Forensic Science International: Genetics 41 (2019) 137–144

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

A comprehensive mutation study in wide deep-rooted R1b Serbian pedigree: mutation rates and male relative differentiation capacity of 36 Y-STR markers

Vladan P. Čokić[a,1], Miljana Kecmanović[b,1], Dragana Zgonjanin Bosić[c], Zlatko Jakovski[d], Aleksandar Veljković[e], Srdjan Katić[f], Milica Keckarević Marković[b], Dušan Keckarević[b,*]

[a] Institute for Medical Research, University of Belgrade, Belgrade, Serbia
[b] Center for Forensic and Applied Molecular Genetics, Faculty of Biology, University of Belgrade, Belgrade, Serbia
[c] Institute of Forensic Medicine, Clinical Center of Vojvodina, Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia
[d] Institute for forensic medicine, criminology and medical deontology, Medical Faculty, University of St. Cyril and Methodius, Skopje, Macedonia
[e] Faculty of Mathematics, University of Belgrade, Belgrade, Serbia
[f] Institute of History, Belgrade, Serbia

ARTICLE INFO

Keywords:
Deep-Rooted pedigree
Y-STR
Mutation rate
Male relative differentiation

ABSTRACT

Haplotyping of Y-chromosomal short tandem repeats (Y-STRs) reflects the paternal lineage, although, the father-son pair profiles may differ due to the germline mutations. In order to discriminate between closely related males in criminal cases, as well as for the correct application of Y-STRs in the paternity/kinship analysis and determination of the most recent common ancestor in the familial searching or genealogy research, the assessment of mutation rates of routinely used Y-STRs is of a great importance. We genotyped 120 males belonging to one wide deep-rooted pedigree separated by 1–20 meiosis. The haplotypes of analyzed males distributed over 12 different families (according to their surnames), with 113 originating from one ancestor, and the remaining 7 from the second, closely related to the previous one, belong to the R1b haplogroup. The analysis was performed using Powerplex® Y23 kit, Yfiler™ plus kit and 13 rapidly mutating (RM13) Y-STRs. In 20,855 allele transmissions, 175 mutations (61% repeat losses and 39% gains) and one gene conversion event were found at 25 out of 36 markers. The medians of locus-specific mutation rates estimated using the Bayesian approach ranged from 1.42×10^{-3} (95% credible interval (CI): 0.05×10^{-3} - 7.56×10^{-3}) for loci with no observed mutations to 130.91×10^{-3} (95% CI: 102.91×10^{-3} - 162.78×10^{-3}) for *DYF399S1,* with a median rate across all 36 markers of 10.06×10^{-3} (95% CI: 8.65×10^{-3} - 11.61×10^{-3}). In 6349 male relative pairs, the 36 Y-STR set distinguished 98.4% relative pairs by at least one mutation, compared to 95.9%, 65.5% and 57.4% for RM13, Yfiler™ plus, and Powerplex® Y23 set, respectively. The extra-pair paternity rate was estimated at 11.9×10^{-3} (95% CI: 4.4×10^{-3} – 25.8×10^{-3}) fitting within the range reported for some European populations. A significant positive correlation was observed between fathers' ages at the time of the Y chromosome transmission and mutability rates ($R^2 = 0.9495$, $p = 0.0256$), with more significant results when analyzing RM markers ($R^2 = 0.9827$, $p = 0.0087$).

1. Introduction

The combined analysis of Y-STRs is a powerful tool that enables the identification of the paternal lineage in forensic caseworks, paternity testing and familial search [1,2]. Even though there are no DNA recombination events, spontaneous germline mutations could lead to a differentiation of Y-STR haplotypes between a father and his sons. The mutations at specific STR loci on the Y chromosome are proposed to facilitate the identification of male individuals in forensic investigations [3], but they could also lead to an erroneous exclusion of biological paternity [4]. Y-STRs are also used in anthropological, genealogical and evolutionary studies [1,2], but for the reconstruction of the Y chromosome phylogeny and the determination of the divergence time of different lineages, single-nucleotide polymorphisms (SNPs) on the Y

* Corresponding author at: University of Belgrade, Faculty of Biology, Center for forensic and applied molecular genetics, Studentski trg 16, Serbia.
E-mail address: duskec@bio.bg.ac.rs (D. Keckarević).
[1] equally contributed.

https://doi.org/10.1016/j.fsigen.2019.04.007
Received 31 December 2018; Received in revised form 5 April 2019; Accepted 29 April 2019
Available online 04 May 2019
1872-4973/

chromosome are more suitable [5–7].

The autosomal STR typing leads to the identification of a perpetrator, which is a crucial requirement in the forensic casework, unlike the Y-STR loci that are still suitable for excluding the suspect as a trace donor only. In order to differentiate a male perpetrator from other males paternally related to him, an increase in the number of markers included in the Y-STR kits could partly contribute to that demand [8]. Over the years, the capillary electrophoresis (CE) has improved and more and more Y-STR markers have been validated [9–13] and included in commercially available kits, reaching up to 25 in Yfiler® Plus kit (*Thermo Fisher Scientific*, CA, USA). Nevertheless, an increased number of markers in kits also has technical limitations in routine forensic analysis.

Most Y-STR markers included in currently available Y-STR kits with low to medium mutation rates $\sim 10^{-3}$ [3,14] are suitable for paternity testing, but not for discriminating paternally related males at the same time. On the other hand, the differentiation of closely related males reaching the goal of forensic genetics in criminal cases, was significantly increased after comprehensive studies by Ballantyne et al. [15,16] and introduction of a set of 13 rapidly mutating (RM) Y-STRs with the median mutation rate of $\sim 1.97 \times 10^{-2}$. However, the usage of markers with high mutation rates could also result in an incorrect exclusion of the biological paternity. Thus, both types of Y chromosome STRs are essential for forensic DNA analysis and a precise establishment of their mutation rates has a great importance for gaining certainty in the Y-STR data interpretation in their forensic, paternity/kinship and evolutionary application. Furthermore, the mutation rates are important for the estimation of the time to the most recent common ancestor (tMRCA) in genealogical studies. It has been recently noticed that even small differences in Y-STR mutation rates could be detected in different Y haplogroups [17–19], and that precise mutation rates data are needed in order to avoid over- or underestimations of tMRCA, which could occur when average mutation rates are used [19].

There are several approaches for the establishment of Y-STR mutation rates including the analysis of father/son pairs [20], the related males from deep-rooted pedigrees [19,21–24], the single or small pools of sperm cells [25], or the use of Y-STR population data in combination with known historical events for the time calibration [5]. So far, the vast majority of studies of the estimation of the Y-STR mutation rates and differentiation of male relatives was conducted on father/son pairs where a large sample size is needed [4,16,26–30], or by using a different number of deep-rooted pedigrees to enable deduction of Y-STR mutation rates with a smaller number of DNA samples [19]. However, both approaches have their own limitations. Father/son pairs analysis does not consider a possible influence of different time scale dependent factors on the mutation rates (*e.g.*, increase in generation time, enormous change in life conditions) and the latter suffers from uncertainty in the ancestral Y-STR haplotype assumption and possible hidden backward mutations that could lead to inaccuracy in the number of detected mutations and their direction (forwards or backwards). The molecular factors associated with Y-STR's mutability rates are the length of the repeat unit, number of the repeats and the complexity of the repetitive sequence [3]. Several studies showed a correlation between the father's age at the time of the son's birth and the mutation rates [3,14,19].

Analysis of wide and deep-rooted pedigrees can provide an estimation of the actual mutation rates in the same DNA background through the time scale, as well as information about the male differentiation capacity in more distant relatives. Although some mutations could be hidden, multiple parallel mutations could be easily detected, thus allowing a more precise estimation of tMRCA, unlike a small deep rooted-pedigrees analysis where Y-STR alleles should be sequenced in order to determine the independent origin of identical alleles obtained after CE [31].

As far as we know, there is no data about the Y-STR mutation rates and male relative differentiation for large wide and deep-rooted pedigrees. In our study, we provided robust data on the mutation rates of 36 markers, combining PPY23, Yfiler + and RM13 loci, in one wide deep-rooted pedigree consisting of 120 male relatives related by 485 meiotic events. The use of different kits with overlapping loci allowed us to conduct a concordance study. We also studied discrimination capacity of the different sets of analyzed loci, the extra-pair paternity rate, and influence of the father's age at the time of transmission to the mutation rates.

