Forensic Science International: Genetics 69 (2024) 102999

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Microsatellites' mutation modeling through the analysis of the Y-chromosomal transmission: Results of a GHEP-ISFG collaborative study

Sofia Antão-Sousa [a,b,c,d,1], Leonor Gusmão [d,1], Nidia M. Modesti [e], Sofía Feliziani [e], Marisa Faustino [a,c], Valeria Marcucci [f], Claudia Sarapura [f], Julyana Ribeiro [d], Elizeu Carvalho [d], Vania Pereira [g], Carmen Tomas [g], Marian M. de Pancorbo [h], Miriam Baeta [h], Rashed Alghafri [i], Reem Almheiri [i], Juan José Builes [j,k], Nair Gouveia [l], German Burgos [m,n], Maria de Lurdes Pontes [o], Adriana Ibarra [p], Claudia Vieira da Silva [q], Rukhsana Parveen [r], Marc Benitez [s], António Amorim [a,b,c], Nadia Pinto [a,b,t,*]

[a] Instituto de Investigação e Inovação em Saúde (i3S), Porto, Portugal
[b] Institute of Molecular Pathology and Immunology, University of Porto (IPATIMUP), Porto, Portugal
[c] Faculty of Sciences of the University of Porto (FCUP), Porto, Portugal
[d] DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil
[e] Centro de Genética Forense, Poder Judicial de Córdoba, Argentina
[f] Laboratorio Regional de Investigación Forense, Tribunal Superior de Justicia de Santa Cruz, Argentina
[g] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
[h] BIOMICs Research Group, Lascaray Research Center, Department of Zoology and Animal Cell Biology, University of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain
[i] International Center for Forensic Sciences, Dubai Police G.H.Q., Dubai, United Arab Emirates
[j] GENES SAS Laboratory, Medellín, Colombia
[k] Institute of Biology, University of Antioquia, Medellín, Colombia
[l] Instituto Nacional de Medicina Legal e Ciências Forenses, I.P. / Serviço de Genética e Biologia Forenses, Delegação do Centro, Portugal
[m] One Health Global Research Group, Facultad de Medicina, Universidad de Las Américas (UDLA), Quito, Ecuador
[n] Grupo de Medicina Xenómica, Universidad de Santiago de Compostela, Santiago de Compostela, Spain
[o] Instituto Nacional de Medicina Legal e Ciências Forenses, I.P. / Serviço de Genética e Biologia Forenses, Delegação do Norte, Portugal
[p] Laboratorio IDENTIGEN, Universidad de Antioquia, Colombia
[q] Instituto Nacional de Medicina Legal e Ciências Forenses, I.P. / Serviço de Genética e Biologia Forenses, Delegação do Sul, Portugal
[r] Forensic Services Laboratory, Centre for Applied Molecular Biology, University of the Punjab, Lahore, Pakistan
[s] Policia de la Generalitat de Catalunya – Mossos d'Esquadra. Unitat Central del Laboratori Biològic, Barcelona, Spain
[t] Centre of Mathematics of the University of Porto, Porto, Portugal

ARTICLE INFO

Keywords:
Y chromosome
Mutation
Mutation modeling
Mutation rate estimation
Microsatellites
Y-STRs

ABSTRACT

The Spanish and Portuguese Speaking Working Group of the International Society for Forensic Genetics (GHEP-ISFG) organized a collaborative study on mutations of Y-chromosomal short tandem repeats (Y-STRs). New data from 2225 father-son duos and data from 44 previously published reports, corresponding to 25,729 duos, were collected and analyzed. Marker-specific mutation rates were estimated for 33 Y-STRs. Although highly dependent on the analyzed marker, mutations compatible with the gain or loss of a single repeat were 23.2 times more likely than those involving a greater number of repeats. Longer alleles (relatively to the modal one) showed to be nearly twice more mutable than the shorter ones. Within the subset of longer alleles, the loss of repeats showed to be nearly twice more likely than the gain. Conversely, shorter alleles showed a symmetrical trend, with repeat gains being twofold more frequent than reductions. A positive correlation between the paternal age and the mutation rate was observed, strengthening previous findings. The results of a machine learning approach, via logistic regression analyses, allowed the establishment of algebraic formulas for estimating the probability of mutation depending on paternal age and allele length for DYS389I, DYS393 and DYS627. Algebraic formulas could also be established considering only the allele length as predictor for DYS19, DYS389I, DYS389II-I, DYS390, DYS391,

* Corresponding author at: Instituto de Investigação e Inovação em Saúde (i3S), Porto, Portugal.
E-mail address: nmgapinto@gmail.com (N. Pinto).
[1] These authors contributed equally to the work.

https://doi.org/10.1016/j.fsigen.2023.102999
Received 8 May 2023; Received in revised form 25 October 2023; Accepted 10 December 2023
Available online 14 December 2023

DYS393, DYS437, DYS439, DYS449, DYS456, DYS458, DYS460, DYS481, DYS518, DYS533, DYS576, DYS626 and DYS627 loci. For the remaining Y-STRs, a lack of statistical significance was observed, probably as a consequence of the small effective size of the subsets available, a common difficulty in the modeling of rare events as is the case of mutations. The amount of data used in the different analyses varied widely, depending on how the data were reported in the publications analyzed. This shows a regrettable waste of produced data, due to inadequate communication of the results, supporting an urgent need of publication guidelines for mutation studies.

1. Introduction

The Y chromosome provides invaluable data to study the biological mechanisms of germinal mutations, with no parallel in any other component of the nuclear genome. This is due to its haploid mode of genetic transmission, which allows the unambiguous identification of which parental allele originated which filial one, whenever father-son duos are analyzed for simple-structure markers [1]. Indeed, when analyzing length polymorphisms for either autosomal or X-chromosomal markers, it is rarely possible to determine with certainty which parental allele originated which filial one, or even if an undetected mutation occurred, as length mutations may not lead to Mendelian incompatibilities. Thus, the standard approach of estimating mutation rates through the number of Mendelian incompatibilities observed necessarily leads to mutation rate underestimation in both autosomal and X-chromosomal markers [2–6]. However, it is noteworthy that this underestimation occurs in a smaller extent for X-chromosomal transmission, since in father-daughter and mother-son transmissions the parental and filial allele, respectively, involved in each allelic transmission is known [2]. Theoretical approaches have been proposed to mitigate this bias for autosomal markers [3–6], and an informatics tool was recently presented [7] aiming for more accurate mutation rates estimation.

Short tandem repeats (STRs) are the most used markers in population and forensic genetics, being highly polymorphic due to their relatively high mutation rates. The primary mutation mechanism is thought to be the polymerase template slippage that leads to the insertion or deletion of single or multiple repeats in the allele transmitted from the parent to the child [8,9].

Concerning STR mutations, and regardless of the mode of genetic transmission considered, it is generally accepted that (*a*) single-step mutations, i.e., the gain or loss of a single repeat, are more frequent than multistep ones [10–12], (*b*) longer alleles are more prone to mutation than shorter ones [13], and (*c*) longer alleles tend to lose repeats [14–17], while shorter ones are more prone to gains than losses [15–17]. Moreover, it is also commonly accepted that (*d*) older fathers are more prone to allele mutations, and, for the autosomal and X-linked transmission modes, that (*e*) paternal mutations are more frequent than maternal ones [16,18]. These premises have been consolidated considering any mode of genetic transmission and genotyping methodologies mostly based on fragment length determination, which may lead to biased interpretations, especially when analyzing autosomal or X-chromosomal STRs. For example, a Mendelian incompatibility caused by a multistep mutation may be explained by a single-step one, depending on the genotypic configuration of the parent(s)-child duo or trio analyzed. In this case, since single-step mutations are assumed to be much more frequent than multistep ones, the first is assumed to have occurred, which necessarily (and artificially) increases the preponderance of single-step over multistep mutations. The weight of this bias was previously studied for both autosomal [3–7] and X-chromosomal STRs [19,20].

