Forensic Science International
87 (1997) 179–184

Editorial

The DNA Commission of the International Society for Forensic Haemogenetics (ISFH) has over the years, published a series of documents providing guidelines and recommendations concerning the application of DNA polymorphisms to problems of identification.

The latest report published here, provides recommendations relating to the nomenclature of STR (short tandem repeat) typing systems which are at the forefront of systems used at present by forensic scientists and are likely to remain so for the immediate future.

Also included in this number of the journal is a paper by the European DNA Profiling Group (EDNAP) which also discusses and provides guidelines concerning the nomenclature of STR systems and as such complements the report of the DNA Commission of the ISFH.

Prof. P.J. Lincoln,
Associate Editor

0379-0738/97/$17.00 © 1997 Elsevier Science Ireland Ltd. All rights reserved
PII S0379-0738(97)00110-2

DNA recommendations – further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems

Introduction

A common nomenclature for STRs is a prerequisite for interlaboratory reproducibility and for the exchange and comparison of data. For many loci there has been an enormous increase of information about sequence and substructure which has repeatedly raised questions relating to nomenclature. While the previous recommendations are still valid, here we describe some additional rules relating to STR allele designation.

1. Sequence and repeat designation of STRs

DNA sequences are read in the 5′ to 3′ direction. The choice of the strand also influences the sequence designation. To avoid confusion, the following guidelines should be followed:

- for STRs within protein coding genes (as well as in the intron of the genes), the coding strand should be used, An example of such a locus is vWA (GenBank: M25716). The same applies to pseudogenes, such as the ACTBP2 locus, where the strand with the sequence similar to the coding strand of the "original" gene should be used (Moss and Gallwitz, 1983).
- for repetitive sequences without any connection to protein coding genes, like many of the D#S## loci, the sequence originally described in the literature or the first public database entry shall become the standard reference (and strand) for nomenclature.

If a nomenclature is already established in the forensic field but not in accordance with the aforementioned guideline, the nomenclature shall be maintained to avoid unnecessary confusion.

It is sometimes possible to define different repetitive motifs, even though the choice of the strand is clear. In the two following examples, the initial point for reading the repeat motif is different for the same sequence:

1. 5′ – GG TCA TCA TCA TGG – 3′ 3 × TCA

2. 5′ – GGT CAT CAT CAT GG – 3′ 3 × CAT

The recommendation is that the repeat sequence motif must be defined so that the first

5′-nucleotides that can define a repeat motif are used. Thus only the first edition is correct.

These two rules together, for the choice of the strand and for the choice of the motif, will allow for the definition of the sequence and repeat designation of any new STR. If a repeat designation of a commonly used STR-system does not follow these guidelines, the established nomenclature for the sequence can continue to be used to avoid new confusion. For those situations where two or more nomenclatures already exist, priority should be given to the nomenclature that more closely adheres to the guidelines described here. If this is not possible, priority shall be given to the nomenclature that was documented first.

2. Allele designation of STRs

Since the polymorphisms concerned are defined by variations in the number of repeats (VNTR's), allele designation basically should observe this structural principle.

- For simple systems like HumFES/FPS, $(ATTT)_{8\text{-}14}$, this is straightforward.
- For systems composed of repeat regions where the sequence may vary, designation of alleles should refer to the total number of full repeats, although the sequence can be different (as for HumVWA or the longer HumCD4 alleles).
- The designation of incomplete repeat motifs should include the number of complete repeats and, separated by a decimal point, the number of basepairs in the incomplete repeat. Examples are allele 9.3 at the HumTH01 locus (Puers et al., 1993) which contains 9 tetranucleotide repeats and one incomplete repeat of three nucleotides and the allele 22.2 at the FGA locus which contains 22 tetranucleotide repeats and one incomplete repeat with two nucleotides only (Barber et al., 1996).
- For complex repeat systems, repeat nomenclature should have a mathematical relationship to the length in bp of the consensus allele. An example is D21S11 (Brinkmann et al., 1996):

repeats	length	relationship
27	213 bp	$27\times4+105=213$
31	229 bp	$31\times4+105=229$
33.2	239 bp	$33\times4+2+105=239$

where 105 bp is the sum of 5′-flanking region, the 3′-flanking region and an interspread 43 bp constant array. This kind of nomenclature allows for the easy conversion of the results from automated typing in bp into the repeat based nomenclature referring to common and consensus alleles in the ladder.