2. Materials and methods

2.1. DNA sampling

Buccal swabs from 128 individuals were collected from 2015 to 2018. The pedigree-based dataset involved individuals from the population of central Serbia, located between the mountains Rudnik and Kosmaj. The samples belonging to 12 families according to their surnames (a specific surname does not imply a unique lineage, since heritable surnames were officially established in Serbia in the middle of the 19th century) were selected on the basis of their documented affiliation to one deep-rooted paternal pedigree based on the gathered genealogies, which were available after a long-standing survey (2008–2018) of a variety of record types, including Ottoman census registers and records from Serbian municipal and parish archives (1741–2017). Furthermore, all families within a pedigree shared the same migration history and patron saint (they celebrated St. George and they migrated from Sjenica to the mountain Rudnik in the middle of the 18th century), which is paternally transferred, hence, co-ancestry of the Y chromosome is expected.

Of 128 males considered for the study, 121 males (11 surnames) were thought to be descendants of one common ancestor (A1) who had three sons giving three sub-branches with 46, 62 and 13 analyzed contemporaries, while the other 7 (1 surname) originated from the ancestor A2 (Fig. 1, and Suppl. Fig. 1). The exact relationship between A1 and A2 was not clear, but it was supposed based on the same migration origin, patron saint and genealogical data that a common ancestor for A1 and A2 might have lived a few generations earlier. A paternal exclusion was observed in 8 samples of the A1 branch, 4 from the first and 4 from the second A1 sub-branch, after an initial analysis with a PPY23 kit. The allelic differences at more than four Y-STR markers relative to the closest pedigree peers [14] were the criteria for exclusion.

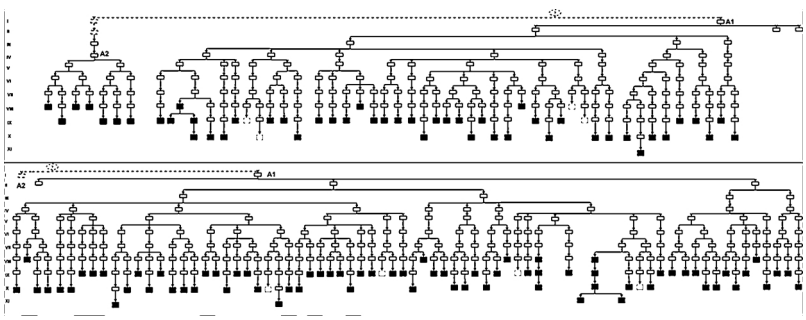


Fig. 1. Pedigree A (120 related subjects (black boxes) of total 128 analyzed) comprising of main branch A1 (113 subjects) and A2 (7 subjects, on the left). Grey boxes represent males excluded from the paternal line. For the better resolution, the entire pedigree is divided in two rows (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

According to the archives, the year of birth of the most recent common ancestor (MRCA) of the pedigree A1 was 1720, while for the MRCA of the pedigree A2, it was 1808 (with a known year of birth of his predecessor being the year 1785). We had the documented date of birth for all the members of the pedigree and we used them to calculate the age of the father at the time of his son's birth. The branch A1 is covered by 461 meiosis (for sub-branches 177, 226, and 58, respectively), and branch A2 by additional 24, comprising a total of 485 meiotic events analyzed in this study (due to the unknown exact relationship of A1 and A2, the meiotic events between them were not used for the calculation). The samples excluded from the paternal line were associated by additional 18 meiosis to the genetically confirmed family tree, and they were not considered for the mutation rate estimation, but for the extra-paternal pair (EPP) rate estimation only. The entire family tree figure with haplotype differences compared to ancestral haplotypes is provided in the supplemental material (Suppl. Fig. 1).

All the participants in this study gave their written informed consent and the approval of the study was granted by the Ethical Committee of the Institute for Medical Research, University of Belgrade. The samples were processed in a linked anonymous form, and the confidentiality of personal information of each study participant was assured.

2.2. Genotyping

DNA was extracted from buccal swabs using QIAamp DNA Mini Kit (*Qiagen*, Hilden, Germany) and quantified using Qubit® dsDNA BR Assay (*Thermo Fisher Scientific*, Foster City, USA). The PCR amplification of 23 Y-STR loci included in the PowerPlex® Y23 System (*Promega Corporation*, Madison, USA) and 27 Y-STR loci included in the Yfiler® Plus kit (*Thermo Fisher Scientific*, Foster City, USA) was carried out using the Applied Biosystems GeneAmp PCR System 9700 thermal cycler, although using half volume (12.5 μl) reactions. The RM panel, including 13 RM Y-STR markers, was amplified as described previously [32]. The amplified fragments were separated on the ABI Prism® 3500 Genetic Analyzer (*Applied Biosystems*, Foster City, CA, USA) along with the allelic ladder and the appropriate internal size standard. All the procedures and protocols were conducted following the manufacturers' instructions. The genotypes were analyzed using the GeneMapperID-X v1.4 software (*Thermo Fisher Scientific*, Foster City, CA, USA). The *in silico* haplogroup assignment was made using the *NevGen* (www.nevgen.org) predictor, which was based on the previously-implemented Athey's Haplogroup Predictor (http://hprg.com/hapest5). A few representative samples were assigned to a specific SnapShotY-SNP assay as previously described [33] to confirm the predicted R1b haplogroup. Further confirmation was obtained by a deep SNP analysis of a commercially available test (R1b: L23 - L51 - U106 - Z381 - Z301 - L48 - Z9 - Z331 - Z330 - Z326 - Z8168 - FGC18842).

2.3. Sequencing analysis

Determination of the structure and the exact number of repeats of allele 20.2 at the *DYS448* locus, the "landmark" of these haplotypes, was accomplished through sequencing analysis, using primers, obtained from the Ensembl genome browser (www.ensembl.org), that flank repeats. PCR fragments were purified and directly sequenced using BigDye Terminator Kit 1.1 (*Thermo Fisher Scientific*, Foster City, USA), and analyzed in the ABI 3130 Genetic analyzer (*Thermo Fisher Scientific*, Foster City, USA). The results were analyzed using the Sequencing Analysis v5.2 (*Thermo Fisher Scientific*, Foster City, USA) and BioEdit v.7.0.9. programs (*Tom Hall, Ibis Biosciences*,CA, USA).

2.4. Quality control

Three independent laboratories conducted the genotyping using different kits with overlapping loci. The quality control was ensured by the successful participation of all three laboratories in GEDNAP proficiency tests and by using internal control standards and kit controls.

2.5. Statistical analysis

The mutation rates were estimated using the Bayesian and the frequentist approaches. The frequentist mutation rates per locus were calculated as the number of mutations observed at a specific locus divided by the number of meiosis. The mutation rate estimated across all 36 Y-STR markers was calculated as a total number of mutations divided by a total number of allele transmissions. The multi-locus markers were considered jointly, because they all provided a possibility for mutations occurrence, which cannot be separately analyzed with the used genotyping assay. The confidence intervals (ci) of the mutation rates were estimated from a frequentist approach assuming a binomial distribution at http://statpages.info/confint.html. The mutation rates for each locus were also estimated via a binominal hierarchical Bayesian model with a uniform prior, which led to a posterior Beta distribution with parameters $\alpha = m + 1$ and $\beta = n+1$, where m is the number of mutational events and n is the number of non-mutational events [34]. The estimated mutation rates were further compared with the mutation rates obtained from the literature and the Y-chromosome haplotype reference database (YHRD) [3; see Web Resource], whereby the statistical significance was tested using the Fisher's exact test. In addition, the mutation rates for individual markers were also estimated using a binomial hierarchical Bayesian model with the Markov chain Monte Carlo (MCMC) Gibbs sampling. Mutation rate hyperparameter θ for individual markers was assumed to be a sample from a normal distribution with hyperparameters μ and τ, $\tau = 1/\sigma$. Hyperparameter μ was estimated as a sample from a non-informative prior normal distribution ($\mu = 0$, σ = 1E-6), and hyperparameter τ was estimated using prior gamma distribution (α = 1E-5, β = 1E-5).

