



## Mutation in Y STRs: Repeat motif gains vs. losses

S. Antão-Sousa<sup>a,b,c,d,\*</sup>, A. Amorim<sup>a,b,c</sup>, L. Gusmão<sup>d</sup>, N. Pinto<sup>b,c,e</sup>

<sup>a</sup> Faculty of Sciences of the University of Porto (FCUP), Portugal

<sup>b</sup> Institute of Pathology and Molecular Immunology from University of Porto (IPATIMUP), Portugal

<sup>c</sup> Institute of Innovation and Research in Health (i3S), University of Porto, Portugal

<sup>d</sup> DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Brazil

<sup>e</sup> Center of Mathematics of the University of Porto (CMUP), Portugal

### ARTICLE INFO

#### Keywords:

Y-STRs  
Mutation's direction  
Repeat motif  
S  
single-step mutation  
Multistep mutation

### ABSTRACT

Y chromosome specific short tandem repeats (Y-STRs) are widely used in population genetics and forensics. Since these markers do not recombine, mutation is the only source of diversity. The primary mutational mechanism leading to length changes in STRs is thought to be polymerase template slippage, and the most common change is the gain or the loss of one repeat motif. In this work, we aim to study 19 Y-STR alleles' contraction and expansion. Alleles were grouped into tertiles: short (1<sup>st</sup> tertile), intermediate (2<sup>nd</sup> tertile) and long alleles (3<sup>rd</sup> tertile). Significant differences between repeat gains and losses were found at four markers - DYS19, DYS439 for intermediate alleles, and DYS570 and DYS626 for long alleles. When the average number is computed for the pooled loci, for short alleles, the number of repeat motif gains is higher than of repeat losses, and the opposite happens for long alleles. For intermediate alleles, the proportion between the number of repeat gains and losses is close to one. Generally, the rate of expansion decreases from the first tertile to the third, and conversely, the rate of contraction increases from the first tertile to the third. The pooled loci tertiles' mutation rate increases from short to long alleles. Our results demonstrate that the mutation direction and rate depend on alleles' length. The longer the allele the greater the mutation and contraction rates.

### 1. Introduction

Short tandem repeat (STR) loci, also known as microsatellites, are tandemly arrayed repeats of DNA fragments of 1–6 base pairs per repeat motif. STRs have been widely used in population and forensic genetics research. The increasing application of these markers requires the study of mutation rates and mechanisms, as this knowledge is an essential prerequisite for the accurate interpretation of experimental data.

The non-recombining region of the Y chromosome acts as a unique tool for forensic investigations as it is inherited through the patrilineal line without recombination. Thus, this haploid system allows the unambiguous detection of any mutation through the identification of both paternal and filial alleles. Other genetic systems, such as the X-chromosomal (haplodiploid) or autosomal (diploid), always entail uncertainty concerning the identification of either the parental or the mutated allele or both [1–3].

The primary mutational mechanism leading to changes in microsatellite length is thought to be polymerase template slippage [4,5] and most of the observed changes in length are due to the gain or loss of a single repeat [6,7]. The stepwise mutation model (SMM) [8] has been

widely used to model STR evolution [9–12]. According to this model, the length of a microsatellite varies at a fixed rate independent of length and with the same probability of expansion and contraction. Xu et al. [7] have indeed observed that the rate of expansion mutations is independent of allele length, but several studies [e.g. 7,13,14] report the increase of the overall mutation rate with allele length.

### 2. Material and methods

We have analyzed 104,083 allele transfers between father-son duos at 19 Y-STR loci.

For each marker, alleles were clustered in tertiles (short, intermediate, and long) considering the number of repeats. The proportion of gains and losses of repeats and the expansion and contraction rates were computed for each tertile for each marker (see Table 1). Overall tertile number of expansions/contractions and mutation rates were also computed.

\* Corresponding author at: Faculty of Sciences of the University of Porto (FCUP), Portugal.

E-mail address: [sofias@ipatimup.pt](mailto:sofias@ipatimup.pt) (S. Antão-Sousa).

<https://doi.org/10.1016/j.fsigss.2019.09.092>

Received 6 September 2019; Accepted 26 September 2019

Available online 27 September 2019

1875-1768/ © 2019 Elsevier B.V. All rights reserved.

