Int J Legal Med (2003) 117 : 287–290
DOI 10.1007/s00414-003-0388-1

ORIGINAL ARTICLE

**Leonor Gusmão · Michael Krawczak
Paula Sánchez-Diz · Cíntia Alves · Alexandra Lopes
Sandra Beleza · Angel Carracedo · António Amorim**

Bimodal allele frequency distribution at Y-STR loci DYS392 and DYS438: no evidence for a deviation from the stepwise mutation model

Received: 5 March 2003 / Accepted: 19 May 2003 / Published online: 12 July 2003

Abstract A deviation from the stepwise mutation model (SMM) has been suggested for the trinucleotide Y-STR locus DYS392, based upon its bimodal allele frequency distribution in various populations. The same type of distribution is also observed for the pentanucleotide Y-STR DYS438. In order to verify whether a departure from an SMM is likely for these two loci, we studied a large number of Portuguese male DNA samples typed for the two loci and in addition, for the Y-STR loci DYS19, DYS389I/II, DYS390, DYS391 and DYS393. The compatibility of the observed allele frequency spectrum with an SMM was assessed by an apportionment of the molecular variance among, and consideration of the molecular distances between, haplotype groups defined according to their allelic state at each of the two markers of interest. For haplotypes carrying either modal alleles 11 or 13 of DYS392, 18.6% of the molecular variance of the remaining Y-STR background could be attributed to variation between the two groups. When all pairwise Φ_{st} values between haplotype groups were compared, group 12 was found to be closer to 11 than to 13, and group 14 was much closer to 13 than to 12 and 11. It may therefore be concluded that DYS392 allele 13 represents an evolutionary lineage with little or no relationship to 11 and 12. Furthermore, allele 14 is a one-step neighbour of 13 and is therefore likely to represent an offshoot from group 13. For haplotypes carrying either modal allele 10 or modal allele 12 of DYS438, 27.7% of the molecular variance of the Y-STR background was found to be due to variation between the two groups. Comparison of the other pairwise Φ_{st} values indicated that group 10 was closer to 9 and 11 than to 12, and that group 12 was closer to 11 and 13 than to 10. The lineages defined by the two modal alleles of DYS438 therefore also seem to be phylogenetically distant. When the two loci were analysed in combination, using the standardised linkage disequilibrium measure (D'), a strong association was noted between alleles DYS392*11 and DYS438*10 (D'=0.70) and between DYS392*13 and DYS438*12 (D'=0.72). Taken together, these results show that the bimodal allele frequency distributions of DYS392 and DYS438 are explicable in terms of (probably the same) historical and demographic causes, rather than a mutational mechanism other than SMM. The loci do therefore not appear to warrant any special attention when applied in population genetic or forensic studies.

Keywords Y-STR · DYS392 · DYS438 · Mutation models

L. Gusmão · C. Alves · A. Lopes · S. Beleza · A. Amorim (✉)
IPATIMUP, Instituto de Patologia e Imunologia Molecular
da Universidade do Porto,
Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal
Tel.: +351-22-5570700, Fax: +351-22-5570799,
e-mail: aamorim@ipatimup.pt

M. Krawczak
Institut für Medizinische Informatik und Statistik,
Christian-Albrechts-Universität Kiel, Germany

P. Sánchez-Diz · S. Beleza · A. Carracedo
Institute of Legal Medicine,
University of Santiago de Compostela, Spain

A. Amorim
Faculdade de Ciências, Universidade do Porto, Portugal

Introduction

Short tandem repeat (STR) loci have been widely used in population genetic and forensic research. In both instances, knowledge of the mutation rates and mechanisms involved is an essential prerequisite for the sensible interpretation of experimental data. Most of the studies undertaken so far have revealed mutation rates for commonly used Y-chromosomal STRs (Y-STR) that are similar to those of their autosomal counterparts (Heyer et al. 1997; Kayser et al. 2000; Holtkemper et al. 2001; Kayser and Sajantila 2001). Like with autosomal STRs, it is also generally accepted that a stepwise mutation model (SMM) explains the observed allelic variation of Y-STRs sufficiently well. For some loci, however, including the trinucleotide Y-STRs DYS392 and DYS388, it has been suggested that the bimodal allele frequency distribution observed in some pop-

