Y-STRS

Y Chromosome-Specific STRs

By Leonor Gusmão[1] and Angel Carracedo[2]
[1]Instituto de Patologia e Imunologia, Molecular da Universidade do Porto, Porto, Portugal and [2]Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain

Y CHROMOSOME-SPECIFIC MARKERS

The Y chromosome is one of the smallest human chromosomes with an average size of 60 million base pairs (Mb). Between X and Y chromosomes, exchange is limited to small pseudoautosomal regions (PAR) of the X-Y pair (see Figure 1). During male meiosis, recombination only takes place at the most distal short arm (PAR 1) and at the tip of the Y chromosome long arm (PAR 2). For most of its length (the nonrecombining portion or NRY) the Y chromosome is male-specific and effectively haploid and is transmitted from father to son unchanged unless a mutational event takes place.

In the last few years, information on the paternally inherited NRY has been extensively applied in population genetics and evolution studies to track male-specific movements and admixture as well as mating behavior. The NRY contains different kinds of polymorphisms with different mutation rates, and consequently, scientists can select appropriate Y polymorphisms to study evolutionary events over different time scales. Short tandem repeats (STRs) on the Y chromosome can be used for analyses on a short evolutionary time scale or at the microgeographic level (1–3).

Information on the paternally inherited nonrecombining region of the Y chromosome has been extensively applied in population genetics and evolution studies to track male-specific movements and admixture as well as mating behavior.

Y CHROMOSOME-SPECIFIC STRS

Y chromosome-specific STRs have proved to be an important tool in paternity cases, especially when the alleged father is deceased. It is possible to determine his complete Y chromosome haplotype by analyzing any relative male in the patrilineage. However, a result based exclusively on Y-STRs does not exclude as the father any relative male in the same patrilineage, and the applicability of these markers is limited to approximately one half of paternity cases, those in which the paternity of a male descendant is in question.

Y-STRs are also useful for analysis of stains in forensic investigations when a male suspect is involved, as is the case in most violent crimes, including sexual offenses. Mixtures of body fluids from different individuals are frequent, and Y chromosome analysis allows detection of a male DNA fraction in stains involving male/female mixtures and the direct determination of the Y haplotype without differential extraction. This is advantageous in cases of rape committed by azoospermic individuals, which represent approximately 1–2% of all current rape cases, a percentage that will increase as vasectomy becomes more frequent.

Y-STRs are the most used Y chromosome markers in the forensic field due to their typing simplicity and high level of diversity. STR typing involves simple and reliable polymerase chain reaction (PCR)[(a)] techniques and is tolerant of very degraded samples. Of all Y chromosome polymorphic STRs described to date, DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 and YCAII have more data accumulated, being the most used in population and forensic genetics. Because of collaborative efforts to construct large databases (see www.ystr.org, www.ystr.org/usa and www.ystr.org/asia), these markers are the best characterized for amplification performance and specificity, multiplex amplification strategies, sequence structure and nomenclature, as well as worldwide allele frequency distributions.

Y-STRS

Additional Y-STRs have been introduced into the forensic field: DYS434, DYS435, DYS436, DYS437, DYS438 and DYS439, reported by Ayub *et al*. (4), and DYS460 (GATA A7.1), DYS461 (GATA A7.2), GATA-10, GATA-C4 and GATA-H4, reported by White *et al*. (5). Although these are not as widespread in the forensic field as the previously mentioned loci, data on sequence, population distribution and multiplex amplification have started to accumulate. More recently, other Y-STRs have been described (6–9), and many more are expected to be reported in the near future.

Table 1 shows a list of Y-STR loci described to date with their structures and allele ranges. Chromosomal locations are represented in Figure 1.

PCR PERFORMANCE AND Y CHROMOSOME SPECIFICITY

Inclusion of Y-STRs in routine forensic work requires studies on PCR performance and Y chromosome specificity. Amplification can be, in some cases, improved by primer redesign to reduce nonspecific amplification or the size of amplified fragments (as was the case for DYS389 and DYS385 primers; 10,11). During the design of new primers, screening for point mutations is important to avoid null alleles resulting from primer mismatches. For Y-STRs currently used in forensic science, point mutations have been described in the flanking regions of DYS391, DYS437 and DYS438 (12).

Y-chromosome specificity can be tested by amplifying female DNA samples, which act as negative amplification controls. An example of non-Y specificity is the amplification of female DNA samples by the first described primers for DYS391 (13).

Multiplex amplification increases the yield of information without increasing the number of amplifications and reduces the necessary sample quantity in forensic casework. A great effort to develop STR multiplex systems with a large number of markers has greatly improved the power of discrimination and minimized costs and labor (14).