The male relative differentiation rates were calculated as the number of pairs of relatives differentiated by at least one Y-STR locus divided by the total number of pairs of relatives grouped by a degree of relationship (i.e. pair members separated by 1–20 meiosis). One way ANOVA and Pearson's correlation tests were applied using the Prism 6 software (GraphPad Software Inc., San Diego, CA, USA), and $p < 0.05$ differences were accepted as statistically significant. The results were expressed as the average ± SD, median and mode. To allow a maximum resolution, multi-copy Y-STRs (*i.e. DYS385a/b, DYF399S1, DYF387S1, DYF404S1,* and *DYF403S1a*) were considered in calculations as separate loci.

To evaluate the influence of the father's age at the time of the son's birth on the Y-STR mutation rate, we performed Pearson's correlation tests using the Prism 6 software, with $p < 0.05$ accepted as statistically significant. To perform this correlation, we plotted the total sum of fathers' ages at the time of transmission against the sum of mutations for each A1 sub-branch, and A2 branch of the pedigree.

The estimation of extra-pair paternity rate per generation was calculated by dividing the number of observed events with the total number of meiosis (503) which also comprised the transmission events that connected the EPP samples to the pedigree (18 additional meiosis). A general assumption was made that only one EPP event occurred in the family line above each observed case.

Since the exact relationship between A1 and A2 was unknown, we quantified their relatedness by the time to the most recent common ancestor estimation. For that purpose, we used *NevGen* tMRCA calculator (www.nevgen.org). It calculates probability distribution of being at an exact number of generation back to the most common ancestor of a certain pair of people and the cumulative probability that actual number of generation is less than a certain value, based on the stepwise model [35], which also considers multi-step and backward mutations and different mutation probabilities based on Marko Heinila's mutation rates [36]. For the estimation we used markers included in PPY23 kit, where A1 and A2 ancestral haplotypes differed in 1 out of the 23 markers.

Table 1
Mutations at 36 Y-STR markers observed in a total of 485 meiotic events in wide deep-rooted pedigree comprised of 120 males of pedigree A (**-observed statistical significant difference compered to YHRD and [3], *f* - frequentist approach, *B* - Bayesian approach, ci- confidence interval, CI- credible interval).

Marker	Panel			No of mutations	Mutation		Mutation rate											
							f ($x10^{-3}$)	95% ci ($x10^{-3}$)	*B* ($x10^{-3}$)	95% CI ($x10^{-3}$)	Per panel/ per markers							
											RM13 ($x10^{-3}$)		Yfiler+ ($x10^{-3}$)		PPY23 ($x10^{-3}$)		Across all 36 markers ($x10^{-3}$)	
					gains	losses					*f*	*B*	*f*	*B*	*f*	*B*	*f*	*B*
DYS526a				3	1	2	**6.18**	1.3-18.0	**7.55**	2.25-17.9								
DYS612				10	5	5	**20.62**	9.9-37.6	**21.93**	11.35-37.51								
DYF399S1				63	25	38	**129.90**	101.3-163.1	**130.91****	102.91-162.78								
DYS526b				5	3	2	**10.31**	3.4-23.9	**11.66**	4.54-23.84								
DYS547				5	2	3	**10.31**	3.4-23.9	**11.66**	4.54-23.84								
DYF404S1				6	4	2	**12.37**	4.6-26.7	**13.71**	5.81-26.67								
DYS626				1	0	1	**2.06**	0.10-11.4	**3.45**	0.50-11.41								
DYF403S1a				9	7	2	**18.56**	8.5-34.9	**19.88**	9.91-34.86								
DYF403S1b				2	1	1	**4.12**	0.5-14.8	**5.50**	1.27-14.79								
DYS627				12	1	11	**24.74**	12.9-42.8	**26.05**	14.32-42.73								
DYS518				10	3	7	**20.62**	9.9-37.6	**21.93**	11.35-37.51								
DYS449				5	3	2	**10.31**	3.4-23.9	**11.66**	4.54-23.84								
DYF387S1	*RM13*			3	3	0	**6.18**	1.3-18.0	**7.55**	2.25-17.9								
DYS460				2	0	2	**4.12**	0.5-14.8	**5.50**	1.27-14.79	**21.25** (95%CI:17.8-25.1)	**21.35** (95%CI:17.98-25.11)						
DYS570				7	3	4	**14.43**	5.8-29.5	**15.77**	7.13-29.45								
DYS576				4	3	1	**8.25**	2.3-21.0	**9.60**	3.35-20.94								
DYS389I				2	1	1	**4.12**	0.5-14.8	**5.50**	1.27-14.79								
DYS635				0	0	0	**0**	0-7.6	**1.42**	0.05-7.56								
DYS389II				1	0	1	**2.06**	0.1-11.4	**3.45**	0.50-11.41								
DYS458				3	0	3	**6.18**	1.3-18.0	**7.55**	2.25-17.9								
DYS19				0	0	0	**0**	0.0-7.6	**1.42**	0.05-7.56								
YGATAH4				0	0	0	**0**	0.0-7.6	**1.42**	0.05-7.56								
DYS448				0	0	0	**0**	0.0-7.6	**1.42**	0.05-7.56								
DYS391				3	1	2	**6.18**	1.3-18.0	**7.55**	2.25-17.9								
DYS456				3	0	3	**6.18**	1.3-18.0	**7.55**	2.25-17.9								
DYS390				0	0	0	**0**	0.0-7.6	**1.42**	0.05-7.56								
DYS438				0	0	0	**0**	0.0-7.6	**1.42**	0.05-7.56								
DYS392				0	0	0	**0**	0.0-7.6	**1.42**	0.05-7.56								
DYS437				1	0	1	**2.06**	0.1-11.4	**3.45**	0.50-11.41								
DYS385				0	0	0	**0**	0.0-7.6	**1.42**	0.05-7.56								
DYS393				0	0	0	**0**	0.0-7.6	**1.42**	0.05-7.56								
DYS439				9	2	7	**18.56**	8.5-34.9	**19.88****	9.91-34.86								
DYS481		*Yfiler +*		1	0	1	**2.06**	0.1-11.4	**3.45**	0.50-11.41			**5.44** (95%CI: 4.2-6.9)	**5.5** (95%CI: 4.28-6.92)				
DYS533				0	0	0	**0**	0.0-7.6	**1.42**	0.05-7.56								
DYS549			*PPY23*	5	1	4	**10.31**	3.4-23.9	**11.66**	4.54-23.84					**3.65** (95%CI: 3.6-5.0)	**3.72** (95%CI: 2.68-4.99)		
DYS643				0	0	0	**0**	0.0-7.6	**1.42**	0.05-7.56							**10.02** (95%CI: 8.6-11.6)	**10.06** (95%: 8.65-11.61)
Σ					**69**	**106**												

3. Results and discussion

In the present study, 120 patrilineally related males distributed over two branches (A1 and A2) of one wide-ranged deep-rooted pedigree were genotyped for 36 Y-STRs, consisting of RM13, Yfiler + and PPY23 loci (Fig. 1, Suppl. Fig. 1). Based on the obtained data and genealogical information, along with Y-STR genotyping, we were able to reconstruct the ancestral haplotypes of the branches of A1 and A2 descendants. Exact relationship of A1 and A2 was unknown, but was indicated by genotyping, with the characteristic rare allele 20.2 at *DYS448* locus, besides migration data and patron saint links. The ancestral A1 and A2 haplotypes differed from each other in 6 markers (*DYS526a, DYS526b, DYS399S1, DYF404S1, DYS627* and *DYS456*) (Suppl. Fig. 1). The obtained haplotypes have been predicted and further confirmed with SNPs to belong to the R1b haplogroup (R-M343). A putative ancestral Y-STR haplotype (in the order *DYS526a-DYS612-DYS399S1-DYS526b*-DYS547-DYS404S1-DYS62-DYF403S1a-DYF403S1b-DYS627-DYS518-DYS449-DYS387S1-DYS460-DYS570-DYS576-DYS389I-DYS635-DYS389II-DYS458-DYS19-YGATAH4-DYS448-DYS391-DYS456-DYS390-DYS438-DYS392-DYS437-DYS385a/b-*DYS393-DYS439-DYS481-DYS481-DYS533-DYS549-DYS643*) was constructed for each branch (the main A1 branch: 15-37-22,26,27.1-38-47-14,16-29-11,13,15-46-22-39-29-35-11-18-17-13-23-29-17-13-23-29-17-14-11-20.2-11-16-23-12-13-15-11,14-13-13-22-12-13-10, and the A2 branch with different alleles: 14..-15,16..-19..-15 at *DYS526a, DYS404S1, DYS627, DYS456* loci). The ancestral haplotypes were inferred after aligning all the obtained haplotypes and determining the most frequent allele at each locus as the allele in the ancestral haplotype, taking into account the minimal number of mutational events. The observed and inferred father-son transmissions were used to estimate the pedigree-specific Y-STR mutation rates. Although several multi-step mutations were observed, all the mutations were considered to be multiple one-step events. When we were considering the mutations at the *DYS389* loci, *DYS389I* was subtracted from *DYS389II*. With a large number of male relatives from one wide deep-rooted pedigree, we were able to precisely resolve majority of mutational events.