The study of Y-chromosomal STRs (Y-STRs), specifically those with simple structure, is thus an invaluable approach for STR mutation modeling, which has recently allowed the detection of a correlation between the structure of the repetitive motif and the ratio between single and multi-step mutations [12]. Indeed, the unambiguous identification of which parental allele originated which filial one, provides unparalleled insights into the strength of premises *(a)* to *(d)* above.

Here, the results of a collaborative study on Y-STR mutations organized by the Spanish and Portuguese Speaking Working Group of the International Society for Forensic Genetics (GHEP-ISFG) are presented, adding new data to the results of a thorough revision of the literature. Besides updating the mutation rate estimates of the most widely used Y-STRs, our approach allowed the quantification of the relative frequencies of single- and multistep mutations, as well as the comparison of the number of mutations involving shorter and longer (relatively to the modal one) paternal alleles. The trend for paternal alleles to either gain or lose repeats depending on their length was also quantified, as well as the correlation between the paternal age and the frequency of mutations. Finally, logistic regression computations considering as predictors either, both paternal age and allele length, or only the allele length (in a greater subset), were also performed.

2. Material and methods

At the 2019 General Assembly, a GHEP-ISFG collaborative study on Y-STR mutations was approved and opened to the participation of all GHEP-ISFG members. The presentation of the certificates of a proficiency test for the previous three years showing correct results for the analyzed kits was required for each participating laboratory.

2.1. Sample collection and genotyping

New data from 2225 father/son duos were obtained from 10 worldwide populations: Argentina (414), Brazil (202), Colombia (222), Denmark (96), Ecuador (102), Greenland (104), Pakistan (110), Portugal (509), Spain (250), and United Arab Emirates (216). Each laboratory ensured the anonymization of the samples and total compliance with the legal and ethical requirements for the use of the corresponding data in this research project. Most of the used samples were originally obtained from paternity investigations, where the biological relationship was confirmed using autosomal markers (likelihood ratio greater than 10^4). Some volunteer duos, with the biological relationship legally recognized, were also analyzed.

Each participating laboratory had to present genotypic data regarding a minimum number of father-son duos, analyzed by either PowerPlex® Y23 System, Promega (PPY23), or Yfiler™ Plus PCR Amplification kit, ThermoFisher Scientific (YFPlus).

2.2. Data collection

Worldwide data regarding father-son genotypic configurations for the 27 Y-STRs included in either (or both) PPY23 or YFPlus kits were gathered from 44 published works [21–64], as well as for 8 further Y-STRs (not included in the previously mentioned kits): DYS388, DYS435, DYS461, DYS526a/b, DYS547, DYS612, DYS626, and GATA A10. Allele and mutation data for DYS389II and DYS526b were obtained subtracting those from DYS389I and DYS526a, respectively, and thus are represented by DYS389II-I and DYS526b-a.

The total number of father-son duos used in each of the analyses performed depended on the information available from published works (Table S1). Namely, marker-specific mutation rates (Results 3.1.) were

obtained using data from the present GHEP-ISFG collaborative study and a set of 44 published reports [21–64], for which the absolute numbers of analyzed meiosis and observed mutations were presented. The correlation between the age of the fathers at the time of the birth of the son and the mutation frequency (Results 3.2.) was obtained by combining the data herein presented with those from two previous GHEP-ISFG collaborative studies [30,40] and from [60], in which paternal age was reported for both the mutated and unmutated transmissions. Age raw data from [40] were used to match age intervals. The correlation between the length of the observed alleles and both the occurrence of mutation, and the repeat gain or loss (Results 3.3.) was assessed (and quantified) using data from the herein presented GHEP-ISFG working commission and from [26,27,29,30,32,33,40,44,48–51,60]. Two markers out of the 33 for which mutation data regarding both the parental and filial alleles were gathered, DYS385a/b and DYS643, were not included in these analyses as the identification of mutations in marker DYS385a/b might be ambiguous and no mutations were observed for marker DYS643. Works for which we could not assess the complete set of observed alleles were not considered for this analysis, as the relative length of the mutated alleles compared to the observed ones could not be ascertained. Finally, for the same set of markers excluding the multi-copy DYS526a/b, logistic regression analyses were computed considering as predictors the paternal age (whenever available) and/or allele length (Results 3.4). Raw data concerning the paternal age, the allele length and the mutation status were gathered from [30,40,60].

2.3. Data analysis

The statistical power of the results was assessed considering a level of significance $\alpha = 0.05$. Marker-specific mutation rates were estimated for 33 Y-STRs, and the corresponding confidence intervals were estimated from the binomial standard deviation. The number of mutations compatible with the gain or loss of a specific number of repeats was also assessed.

The correlation between the paternal age and the occurrence of mutation was assessed through Chi-square tests, as well as the correlation between the length of the alleles and both (a.) the occurrence of a mutation, and (b.) the gain or loss of repeats for the 31 Y-STRs with simple structure. In both cases, and for the sake of obtaining statistical significance, age and allele data were gathered into classes. For analysis (b.), the modal paternal allele, i.e., the allele with the greatest number of observations, was established for each marker. For each marker, paternal alleles were then included into one of the following three categories: *i.* modal allele, *ii.* shorter, and *iii.* longer allele (relative to the modal one). The correlation between the occurrence of mutations compatible with either the gain or loss of repeats, and the length of the alleles was then assessed for each category.

Logistic regression analyses were computed using RStudio [65] and dplyr package [66], considering as dependent variable the occurrence of mutation (coded as '1', and as '0' otherwise) and as predictors either *i.* both the paternal age and allele length, or *ii.* only the allele length. Analysis *ii.* was performed separated from *i.* as the effective size of the subsets of data available for one and another case ("paternal age at the time of birth" + "paternal allele length", or "paternal allele length" only) greatly differ. The parameter values of the model were obtained via maximum likelihood estimation for each marker and due to the little amount of existing data, whole of it were used as training dataset. To avoid an increased level of complexity, the subset of markers analyzed were those single copy, most with simple structure.

3. Results

3.1. Average mutation rates

The number of father-son duos analyzed for each of the 36 studied markers varied from 26,372 for DYS19, to 161 for DYS435, the median and the average number of pairs of subjects equating to 9178 and 13,297, respectively (Table S2). Due to the low number of allelic transmissions analyzed, DYS435 and DYS461 (N = 161 and N = 873, respectively) were disregarded henceforth. DYF387S1 was also not further considered, since it is a multi-copy marker with a variable number of loci amongst individuals, which does not allow for the correct counting of allelic transmissions. A total of 469,611 allelic transmissions were then analyzed for the remaining 33 Y-STRs, and 1863 mutations were observed. Average mutation rates varied between 0.0005 (for DYS438 and DYS643) and 0.0170 (for DYS547) (Table 1). Discrimination between allelic transfers and mutations observed in previous works [21–64] and those of this study are presented in Table S2. Mutations observed in DYF387S1 are presented in Table S3. Of the 233 mutations detected in this study, 22 correspond to co-occurrences in the same father/son duo, 8 duos showing mutations in two loci and two duos in three loci (Table S4).