Y-STR NOMENCLATURE

The use of a consensus nomenclature is crucial to allow for second opinions, proficiency testing, exchange of data and databasing. Sequence data are important to establish allele nomenclature. For STRs with simple repeat structure, it is easy to find a consensus nomenclature, but for others with complex structure, it becomes more difficult. For example, new sequence data for DYS19, DYS390 and DYS389 led to nomenclature changes to include nonrepetitive or variable motifs. International Society for Forensic Genetics guidelines for Y-STR nomenclature have been summarized (15). To prevent future nomenclature changes, these guidelines state

Table 1. Y-STR consensus structures and allele ranges. Segments that are not included in the allele nomenclature are shown in bold

Marker	Repeat Structure	Allele Range (# of repeats)
DYS19/DYS394	$(TAGA)_3$**TAGG**$(TAGA)_n$	10–19
DYS385	**$(AAGG)_{6-7}$**$(GAAA)_n$	7–25,28
DYS388	$(ATT)_n$	8–18
DYS389 I	$(TCTG)_3(TCTA)_n$	9–17
DYS389 II	$(TCTG)_n(TCTA)_n$**N_{28}**$(TCTG)_3(TCTA)_n$	23–34
DYS390	**$(TCTA)_2$**$(TCTG)_n(TCTA)_n(TCTG)_n(TCTA)_n$**TCA$(TCTA)_2$**	17–28
DYS391	**$(TCTG)_3$**$(TCTA)_n$	6–14
DYS392	$(TAT)_n$	6–18
DYS393	$(AGAT)_n$	8–17
YCAII	$(CA)_n$	11–25
YCAIII	$(CA)_n$	19–25
DYS426	$(GTT)_n$	9–14
DYS434	$(TAAT)_{1-2}(CTAT)_n$	10–13
DYS435	$(TGGA)_n$	9–13
DYS436	$(GTT)_n$	9–15
DYS437	$(TCTA)_n(TCTG)_{1-3}(TCTA)_4$	13–17
DYS438	$(TTTTC)_1(TTTTA)_{0-1}(TTTTC)_n$	6–14
DYS439/GATA A4	$(GATA)_n$	8–15
DYS441	$(TTCC)_n$	13–19
DYS442	$(TCTA)_n$	10–14
DYS443	$(TTCC)_n$	12–17
DYS444	$(TAGA)_n$	11–15
DYS445	$(TTTA)_n$	10–13
DYS446	$(TCTCT)_n$	10–18
DYS447	$(TAATA)_n(TAAAA)_1(TAATA)_n(TAAAA)_1(TAATA)_n$	22–29
DYS448	$(AGAGAT)_n$**N_{42}**$(AGAGAT)_n$	17–23
DYS449	$(TTTC)_n$**N_{50}**$(TTTC)_n$	26–33,35–36
DYS450	$(TTTTA)_n$	8–11
DYS452	$(TATAC)_2(TGTAC)_2(TATAC)_n(CATAC)_1(TATAC)_1(CATAC)_1$ $(TATAC)_3(CATAC)_2(TATAC)_3(CATAC)_1(TATAC)_3$	27–33
DYS453	$(AAAT)_n$	9–13
DYS454	$(AAAT)_n$	10–12
DYS455	$(AAAT)_n$	8–12
DYS456	$(AGAT)_n$	14–19
DYS458	$(GAAA)_n$	13–20
DYS459	$(TAAA)_n$	7–10
DYS460/GATA A7.1	$(ATAG)_n$	6–13
DYS461/GATA A7.2	$(TAGA)_n(CAGA)$	8–15
DYS462	$(TATG)_n$	8–14
DYS463	$(AAAGG)_n(AAGGG)_n(AAGGA)_2$	18,20–27
DYS464	$(CCTT)_n$	11–19
GATA-A10	$(TCCA)_2(TATC)_n$	11–18
GATA-C4	$(TCTA)_4(TGTA)_2(TCTA)_2(TGTA)_2(TCTA)_2$ $(TGTA)_{0,2}(TCTA)_n$	17,19–26
GATA-H4	$(AGAT)_4CTAT(AGAT)_2(AGGT)_3(AGAT)_n$**$N_{24}$** $(ATAG)_4(ATAC)_1(ATAG)_2$	24–30

Y-STRS

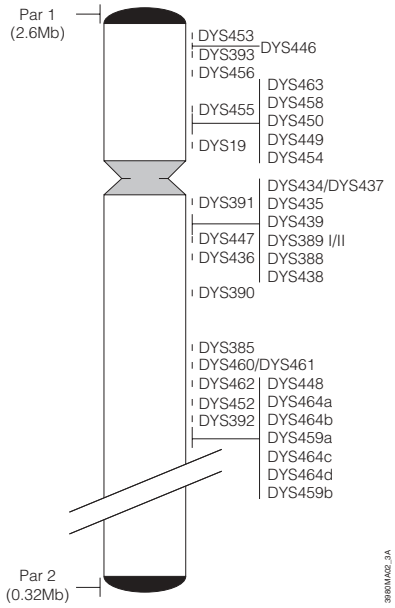

Figure 1. The Y chromosome. The relative positions of PAR 1 and 2 and Y-STRs are shown.