A recent study by Claerhout et al. [31], who performed the Y-STR analysis by capillary electrophoresis on multiple deep-routing family pedigrees comprising a minimal set of DNA samples, showed that certain Y-STR mutations could remain hidden, which could lead to the underestimation of the tMRCA. In the present study, we performed mutation rate estimates of Y-STRs by CE and a wide-deep rooted pedigree-based design. In our dataset, we were able to observe numerous parallel mutations, and to resolve their origin more precisely, since we have studied the transmission of one single Y-chromosome through generations (Suppl. Fig. 1). Eventually, it may be interesting to sequence all these parallel modifications, in order to examine if the same mechanism is responsible for their occurrence, as well as to sequence the mutations that were designated as identical by descent for several members in certain lines of the pedigree.

Two commercially available kits used in this study, the PPY23 and Yfiler+, as well as the RM13 panel [15], included overlapping loci, but they were all used to expand the number of analyzed markers and for traceability purposes. Six new Y-STR markers were included in the PPY23 (*DYS570, DYS576, DYS549, DYS481, DYS533*, and *DYS643*) besides the existing 17 markers of the original Yfiler kit (*Thermo Fisher Scientific*, CA, USA). Two of the new markers were categorized as rapidly mutating [3] and were present also in succeeding Thermo Fisher Yfiler+ (*DYS576* and *DYS570)*. The RM13 panel contained 4 Y-STR markers that were a part of the Yfiler+ (*DYS627, DYS518, DYS449 and DYF387S1*), while *DYS460* was unique to this kit, as *DYS549* and *DYS643* were unique for PPY23. The fact that some markers resided in two or all three kits also enabled a concordance testing. In our study, we found no discordance, which was in contrast to the results of Westen et al. [37], who genotyped 2085 Dutch male donors with 36 Y-STR marker units (the same that were used in our study but excluding *DYS460*) and found discordance in two samples. However, our results were in line with the studies of Davis et al. [38] and Larmuseau et al. [39] that showed no discordance in the 17 overlapping loci between Yfiler and PPY23 in a set of 951 American and 535 Belgian donors.

3.1. Mutation rates' estimation

The results of genotyping using 36 Y-STR markers across 20,855 allele transmissions revealed the presence of 175 mutations in 485 meiotic events at 25 markers (*DYS526a, DYS612, DYF399S1, DYS526b, DYS547, DYF404S1, DYS626, DYF403S1a, DYF403S1b, DYS627, DYS518, DYS449, DYF387S1, DYS460, DYS570, DYS576, DYS389I, DYS389II, DYS458, DYS391, DYS456, DYS437, DYS439, DYS481, DYS549*) (Table 1). In 11 Y-STR markers (*DYS643*, *DYS533*, *DYS393*, *DYS385a/b*, *DYS392*, *DYS438*, *DYS390*, *DYS448*, *YGATAH4*, *DYS19*, *DYS635*), included in the PPY23 and Yfiler+ (except *DYS643*), no mutations were detected across all samples. Medians from the Bayesian estimation of locus-specific mutation rates ranged from 3.45×10^{-3} (95% CI: 0.50×10^{-3}-11.41×10^{-3}) for 11 loci where no mutation was observed, to 130.91×10^{-3} (95% CI: 102.91×10^{-3}-162.78×10^{-3}) for the *DYF399S1* marker. The locus specific mutation rates and mutation rates per each panel, calculated by the frequentist and the Bayesian approach, are listed in Table 1.

With regard to each panel separately, 134, 66 and 39 mutation events were observed with mutation rates estimated using the Bayesian approach 21.35×10^{-3} (95% CI, 17.98×10^{-3}- 25.11×10^{-3}), 5.5×10^{-3} (95% CI, 4.28×10^{-3}- 6.92×10^{-3}) and 3.72×10^{-3} (95% CI, 2.68×10^{-3} – 4.99×10^{-3}) for the RM, Yfiler + and PPY23, respectively (Table 1). Unlike the PPY23 and Yfiler+, we detected mutations in all loci of the RM panel. As expected due to its multi-copy nature, the highest mutation rate was 130.91×10^{-3} (95% CI, 102.91×10^{-3} – 162.78×10^{-3}) for *DYF399S1* locus, which is in line with the previous studies [4,27]. Across all 36 markers combining the PPY23, Yfiler+, and RM13, the mutation rate was 10.02×10^{-3} (95% CI: 8.6×10^{-3} - 11.6×10^{-3}) using the frequentist approach and 10.06×10^{-3} (95% CI: 8.65×10^{-3} - 11.61×10^{-3}) using the Bayesian approach in the pedigree (Table 1). The average mutation rate for the 36 analyzed Y-STRs between DNA donors and MRCA was 9.995×10^{-3} (95% CI: 8.8–11.2×10^{-3}) obtained by Bayesian model with MCMC.

Of the 175 detected mutations, 106 (60.57%) were repeat losses, while 69 (39.43%) were repeat gains. The observed 175 single mutations included 158 single-step mutations, with repeat losses clearly being favored over repeat gains (91 one repeat losses, 67 one repeat gains), 6 two repeats losses, 1 two repeats gain and 1 three repeats loss. Among the mentioned single-step mutations, we observed step-by-step mutations in 6 cases, which we could elucidate due to a dense, wide-range pedigree. Otherwise, these mutations would be considered as multi-step ones. Furthermore, we were able to detect reverse mutations in 4 cases. In one sample, we observed a gene conversion at the multi-copy *DYF399S1* locus, revealed by one of the three copies missing and one allele (22) present with double intensity (22, 27.1 instead of 22, 26, 27.1). At the same time, the closest male relatives of the previously mentioned sample, 7 meiosis apart, had all three ancestral alleles.

No mutations were found at the *DYS533* locus designated as fast mutating [3] in a total of 485 meiosis. In our previous study performed on father-son pairs, we detected a mutation at the *DYS533* locus [40]. With the exception of *DYS533* and *DYS643,* we detected mutations in the remaining four new Y-STRs of the PPY23, in contrast to the results of Turrina et al. [28] who observed no mutations in the PPY23 newly added markers, with exception of *DYS576*.

The same mutation could be considered multiple times in different pairs of relatives within a limited pedigree (except father-son pairs), and thus may lead to an overestimation of the mutation rates. Generally, this could be resolved by including all the pairs of relatives from the examined branches of pedigree [29], available in the presented study. The mutation rates could be one order of magnitude lower than the father-son based rates in deep-rooted pedigrees, as a consequence of non-inclusion of multi-step and particularly back mutations [29], and may be considered as a lower limit of the actual "lineage" mutation rates. The locus-specific mutation rates determined in our study showed no significant differences (p-value > 0.05, Fisher's exact test) from the ones obtained in the study of Ballantyne et al. [3] and YHRD, except for *DYS439* (p = 0.002) and *DYF399S1* (p = 0.00001) loci that showed higher mutation rates. In line with the previous, the examined binomial confidence and credible intervals also suggested that there was no substantial difference, except for the mentioned loci. Hence, our results on the Y-STR mutation rates obtained by analyzing the transmission of one single Y chromosome in a single wide deep-rooted and relatively dense pedigree are representative of mutation rates obtained through large numbers of direct father-son Y-chromosome transmissions, as suggested by Balanovsky [18].