Out of the 1863 mutations observed, 1786 were compatible with single-step mutations, 74 with multistep ones, and 3 were mutations not compatible with changes involving an integer number of repeats (Table 1), the first being 24.1 times more frequent than the multistep ones. It is however noteworthy that this ratio showed to be highly variable between markers, varying between 1.25 (i.e., essentially equifrequent) for DYS438 and 98 for DYS449, which is correlated with the structure of the repetitive motif as recently shown [12].

3.2. Correlation between the age of the fathers and the occurrence of mutation

Information on the age of the father at the time of the birth of the son was gathered for 84,715 allelic transmissions (Table 2 and S5). The findings support that the mutation rate and the age of the father are positively correlated. The 10-year age class with more subjects was the one for ages between 21 and 30 years old, gathering 46.5 % of the analyzed individuals. The age class 51–60 was the one showing the highest mutation rate (Table 2). Considering dichotomous classes, statistically significant differences were found between individuals aged <31 and >30 (p-value = 0.00302), and <41 and >40 (p-value = 0.00018), and a nearly significant difference was reached between individuals aged <51 and >50 (p-value = 0.05830) (Table S6). In all the cases, the older individuals were associated with higher mutation rates.

3.3. Correlation between the length of the analyzed alleles and mutation occurrence

The number of allelic transmissions, mutations, and mutations compatible with either the gain or loss of one or more repeats were analyzed for each marker, considering the following categories for the observed paternal alleles: modal allele, either shorter or longer than the modal allele (Table 3). The presented results were reached for 31 Y-STRs through the mutation matrices presented in Supplementary Material that were elaborated considering the data obtained in this study and from [26,27,29,30,32,33,40,44,48–51,60]. Two markers from the set of those analyzed in Section 3.1 were excluded from the analysis: DYS385a/b because the identification of mutations would result ambiguous, and DYS643 because no mutations were reported for this marker in the data used for this analysis.

Considering the subset of 62,380 paternal alleles shorter than the modal one, 241 suffered mutation, which resulted in an overall mutation rate of 3.9E-03. This figure increases to 9.8E-03 if considering the 621 mutations observed among the 63,116 transmissions involving paternal alleles greater than the modal one. Indeed, longer alleles showed to be 2.55 times more prone to mutation than shorter ones, and this difference showed strong statistical significance (p = 1.36E-37).

On the other hand, when considering the gain or loss of repeats, most of the mutations involving shorter paternal alleles (241) involved the

Table 1
Mutation rates estimation for 33 Y-STRs, and corresponding confidence intervals. The number of mutations compatible with either the gain or loss of 1 to 6 repeats, and corresponding ratios between the number of single and multistep mutations. Data include both those generated in the present study and those obtained from the literature [21–64]. NC: Not compatible with changes involving an integer number of repeats.

Markers	Observations (N)		Mut rate	CI (95 %)	Number of repeats compatible to be involved in the mutation							Single/ Multistep
	Muts	Allelic transmissions			1	2	3	4	6	>1*	NC	
DYS19	57	26,372	0.0022	0.00164–0.00280	56	0	0	0	0	0	1	–
DYS385a/b	112	46,659	0.0024	0.00198–0.00289	104	7	1	0	0	0	0	13.00
DYS388	2	3612	0.0006	0.00007–0.00200	2	0	0	0	0	0	0	–
DYS389I	69	26,154	0.0026	0.00205–0.00334	69	0	0	0	0	0	0	–
DYS389II-I	91	26,113	0.0035	0.00281–0.00428	88	2	1	0	0	0	0	29.33
DYS390	63	25,103	0.0025	0.00193–0.00321	62	1	0	0	0	0	0	62.00
DYS391	66	25,789	0.0026	0.00198–0.00325	64	2	0	0	0	0	0	32.00
DYS392	14	24,266	0.0006	0.00032–0.00097	13	0	0	1	0	0	0	13.00
DYS393	31	24,332	0.0013	0.00087–0.00181	31	0	0	0	0	0	0	–
DYS437	22	21,018	0.0010	0.00066–0.00158	22	0	0	0	0	0	0	–
DYS438	10	21,033	0.0005	0.00023–0.00087	5	3	0	1	0	0	1	1.25
DYS439	110	21,111	0.0052	0.00428–0.00628	110	0	0	0	0	0	0	–
DYS448	20	16,577	0.0012	0.00074–0.00186	19	1	0	0	0	0	0	19.00
DYS449	99	8480	0.0117	0.00950–0.01420	98	0	1	0	0	0	0	98.00
DYS456	81	17,892	0.0045	0.00360–0.00562	80	0	1	0	0	0	0	80.00
DYS458	126	17,920	0.0070	0.00586–0.00837	122	3	1	0	0	0	0	30.50
DYS460	23	5605	0.0041	0.00260–0.00615	23	0	0	0	0	0	0	–
DYS481	25	5504	0.0045	0.00294–0.00670	22	3	0	0	0	0	0	7.33
DYS518	116	7795	0.0149	0.01231–0.01782	105	5	3	2	0	0	1	10.50
DYS526a	15	4425	0.0034	0.00190–0.00559	11	4	0	0	0	0	0	2.75
DYS526b-a	50	4401	0.0114	0.00843–0.01498	49	1	0	0	0	0	0	49
DYS533	9	6975	0.0013	0.00059–0.00245	8	1	0	0	0	0	0	8.00
DYS547	69	4053	0.0170	0.01327–0.02150	67	0	0	1	0	1	0	33.50
DYS549	10	2617	0.0038	0.00183-0.00702	10	0	0	0	0	0	0	–
DYS570	87	9976	0.0087	0.00699–0.01075	80	6	1	0	0	0	0	11.43
DYS576	140	9876	0.0142	0.01194–0.01671	135	3	2	0	0	0	0	27.00
DYS612	66	4056	0.0163	0.01261–0.02066	60	6	0	0	0	0	0	10.00
DYS626	41	4441	0.0092	0.00663–0.01250	39	1	0	0	1	0	0	19.50
DYS627	133	8028	0.0166	0.01389–0.01960	127	5	1	0	0	0	0	21.17
DYS635	59	18,813	0.0031	0.00239–0.00404	59	0	0	0	0	0	0	–
DYS643	1	1978	0.0005	0.00001–0.00281	1	0	0	0	0	0	0	–
GATA A10	4	1026	0.0039	0.00106–0.00995	4	0	0	0	0	0	0	–
GATA H4	42	17,611	0.0024	0.00172–0.00322	41	1	0	0	0	0	0	41.00
Total	1863	469,611	–	–	1786	55	12	5	1	1	3	24.1

* Multistep mutation reported without specifying the number of steps [61].

Table 2
Number of father-son allelic transmissions analyzed considering paternal age intervals (at the time of the son's birth) and the corresponding mutation rate. Included data are from the presented study and previous reports [30,40,60]. See Tables S5 and S6 for more details.