"alleles should be named taking into account both variant and nonvariant repeats". However, the use of recently described STRs introduces problems, since different nomenclatures are being used. A comparative analysis of sequence structure of Y-STR loci in humans and other primates may improve nomenclature as we gain a deeper insight into the variability of the concerned polymorphisms (16).

Y-STR MUTATION RATES

According to data on autosomal STRs, the frequency of mutation events in the male germ line is higher than in the female germ line (17,18). In a survey on Y- and X-linked loci, Scozzari *et al*. report an overall higher diversity for Y-linked loci and suggest a higher rate of accumulation of variants on this chromosome (19). These results can be explained by the higher number of divisions involved in male gametogenesis, which should be reflected in a higher mutation rate. However, no significant differences were found between the average Y-STR mutation rates (2.8×10^{-3}) and those previously found for autosomal STRs (20,21). Data also support the idea that slippage is the mutational mechanism involved and agree with the single-step mutation model.

POPULATION GENETICS AND DATABASING

The use of Y-STRs as inclusion evidence involves population genetic profile definition with the elaboration of a large number of databases. Construction of Y-specific STR databases seems to be more complex than that of unlinked autosomal markers, since the whole haplotype must be typed for each sample. The suitability of Y-STR databases for practical use will be greatly increased with the typing of each individual at as many loci as possible, as opposed to typing a great number of individuals at a small number of Y-STRs. Population substructuring seems to be more extreme in the case of the Y chromosome than for unlinked autosomal markers. This interpopulational variability of Y profiles makes the definition of local databases crucial for the practical application of Y-specific markers.

Y-STR haplotype distributions in populations worldwide have been made available through publications and, more recently, through large-scale forensic databases. The Y-STR Haplotype Reference Databases (YHRD) are the most extensive surveys available online (see www.ystr.org). The development of these databases is important not only for haplotype frequency estimation and subsequent match probability calculations in forensic studies but also for comparative population analysis. Table 2 shows the diversity levels of the most used Y-STR markers in three major human population groups (North Portugal Caucasians, Mozambique Africans and Macao Asians).

REPORTING GUIDELINES

Interlaboratory reproducibility and forensic validation has been demonstrated and is widely accepted in European courts (22). Y-STRs have also been included in some international proficiency testing schemes (23).

During interpretation, it is not valid to multiply individual allele frequencies—a set of Y-chromosome STR data is considered to be a single haplotype and its frequency is assessed relative to the relevant population in large haplotype databases.

As with other DNA evidence, the value of Y-STR evidence should be communicated using likelihood ratios. The Bayesian approach of Roewer *et al*. (24) and Krawczak (25), which estimates a posterior frequency distribution of haplotypes, is a potential way forward and is included as a facility in the European database (www.ystr.org).

CONCLUSION

The application of Y-chromosome markers in genetic investigations requires the (I) identification of polymorphic loci, (II) study of amplification performance and specificity, sequence variation and nomenclature, diversity levels and determination of mutation rates, (III) development of typing methods to improve amplification of low quantity or degraded samples and (IV) the construction of large databases.

The use of Y-STRs in the forensic field can be greatly improved by coordinating efforts to develop typing methodologies, standardize nomenclature and expand databases. The availability of commercial multiplex kits for a large number of Y-STRs will be very helpful in these efforts.

Table 2. Y-STR gene and haplotype diversities (%) in Caucasians (N. Portugal), Africans (Mozambique) and Asians (Macao).

Marker	N. Portugal (N=208)	Mozambique (N=112)	Macao (N=63)
DYS19	56.4	69.1	62.3
DYS385	83.6	92.2	95.1
DYS389 I	57.8	49.5	60.3
DYS389 II	77.4	71.4	76.8
DYS390	58.4	52.8	68.5
DYS391	56.3	30.5	46.8
DYS392	57.6	1.8	68.4
DYS393	46.7	63.0	66.6
"minimal haplotype" diversity	99.25	98.84	99.90
DYS434	4.7	7.0	30.5
DYS437	57.7	5.3	65.9
DYS438	60.5	40.1	37.6
DYS439	68.4	61.2	62.5
DYS460	56.8	55.5	66.5
DYS461	57.1	61.4	58.5
Y-GATA-A10	63.4	64.7	71.6
Y-GATA-C4	65.7	69.1	79.1
Y-GATA-H4	58.3	54.9	62.0
"extended haplotype" diversity	99.96	99.73	No shared haplotypes

ACKNOWLEDGMENTS

L.G. is partially supported by Fundação para a Ciência e a Tecnologia (POCTI, Programa Operacional Ciência, Tecnologia e Inovação).

