPs and indicated that these differences were probably due to a difference in allele size [9]. This would indicate that not only locus-specific mutation rates should be used for dating Y-chromosome lineages or for paternity cases [1], but also haplogroup specific mutation rates. Unfortunately, the paper of Dupuy et al. was only based on nine Y-STRs and 'broadly' defined haplogroups. Until now, this research question has not yet been further investigated in detail since most studies focus on Y-STR mutation rates without including Y-SNP analysis for determination of the (sub)haplogroup. Furthermore, Balanovsky recently stated that haplogroup specific mutation rates could be possible 'in theory' and can therefore not be excluded [10]. Y-STR mutation rates based on one haplogroup were already questioned in a few studies, but until now no significant differences could be observed [3,11,12]. More precise knowledge of individual Y-STR mutation rates provides the opportunity to improve estimations for the time to the most recent common ancestor (tMRCA) between two paternally related men [10,13]. Timescales of tens to hundreds of generations are highly dependent on the accuracy of Y-STR mutation rates and the correct knowledge concerning the different molecular factors influencing their mutability rates. When a difference in Y-STR mutation rates between different Y-SNP (sub)haplogroups is present, there could be consequences for population studies where average Y-STR mutation rates are used. Correspondingly, the tMRCA estimations based on the average Y-STR mutation rates used for familial searching analysis could be over- or underestimated between paternally related couples of namesakes.

In this study, 540 patrilineally related namesakes distributed over 212 deep-routing genealogies were used to estimate individual mutation rates for 42 Y-STRs, including currently used Y-STRs in commercial kits and 15 Y-STRs commonly used in genetic genealogy studies. This dataset was used to investigate molecular factors influencing Y-STR mutability rates and to identify differences in individual Y-STR mutation rates between the two most prominent haplogroups in the dataset: subhaplogroup 'R1b' and sister haplogroups 'I & J'.

2. Materials and methods

2.1. DNA sample collection

The study includes 321 different paternal pedigrees with in total 723 males for which the legal kinship was confirmed through archival data. The collected namesakes are all male patrilineal relatives from the Low Countries with an MRCA between 1400 and 1850. The related participants needed to be separated by at least seven meiotic divisions (or generations) for ethical purposes. Written informed consents were obtained for permission on both DNA analysis and on the scientific publication of the anonymized results. DNA samples were collected through buccal swabs (Whatman™ OmniSwab). Ethical approval (S55864, S59085) for this study has been given by the Ethical Commission of University Hospital Leuven.

2.2. DNA extraction and Y-chromosome genotyping

DNA from the buccal swabs was extracted by using 0.5 ml of SwabSolution™ Reagent (Promega, Madison, WI, USA) with a quarter of a buccal swab. The extracted DNA samples were genotyped for 38 Y-STRs, including 32 single-copy Y-STRs and six multi-copy Y-STRs, which mostly resulted in a 46 Y-STR loci haplotype. 27 of the 46 Y-STR loci are included in the commercially available Yfiler® Plus genotyping kit, namely DYF387S1a/b, DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS449, DYS456, DYS458, DYS460, DYS481, DYS518, DYS533, DYS570, DYS576, DYS627, DYS635 and GATA H4.1. In addition, 19 other Y-STR loci were included, i.e. DYS388, DYS426, DYS442, DYS447, DYS454, DYS455, DYS459a/b, DYS464a/b/c/d, DYS549, DYS607, DYS643, DYS724a/b and YCAIIa/b (Supplementary Fig. 1). These 46 Y-STR loci were distributed over four multiplex (MP) kits, whereby MP2, MP4 and MP5p were developed in previous research [14,15] and a new multiplex kit MP7 was developed using primer sequences obtained from literature [5]. The MP7 kit contains two RM single-copy Y-STR markers (DYS518 and DYS627), one RM multi-copy Y-STR marker (DYF387S1a/b) and two previously used Y-STR markers (DYS576 and DYS533) as an internal control.

DNA was amplified using 6.25 µl Qiagen® Multiplex PCR Kit (Qiagen), 1.75 µl AmpSolution™ Reagent (Promega, Madison, WI, USA), 2.5 µl primer mix and 3 µl of DNA extract. PCR conditions were as described by Jacobs et al. for MP2, MP4 and MP5p with the exception of the final extension being extended to 45' at 60 °C [14]. Optimized MP7 PCR conditions were the following: 15' at 94 °C, 5 cycles of 30" at 94 °C, 30" at 65 °C and 1' at 72 °C followed by 24 cycles of 30" at 94 °C, 30" at 62 °C and 1' at 72 °C, and at the end a final extension of 45' at 60 °C. The amplified DNA samples were purified with the BigDye XTerminator® Purification Kit (Applied Biosystems). The purified DNA samples were plated out in Hi-Di™ Formamide on a 96 well-plate for analysis on an ABI PRISM 3130 XL Genetic Analyzer with POP7 and a 50 cm capillary (Applied Biosystems) using GeneScan™ 500 LIZ™ Dye Size Standard (Applied Biosystems) as a size standard. Fragment length analysis of the 46 Y-STRs was done using GeneMapper® v3.2.1 (Applied Biosystems).

Whit Athey's Haplogroup Predictor (http://hprg.com/hapest5) was used to predict the most likely haplogroup in Northwest Europe based on their Y-STR haplotype, making it convenient to find the most optimal multiplex kit for Y-SNP analysis [16]. When a haplogroup prediction was not possible for a specific DNA sample, we genotyped it with a general Y-SNP multiplex kit containing all phylogenetically important Y-SNPs to find the main haplogroup. A total of 81 Y-SNPs were divided in 13 MP kits to confirm the haplogroup and to define the subhaplogroup of the DNA sample. For amplification, two µl of DNA extract was added to 23 µl of amplification mix containing 12.5 µl 2x QIAGEN Multiplex PCR Master Mix, 0.5 µl of Nuclease-Free H_2O (Promega, Madison, WI, USA), 5 µl of 5X AmpSolution™ Reagent, and 5 µl of the relevant multiplex primer mix. PCR conditions were: 15' at 95 °C, 30 cycles of 30" at 94 °C, 30" at 55 °C and 90" at 72 °C followed by a final extension of 10' at 72 °C. Analysis of the amplified DNA samples was done with SNaPshot® Multiplex System for SNP genotyping (Applied Biosystems) as previously described [17–19].

2.3. Data analysis

The pedigree-based dataset gives the opportunity to confirm the biological kinship between namesakes through Y-haplotype