Paternal age classes		Number of mutations	Number of allelic transmissions	Mutation rate	Confidence interval (95 %)
10-years	<21	31	7995	0.00388	0.00264–0.0055
	21–30	112	39,408	0.00284	0.00234–0.00342
	31–40	96	25,995	0.00369	0.00299–0.00451
	41–50	45	8076	0.00557	0.00407–0.00745
	51–60	14	2407	0.00582	0.00318–0.00974
	>61	4	834	0.00480	0.00131–0.01223
Dichotomous	<31	143	47,403	0.00302	0.00254–0.00355
	>30	159	37,312	0.00426	0.00363–0.00498
	<41	239	73,398	0.00326	0.00286–0.00370
	>40	63	11,317	0.00557	0.00428–0.00712
	<51	284	81,474	0.00349	0.00309–0.00392
	>50	18	3241	0.00555	0.00329–0.00878

gain of repeats (161), and this difference between gains and losses showed statistical significance (p = 1.82E-04). Conversely, 394 out of the 621 mutations involving longer paternal alleles corresponded to the loss of repeats, this difference between gains and losses also showing statistical significance (p = 8.4E-06). The trend for shorter alleles to gain repeats and longer ones to lose repeats showed to be highly significant (p = 8.92E-15). Finally, it should be remarked that mutations involving the modal allele did not show statistically supported evidence to either gain or lose repeats (p = 0.11).

3.4. Logistic regression computations

A logistic regression was computed to model the probability of a mutation to occur, considering as predictors both the paternal age and allele length, or only the paternal allele length (Tables 4, 5 and S7), for the subset of single copy markers, most with simple structure. Markers DYS388, DYS612, and DYS626 were not considered for age analyses due to lack of data.

For the subset of 25 markers for which both paternal age and allele data were available, only three: DYS389I, DYS393, and DYS627, showed statistical significance (α = 0.05) for both predictors (Table 4). For these

Table 3
Number of allelic transmissions (Transm) and mutations (Mut) observed compatible with changes involving an integer number of repeats (Mut), as well as the number of Mut compatible with either gains or losses, considering the following categories for the paternal allele: modal allele, shorter or longer than de modal allele. Data include both those generated in the presented GHEP-ISFG collaborative study and those obtained from [26,27,29,30,32,33,40,44,48–51,60].

Markers	Transm	Mut	Modal allele	Length of Paternal allele analyzed (relative to the modal)											
				Shorter than modal				Modal				Longer than modal			
				Transm	Mut	Loss	Gain	Transm	Mut	Loss	Gain	Transm	Mut	Loss	Gain
DYS19	11,688	26	14	1463	2	0	2	5333	10	1	9	4892	14	7	7
DYS388	101	1	12	19	1	0	1	73	0	0	0	9	0	0	0
DYS389I	10,616	37	13	2426	6	1	5	6011	14	4	10	2179	17	15	2
DYS389II-I	10,593	48	16	886	4	0	4	5456	10	0	10	4251	34	23	11
DYS390	11,826	29	24	4644	5	1	4	5174	15	8	7	2008	9	9	0
DYS391	11,827	35	10	632	1	0	1	6775	10	1	9	4420	24	14	10
DYS392	11,810	7	13	5537	1	1	0	4794	3	0	3	1479	3	1	2
DYS393	10,577	17	13	2131	2	0	2	6848	8	4	4	1598	7	5	2
DYS437	10,057	15	14	37	0	0	0	4720	3	0	3	5300	12	8	4
DYS438	10,119	6	12	6340	4	3	1	3435	2	2	0	344	0	0	0
DYS439	10,078	63	12	4330	13	3	10	4258	18	6	12	1490	32	23	9
DYS448	8756	13	19	1083	1	0	1	3436	6	4	2	4237	6	6	0
DYS449	5538	69	30	1769	12	4	8	952	11	8	3	2817	46	25	21
DYS456	8757	53	15	1213	3	1	2	4005	14	2	12	3539	36	25	11
DYS458	8765	79	17	3884	23	8	15	2620	24	12	12	2261	32	18	14
DYS460	3411	14	11	1797	5	1	4	1466	5	4	1	148	4	4	0
DYS481	3707	21	22	276	0	0	0	1221	5	1	4	2210	16	6	10
DYS518	5493	93	38	1425	14	4	10	1166	15	7	8	2902	64	32	32
DYS526a	3119	14	14	890	6	3	3	1000	4	2	2	1229	4	2	2
DYS526b-a	3095	38	22	842	6	2	4	773	8	1	7	1480	24	12	12
DYS533	3705	5	12	1718	1	0	1	1714	2	1	1	273	2	2	0
DYS547	3259	57	48	1223	21	8	13	778	9	4	5	1258	27	22	5
DYS549	1317	7	12	240	0	0	0	491	3	2	1	586	4	4	0
DYS570	7015	64	17	899	7	0	7	2060	18	6	12	4056	39	25	14
DYS576	6914	108	18	3111	20	3	17	2191	35	20	15	1612	53	37	16
DYS612	3288	57	36	884	21	5	16	933	6	3	3	1471	30	15	15
DYS626	3237	35	30	1241	6	3	3	610	2	1	1	1386	27	20	7
DYS627	5663	89	21	2811	28	10	18	1152	15	9	6	1700	46	29	17
DYS635	9635	35	23	4424	20	16	4	4141	9	6	3	1070	6	3	3
GATA A10	874	4	15	316	2	1	1	433	2	2	0	125	0	0	0
GATA H4	9848	22	12	3889	6	2	4	5173	13	9	4	786	3	2	1
Overall average	**214,688**	**1161**	–	**62,380**	**241**	**80**	**161**	**89,192**	**299**	**130**	**169**	**63,116**	**621**	**394**	**227**

Table 4
Summary of logistic regression computations considering as predictors of mutation (coded as "1", and as "0" otherwise) both the paternal age and allele, for the three markers (out of the 25 analysed) for which statistical significance ($\alpha = 0.05$) was reached for both predictors. Summary data for the complete set of markers analysed can be found in Table S7. Mut represents the number of mutations observed in N allele transmissions analysed. Data rely on the consideration of both the paternal allele and age at time of son's birth of each one of the subjects and include those generated in the presented GHEP-ISFG collaborative study and those obtained from raw data of [30], [40] and [60].

Marker	Mut	N	Coefficients	Estimate	Std. Error	p-value
DYS389I	12	5128	Intercept (b)	–18.1927	5.3056	0.000606
			Age (a1)	0.05199	0.02328	0.025517
			Allele (a2)	0.78672	0.38916	0.043216
DYS393	5	5122	Intercept (b)	–27.2037	6.62249	3.99E-05
			Age (a1)	0.08083	0.032	0.01155
			Allele (a2)	1.30499	0.45972	0.00453
DYS627	21	1081	Intercept (b)	–10.9578	2.793	8.73E-05
			Age (a1)	0.05134	0.02065	0.0129
			Allele (a2)	0.25196	0.12522	0.0442

markers, the estimates associated with the coefficients of both predictors are positive, which means that greater alleles and older individuals are associated with an increased probability of mutation, as expected. Also, the estimate associated with the predictor allele showed to be greater (15.1x, 16.1x, and 4.9x, for DYS389I, DYS393, and DYS627, respectively), than the one associated with age, which supports a greater influence of the first on the probability of mutation. Algebraic expressions for the probability of mutation depending on both the paternal age and allele length can be inferred from the results presented in Table 4 for the markers for which statistical significance was reached. Indeed, through the logistic regression model, the probability of mutation of the marker M, depending on both paternal age and allele, can be estimated by

$p(mutation \mid age, allele) = \frac{1}{1+e^{-(b+a1\times age+a2\times allele)}}$, where b, $a1$, and $a2$, are the estimates for the coefficients of intercept, age, and allele, respectively, presented in Table 4. For example, through this model the probability of occurring a mutation in the marker DYS389I involving a father with 29 years old and genotype 15 is estimated as

$p(mutation\ |29, 15) = \frac{1}{1+e^{-(-18.1927+0.05199\times 29+0.78672\times 15)}} = 0.007508$, changing this estimation to

$p(mutation\ |43, 12) = 0.001477$ in the case of a father with 43 years old and genotype 12.