**Table 1**

Allelic transfers in father-son duos; repeat number gains and losses, and rates of expansion/contraction, per marker and per allelic length tertile.

Marker	N	1st tertile		2nd tertile		3rd tertile		TOTAL	1st tertile		2nd tertile		3rd tertile	
		No. of repeat gains	No. of repeat losses	No. of repeat gains	No. of repeat losses	No. of repeat gains	No. of repeat losses		Rate of expansion	Rate of contraction	Rate of expansion	Rate of contraction	Rate of expansion	Rate of contraction
DYS19	9,074	2	0	13 <sup>a</sup>	2 <sup>a</sup>	1	5	23	1.00	0.00	0.87	0.13	0.17	0.83
DYS391	9,108	0	0	13	9	0	2	24			0.59	0.41	0.00	1.00
DYS389I	8,084	0	0	10	10	0	2	22			0.50	0.50	0.00	1.00
DYS460	2,002	1	0	3	4	0	2	10	1.00	0.00	0.43	0.57	0.00	1.00
DYS533	1,085	0	0	1	1	0	1	3			0.50	0.50	0.00	1.00
GATA H4	7,128	0	0	4	7	0	2	13			0.36	0.64	0.00	1.00
DYS439	7,355	2	0	10 <sup>a</sup>	1 <sup>a</sup>	6	14	33	1.00	0.00	0.91	0.09	0.30	0.70
DYS549	104	0	0	0	1	0	0	1			0.00	1.00		
DYS393	7,861	0	0	5	2	1	3	11			0.71	0.29	0.25	0.75
GATA A10	874	1	1	0	2	0	0	4	0.50	0.50	0.00	1.00		
DYS456	6,036	1	0	11	8	0	2	22	1.00	0.00	0.58	0.42	0.00	1.00
DYS437	7,333	2	0	2	4	1	1	10	1.00	0.00	0.33	0.67	0.50	0.50
DYS461	873	0	0	0	0	0	0	0						
DYS385	15,341	8	2	11	5	0	1	27	0.80	0.20	0.69	0.31	0.00	1.00
DYS458	6,043	3	3	16	15	6	2	45	0.50	0.50	0.52	0.48	0.75	0.25
DYS570	4,293	2	0	16	20	0 <sup>a</sup>	5 <sup>a</sup>	43	1.00	0.00	0.44	0.56	0.00	1.00
DYS576	4,194	1	0	31	24	3	6	65	1.00	0.00	0.56	0.44	0.33	0.67
DYS626	3,138	0	2	9	11	2 <sup>a</sup>	10 <sup>a</sup>	34	0.00	1.00	0.45	0.55	0.17	0.83
DYS627	4,157	0	0	21	27	6	12	66			0.44	0.56	0.33	0.67

<sup>a</sup> Significant differences between gains and losses in the tertile.

### 3. Results

Statistically significant differences between gains and losses were observed in the first tertile and in the third only at DYS570 and DYS626 markers, while in the second tertile differences were verified at DYS19 and DYS439 loci.

Xu et al. [4] observed an increase in the rate of contraction with the increase in allele length, the rate of expansion remaining constant. Our observations support these findings for the rate of contraction, with the exception of markers DYS437, DYS458 and DYS626. However, our data do not support the conclusions concerning the rate of expansion, as it decreases with allele length for 16 out of the 19 markers (excluding DYS437, DYS458 and DYS626) (see Table 1).

Several studies [7,13,14] report that overall mutation rate increase with allele length. Indeed, when the overall mutation rate per tertile was computed, an increase from short (0.00150) to intermediate (0.00531) and long alleles (0.01468) was found.

### 4. Discussion and conclusions

Only DYS570 and DYS626 markers have shown statistically different number of losses over gains (in the third tertile). Markers DYS19 and DYS439 have shown significantly different number of gains over losses (in the second tertile). No other markers have shown significant differences between gains and losses. When the average number of repeat motif gains and losses for the tertile-pooled loci is computed, for short alleles, the average number of repeat motif gains is higher than the average number of repeat losses ( $p = 0.000072$ ), and the opposite happens for long alleles ( $p = 0.000144$ ). For intermediate alleles, the proportion between the number of repeat gains and losses is close to one ( $p = 0.624137$ ).