ulations is not consistent with an SMM (Thomas et al. 2000; Forster et al. 2000; Nebel et al. 2001), despite the notion that DYS392 has a unimodal allele frequency distribution within Y-SNP (single nucleotide polymorphism) haplogroups (de Knijff 2000). Furthermore, the heterogeneity of DYS388 and DYS392 allele frequencies within and between SNP haplogroups has been held as evidence in favour of a haplogroup-specific deviation from the SMM (Thomas et al. 2000; Nebel et al. 2001), although genealogy-based explanations have not been ruled out (Thomas et al. 2000).

In European populations, DYS392 allele frequencies undoubtedly have a bimodal distribution (e.g. Gusmão et al. 2002) whilst a unimodal distribution has been observed in Asia and Africa, and the locus is almost monomorphic in sub-Saharan populations (Kayser et al. 2001). A similar pattern emerges for the pentanucleotide Y-STR DYS438 which has a bimodal allele frequency distribution in northern Portugal, but not in Mozambique (SE Africa) or in Asians from Macao (Gusmão et al. 2001; Uchihi et al. 2003).

In an attempt to further clarify this matter, we sought independent evidence for or against the supposition of de Knijff (2000) that founder haplotypes differing by two repeat units have coexisted in several populations for a period of time too short to generate intermediate allelic states of sufficiently high frequency. To this end, Y-STR haplotype data from a large number of Portuguese males were analysed with respect to their molecular variance between groups of haplotypes, defined according to the allelic state at each of the two loci in question. The rationale underlying this strategy is the assumption that the distribution of haplotype frequencies within Y-STR-defined haplotype groups should reflect more of the recent population history than SNP-based haplogroups, and that the former grouping should therefore discriminate better between mutational and demographic events than the latter. In order to distinguish further between the effects of mutational mechanism and population history, we also assessed formally the allelic association of DYS392 and DYS438 and analysed the allele frequency distribution of each locus within the allelic classes defined by the other locus.

Materials and methods

A sample of 479 unrelated Portuguese males was typed for 7 Y-STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393) as described by Gusmão et al. (2002). These samples have been logged in the Y-STR Haplotype Reference Database (Roewer et al. 2001) where they were assigned to one of three geographical regions, namely northern Portugal (182 samples from north of the Douro river), central Portugal (185 samples from between the Douro and Tejo rivers) or southern Portugal (112 samples from south of the Tejo river). Portugal is divided into 18 administrative districts, and between 20 and 50 unrelated samples have been included from each district. A subset of 374 samples was also typed for DYS438 as described by Gusmão et al. (2002).

Analysis of molecular variance (AMOVA) was performed for DYS19, DYS389I/II, DYS390, DYS391 and DYS393, with haplotypes classified according to the allelic state of either DYS392 or DYS438, respectively. Locus DYS392 was also included in the AMOVA upon stratification by DYS438. The AMOVA results were summarised in the form of global/pairwise Φ_{st} values and assessed for statistical significance using a Monte-Carlo test as implemented in the Arlequin software (Schneider et al. 2000). Arlequin was also used to measure the degree of association between alleles at DYS392 and DYS438 in the form of standardised linkage disequilibrium (D') values.

Results and discussion

The bimodal allele frequency distribution observed for some STRs in some populations has been held as evidence against the general validity of the SMM. Thus, Thomas et al. (2002) claimed that the distribution of DYS388 alleles across haplogroups indicates that the locus does not consistently mutate in a stepwise fashion over the entire range of repeat lengths. A similar putative deviation from the SMM was reported for DYS392 by Forster et al. (2000) and Nebel et al. (2001), who also suggested that, since DYS392 has a bimodal allele frequency distribution in two of three haplogroups, the deviation from SMM may be haplogroup-specific. However, when haplogroups were further refined by the inclusion of additional SNPs, DYS392 alleles segregated almost completely between haplogroups. Furthermore, de Knijff (2000) had shown previously that the overall bimodal distribution of DYS392 resulted from a strong association between its two modal alleles, 11 and 13, and two different but frequent haplogroups. The distribution inside each group was unimodal. From the data published by Weale et al. (2001), an association can be inferred between DYS392 allele 11 and haplogroups 2, 3 and 21, and between allele 13 and haplogroups 1, 26 and 28.