When the paternal allele length was considered as the only predictor of mutation, greater subsets of subjects were able to be considered, and statistical significance was reached for 18 out of the 28 analysed markers (Table 5 and Table S7). As before, the estimate associated with the allele showed to be positive for all the cases, reinforcing that greater alleles are generally associated with greater probabilities of mutation. Algebraic expressions for the probability of mutation depending only on the paternal allele length can be inferred from the results presented in Table 5 for the markers that showed statistical significance – for data for the complete set of analysed markers see Table S7. Through the logistic regression model, the probability of mutation on the marker M, depending on the paternal allele, is estimated by $P(mutation \mid allele) = \frac{1}{1+e^{-(b+a\times allele)}}$, where b, and a, are the estimates for coefficients of intercept, and allele, respectively, presented in Table 5. For example, through this model the probability of occurring a mutation in the marker DYS19 involving a father with genotype 12 is estimated as $P(mutation\ |12) = \frac{1}{1+e^{-(-13.1024+0.4762\times 12)}} = 0.000618$, changing this estimation to

Table 5
Summary of logistic regression computations considering as predictor of mutation (coded as "1", and as "0" otherwise) the length of the paternal allele, for the 18 markers (out of the 28 analysed) for which statistical significance ($\alpha = 0.05$) was reached. Summary data for the complete set of markers analysed can be found in Table S7. Mut represents the number of mutations observed in N allele transmissions analysed. Data rely on those generated in the presented GHEP-ISFG collaborative study and those obtained from [26,27,29,30,32,33,40,44,48–51,60].

Marker	Mut	N	Coefficients	Estimate	Std. Error	p-value
DYS19	27	11,687	Intercept (b)	–13.1024	2.5184	1.97E-07
			Allele (a)	0.4762	0.1666	0.00426
DYS389I	37	10,616	Intercept (b)	–14.9595	2.6134	1.04E-08
			Allele (a)	0.7071	0.195	0.000288
DYS389II-I	48	10,590	Intercept (b)	–16.982	2.4474	3.95E-12
			Allele (a)	0.6933	0.1434	1.33E-06
DYS390	29	11,826	Intercept (b)	–18.9412	4.8062	8.12E-05
			Allele (a)	0.5415	0.199	0.00651
DYS391	35	11,827	Intercept (b)	–16.0732	2.7696	6.50E-09
			Allele (a)	0.9748	0.2581	0.000159
DYS393	17	10,577	Intercept (b)	–20.7473	3.5596	5.59E-09
			Allele (a)	1.0788	0.2603	3.41E-05
DYS437	15	10,056	Intercept (b)	–18.8974	5.1514	0.000244
			Allele (a)	0.835	0.3415	0.014484
DYS439	63	10,075	Intercept (b)	–16.8071	1.69E+00	<2.00E-16
			Allele (a)	0.9752	0.135	5.11E-13
DYS449	69	5538	Intercept (b)	–11.789	1.81123	7.57E-11
			Allele (a)	0.2385	0.05715	3.01E-05
DYS456	53	8755	Intercept (b)	–14.2928	1.9923	7.27E-13
			Allele (a)	0.586	0.1241	2.35E-06
DYS458	78	8764	Intercept (b)	–10.2435	1.39083	1.77E-13
			Allele (a)	0.32583	0.08018	4.83E-05
DYS460	14	3411	Intercept (b)	–14.3873	3.9836	0.000304
			Allele (a)	0.8373	0.3662	0.022223
DYS481	21	3707	Intercept (b)	–12.1844	2.41783	4.67E-07
			Allele (a)	0.29233	0.09767	0.00276
DYS518	94	5494	Intercept (b)	–12.3526	1.92084	1.27E-10
			Allele (a)	0.21177	0.04836	1.19E-05
DYS533	5	3705	Intercept (b)	–15.8535	4.8279	0.00102
			Allele (a)	0.7821	0.3943	0.0473
DYS576	107	6913	Intercept (b)	–13.6211	1.3967	<2E-16
			Allele (a)	0.5251	0.0755	3.52E-12
DYS626	35	3237	Intercept (b)	–15.0473	2.42582	5.54E-10
			Allele (a)	0.3427	0.07698	8.51E-06
DYS627	89	5663	Intercept (b)	–10.6755	1.24828	<2.00E-16
			Allele (a)	0.31322	0.05801	6.68E-08

$p(mutation\ |14) = 0.001601$ in the case of a male with genotype 14 for the same marker.

4. Discussion and conclusions

The analysis of the transmission of Y-chromosomal markers provides invaluable insights into mutation mechanisms, especially when genotyping is based on fragment length determination, as it is still the common practice in forensic genetics routine. Indeed, the unambiguous identification in simple-structure STRs of which parental allele originated which filial one, prevents the biases inherent to other modes of genetic transmission, such as the occurrence of hidden mutations or the overestimation of single-step mutations compared to multi-step ones [3,7,19]. The insights provided by these haploid markers may be analyzed under the scope of autosomal and X-chromosomal modes of genetic transmission, and conclusions eventually transferred and included in specifically devoted software. It must be said, however, that the absence of recombination at the Y-specific regions may limit the transferability of the model.

The average mutation rates of the 33 different Y-STRs vary significantly, from 0.0005 for DYS438 and DYS643 to 0.0170 for DYS547 (i.e., 34 times), which strengthens the recommendation on the use of marker-specific estimates. For markers with a complex structure, the presented mutation rate may be underestimated if 'compensating' mutations occur within the markers (that is, a gain in one region and an equal length loss in another).

Our work also supports that single-step mutations are more frequent than multistep ones, although the ratio between the two varies from marker to marker – from slightly above one (DYS438) to almost one hundred (DYS449), which is correlated with the structure of the repetitive motif such as recently shown [12].

Longer alleles (relatively to the modal one) showed ~two times more mutations than the shorter ones. Within the subset of longer alleles, the trend to lose repeats showed to be greater (1.7 times) than for gains, the opposite trend being observed within the subset of shorter alleles (2.0 times more mutations involving gains than losses). When considering the mutations involving the modal paternal allele, no statistically significant differences were found between the numbers of the two types of mutations.

In agreement with previous reports [16,17,67,68], the data analyzed in this paper support a positive correlation between the age of the father and the occurrence of mutation.

In summary, our results provide quantified evidence supporting the generally accepted premises that (*i.*) single-step mutations are more common than multistep ones, but showed that the magnitude of the difference is highly variable across the analyzed markers, (*ii.*) long alleles are more prone to mutation than short ones (~ twice), and (*iii.*) long alleles are more prone to lose repeats (~ twice) while short alleles show the opposite tendency (also ~ twice).

In any case, it is noteworthy that the statistical strength of the conclusions described above is greatly dependent on the level of detail under which the data are published. Indeed, 44 published reports were used to estimate the average mutation rates presented in Table 1 (467,073 father-son allelic transmissions), as all of them reported for each marker both the number of allelic transmissions and the number of mutations observed [21–64]. The age of the father at the time of the

birth of the child was obtained only for 84,715 allelic transmissions out of the 467,073 produced and analyzed (18.1 %). On the other hand, the complete set of analyzed allelic transmissions, including non-mutated alleles which allowed us to present Table 3, was only possible to gather for 214,685 out of the 467,073 allelic transmissions produced and analyzed. This represents a huge amount of wasted, or at least under-exploited, data (46.0 %), which would be useful to advance the state of the art of microsatellites' mutation modeling. Specifically, a proper communication of the produced data would allow a more robust estimation of bi-allele mutation rates, essential for the weighing of the evidence in kinship problems [1].

