Y Chromosome STR Analysis in Forensic Practice

Lutz Roewer[1], Peter de Knijff [2] and Manfred Kayser[1,3]
[1] Institute for Legal Medicine, Humboldt-University, Berlin, Germany
[2] Forensic Laboratory for DNA Research, Leiden University, The Netherlands
[3] Department of Anthropology, The Pennsylvania State University, USA

THE PERSPECTIVE OF FORENSIC Y-STR ANALYSIS

Currently, the gradual introduction and routine application of validated Y chromosomal STR systems in stain analysis is being experienced. These systems are powerful forensic tools especially for the area of sexual assault evidence and are superior to autosomal systems for two reasons: the differential cell lysis step *prior* to the PCR is replaced by a sex chromosome selection *during* PCR and the recognition and interpretation of perpetrator profile(s) is facilitated because a female profile does not occur.

However, prior to the implementation of known Y-STR systems in the core marker set of forensic routine laboratories we think that two important questions have to be answered:

1. How can national haplotype databases be established allowing valid Y-chromosome related statistics in cases of non-exclusion?

2. Can Y-STR systems be efficiently and safely analyzed by high-throughput techniques?

The answer to question 1 is rather puzzling. Because of the enormous variability of Y-STR haplotypes the population database must be very large with at least 10,000 entries. In a country like Germany, with a number of modern forensic science laboratories, these data sets could be collected at different places in a reasonable time span, provided the data are allowed to be pooled. However, due to its inheritance mode it is clear that Y chromosome haplotypes will have the tendency to show up a population or region specific distribution (Jobling *et al.* 1997, de Knijff *et al.* 1997). It is the (solvable) problem to analyze the extent of this population sub-structuring in a statistically correct way. This problem cannot be addressed theoretically but by a comparison of representative region-specific population samples typed for the standard Y-STR haplotype format including DYS19, DYS385, DYS389, DYS390, DYS391, DYS392, DYS393 (Kayser *et al.* 1997). The number and choice of the regions and the number of samples collected for each region depends on the demographic and geographic peculiarities of a territory or country. Once the haplotypes have been collected, known statistical methods (AMOVA i.e. Analysis of Molecular Variance) can be applied to describe the extent of between population variability of haplotypes represented by F_{st} values (Roewer *et al.* 1996). The freeware computer program ARLEQUIN ver.1.1 can be used to perform these calculations (S. Schneider *et al.* 1997). By means of an AMOVA based significance test, relatedness between two regional male gene pools can then be established or denied. Only those haplotype data from regional pools are allowed to be combined where the differences between F_{st} values turn out to be insignificant.

It has been speculated that regional clustering at least in some urban European countries could be neglected. If this is true, existing regional databases could be pooled to get in a time-saving way the large data amounts which allow meaningful forensic probability calculations on the basis of haplotype frequencies. It should be emphasized that as long as this hypothesis has not been proven by a careful collection of standardized haplotypes subsequently treated by an adequate statistical analysis this idea is what it is: just a speculation.

However, from the author's point of view, the Y-STR databasing problem can be solved in a reasonable time on the basis of two essential prerequisites: inter-laboratory standardization and collaboration. National projects to collect and compare representative regional population samples are under way in Holland, Germany, Finland, the U.S.A. and elsewhere.

The answer to the second question is "yes," since actual high-throughput techniques (multiplex electrophoresis using fluorescent-tagged STR amplicons) as well as progressive new technologies (e.g. mass spectrometry, DNA microchip technology) can be applied to all STRs regardless of their chromosomal location. Some features apply to the above mentioned Y-STRs which qualify them for automatization: low or zero microvariation, consecutive repeats, tetrameric structure (with the exception of DYS392 which is trimeric). The number of highly informative Y-STRs fulfilling this requirements is still stagnant and is confined so far to the 7 above mentioned systems, an astonishing small number compared to the thousands of systems available on the autosomes. However, this limitation could be seen as an advantage since

these 7 systems have been extensively validated both with respect to molecular properties (sequence structure, chromosomal localization, mutation rate, etc.) as well as to forensic demands (population studies, PCR optimization and primer design, multiplex conditions, stain mixture studies, etc.). Several laboratories are working on further optimization (Kayser *et al.* 1997, de Knijff *et al.* 1997, and Prinz *et al.* 1997). We think that there is no need to complement the large panel of forensically applied autosomal markers by more than these 7 Y-STR systems.