We tested the compatibility of the SMM with the bimodal allele frequency distributions of DYS392 and DYS438 by a different approach, including (i) an analysis of molecular variance (AMOVA) of Y-STR haplotype groups defined by the allelic state at either of the two loci in question, and (ii) the consideration of the allele frequency distribution of one locus in sub-samples carrying different alleles of the other locus.

Analysis of molecular variance (AMOVA) of haplotype groups

Haplotypes were first grouped according to their allelic state at DYS392, resulting in a total of four groups ("11" – "14"). When only modal groups 11 and 13 were considered, AMOVA of the 6 other Y-STRs (DYS19, DYS389I/II, DYS390, DYS391, DYS393) revealed that 18.6% of their molecular variation was due to variation *between* groups (pairwise Φ_{st}=0.186). A similarly high degree of differentiation became apparent in an AMOVA of all four groups combined (global Φ_{st}=0.169). Haplotype grouping according to DYS438 resulted in five groups ("9" – "13"). For modal groups 10 and 12, AMOVA of DYS19, DYS389I/II, DYS390, DYS391, DYS392 and DYS393 revealed that 27.7% of the molecular variation was due to variation *between* groups (pairwise Φ_{st}=0.277). Again, joint AMOVA

Table 1 Pairwise Φ_{st} values between haplotype groups of Y-STR loci DYS19 DYS389I/II, DYS390, DYS391 and DYS393, defined by their DYS392 allelic state

Haplotype group	11	12	13
12	0.032		
13	0.186	0.230	
14	0.122	0.163	0.006

Table 2 Pairwise Φ_{st} values between haplotype groups of Y-STR loci DYS19, DYS389I/II, DYS390, DYS391, DYS392 and DYS393, defined by their DYS438 allelic state

Haplotype group	9	10	11	12
10	0.053			
11	0.116	0.071		
12	0.258	0.277	0.218	
13	0.174	0.211	0.155	–0.006

of all five groups combined indicated a high degree of overall differentiation between them (global Φ_{st}=0.223). The observed Φ_{st} values, which were statistically significant at the 0.1% level, appear exceptionally high since AMOVA of the same Y-STRs, grouped according to geographical sample origin in either northern, central or southern Portugal, left 99.96% of the variation attributable to variation *within* populations (global Φ_{st}=0.0004). Therefore, DYS392 alleles 11 and 13 and DYS438 alleles 10 and 12 are likely to represent evolutionary very distant Y chromosomal lineages.

Pairwise Φ_{st} values were next calculated in order to evaluate in more detail the molecular distance between the modal and the other haplotype groups, as defined by the DYS392 or DYS438 allelic state, respectively. The results revealed that for DYS392, group 12 is closer to group 11 than to 13, and group 14 is much closer to 13 than to 12 or 11 (Table 1). It may therefore be concluded (i) that group 13 represents a distinct lineage with little or no evolutionary relationship to both 11 or 12, and (ii) that group 14, a one-step neighbour of 13, is likely to represent an offshoot from group 13. For DYS438, group 11 is closer to 10 than to 9 or 12, and group 13 is much closer to 12 than to 11 (Table 2). Here, group 13 is most probably a recent offshoot from its one-step neighbour group 12.