- For some highly variable systems (such as the ACTBP2 locus), the repetitive

structure can be very complex and the definition of a consensus repeat structure can be difficult. In such cases, alleles should be identified according to their size in bp, by comparison with a sequenced ladder. Sequence variants of the same length in these highly variable systems can have different electrophoretic mobilities under certain conditions. Thus the type of the allele determined will depend on the electrophoretic system used. To minimise electrophoretic mobility differences for alleles, these systems should be typed only under denaturing conditions.

3. Use of allelic ladders

Allelic ladders are used as a reference for allele designation. All alleles in an allelic ladder should be sequenced and nominated according to the aforementioned rules. They can be obtained commercially or prepared in house. Ladders should contain all common alleles, so that the spacing allows for exact typing of the samples (this is a 4 bp spacing for many of the common tetranucleotide repeats).

Typing can be carried out using manual or automated methods. Alleles in a database or in identity testing are named according to the visual comparison with an allelic ladder or by automated sizing using standards; sequencing is not necessary. The base composition can influence the electrophoretic mobility, alleles of the respective system should be used as standard for automated sizing whenever possible. Results in bp must be converted into the repeat-based names (when possible, see above), A prerequisite for typing is that the resolution of the electrophoretic system used allows for the detection and precise identification of all common alleles. In some highly variable systems, alleles with 2 bp differences to the next common larger and smaller allele are observed (e.g. allele 22, 22.2 and 23 at the FGA locus or many of the ACTBP2 alleles). Manual typing of this locus should allow for the reproducible detection of these alleles. Automated fragment sizing should have a typing error smaller than half of the distance between common alleles (otherwise a continuous allele distribution would result).

The separation and detection system can influence the number of variants/alleles detected. For example, some laboratories distinguish between alleles 9.3 and 10 at the TH01 locus, some do not. The same is true for variants of allele 10 at the FES locus, which have either an A or a C in the 5′-flanking region. These differences between laboratories are relevant to questions of reproducibility and result sharing. Allele frequencies depend on the correct definition of alleles as distinguishable entities. It should be clear whether variants have been detected or not. Applying the above rules for the TH01 locus, the pooled allele frequencies in the database must be nominated 9.3/10. At the FES locus, allele 10 includes both variants, with the A and the C in the 5′-flanking region. If the variants are detected, they should be designated 10^{Ade} or 10^{Cyt}. (10A or C would confuse people using this abbreviations for electrophoretic shifts to the anode or the cathode).

Information on electrophoretic conditions, the ladders used, the primer sequences and the detection system should be made available upon the publication of data. Authors should follow these recommendations and state that they did so.

References

Barber, M.D., McKeown, B.J. and Parkin, B., Structural variation in the alleles of a short tandem repeat system at the human alpha fibrinogen locus, Int. J. Legal. Med., 108 (1996) 180–185.

Brinkmann, B., Meyer, E. and Junge, A., Complex mutational events at the HumD21S11 locus, Hum. Genet. 98 (1996) 60–64.

Moss, M. and Gallwitz, D., Structure of two human β-actin-related processed genes one of which is located next to a simple repetitive sequence, EMBO J. 2 (1983) 757–761.

Puers, C., Hammond, H.A., Jin, L., Caskey, C.T. and Schumm, J.W., Identification of repeat sequence heterogeneity at the polymorphic short tandem rapeat locus HUMTH01 $(AATG)_n$ and reassignment of alleles in population analysis by using a locus-specific allelic ladder, Am. J. Hum. Genet., 53 (1993) 953–958.

Members of the DNA Commission: Board members of the ISFH (B. Olaisen, W. Bär, W. Mayr, P. Lincoln and A. Carracedo) together with B. Brinkmann, B. Budowle and P. Gill.

Reprints from: Dr. Patrick J. Lincoln
Dept. of Haematology
St. Bartholomews and the Royal London School
of Medicine and Dentistry, Turner Street
London, E1 2AD