For the systems DYS385 (P.M. Schneider *et al.*, Foren. Sci. Int., in press) and DYS19 (Szibor *et al.* submitted for publication) radical new primer sets (compared to those reported by Kayser *et al.* 1997) have been designed which reduce the overall lengths of the PCR products by approximately 100 bps and improve the applicability of these markers to minor or degraded stains. An optimized primer design will also improve multiplex approaches, a number of which have been already proposed (Kayser *et al.* 1997, Prinz *et al.* 1997, Redd *et al.* 1997, see Table 1).

MUTATION RATE REVISITED

The prevalent mutational mechanism acting on STRs during the meiosis is the so-called slipped strand mispairing leading to new length alleles. A gain of repeats is dominant over their loss and mutations occur mainly in a stepwise manner. Since spontaneous mutations can severely influence the result of a forensic STR analysis (e.g. exclude a true father from paternity) the estimation of locus-specific slippage mutation rates is an obligate prerequisite for the application of STRs, at least for paternity testing.

Among 844 meioses analyzed for DYS19 (Kayser *et al.* 1997, Heyer *et al.* 1997) 2 slippage mutations have been observed ($\mu = 2.4 \times 10^{-3}$; CIL 3.0- 8.1 x 10^{-3}). Based on 2790 meioses altogether, 12 slippage mutations have been observed for 12 Y-chromosomal STRs (average $\mu = 4.3 \times 10^{-3}$). This places the average mutation rates for known di-, tri-, tetra- and pentameric Y chromosomal STRs between higher values for SE33 (ACTBP2) and lower values for HUMTH01 and HUMFES/FPS, just in the range of a system like HUMFIBRA/FGA (Brinkmann *et al.* 1998). Further investigations are under way to evaluate truly reliable locus-specific rates.

Besides the predominant slippage mechanism other mutational forces act on STRs, which are often considered as "mutational hotspots". For example, a loss of larger numbers of repeat units (Forster *et al.* in press) as well as the gain of a complete multi-repeat block including flanking sequences (Kayser *et al.* 1995) has been reported.

For some STRs on the Y chromosome, a comparable frequent occurrence of aberrant numbers of alleles per locus has been observed (Table 2). In all cases, the additional alleles fall in regular length classes. From the two possible explanations for this phenomenon - aberrant karyotypes or aberrant locus numbers - we favor the second hypothesis. Since it is known, that large parts of the Y chromosome consist of repeated elements with units of hundreds and thousands of basepairs in length, an insertion of a Y-STR sequence into a major repeat could result in a constitutional locus multiplication. This situation - a constitutional locus duplication - has been postulated for the Y-STR systems DYS385, YCAI, YCAII and YCAIII/DYS413 (Kayser *et al.* 1997). Since the number of major repeat units on the Y chromosome will tend to be variable, even a higher number of alleles per STR locus cannot be excluded. As shown in Table 2, altogether 7 unusual diallelic or triallelic genotypes have been found for the Y-STR loci DYS19, DYS390 and DYS385.

A rate for locus multiplication events has been calculated for DYS19 with 7.6 x 10^{-4} (n = 5232), roughly four times lower than the slippage mutation rate at this locus.

If male relatives were available for genotyping the di- and triallelic patterns are found to be inherited from father to son.

Another mutation-associated phenomenon has been observed in our casework material: in two paternity cases we found two slippage mutations in two different Y-STR systems, respectively. Following conventional analysis with SLS and MLS systems, probabilities of paternity well above 99.999 % were calculated. Theoretically, the occurrence of two independent exclusions of paternity would be sufficient for a declaration of non-paternity to be made. However, the strength of evidence from the other loci renders this conclusion unlikely.

Under the assumption of a confirmed paternity two hypotheses could be applied to explain the striking Y-STR results:

1. The aberrant paternal alleles found in the sons stem from the true father and were generated by independent slippage mutations during spermatogenesis (average rate $\mu_{Y\text{-}STR} = 4.3 \times 10^{-3}$/locus).