DYS392 and DYS438 allele distributions

The trinucleotide repeat structure of DYS392 has been used as an additional argument in favour of a putatively abnormal mutational mechanism (Nebel et al. 2001). However, DYS438 is a pentanucleotide repeat yet shows a similar bimodal allele frequency distribution as DYS392. In order to seek additional evidence for population rather than mutational events underlying this phenomenon, we assessed the joint haplotype distribution of the two markers (Table 3). If two-step mutations were indeed common at one of the two loci, then a bimodal allele frequency distribution should have become apparent inside at least some of the allelic classes defined by the other locus. If the observed bimodality resulted mainly from population history, such a coincidence would have been unlikely. As was to be expected for non-recombining markers, the two loci were highly associated. Almost half of the haplotypes (173=46%) carried allele combination DYS392*13/DYS438*12 (D'= 0.72) and some 23% carried DYS392*11/DYS438*10 (D'=0.70). Nevertheless, none of the allelic sub classes of either locus showed any signs of bimodality, a finding which again favours demographic over mutational events as the major factor determining the observed allele frequency distributions.

Table 3 Joint haplotype distribution of DYS392 and DYS438

Allele DYS392	DYS438					
	9	10	11	12	13	Total
10	0	1	0	0	0	1
11	29	86	8	4	0	127
12	1	16	1	3	0	21
13	9	3	4	173	10	199
14	3	0	0	20	0	23
15	0	1	0	2	0	3
Total	42	107	13	202	10	374

Conclusions

We have shown that the bimodal allele frequency distribution of Y-STRs DYS392 and DYS438 is not incompatible with an SMM. On the contrary, AMOVA as well as an in-depth analysis of the highly associated allelic spectrum of the two markers support the hypothesis that the allele frequency distributions reflect demographic events rather than alternative mutation mechanisms. In contrast with previous claims (Nebel et al. 2001) of the opposite, but in agreement with an earlier SNP haplogroup analysis of DYS392 (de Knijff 2000), the two microsatellites do not therefore appear to warrant any special attention when applied in population genetic or forensic studies.

Acknowledgements This work was partially supported by Fundação para a Ciência e a Tecnologia (through grants SFRH/BD/860/2000 and SFRH/BD/7006/2001, and POCTI, Programa Operacional Ciência, Tecnologia e Inovação).

References

Forster P, Rohl A, Lunnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B. (2000) A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet 67: 182–196

Gusmão L, Alves C, Amorim A (2001) Molecular characterisation of four Y-specific microsatellites (DYS434, DYS437, DYS438, DYS439) for population and forensic studies. Ann Hum Genet 65:285–291

Gusmão L, Alves C, Beleza S, Amorim A (2002) Forensic evaluation and population data on the new Y-STRs DYS434, DYS437, DYS438, DYS439 and GATA A10. Int J Legal Med 116:139–147

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

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

Kayser M, Sajantila A (2001) Mutation at Y-STR loci: implications for paternity testing and forensic analysis. Forensic Sci Int 118:116–121

Kayser M, Roewer L, Hedman M et al. (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet 66:1580–1588

Kayser M, Krawczak M, Excoffier L et al. (2001) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet 68:990–1018

Knijff P de (2000) Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am J Hum Genet 67:1055–1061

Nebel A, Filon D, Hohoff C, Faerman M, Brinkmann B, Oppenheim A (2001) Haplogroup-specific deviation from the stepwise mutation model at the microsatellite loci DYS388 and DYS392. Eur J Hum Genet 9:22–26

Roewer L, Krawczak M, Willuweit S et al. (2001) Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. Forensic Sci Int 118:106–113

Schneider S, Roessli D, Excoffier L (2000) Arlequin ver. 2.000. A software for population genetics data analysis. University of Geneva

Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, Roux M le, Bradman N, Goldstein DB (2000) Y Chromosomes travelling south: the Cohen modal haplotype and the origins of the Lemba – the "Black Jews of Southern Africa". Am J Hum Genet 66:674–686

Uchihi R, Yamamoto T, Usuda K et al. (2003) Haplotype analysis with 14 Y-STR loci using 2 multiplex amplification and typing systems in 2 regional populations in Japan. Int J Legal Med 117: 34–38

Weale ME, Yepiskoposyan L, Jager RF, Hovhannisyan N, Khudoyan A, Burbage-Hall O, Bradman N, Thomas MG (2001) Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. Hum Genet 109: 659–674