The hallmark of all haplotypes is allele 20.2 at *DYS448* locus, indicating that no mutation at this locus occurred in all 485 meiosis. Generally, more repeats in a non-interrupted stretch lead to more DNA slippage during replication and those loci are more prone to mutations [1,3]. At the *DYS448* locus, two stretches of six-nucleotide repeats were separated by the stretch of 42 nucleotides $[AGAGAT]_nN_{42}[AGAGAT]_m$. The sequencing showed that both stretches had 9 repeats interrupted by 56 nucleotides $[AGAGAT]_9N_{56}[AGAGAT]_9$. Our result could be expected, especially among relatives, since this locus is slow mutating [3], although in some previous studies mutations at *DYS448* locus have been detected [28]. Besides allele 20.2 at *DYS448* locus, the persistence of the ancestral Y-STR haplotype in 12 present-day descendants of A1 in distant parts of the pedigree, separated by 9–18 meiosis, was an additional confirmation of the relationship between the analyzed males.

3.2. Male relatives differentiation

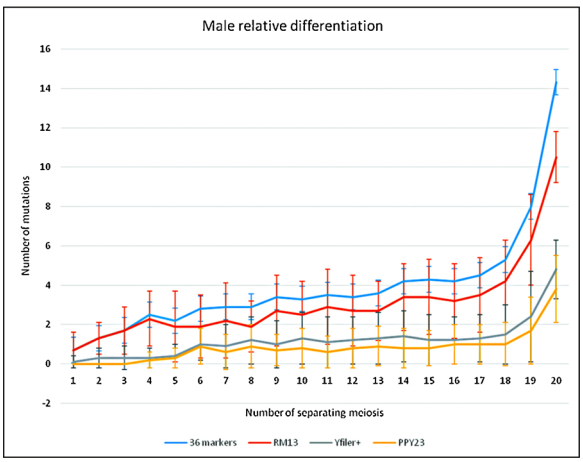
In the current study, we have genotyped 6349 pairs of relatives separated by 1–20 meiotic events. The average number of mutations in male relative pairs separated by 1–20 meiosis ranged from 0.7–14.3, 0.7–10.5, 0.1–4.8, and 0–3.8 when considering 36 markers, RM13 Y-STRs, Yfiler+, and PPY23, respectively (Fig. 2, Table 2). The overall discrimination capacity of 36 markers, combining PPY23, Yfiler+, and RM13 loci, was 98.4% (Fig. 2, Table 2). Only 1.6% of relative pairs displayed no difference. The 36 Y-STR set distinguished 44.4% of father-son pairs, and 85.7% of relatives separated by 2 meiosis (brothers and grandfather/grandson pairs). The male relatives, with more than 7 meiosis apart, demonstrated difference in at least one marker in a minimum of 94.8% cases. The percentage of differentiated relatives with more than 17 meiosis apart was higher than 99% (Fig. 2; Table 2). The complete results with differences in 1, 2, 3, 4, 5 and more than 5 loci, are shown in the Supplemental Tables 2A-D.

The differentiation rate of RM13 Y-STRs was 95.9%, with 44.4% father-son pairs separated. Our results are in line with some previous studies that showed a great value of RM13 markers in differentiating closely related male relatives, even in endogamous populations [30,31]. However, in the above mentioned studies, 26.5% and 20.40% of father-son pairs were separated by RM Y-STRs. The higher rate of father-son differentiation in our study could be attributed to the *DYF399S1* locus with a significantly higher mutation rate than in previous studies [3], but also to stochastic effects due to a small sample size.

The discrimination capacities of Yfiler + and PPY23 were 65.5% and 57.4%, respectively. The PPY23 set, with the lowest male differentiation capacity, was not able to differentiate any of the father-son pairs, nor relatives separated by 2 meiosis (brothers and grandfather/grandson pairs), due to a lesser number of pairs in this category. Considering the relative pairs separated by more than 7 meiosis, 40.0% of cousins could be differentiated with PPY23, with 57.1% differentiation for relatives that are more than 17 meiosis apart (Fig. 2; Table 2).

The biological kinship among namesakes could be confirmed through a Y-haplotype comparison based on a correspondence of the Y-SNP sub-haplogroup and the presence of no more than seven mutated Y-STR loci [19]. The latter is based on the formula of Walsh, indicating that it is highly implausible that more than seven mutations on 46 Y-

A)



B)

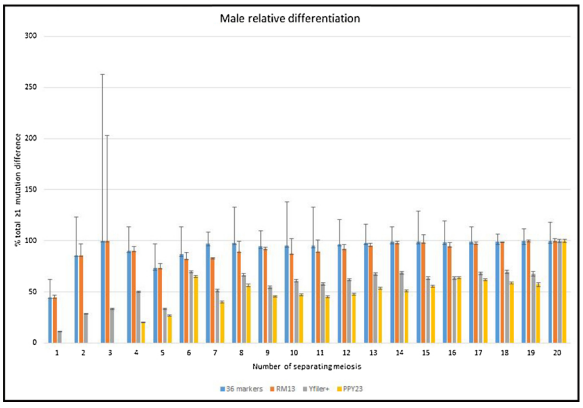


Fig. 2. Male relative differentiation through generations. **A)** The number of mutations (± SD) between relative pairs separated by 1–20 meiosis observed using 36 markers, RM13, Yfiler+, and PPY23 **B)** Male relative differentiation observed – comparison of all sets; Error bars represent 95% binomial credible intervals.

STRs would occur among namesakes separated by less than 30 generations [35]. The number of mutations in our study that included RM13 markers ranged from 0 to 16 in related males separated by, for example, 18 meiosis, with the average number of 5.3 when considering the entire 36-marker set (Table 2). At the same time, the number of detected mutations when using RM13, Yfiler + and PPY23 panels ranged from 0 to 12, 0–8, and 0–5 respectively, with the average of 4.2, 1.5, and 1 mutation. Male relatives separated by 20 meiosis were fully differentiated regardless of the marker set.

3.3. Estimation of the EPP rate

Out of 128 participants, 121 males were considered as A1 pedigree descendants, with 8 that were excluded from further analysis due to a disruption in their paternal genealogy based on their Y-haplotype. Among the excluded males, 2 pairs shared the same PPY23 haplotype with their closest relative based on the number of meiosis, indicating that the change occurred earlier in the family tree. The remaining 4 males had completely different Y-STR haplotypes compared to the nearest relative. An interruption in the paternal genealogy could be caused by a hidden adoption, or by an extra-pair paternity event. Taking into account the total number of meiosis, we estimated the EPP rate in our wide-deep rooted pedigree at 11.9×10^{-3} (95% CI: 4.4×10^{-3} - 25.8×10^{-3}). Our results are in line with the low rates estimated for Western European populations [41–43] and are in

Table 2
Male relative differentiation in pedigree A using 36 marker set, RM13, Yfiler+, and PPY23 (Ped – pedigree; GD - Generation distance; Avg – Average difference. SD – standard deviation; Med – Median; R – range; Mo – Mode in Supplemental Tables 2A-D).