Mutation models have been proposed considering the relationship between the mutation rates of the markers and the allele length, including a logistic one that showed a general best fit according to Akaike's information criterion [69]. In this work, the mutation rate was modelled through logistic regression, a machine learning algorithm that uses the Maximum Likelihood for parameter estimation, considering either one (paternal allele length) or two (paternal age and allele length) predictors. The number of markers that reached statistical significance was much greater in the first case (18 out of the 28 analysed) than in the second (3 out of 25) likely due to the larger datasets considered. For these cases, algebraic formulae for the estimation of marker specific mutation rates depending on either, both paternal allele length and age, or only paternal allele, are provided.

Proper estimation of mutation parameters is crucial for a wide range of forensic genetic problems as well as in evolutionary and phylogenetic studies. Since mutation is a rare event, proper modeling necessarily depends on the analysis of a prohibitively large number of individuals to be collected and analyzed by a single laboratory. The organization of collaborative studies gathering efforts and synergies from several laboratories, such as the one here presented from the GHEP-ISFG, seems a good strategy to overcome the difficulty of recruiting such a large amount of genetic data. Properly produced, analyzed, and peer-reviewed data must be suitably communicated in a way that allows it to be reused, verified, and re-analyzed later for further investigations. The only way to report data allowing the study of the co-occurrence of mutations in the same meiosis (an overlooked possibility deserving investigation) would require data release in a haplotypic format, as the non-mutated loci should also be known. We recognize however that the public release of individual haplotypes, even anonymized, may raise ethical concerns. Anyhow, we urge the forensic community and their representative bodies to undertake the elaboration of recommendations concerning the publication of results on this topic. In accordance, we dare to suggest as a minimum standard the mutation matrix format used in this work and presented as supplementary material which, although not allowing the investigation of co-occurrence of mutations, provides nonetheless the information required for the estimation of allele specific mutation rates per marker.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We deeply thank the support of Promega and ThermoFisher Scientific as reduced prices were provided for the participants in the purchase of the kits PowerPlex® Y23 System, Promega (PPY23), and Yfiler™ Plus PCR Amplification kit, ThermoFisher Scientific (YFPlus), respectively. This work was partially financed by FEDER- Fundo Europeu de Desenvolvimento Regional funds through the COMPETE - Operational Program for Competitiveness and Internationalization POCI, Portugal, and by Portuguese funds through FCT- Fundação para a Ciência e a Tecnologia, Ministério da Ciência, Tecnologia e Inovação in the framework of the projects "Institute for Research and Innovation in Health Sciences" POCI-01–0145- FEDER-007274). NP (2022.04997. CEECIND), SAS (SFRH/BD/136284/2018) and MF (2021.08783. BD) are funded by FCT. LG is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq ref. 306342/2019-7), and Fundação de Amparo Pesquisa do Estado do Rio de Janeiro - FAPERJ (CNE-2022). GB is supported by MED.GBF.20.07 funded by DGIV from Universidad de Las Américas; Quito, Ecuador. Funds for MMP and MB were provided by the Basque Government (Grupo Consolidado IT-1271-19 and IT-1633-22). Centro de Genética Forense (NM and SF) is financed by Poder Judicial de Córdoba.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2023.102999.

References

[1] N. Pinto, L. Gusmão, A. Amorim, Mutation and mutation rates at Y chromosome specific short tandem repeat polymorphisms (STRs): a reappraisal (pp), Forensic Sci. Int. Genet. 9 (1) (2014) 20–24, https://doi.org/10.1016/J.FSIGEN.2013.10.008.
[2] S. Antão-Sousa, E. Conde-Sousa, L. Gusmão, A. Amorim, N. Pinto, Underestimation and misclassification of mutations at X chromosome STRs depend on population's allelic profile (pp), Forensic Sci. Int. Genet. Suppl. Ser. 7 (1) (2019) 718–720, https://doi.org/10.1016/J.FSIGSS.2019.10.150.
[3] K. Slooten, F. Ricciardi, Estimation of mutation probabilities for autosomal STR markers (pp), Forensic Sci. Int. Genet 7 (3) (2013) 337–344, https://doi.org/10.1016/J.FSIGEN.2013.01.006.
[4] R. Chakraborty, D.N. Stivers, Y. Zhong, Estimation of mutation rates from parentage exclusion data: applications to STR and VNTR loci (pp.), Mutat. Res. 354 (1) (1996) 41–48, https://doi.org/10.1016/0027-5107(96)00014-0.
[5] P. Vicard, A.P. Dawid, A statistical treatment of biases affecting the estimation of mutation rates (pp.), Mutat. Res. 547 (1–2) (2004) 19–33, https://doi.org/10.1016/J.MRFMMM.2003.11.005.
[6] P. Vicard, A.P. Dawid, J. Mortera, S.L. Lauritzen, Estimating mutation rates from paternity casework (pp.), Forensic Sci. Int. Genet 2 (1) (2008) 9–18, https://doi.org/10.1016/J.FSIGEN.2007.07.002.
[7] S. Antão-Sousa, E. Conde-Sousa, L. Gusmão, A. Amorim, N. Pinto, Estimations of mutation rates depend on population allele frequency distribution: the case of autosomal microsatellites (Jul), Genes (Basel) 13 (7) (2022), https://doi.org/10.3390/genes13071248.
[8] M. Strand, T.A. Prolla, R.M. Liskay, T.D. Petes, Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair, vol. 365, no. 6443, pp, Nature 365 (6443) (1993) 274–276, https://doi.org/10.1038/365274a0.
[9] C. Schlötterer, D. Tautz, Slippage synthesis of simple sequence DNA (p.), Nucleic Acids Res. 20 (2) (1992) 211, https://doi.org/10.1093/NAR/20.2.211.
[10] X. Xu, M. Peng, Z. Fang, X. Xu, The direction of microsatellite mutations is dependent upon allele length (pp.), Nat. Genet. 24 (4) (2000) 396–399, https://doi.org/10.1038/74238.
[11] J.L. Weber, C. Wong, Mutation of human short tandem repeats (pp.), Hum. Mol. Genet. 2 (8) (1993) 1123–1128, https://doi.org/10.1093/HMG/2.8.1123.
[12] S. Antão-Sousa, N. Pinto, P. Rende, A. Amorim, L. Gusmão, The sequence of the repetitive motif influences the frequency of multistep mutations in short tandem repeats, vol. 13, no. 1, pp. 1–9, Jun. 2023, Sci. Rep. 13 (1) (2023), https://doi.org/10.1038/s41598-023-32137-y.
[13] M. Wierdl, M. Dominska, T.D. Petes, Microsatellite instability in yeast: dependence on the length of the microsatellite (pp.), Genetics 146 (3) (1997) 769–779, https://doi.org/10.1093/GENETICS/146.3.769.
[14] H. Ellegren, "Heterogeneous mutation processes in human microsatellite DNA sequences," 2000. [Online]. Available: http://www.ebc.uu.se/evbiol/index.shtml.
[15] S. Antão-Sousa, A. Amorim, L. Gusmão, N. Pinto, Mutation in Y STRs: repeat motif gains vs. losses (pp.), Forensic Sci. Int. Genet Suppl. Ser. 7 (1) (2019) 240–242, https://doi.org/10.1016/j.fsigss.2019.09.092.
[16] J.X. Sun, et al., A direct characterization of human mutation based on microsatellites (pp.), Nat. Genet. 44 (10) (2012) 1161–1165, https://doi.org/10.1038/NG.2398.
[17] K.N. Ballantyne, et al., Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications (p.), Am. J. Hum. Genet. 87 (3) (2010) 341, https://doi.org/10.1016/J.AJHG.2010.08.006.
[18] N. Pinto, et al., Paternal and maternal mutations in X-STRs: a GHEP-ISFG collaborative study, Forensic Sci. Int. Genet. 46 (2020), https://doi.org/10.1016/J.FSIGEN.2020.102258.
[19] S. Antão-Sousa, E. Conde-Sousa, L. Gusmão, A. Amorim, N. Pinto, How frequently are autosomal and X-STRs multistep mutations perceived as single-step? Forensic Sci. Int. Genet. Suppl. Ser. (2022) https://doi.org/10.1016/j.fsigss.2022.10.022.
[20] S. Antão-Sousa, E. Conde-Sousa, L. Gusmão, A. Amorim, N. Pinto, How often have X- and autosomal-STRs mutations equivocal parental origin been assigned?