Generally, the rate of expansion decreases from the first tertile (0.80) to the second (0.49) and from the second to the third (0.18). Conversely, the rate of contraction increases from the first tertile (0.20) to the second (0.51) and from the second to the third (0.83).

Our results indicate that the proportion of repeat gains and losses varies with allele length. Namely, shorter alleles tend to expand, longer alleles tend to contract, and intermediate alleles have approximately equal tendency to contract or expand.

Furthermore, mutation rate increases as the alleles' length increases.

The highest mutation rate was found for long alleles, followed by intermediate alleles and, finally, short alleles.

Our results support the observations that the mutation direction and rate depend on alleles' length. Long alleles have higher mutation and contraction rates and short alleles have lower mutation and higher expansion rates.

### Declaration of Competing Interest

The authors declare no conflict of interests.

### Acknowledgements

This work was partially financed by FEDER—Fundo Europeu de Desenvolvimento Regional - funds through the COMPETE 2020—Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020; Portuguese funds through FCT—Fundação para a Ciência e a Tecnologia/Ministério da Ciência, Tecnologia e Inovação in the framework of the Decreto-Lei nº 57/2016 de 29 de Agosto (NP work contract) and through the doctoral grant SFRH/BD/136284/2018; and the projects “Institute for Research and Innovation in Health Sciences” (POCI-01-0145-FEDER-007274) and “Center of Mathematics of the University of Porto” (UID/MAT/00144/2019).

LG (ref. 305330/2016-0) was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ (CNE-2018).

### References

- [1] N. Pinto, C. Alves, L. Gusmão, A. Amorim, Theory and statistics of mutation rates: a mathematical framework reformulation for forensic applications, *Forensic Sci. Int. Genet. Suppl. Ser.* 5 (2015) e131–e132.
- [2] N. Pinto, L. Gusmão, A. Amorim, Mutation and mutation rates at Y chromosome specific Short Tandem Repeat Polymorphisms (STRs): a reappraisal, *Forensic Sci. Int. Genet.* 9 (2014) 20–24.
- [3] K. Slooten, F. Ricciardi, Estimation of mutation probabilities for autosomal STR markers, *Forensic Sci. Int. Genet.* 7 (3) (2013) 337–344.
- [4] C. Schlötterer, D. Tautz, Slippage synthesis of simple sequence DNA, *Nucleic Acids Res.* 20 (2) (1992) 211–215.
- [5] 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, *Nature* 365 (6443) (1993) 274.

- [6] J.L. Weber, C. Wong, Mutation of human short tandem repeats, *Hum. Mol. Genet.* 2 (8) (1993) 1123–1128.
- [7] X. Xu, M. Peng, Z. Fang, X. Xu, The direction of microsatellite mutations is dependent upon allele length, *Nat. Genet.* 24 (4) (2000) 396.
- [8] T. Ohta, M. Kimura, A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population, *Genet. Res.* 22 (2) (1973) 201–204.
- [9] A. Di Rienzo, A.C. Peterson, J.C. Garza, A.M. Valdes, M. Slatkin, N.B. Freimer, Mutational processes of simple-sequence repeat loci in human populations, *Proc. Natl. Acad. Sci.* 91 (8) (1994) 3166–3170.
- [10] M. Kimmel, R. Chakraborty, Measures of variation at DNA repeat loci under a general stepwise mutation model, *Theor. Popul. Biol.* 50 (3) (1996) 345–367.
- [11] A.M. Valdes, M. Slatkin, N.B. Freimer, Allele frequencies at microsatellite loci: the stepwise mutation model revisited, *Genetics* 133 (3) (1993) 737–749.
- [12] L.A. Zhivotovsky, M.W. Feldman, Microsatellite variability and genetic distances, *Proc. Natl. Acad. Sci.* 92 (25) (1995) 11549–11552.
- [13] B. Brinkmann, M. Klitsch, F. Neuhuber, J. Hühne, B. Rolf, Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat, *Am. J. Hum. Genet.* 62 (6) (1998) 1408–1415.
- [14] C.R. Primmer, H. Ellegren, N. Saino, A.P. Møller, Directional evolution in germline microsatellite mutations, *Nat. Genet.* 13 (4) (1996) 391.