Ped A		36 set					RM13				Yfiler +				PPY23				
GD	Pairs of relatives	0 diff (%)	% tot. ≥1 diff	R	Avg	Mo	0 diff (%)	% tot. ≥1 diff	R	Avg	Mo	0 diff (%)	% tot. ≥1 diff	R	Avg	Mo	0 diff (%)	% tot. ≥1 diff	R	Avg	Mo
1	9	5 (556)	44.4	0-2	0.7	0	5 (556)	44.4	0-2	0.7	0	8 (889)	11.1	0-1	0.1	0	9 (100)	0	0	0	0
2	7	1 (143)	85.7	0-2	1.3	1	1 (143)	85.7	0-2	1.3	1	5 (100)	28.6	0-1	0.3	0	9 (100)	0	0	0	0
3	3	0	100	1-3	1.7	1	0	100	1-3	1.7	1	2 (667)	33.3	0-1	0.3	0	9 (100)	0	0	0	0
4	10	1 (10)	90	0-4	2.5	3	1 (10)	90	0-4	2.3	2	5 (500)	50.0	0-1	0.3	0	8 (80)	20	0-1	0.2	0
5	15	4 (267)	73.3	0-5	2.2	0	4 (267)	73.3	0-5	1.9	0	10 (667)	33.3	0-2	0.4	0	11 (733)	26.7	0-1	0.3	0
6	23	3 (13)	87.0	0-6	2.8	2	4 (174)	82.6	0-5	1.9	1	7 (304)	69.6	0-2	1	1	8 (348)	65.2	0-3	0.9	1
7	35	1 (2,9)	97.	0-9	2.9	1	6 (171)	82.9	0-8	2.2	1	17 (486)	51.4	0-5	0.9	0	21 (60)	40	0-3	0.6	0
8	48	1 (2,1)	97.9	0-7	2.9	1	5 (104)	89.6	0-5	1.9	1	16 (333)	66.7	0-5	1.2	0	21 (437)	56.3	0-4	0.9	0
9	77	4 (5,2)	94.8	0-11	3.4	2	6 (7,8)	92.2	0-9	2.7	2	35 (455)	54.5	0-4	1.0	0	42 (545)	45.5	0-3	0.7	0
10	112	5 (4,5)	95.5	0-10	3.3	2	14 (125)	87.5	0-7	2.5	2	44 (393)	60.7	0-5	1.3	0	59 (527)	47.3	0-5	0.8	0
11	188	9 (4,8)	95.,2	0-13	3.5	2	20 (106)	89.4	0-9	2.9	2	79 (420)	58.0	0-8	1.1	0	103 (548)	45.2	0-4	0.6	0
12	348	11 (3,2)	96.8	0-12	3.4	3	28 (8,0)	92.0	0-9	2.7	2	133 (382)	61.8	0-6	1.2	0	181 (52)	48.0	0-4	0.8	0
13	492	10 (2)	98.0	0-12	3.6	4	22 (4,5)	95.5	0-9	2.7	2	160 (325)	67.5	0-6	1.3	0	229 (465)	53.5	0-5	0.9	0
14	498	5 (1)	99.0	0-13	4.2	3	10 (2,0)	98.0	0-8	3.4	4	155 (311)	68.9	0-7	1.4	0	243 (488)	51.2	0-6	0.8	0
15	746	5 (0,7)	99.3	0-14	4.3	4	9 (1,2)	98.8	0-11	3.4	3	273 (366)	63.4	0-7	1.2	0	332 (445)	55.5	0-4	0.8	0
16	1391	18 (1,3)	98.7	0-16	4.2	4	74 (5,3)	94.7	0-10	3.2	3	506 (364)	63.6	0-7	1.2	1	506 (364)	63.6	0-6	1	1
17	1515	14 (0,9)	99.1	0-13	4.5	5	41 (2,7)	97.3	0-10	3.5	3	483 (319)	68.1	0-7	1.3	1	574 (379)	62.1	0-5	1	1
18	716	5 (0,7)	99.3	0-16	5.3	4	8 (1,1)	98.9	0-12	4.2	4	219 (306)	69.4	0-8	1.5	0	295 (412)	58.8	0-6	1	0
19	112	0	100	1-15	8.0	10	0	100	1-11	6.3	8	36 (321)	67.9	0-7	2.4	0	48 (429)	57.1	0-5	1.7	0
20	4	0	100	12-18	14.3	14	0	100	9-12	10.5	n/a	0	100	4-7	4.8	3	0	100	3-6	3.8	3
Tot	6349	102 (1,6)	98,4				258 (4,1)	95,9				2193 (345)	65,5				2708 (426)	57,4			

contrast to the previously published EPP rates in a range up to 30% [44]. Although there are no official data about EPP rates for the Serbian population, nor results of a systematic EPP study in Serbia, due to the robustness of the analyzed pedigree, our data could be representative for this region.

3.4. Influence of the age of the father at the time of the transmission on the Y-STR mutation rates

Unlike most of the previous studies dealing with deep-rooted pedigrees, but with a relatively small number of contemporaries, we were able to examine the influence of the father's age at the time of son's birth in a more homogenous genetic background. There were three subbranches of the A1 and A2 part in our pedigree with 42, 58, 13 and 7 analyzed contemporaries, respectively. The average age of fathers at their sons' birth in these four lines were 28.06, 29.12, 29.28, and 31.58, with a total sum of 4967, 6582, 1698, and 758 years, spanning 177, 226, 58, and 24 meiotic transmissions, respectively. The average generational age in the total pedigree was 28.88 years, which is in line with the ages (20–30) most frequently used in the field of evolutionary genealogical studies [45–47]. The number of the observed mutations in the aforementioned parts of the pedigree were 52, 94, 19, 10 (with 47, 74, 16, 8 mutations in the Y-STR markers designated as rapidly mutating). When plotted against the number of mutations, a significant positive correlation was observed ($R^2 = 0.9495$, $p = 0.0256$) for the total sum of the fathers' ages at the time of the transmission. We obtained even more significant results when we plotted the sum of the fathers' ages against the mutations in the RM markers ($R^2 = 0.9827$, $p = 0.0087$), which indicated a higher influence of the father's age at the time of the son's birth on markers already prone to mutation. When we considered each marker independently, a positive correlation was detected for *DYF399S1* ($R^2 = 0.9569$, $p = 0.0218$), *DYS627* ($R^2 = 0.9801$, $p = 0.01$), *DYS576* ($R^2 = 0.9222$, $p = 0.0397$), *DYS391* ($R^2 = 0.9436$, $p = 0.0286$), *DYS456* ($R^2 = 0.9436$, $p = 0.0286$) and *DYS439* ($R^2 = 0.9436$, $p = 0.0286$)

3.5. Time to the most recent common ancestor

Reconstructed A1 and A2 ancestral haplotypes were aligned in the reduced PPY23 marker format in *NevGen* tMRCA calculator in which they differed at *DYS456* locus. The obtained results gave the distribution of probabilities with mode at 7 generations and median value of 12 generations (reaching 95% of cumulative probability at 37 generation). According to the documented genealogy data of the pedigree branch A1, the MRCA was born in the year 1720, and he got his first son when he was 30 years old, or in the year 1750, while the MRCA of the A2 branch was born in the year 1808, thus approximately 3 generations later (the average generation time in the entire pedigree was estimated at 28.88 years). Given the obtained data for tMRCA estimation probability distribution and above mentioned genealogy data, the MRCA of the entire pedigree could have lived not earlier than 2–4 generations before A1 ancestor (*i.e.* the A1 and A2 MRCA could have been separated by at least 7–12 generations). That estimation fits well in the male relative differentiation capacity distribution for all markers that were analyzed in our study (Suppl. Table 2A).

The previous studies used different approaches to demonstrate that families can be traced back to a common ancestor, based on the Y-STR evidence, even if these individuals lived a couple of hundred years earlier [19,22,31]. As shown above, the results of our tMRCA estimation in a wide deep-rooted pedigree with known and judicially confirmed relationship, using the previously described model, correspond to the results of male relatives differentiation obtained by analysis of whole pedigree, and confidently revealed that the families of one pedigree, even carrying different surnames and inhabiting the different region, could be traced back to the MRCA.

4. Conclusion

The mutation rates obtained by analyzing a large number of meiosis in one wide deep-rooted pedigree correspond to the mutation rates obtained through large numbers of direct father-son Y-chromosome transmissions. The discrimination capacity of close relatives increased when moving from the PPY23, Yfiler+ to the RM Y-STRs, as expected. Given their higher mutation rates, the RM Y-STRs allow forensic DNA analysis to individualize the relatives. However, we showed that the related males could differ in up to 18 mutations when considering the set of 36 Y-STR markers, including a RM13 set, which may be important when analyzing random forensic samples, searching through families or conducting a genealogical study. We found positive correlation of the fathers' ages at the time of the transmission and mutation rates, which is more prominent for markers (RM) already prone to mutation. Although our pedigree was very dense, eventual further detailed sequencing analysis of all samples by a massively parallel sequencing could possibly even increase the information about allelic variations in our dataset.

Conflict of interest

The authors declare that they have no conflict of interest.

Funding

none.