Forensic Sci. Int. Genet Suppl. Ser. (2022) https://doi.org/10.1016/j.fsigss.2022.09.035.
[21] R. Lessig, J. Edelmann, Y chromosome polymorphisms and haplotypes in west Saxony (Germany) (pp.), Int. J. Leg. Med. 111 (4) (1998) 215–218, https://doi.org/10.1007/S004140050155.
[22] N.O. Blanchi, et al., Characterization of ancestral and derived Y-chromosome Haplotypes of new world native populations (pp.), Am. J. Hum. Genet. 63 (6) (1998) 1862–1871, https://doi.org/10.1086/302141.
[23] M. Kayser, et al., Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs (p.), Am. J. Hum. Genet. 66 (5) (2000) 1580, https://doi.org/10.1086/302905.
[24] B.M. Dupuy, et al., Y-chromosome variation in a Norwegian population sample (pp.), Forensic Sci. Int. 117 (3) (2001) 163–173, https://doi.org/10.1016/S0379-0738(00)00397-2.
[25] B.M. Dupuy, M. Stenersen, T. Egeland, B. Olaisen, Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci (pp.), Hum. Mutat. 23 (2) (2004) 117–124, https://doi.org/10.1002/HUMU.10294.
[26] L.C. Tsai, et al., Haplotype frequencies of nine Y-chromosome STR loci in the Taiwanese Han population (pp.), Int. J. Leg. Med. 116 (3) (2002) 179–183, https://doi.org/10.1007/S004140100236.
[27] R. Kurihara, et al., Mutations in 14 Y-STR loci among Japanese father-son haplotypes (pp.), Int. J. Leg. Med. 118 (3) (2004) 125–131, https://doi.org/10.1007/S00414-003-0422-3.
[28] B. Budowle, et al., Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America (pp.), Forensic Sci. Int. 150 (1) (2005) 1–15, https://doi.org/10.1016/J.FORSCIINT.2005.01.010.
[29] B. Berger, A. Lindinger, H. Niederstätter, P. Grubwieser, W. Parson, Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay (pp.), Int. J. Leg. Med. 119 (4) (2005) 241–246, https://doi.org/10.1007/S00414-005-0546-8.
[30] L. Gusmão, et al., Mutation rates at Y chromosome specific microsatellites (pp.), Hum. Mutat. 26 (6) (2005) 520–528, https://doi.org/10.1002/HUMU.20254.
[31] D.J. Ballard, et al., A study of mutation rates and the characterisation of intermediate, null and duplicated alleles for 13 Y chromosome STRs (pp.), Forensic Sci. Int. 155 (1) (2005) 65–70, https://doi.org/10.1016/J.FORSCIINT.2004.12.012.
[32] S. Turrina, R. Atzei, D. de Leo, Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay (pp.), Int. J. Leg. Med. 120 (1) (2006) 56–59, https://doi.org/10.1007/S00414-005-0054-X.
[33] H.Y. Lee, et al., Haplotypes and mutation analysis of 22 Y-chromosomal STRs in Korean father-son pairs (pp.), Int. J. Leg. Med. 121 (2) (2007) 128–135, https://doi.org/10.1007/S00414-006-0130-X.
[34] C. Hohoff, K. Dewa, U. Sibbing, K. Hoppe, P. Forster, B. Brinkmann, Y-chromosomal microsatellite mutation rates in a population sample from northwestern Germany (pp.), Int. J. Leg. Med. 121 (5) (2007) 359–363, https://doi.org/10.1007/S00414-006-0123-9.
[35] R. Kumagai, A. Kumagai, K. Saigusa, Y. Aoki, Haplotype analysis of 17 Y-STR loci in a Japanese population (pp.), Forensic Sci. Int. 172 (1) (2007) 72–78, https://doi.org/10.1016/J.FORSCIINT.2006.07.022.
[36] M.L. Pontes, L. Cainé, D. Abrantes, G. Lima, M.F. Pinheiro, Allele frequencies and population data for 17 Y-STR loci (AmpFlSTR Y-filer) in a Northern Portuguese population sample (pp.), Forensic Sci. Int. 170 (1) (2007) 62–67, https://doi.org/10.1016/J.FORSCIINT.2006.04.008.
[37] M. sen Shi, J.P. Tang, R.F. Bai, X.J. Yu, J.Y. Lv, B. Hu, Haplotypes of 20 Y-chromosomal STRs in a population sample from southeast China (Chaoshan area) (pp.), Int. J. Leg. Med. 121 (6) (2007) 455–462, https://doi.org/10.1007/S00414-007-0186-2.
[38] P.M. Domingues, L. Gusmão, D.A. da Silva, A. Amorim, R.W. Pereira, E.F. de Carvalho, Sub-Saharan Africa descendents in Rio de Janeiro (Brazil): population and mutational data for 12 Y-STR loci (pp.), Int. J. Leg. Med. 121 (3) (2007) 238–241, https://doi.org/10.1007/S00414-007-0154-X.
[39] A.E. Decker, M.C. Kline, J.W. Redman, T.M. Reid, J.M. Butler, Analysis of mutations in father-son pairs with 17 Y-STR loci, Forensic Sci. Int. Genet. 2 (3) (2008), https://doi.org/10.1016/J.FSIGEN.2007.08.016.
[40] P. Sánchez-Diz, et al., Population and segregation data on 17 Y-STRs: results of a GEP-ISFG collaborative study (pp.), Int. J. Leg. Med. 122 (6) (2008) 529–533, https://doi.org/10.1007/S00414-008-0265-Z.
[41] J.R. Padilla-Gutiérrez, et al., Population data and mutation rate of nine Y-STRs in a mestizo Mexican population from Guadalajara, Jalisco, México (pp.), Leg. Med. (Tokyo) 10 (6) (2008) 319–320, https://doi.org/10.1016/J.LEGALMED.2008.03.004.
[42] J.A. Soares-Vieira, A.E.C. Billerbeck, E.S.M. Iwamura, B.B. Mendonca, L. Gusmão, P.A. Otto, Population and mutation analysis of Y-STR loci in a sample from the city of São Paulo (Brazil) (pp.), Genet. Mol. Biol. 31 (3) (2008) 651–656, https://doi.org/10.1590/S1415-47572008000400009.
[43] M. Goedbloed, et al., Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR Yfiler PCR amplification kit (pp.), Int. J. Leg. Med. 123 (6) (2009) 471–482, https://doi.org/10.1007/S00414-009-0342-Y.
[44] J. Ge, B. Budowle, X.G. Aranda, J. v Planz, A.J. Eisenberg, R. Chakraborty, Mutation rates at Y chromosome short tandem repeats in Texas populations (pp), Forensic Sci. Int. Genet. 3 (3) (2009) 179–184, https://doi.org/10.1016/J.FSIGEN.2009.01.007.
[45] S.H. Kim, et al., Population genetics and mutational events at 6 Y-STRs in Korean population, Forensic Sci. Int. Genet. 3 (2) (2009), https://doi.org/10.1016/J.FSIGEN.2008.06.004.