2. A reduced mismatch repair activity due to gene mutations of repair enzymes potentially results in an overall increased mutability of STR sequences of the affected individual (Strand *et al.* 1993).

To test the latter hypothesis, 39 informative autosomal STRs (a number of these were kindly provided by J.T.

Epplen, Ruhr-Universität Bochum and P. Nürnberg, Humboldt-University Berlin) together with 11 Y chromosomal STRs have been tested in the respective cases. In case #1 with mutations at the loci DYS389 and DYS390, no further mutation has been found, in case #2 with mutations at DYS390 and DYS413 one additional mutation at the locus HUMFIBRA/FGA ($\mu = 0.004$ according to Brinkmann *et al.* 1998) occurred. From this we favour independent mutations rather than a reduction of the mismatch repair activity as causative for the observed constellations.

Taking into account a general estimate of STR mutation rates in a range of about 10^{-3} (Weber & Wong 1993) - which is surely not an overestimation for sperm - the chance of observing at least one such slippage mutation in a child would be about 5% when 50 STR systems are examined. In a recent paper, Gunn *et al.* (1997) report a similar case with two independent mutations at TPOX and HUMFES/FPS, with another 17 systems as well as MLS typing, providing strong evidence in favor of paternity. The authors state that on the basis of an even higher STR mutation rate of 10^{-4} - 10^{-5} double mutations leading to false exclusions from paternity would be expected to occur about once in every nine thousand cases.

As a result of our casework examples, as well as of those of other laboratories, we strongly recommend complementing STR analysis in paternity casework with classical robust methods such as SLS- or MLS analysis.

According to Gunn *et al.* (1997), non-paternity demonstrated only on the basis of two exclusions at STR loci is not sufficient evidence. As shown in our example, there is a probability that even more than two STR slippage mutations can occur during gametogenesis.

CONCLUSIONS

At a second sight, Y-STR results generated in forensic practice must be interpreted carefully with respect to the following points:

1. The establishment of national Y databases must be done on the basis of standard haplotypes using adequate samples (quality and quantity) and adequate statistical analysis methods.

2. The interpretation of biallelic/triallelic patterns in stain mixtures must include the knowledge of (rare) mutations potentially affecting the number of loci.

3. The paternal mutation rate for STRs cannot be neglected, thus an exclusion of paternity should be confirmed by at least two differing father/son genotypes by using the Y-STR systems mentioned above. This applies for deficiency cases (male offspring/patrilinear relative) where no other efficient method than Y-STR typing exists. For all normal trio's the STR analysis should be complemented by classical highly informative systems.

REFERENCES

1. Brinkmann B., Klintschar M., Neuhuber F., Hühne J., Rolf B., (1998) Mutation Rate in Human Microsatellites: Influence of the Structure and Length of the Tandem Repeat. *American Journal Human Genetics*, **62**: 1408-1415.

2. de Knijff P., Kayser M., Caglià A., Corach D., Fretwell N., Gehrig C., Graziosi G., Heidorn F., Herrmann S., Herzog B., Hidding M., Honda K., Jobling M., Krawczak M., Leim K., Meuser S., Meyer E., Oesterreich W., Pandya A., Parson W., Piccinini A., Perez-Lezaun A., Prinz M., Schmitt C., Schneider PM., Szibor R., Teifel-Greding J., Weichhold G., Roewer L., (1997) Chromosome Y Microsatellites: Population Genetic and Evolutionary Aspects. *International Journal of Legal Medicine*, **110**: 134-140.

3. Gunn P.R., Trueman K., Stapleton P., Klarkowski D.B., (1997) DNA Analysis in Disputed Parentage: The Occurrence of Two Apparently False Exclusions of Paternity, Both at Short Tandem Repeat (STR) Loci, in the One Child. *Electrophoresis*, **18**: 1650-1652.

4. Heyer E., Puymirat J., Dieltjes P., Bakker E., de Knijff P. (1997) Estimating Y Chromosome Specific Microsatellite Mutation Frequencies Using Deep Rooting Pedigrees. *Human Molecular Genetics*, **6**: 799-803.

5. Jobling M., Pandya A., Tyler-Smith C., (1997) The Y Chromosome in Forensic Analysis and Paternity Testing. *International Journal of Legal Medicine, 110*: 118-124.