Author contributions

Idea and supervision: DK.
Sampling and genealogies: VPČ, SK.
Genotyping: DK, MK, DZ, ZJ.
Analysis: DK, MK, MKM, VPČ, AV.
Writing: MK, VPČ, DK.
Web Resources
The URL for data presented herein is as follows:
http://statpages.info/confint.html
NevGen: www.nevgen.org
Athey's Haplogroup Predictor: http://hprg.com/hapest5
YHRD: Y-STR Haplotype Reference Database, www.yhrd.org/

Acknowledgments

We are grateful to the contributing members of the deep-rooting pedigrees for their cooperation. We thank all DNA donors for their participation in this genetic genealogy project. We acknowledge Saša Rakić, Zoran Velikić i Miloš Ivanović for their technical assistance.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fsigen.2019.04.007.

References

[1] M. Kayser, Forensic use of Y-chromosome DNA: a general overview, Hum. Genet. 136 (2017) 621–635, https://doi.org/10.1007/s00439-017-1776-9.
[2] G.C. Calacal, F.C. Delfin, M.M. Tan, L. Roewer, D.L. Magtanong, M.C. Lara, Rd. Fortun, M.C. De Ungria, Identification of exhumed remains of fire tragedy victims using conventional methods and autosomal/Y-chromosomal short tandem repeat DNA profiling, Am. J. Forensic Med. Pathol. 26 (2005) 285–291, https://doi.org/10.1097/01.paf.0000177338.21951.82.
[3] K.N. Ballantyne, M. Goedbloed, R. Fang, O. Schaap, O. Lao, A. Wollstein, Y. Choi, K. van Duijn, M. Vermeulen, S. Brauer, R. Decorte, M. Poetsch, N. von WurmbSchwark, P. de Knijff, D. Labuda, H. Vézina, H. Knoblauch, R. Lessig, L. Roewer, R. Ploski, T. Dobosz, L. Henke, J. Henke, M.R. Furtado, M. Kayser, Mutability of Ychromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications, Am. J. Hum. Genet. 87 (2010) 341–353, https://doi.org/10.1016/j.ajhg.2010.08.006.

[4] Y.N. Oh, H.Y. Lee, E.Y. Lee, E.H. Kim, W.I. Yang, K.J. Shin, Haplotype and mutation analysis for newly suggested Y-STRs in Korean father-son pairs, Forensic Sci. Int. Genet. 15 (2015) 64–68, https://doi.org/10.1016/j.fsigen.2014.09.023.

[5] L.A. Zhivotovsky, P.A. Underhill, C. Cinnioğlu, M. Kayser, B. Morar, T. Kivisild, R. Scozzari, F. Cruciani, G. Destro-Bisol, G. Spedini, G.K. Chambers, R.J. Herrera, K.K. Yong, D. Gresham, I. Tournev, M.W. Feldman, L. Kalaydjieva, The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time, Am. J. Hum. Genet. 74 (2004) 50–61, https://doi.org/10.1086/380911.

[6] P. Francalacci, L. Morelli, A. Angius, R. Berutti, F. Reinier, R. Atzeni, R. Pilu, F. Busonero, A. Maschio, I. Zara, D. Sanna, A. Useli, M.F. Urru, M. Marcelli, R. Cusano, M. Oppo, M. Zoledziewska, M. Pitzalis, F. Deidda, E. Porcu, F. Poddie, H.M. Kang, R. Lyons, B. Tarrier, J.B. Gresham, B. Li, S. Tofanelli, S. Alonso, M. Dei, S. Lai, A. Mulas, M.B. Whalen, S. Uzzau, C. Jones, D. Schlessinger, G.R. Abecasis, S. Sanna, C. Sidore, F. Cucca, Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny, Science 341 (2013) 565–569, https://doi.org/10.1126/science.1237947.

[7] G.D. Poznik, B.M. Henn, M.C. Yee, E. Sliwerska, G.M. Euskirchen, A.A. Lin, M. Snyder, L. Quintana-Murci, J.M. Kidd, P.A. Underhill, C.D. Bustamante, Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females, Science 341 (2013) 562–565, https://doi.org/10.1126/science.1237619.

[8] M. Vermeulen, A. Wollstein, K. van der Gaag, O. Lao, Y. Xue, Q. Wang, L. Roewer, H. Knoblauch, C. Tyler-Smith, P. de Knijff, M. Kayser, Improving global and regional resolution of male lineage differentiation by simple single-copy Y-chromosomal short tandem repeat polymorphisms, Forensic Sci. Int. Genet. 3 (2009) 205–213, https://doi.org/10.1016/j.fsigen.2009.01.009.

[9] B.E. Krenke, L. Viculis, M.L. Richard, M. Prinz, S.C. Milne, C. Ladd, A.M. Gross, T. Gornall, J.R. Frappier, A.J. Eisenberg, C. Barna, X.G. Aranda, M.S. Adamowicz, B. Budowle, Validation of a male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex, Forensic Sci. Int. 148 (2005) 111–124.

[10] J. Mulero, C. Chang, L. Calandro, R. Green, Y. Li, C. Johnson, L. Hennessy, Development and validation of the AmpFlSTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system, J. Forensic Sci. 51 (2006) 64–75, https://doi.org/10.1111/j.1556-4029.2005.00016.x.

[11] L. Roewer, Y chromosome STR typing in crime casework, Forensic Sci. Med. Pathol. 2 (2009) 77–84, https://doi.org/10.1007/s12024-009-9089-5.

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

[13] S. Gopinath, C. Zhong, V. Nguyen, J. Ge, R.E. Lagacé, M.L. Short, J.J. Mulero, Developmental validation of the Yfiler(®) plus PCR Amplification Kit: an enhanced Y-STR multiplex for casework and database applications, Forensic Sci. Int. Genet. 24 (2016) 164–175, https://doi.org/10.1016/j.fsigen.2016.07.006.

[14] M. Goedbloed, M. Vermeulen, R.N. Fang, M. Lembring, A. Wollstein, K. Ballantyne, O. Lao, S. Brauer, C. Krüger, L. Roewer, R. Lessig, R. Ploski, T. Dobosz, L. Henke, J. Henke, M.R. Furtado, M. Kayser, Comprehensive mutation analysis of 17 Ychromosomal short tandem repeat polymorphisms included in the AmpFSTR® Yfiler® PCR amplification kit, Int. J. Legal Med. 123 (2009) 471–482, https://doi.org/10.1007/s00414-009-0342-y.

[15] K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, A. Ralf, M. Vermeulen, P. de Knijff, M. Kayser, A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages, Forensic Sci. Int. Genet. 6 (2012) 208–218, https://doi.org/10.1016/j.fsigen.2011.04.017.

[16] K.N. Ballantyne, A. Ralf, R. Aboukhalid, N.M. Achakzai, M.J. Anjos, Q. Ayub, J. Ballantyne, D.J. Ballard, B. Berger, C. Bobillo, M. Bouabdellah, H. Burri, S.C. Gomes, L. Grol, C. Haas, E. Hanson, L. Henke, F. Herrera-rodr, C.R. Hill, G. Holmlund, K. Honda, U. Immel, S. Inokuchi, M.A. Jobling, B. Rerkamnuaychoke, D.R. Gonzalez, C. Robino, L. Roewer, A. Rosa, A. Sala, J.M. Salvador, P. Sanz, C. Schmitt, A.K. Sharma, D.A. Silva, M. Kayser, Toward male individualization with rapidly mutating Y-chromosomal short tandem repeats, Hum. Mutat. 35 (2014) 1021–1032, https://doi.org/10.1002/humu.22599.

[17] B.M. Dupuy, M. Stenersen, T. Egeland, B. Olaisen, Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci, Hum. Mutat. 23 (2004) 117–124, https://doi.org/10.1002/humu.10294.

[18] O. Balanovsky, Toward a consensus on SNP and STR mutation rates on the human Y-chromosome, Hum. Genet. 136 (2017) 575–590, https://doi.org/10.1007/s00439-017-1805-8.

[19] S. Claerhout, M. Vandenbosch, K. Nivelle, L. Gruyters, A. Peeters, M.H.D. Larmuseau, R. Decorte, Determining Y-STR mutation rates in deep-routing genealogies: identification of haplogroup differences, Forensic Sci. Int. Genet. 34 (2018) 1–10, https://doi.org/10.1016/j.fsigen.2018.01.005.