[46] C. Vieira-Silva, P. Dario, T. Ribeiro, I. Lucas, H. Geada, R. Espinheira, Y-STR mutational rates determination in South Portugal Caucasian population (pp.), Forensic Sci. Int. Genet. Suppl. Ser. 2 (1) (2009) 60–61, https://doi.org/10.1016/J.FSIGSS.2009.08.159.
[47] A. Laouina, et al., Mutation rate at 17 Y-STR loci in 'Father/Son' pairs from moroccan population (pp.), Leg. Med. 15 (5) (2013) 269–271, https://doi.org/10.1016/J.LEGALMED.2013.03.003.
[48] K.N. Ballantyne, et al., Toward male individualization with rapidly mutating y-chromosomal short tandem repeats (pp), Hum. Mutat. 35 (8) (2014) 1021–1032, https://doi.org/10.1002/HUMU.22599.
[49] C. Robino, et al., Development of an Italian RM Y-STR haplotype database: results of the 2013 GEFI collaborative exercise (pp), Forensic Sci. Int. Genet. 15 (2015) 56–63, https://doi.org/10.1016/J.FSIGEN.2014.10.008.
[50] Y. Wang, et al., Genetic polymorphisms and mutation rates of 27 Y-chromosomal STRs in a Han population from Guangdong Province, Southern China (pp), Forensic Sci. Int. Genet. 21 (2016) 5–9, https://doi.org/10.1016/J.FSIGEN.2015.09.013.
[51] S. Antão-Sousa, et al., Mutation rates and segregation data on 16 Y-STRs: an update to previous GHEP-ISFG studies (pp), Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) e601–e602, https://doi.org/10.1016/J.FSIGSS.2017.10.008.
[52] A. Adnan, A. Rakha, O. Lao, M. Kayser, Mutation analysis at 17 Y-STR loci (Yfiler) in father-son pairs of male pedigrees from Pakistan (pp.), Forensic Sci. Int. Genet. 36 (2018) e17–e18, https://doi.org/10.1016/J.FSIGEN.2018.07.001.
[53] F.C. Bugoye, E. Mulima, G. Misinzo, Analysis of mutation rate of 17 Y-chromosome short tandem repeats loci using Tanzanian father-son paired samples (vol), Genet. Res. Int. 2018 (2018), https://doi.org/10.1155/2018/8090469.
[54] E. Mertoglu, G. Filoglu, T. Zorlu, O. Bulbul, Estimation of the Y-chromosomal short tandem repeat (Y-STR) mutation rates in Turkey, Turk. J. Biochem. 43 (2) (2018) 142–150, https://doi.org/10.1515/TJB-2017-0240/PDF.
[55] Y. Yang, et al., Haplotypic polymorphisms and mutation rate estimates of 22 Y-chromosome STRs in the Northern Chinese Han father–son pairs, vol. 8, no. 1, pp, Sci. Rep. 8 (1) (2018) 1–6, https://doi.org/10.1038/s41598-018-25362-3.
[56] W. Wu, et al., Mutation rates at 42 Y chromosomal short tandem repeats in Chinese Han population in Eastern China (pp.), Int. J. Leg. Med. 132 (5) (2018) 1317–1319, https://doi.org/10.1007/S00414-018-1784-X.
[57] V. Petrovic, M. Kecmanovic, M. Keckarevic Markovic, D. Keckarevic, Assessment of mutation rates for PPY23 Y chromosome STR loci in Serbian father-son pairs (pp.), Forensic Sci. Int. Genet. 39 (2019) e5–e9, https://doi.org/10.1016/J.FSIGEN.2018.11.014.
[58] J. Zhang, et al., Mutation rates in father-son pairs of the 27 Y-STR loci in the Dezhou Han population from Shandong province, eastern China (pp), J. Forensic Leg. Med. 67 (2019) 61–63, https://doi.org/10.1016/J.JFLM.2019.06.015.
[59] M. Ay, A. Serin, H. Sevay, C. Gurkan, and H. Canan, "Genetic characterisation of 13 rapidly mutating Y-STR loci in 100 father and son pairs from South and East Turkey," https://doi.org/10.1080/03014460.2018.1559353, vol. 45, no. 6–8, pp. 506–515, Nov. 2019, doi: 10.1080/03014460.2018.1559353.
[60] I.B. Ambrosio, et al., Mutational data and population profiling of 23 Y-STRs in three Brazilian populations, Forensic Sci. Int. Genet. 48 (2020), https://doi.org/10.1016/J.FSIGEN.2020.102348.
[61] H. Lin, Q. Ye, P. Tang, T. Mo, X. Yu, J. Tang, Analyzing genetic polymorphism and mutation of 44 Y-STRs in a Chinese Han population of Southern China, Leg. Med. (Tokyo) 42 (2020), https://doi.org/10.1016/J.LEGALMED.2019.101643.
[62] J. Fu, J. Cheng, C. Wei, M.A. Khan, Z. Jin, J. Fu, Assessing 23 Y-STR loci mutation rates in Chinese Han father-son pairs from southwestern China (pp.), Mol. Biol. Rep. 47 (10) (2020) 7755–7760, https://doi.org/10.1007/S11033-020-05851-3.
[63] S. Bredemeyer, L. Roewer, S. Willuweit, Next generation sequencing of Y-STRs in father-son pairs and comparison with traditional capillary electrophoresis, Forensic Sci. Res. (2021), https://doi.org/10.1080/20961790.2021.1898078/SUPPL_FILE/TFSR_A_1898078_SM7293.XLSX.
[64] L. Yuan, et al., Mutation analysis of 13 RM Y-STR loci in Han population from Beijing of China (pp.), Int. J. Leg. Med. 133 (1) (2019) 59–63, https://doi.org/10.1007/S00414-018-1949-7.
[65] "RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/."
[66] Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2018). dplyr: A Grammar of Data Manipulation. R package version 0.7.6. https://CRAN.R-project.org/package=dplyr."
[67] S. Claerhout, et al., Determining Y-STR mutation rates in deep-routing genealogies: Identification of haplogroup differences (pp.), Forensic Sci. Int. Genet. 34 (2018) 1–10, https://doi.org/10.1016/J.FSIGEN.2018.01.005.
[68] A. Ralf, et al., Identification and characterization of novel rapidly mutating Y-chromosomal short tandem repeat markers (pp.), Hum. Mutat. 41 (9) (2020) 1680–1696, https://doi.org/10.1002/humu.24068.
[69] A. Jochens, A. Caliebe, U. Rösler, M. Krawczak, Empirical evaluation reveals best fit of a logistic mutation model for human Y-chromosomal microsatellites (pp.), Genetics 189 (4) (2011) 1403–1411, https://doi.org/10.1534/GENETICS.111.132308.