6. Kayser M., Caglià A., Corach D., Fretwell N., Gehrig C., Graziosi G., Heidorn F., Herrmann S., Herzog B., Hidding M., Honda K., Jobling M., Krawczak M., Leim K., Meuser S., Meyer E., Oesterreich W., Pandya A., Parson W., Penacino G., Perez-Lezaun A., Piccinini A., Prinz M., Schmitt C., Schneider PM., Szibor R., Teifel-Greding J., Weichhold G., de Knijff P., Roewer L., (1997) Evaluation of Y-Chromosomal STRs: A Multicenter Study. *International Journal of Legal Medicine*, **110**: 125-133; 141-149.

7. Kayser M., Nürnberg P., Bercovitch F., Nagy M., Roewer L., (1995) Increased Microsatellite Variability in *Macaca Mulatta C*ompared to Humans Due to a Large Scale Deletion / Insertion Event During Primate Evolution. *Electrophoresis*, **16**: 1607-1611.

8. Prinz M., Boll K., Baum H., Shaler B., (1997) Multiplexing of Y Chromosome Specific STRs and Performance for Mixed Samples. *Forensic Science International*, **85**: 209-218.

9. Redd A.J., Clifford SL., Stoneking M., (1997) Multiplex DNA Typing of Short-Tandem-Repeat Loci on the Y Chromosome. *Biological Chemistry*, **378**: 923-927.

10. Roewer L., Kayser M., Dieltjes P., Nagy M., Bakker E., Krawczak M., de Knijff P., (1996) Analysis of Molecular Variance (AMOVA) of Y-Chromosome Specific Microsatellites in Two Closely Related Human Populations. *Human Molecular Genetics*, **5**: 1029-1033.

11. Santos F.R., Gerelsaikhan T., Munkhtuja B., Oyunsuren T., Epplen J.T., Pena S.D.J., (1996) Geographic Differences in the Allele Frequencies of the Human Y-Linked Tetranucleotide Polymorphism DYS19. *Human Genetics*, **97**: 309-313.

12. Santos F.R., Rodriguez-Delfin L., Pena S.D.J., Moore J., Weiss K.M., (1996a) North and South Amerindians May Have the Same Major Founder Y Chromosome Haplotype. *American Journal of Human Genetics*, **58**: 1369-1370.

13. Schneider S., Kueffer J.-M., Roessli D., Excoffier L. (1997) Arlequin ver 1.1: A Software for Population Genetic Analysis. *Genetic and Biometry Laboratory, University of Geneva, Switzerland*.

14. Strand M., Prolla T.A., Liskay R.M., Petes TD., (1993) Destabilization of Tracts of Simple Repetitive DNA in Yeast by Mutations Affecting DNA Mismatch Repair. *Nature*, **365**: 274-276.

15. Weber J.L. and Wong C., (1993) Mutation of Human Short Tandem Repeats. *Human Molecular Genetics*, **2**: 1123-1128.

Table 1. Multiplexing Y-STRs

Device	Markers	Size ranges (in bp)
ALF	DYS19, DYS389I, DYS389II (Kayser *et al.* 1997) DYS393, DYS390, DYS391 (Kayser *et al.* 1997)	174-206, 243-263, 353-381 116-132, 187-227, 275-291
ABI	DYS393, DYS390, DYS394 (i.e. DYS19), DYS391 (Redd *et al.* 1997) DYS19, DYS390, DYS389I, DYS389II, DYS385 (M. Prinz, personal communication) DYS393, DYS392, DYS391 (Prinz *et al.* 1997)	116-132, 187-227, 242-258, 275-291 174-206, 187-227, 243-263, 353-381, 360-412 116-132, 245-263, 275-291

Table 2. Aberrant di/triallelic genotypes at Y-STR loci

	DYS19	**DYS390**	**DYS385**
2 alleles/locus	1 observation (Santos *et al.* 1996a) 2 observations (this study)	1 observation (Kayser *et al.* 1997)	constitutional
3 alleles/locus	1 observation (Santos *et al.* 1996)	1 observation (Redd *et al.* 1997)	1 observation (this study)