[20] M. Kayser, L. Roewer, M. Hedman, L. Henke, J. Henke, S. Brauer, C. Krüger, M. Krawczak, M. Nagy, T. Dobosz, R. Szibor, P. de Knijff, M. Stoneking, A. Sajantila, Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs, Am. J. Hum. Genet. 66 (2000) 1580–1588, https://doi.org/10.1086/302905.

[21] E. Heyer, J. Puymirat, P. Dieltjes, E. Bakker, P. de Knijff, Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees, Hum. Mol. Genet. 6 (1997) 799–803.

[22] M. Kayser, M. Vermeulen, H. Knoblauch, H. Schuster, M. Krawczak, L. Roewer, Relating two deep-rooted pedigrees from Central Germany by high-resolution Y-STR haplotyping, Forensic Sci. Int. Genet. 1 (2007) 125–128, https://doi.org/10.1016/j.fsigen.2007.02.004.

[23] T. King, M. Jobling, What's in a name? Y chromosomes, surnames and the genetic genealogy revolution, Trends Genet. 25 (2009) 351–360, https://doi.org/10.1016/j.tig.2009.06.003.

[24] A. Boattini, S. Sarno, P. Pedrini, C. Medoro, M. Carta, S. Tucci, G. Ferri, M. Alù, D. Luiselli, D. Pettener, Traces of medieval migrations in a socially stratified population from Northern Italy. Evidence from uniparental markers and deep-rooted pedigrees, Heredity 114 (2015) 155–162, https://doi.org/10.1038/hdy.2014.77.

[25] U. Holtkemper, B. Rolf, C. Hohoff, P. Forster, B. Brinkmann, Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques, Hum. Mol. Genet. 10 (2001) 629–633.

[26] A. Laouina, S. Nadifi, R. Boulouiz, M. El Arji, J. Talbi, B. El Houate, H. Yahia, F. Chbel, Mutation rate at 17 Y-STR loci in "Father/Son" pairs from moroccan population, Leg. Med. Tokyo (Tokyo) 15 (2013) 269–271, https://doi.org/10.1016/j.legalmed.2013.03.003.

[27] N.N. Da Fré, R. Rodenbusch, A.Z. Gastaldo, E. Hanson, J. Ballantyne, C.S. Alho, Genetic data and de novo mutation rates in father-son pairs of 23 Y-STR loci in Southern Brazil population, Int. J. Legal Med. 129 (2015) 1221–1223, https://doi.org/10.1007/s00414-014-1101-2.

[28] S. Turrina, S. Caratti, M. Ferrian, D. De Leo, Are rapidly mutating Y-short tandem repeats useful to resolve a lineage? Expanding mutability data on distant male relationships, Transfusion 56 (2016) 533–538, https://doi.org/10.1111/trf.13368.

[29] A. Adnan, A. Ralf, A. Rakha, N. Kousouri, M. Kayser, Improving empirical evidence on differentiating closely related men with RM Y-STRs: a comprehensive pedigree study from Pakistan, Forensic Sci. Int. Genet. 25 (2016) 45–51, https://doi.org/10.1016/j.fsigen.2016.07.005.

[30] A. Rakha, Y.N. Oh, H.Y. Lee, S. Hussain, A.M. Waryah, A. Adnan, K.J. Shin, Discriminating power of rapidly mutating Y-STRs in deep rooted endogamous pedigrees from Sindhi population of Pakistan, Leg. Med. Tokyo (Tokyo) 34 (2018) 17–20, https://doi.org/10.1016/j.legalmed.2018.08.001.

[31] S. Claerhout, M. Van der Haegen, L. Vangeel, M.H.D. Larmuseau, R. Decorte, A game of hide and seq: identification of parallel Y-STR evolution in deep-rooting pedigrees, Eur. J. Hum. Genet. 27 (2019) 637–646, https://doi.org/10.1038/s41431-018-0312-2.

[32] R. Alghafri, W. Goodwin, A. Ralf, M. Kayser, S. Hadi, A novel multiplex assay for simultaneously analysing 13 rapidly mutating Y-STRs, Forensic Sci. Int. Genet. 17 (2015) 91–98, https://doi.org/10.1016/j.fsigen.2015.04.004.

[33] M. Geppert, L. Roewer, DNA electrophoresis protocols for forensic genetics, Methods Mol. Biol. 830 (2012) 127–140, https://doi.org/10.1007/978-1-61779-461-2_9.

[34] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, 2nd edn., Chapman & Hall, Boca Raton, FL, 2003.

[35] B. Walsh, Estimating the time to the most recent common ancestor for the Y chromosome or mitochondrial DNA for a pair of individuals, Genetics 158 (2001) 897–912.

[36] M. Heinila, Helsinki University of Technology, Unpublished, 2012 [Rootsweb: GENEALOGY-DNA/2014-03].

[37] A.A. Westen, T. Kraaijenbrink, L. Clarisse, L.J. Grol, P. Willemse, S.B. Zuniga, E.A. Robles de Medina, R. Schouten, K.J. van der Gaag, N.E. Weiler, A.J. Kal, M. Kayser, T. Sijen, P. de Knijff, Analysis of 36 Y-STR marker units including a concordance study among 2085 Dutch males, Forensic Sci. Int. Genet. 14 (2015) 174–181, https://doi.org/10.1016/j.fsigen.2014.10.012.

[38] C. Davis, J. Ge, C. Sprecher, A. Chidambaram, J. Thompson, M. Ewing, P. Fulmer, D. Rabbach, D. Storts, B. Budowle, Prototype PowerPlex(R) Y23 System: a concordance study, Forensic Sci. Int. Genet. 7 (2013) 204–208, https://doi.org/10.1016/j.fsigen.2012.06.005.

[39] M.H.D. Larmuseau, N. Vanderheyden, A. Van Geystelen, R. Decorte, A substantially lower frequency of uninformative matches between 23 versus 17 Y-STR haplotypes in north Western Europe, Forensic Sci. Int. Genet. 11 (2014) 214–219, https://doi.org/10.1016/j.fsigen.2014.04.002.

[40] V. Petrovic, M. Kecmanovic, M. Keckarevic Markovic, D. Keckarevic, Assessment of mutation rates for PPY23 Y chromosome STR loci in Serbian father-son pairs, Forensic Sci. Int. Genet. 39 (2019) e5–e9, https://doi.org/10.1016/j.fsigen.2018.11.014.

[41] M.H.D. Larmuseau, J. Vanoverbeke, A. Van Geystelen, G. Defraene, N. Vanderheyden, K. Matthys, T. Wenseleers, R. Decorte, Low historical rates of cuckoldry in a Western European human population traced by Y-chromosome and genealogical data, Proc. Biol. Sci. 280 (2013) (2013) 2400, https://doi.org/10.1098/rspb.2013.240020132400–20132400.

[42] M.H.D. Larmuseau, K. Matthijs, T. Wenseleers, Cuckolded fathers rare in human populations, Trends Ecol. Evol. (Amst.) 31 (2016) 327–329, https://doi.org/10.1016/j.tree.2016.03.004.

[43] M.H.D. Larmuseau, S. Claerhout, L. Gruyters, K. Nivelle, M. Vandenbosch, A. Peeters, P. van den Berg, T. Wenseleers, R. Decorte, Genetic-genealogy approach reveals low rate of extrapair paternity in historical Dutch populations, Am. J. Hum. Biol. 29 (2017), https://doi.org/10.1002/ajhb.23046.

[44] S.M. Platek, T.K. Shackelford, Female Infidelity and Paternal Uncertainty, Cambridge University Press, New York, 2006.

[45] C.C. Wang, H. Li, Evaluating the Y chromosomal STR dating in deep-rooting pedigrees, Investig. Genet. 6 (2015) 8, https://doi.org/10.1186/s13323-015-0025-z.

[46] M.G.B. Blum, M. Jakobsson, Deep divergences of human gene trees and models of human origins, Mol. Biol. Evol. 28 (2011) 889–898, https://doi.org/10.1093/molbev/msq265.

[47] N. Fenner, Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies, Am. J. Phys. Anthropol. 128 (2005) 415–423, https://doi.org/10.1002/ajpa.20188.