Editor's Note: *The Y-STR minimal haplotype consists of nine loci: DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393. A commercially available, single-amplification assay for these loci has yet to be offered. To this end, Promega will soon offer a fluorescent multiplex that includes the Y-STR minimal haplotype plus the DYS437, DYS438 and DYS439 loci. This new PowerPlex® Y System(a) will use four-color chemistry, allowing analysis on the ABI PRISM® 377 DNA Sequencer and ABI PRISM® 310 and ABI PRISM® 3100 Genetic Analyzers.*

REFERENCES

1. Ploski, R. *et al.* (2002) Homogeneity and distinctiveness of Polish paternal lineages revealed by Y chromosome microsatellite haplotype analysis. *Hum. Genet.* **110**, 592–600.
2. Qamar, R. *et al.* (2002) Y-chromosomal DNA variation in Pakistan. *Am. J. Hum. Genet.* **70**, 1107–24.
3. Weale, M. *et al.* (2001) Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. *Hum. Genet.* **109**, 659–74.
4. Ayub, Q. *et al.* (2000) Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucl. Acids Res.* **28**, e8.
5. White, P. *et al.* (1999) New, male-specific microsatellite markers from the human Y chromosome. *Genomics* **57**, 433–7.
6. Iida, R. *et al.* (2001) Identification and characterization of two novel human polymorphic STRs on the Y chromosome. *Intl. J. Legal Med.* **115**, 54–6.
7. Iida, R. *et al.* (2002) Characterization and haplotype analysis of the polymorphic Y-STRs DYS443, DYS444 and DYS445 in a Japanese population. *Intl. J. Legal Med.* **116**, 191–4.
8. Bosch, E. *et al.* (2002) High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. *Forensic Sci. Intl.* **125**, 42–51.
9. Redd, A. *et al.* (2002) Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci. Intl.* **130**, 97–111.
10. Schultes, T. *et al.* (1999) Amplification of Y-chromosomal STRs from ancient skeletal material. *Hum. Genet.* **104**, 164–6
11. Schneider, P.M. (1998) Tandem repeat structure of the duplicated Y chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. *Forensic Sci Intl.* **97**, 61–70
12. Gusmão, L. *et al.* (2002) Point mutations in the flanking regions of the Y-chromosome specific STRs DYS391, DYS437 and DYS438. *Intl. J. Leg. Med.* **116**, 322–6.
13. Gusmão, L. *et al.* (2000) Alternative primers for DYS391 typing: advantages of their application to forensic genetics. *Forensic Sci. Intl.* **112**, 49–57.
14. Butler, J. *et al.* (2002) A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci. Intl.* **129**, 10–24.
15. Gill, P. *et al.* (2001) DNA commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. *Intl. J. Legal Med.* **114**, 305–9.
16. Gusmão, L. *et al.* (2002) Chimpanzee homologous of human Y specific STRs. A comparative study and a proposal for nomenclature. *Forensic Sci. Intl.* **126**, 129–36.
17. Brinkmann, B. *et al.* (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.* **62,** 1408–15.
18. Weber, J. and Wong, C. (1993) Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2**, 1123–8.
19. Scozzari, R. *et al.* (1997) Differential structuring of human populations for homologous X and Y microsatellite loci. *Am. J. Hum. Genet.* **61**, 719–33.
20. Kayser, 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,** 1580–8.
21. Kayser, M. and Sajantila, A. (2001) Mutations at Y-STR loci: implications for paternity testing and forensic analysis. *Forensic Sci. Intl.* **118**, 116–21.
22. Carracedo, A. *et al.* (2001) Results of a collaborative study of the EDNAP group regarding the reproducibility and robustness of the Y-chromosome STRs DYS19, DYS389 I and II, DYS390 and DYS393 in a PCR pentaplex format. *Forensic Sci. Intl.* **119**, 28–41.
23. Sánchez-Diz *et al.*, *Forensic Sci Intl.*, in press
24. Roewer, L. *et al.* (2000) A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. *Forensic Sci. Intl.* **114**, 31–43.
25. Krawczak, M. (2001) Forensic evaluation of Y-STR haplotype matches: a comment. *Forensic Sci. Intl.* **118**,114–5