Tissue Antigens ISSN 0001-2815

Nomenclature for HLA microsatellites

P. A. Gourraud[1], A. Cambon-Thomsen[1], E. M. Dauber[2], M. Feolo[3], J. Hansen[4], E. Mickelson[4], R. M. Single[5], M. Thomsen[1] & W. R. Mayr[2]

1 Inserm, Unit 558, Department of Epidemiology and Public Health, Faculty of Medicine, Toulouse, France
2 Division of Blood Group Serology, Medical University Vienna, Wien, Austria
3 National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA
4 Division of Clinical Research, Fred Hutchinson Cancer Research Center and Department of Medicine, University of Washington, Seattle, WA, USA
5 The Department of Mathematics and Statistics, University of Vermont, Burlington, VT, USA

Key words
HLA; major histocompatibility complex; microsatellite; nomenclature; short tandem repeat; standardization; variable number of tandem repeat

Correspondence
Prof. Dr Dr h.c. Wolfgang R. Mayr
Klinische Abteilung für Blutgruppenserologie
Medizinische Universität Wien
Währinger Gürtel 18-20
A-1090 Wien
Austria
Tel: 431 40400 5338
Fax: 431 40400 5321
e-mail: wolfgang.mayr@meduniwien.ac.at

doi: 10.1111/j.1399-0039.2006.00771.x

Abstract

A proposal for a standardized nomenclature for human leukocyte antigen (HLA) microsatellites is presented. It provides recommendations for Microsatellites as regards to locus name, primer names, and denominations for alleles.

Microsatellite markers also referred to as short tandem repeats (STR) are used increasingly in order to characterize human leukocyte antigen (HLA) haplotypes (1). The nomenclature regarding these markers and their alleles is, however, not standardized and does not cover all possibilities.

The following problems are encountered:

- The names of the loci are not uniform and, in many cases, do not correspond to the guidelines for human gene nomenclature (2). Even the genome-wide used 'D#S####' names provided by the Genome Data Base (GDB) are sometimes redundant in the major histocompatibility complex (MHC) region (3).
- The designation for alleles is erratic as several nomenclature schemes are used, for example the length of the amplicon or an arbitrary numerical value. Many problems emerge from these procedures: the calculated length of amplicons depends on the primers used, on the dyes used for labeling the primers, and on the equipment used to ascertain the alleles (Table 1); in addition, the arbitrary notation cannot account for all possible new alleles.

For this reason and in order to avoid a 'Tower of Babel' in the designation of loci and alleles, it is necessary to agree on a unified nomenclature similar to the HLA nomenclature which has been proposed by an international committee and which is commonly used throughout the world [see e.g. (4)].

In order to agree on such a nomenclature, a discussion took place on 1 December 2005, at the 14th International HLA and Immunogenetics Workshop in Melbourne, Australia. The discussion largely benefited from a previous discussion that occurred during the preworkshop meeting 'Microsat-HLA 2005' in Toulouse (June 28–1 July 2005).

The guidelines presented in this article correspond to the recommendations of the DNA commission of the International Society for Forensic Genetics, specifically the use of the number of repeats as the designation for an allele (5). These rules have been proven to be very reliable as they are used for the characterization of microsatellite alleles in the national and international databases for identification of individuals involved in forensic cases.

Table 1 ACTBP2 (SE33): real length *vs* calculated length

		Dye			
		6-FAM™		Fluorescein	
Alleles	Real length (bp)	Calculated length (bp)	Difference (bp)	Calculated length (bp)	Difference (bp)
Short (*12)	233	226.03	6.97	230.42	2.58
Medium (*22)	273	264.63	8.37	271.55	1.45
Long (*36)	329	319.46	9.54	328.07	0.93

This example is taken from forensics. The human actin beta pseudogene 2 (gene ID: ACTBP2; N°62) short tandem repeat (STR) locus is one of the most informative tetra-nucleotide STR systems for personal identification and paternity testing. The STR is located in chromosome 6 (6q14).

Conclusions of the discussion

The proposed nomenclature has three levels which are

1. Locus level. The unified designation of the microsatellite loci.
2. Primer level. The use of reference name for the primers.
3. Allele level. Denominations of alleles that would facilitate the comparison between different studies.

Nomenclature of the microsatellite locus name: the lowest D6S#### identifier

The D6S#### nomenclature should be used. If two different D6S#### designations correspond to the same locus, the lower figure (which has been assigned earlier) should be used.

Resource:

The reference D6S#### names are listed in the microsatellite dedicated area of MHC database (dbMHCms):

http://www.ncbi.nlm.nih.gov/projects/mhc/MHC.fcgi?cmd=init.

Consequences:

- Designations of loci without a D6S#### name [e.g. tumor necrosis factor (TNF-A), …] should be changed to the D6S#### scheme (e.g. TNF-A → D6S2792).
- Designations of loci with a redundant D6S#### name should be changed to the lower D6S#### scheme (e.g. D6S2810 → D6S2673).
- All potential microsatellite markers will be given a D6S#### name on submission to the Human GDB http://www.gdb.org/gdb/SubmissionDirections.html, and referenced in dbMHC. Thus, the case 'no D6S#### name is available' will not happen anymore.

Nomenclature of primers: UniSTS designations

A pair of primers used to genotype a microsatellite should be referenced using the UniSTS number.

Resources:

- UniSTS database: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists.
- The dbMHCms provides links to the UniSTS number of each primer pair:

http://www.ncbi.nlm.nih.gov/projects/mhc/MHC.fcgi?cmd=init.

Consequences:

- The extended list of primer sequences is not required in publications, the UniSTS number is sufficient.
- dbMHCms can provide UniSTS numbers in the case of a new primer pair being submitted.

Nomenclature of alleles

- *The nomenclature of alleles should be based on the number of repeats. The group recommends having two designations for the microsatellite alleles: (1) figure based on the number of repeats and (2) the fragment size.*

1 – DNA sequences are read in the 5′ to 3′ direction. The choice of the strand influences the sequence designation – to avoid confusion, the + strand and the repeat motifs shown in dbMHCms should be used.

Because the polymorphisms concerned are defined by variations in the number of repeats, the allele designation should observe this structural principle. For simple systems without microvariation within the repeats, this is straightforward – e.g. six repeats correspond to allele *6. The designation of incomplete repeat motifs should include the number of complete repeats, and, separated by a decimal point the number of base pairs (bp) in the incomplete repeat – e.g. nine complete repeats with four nucleotides and one incomplete repeat with three nucleotides correspond to the allele *9.3.

Allelic ladders are used as a reference for allele designation. All alleles in an allelic ladder should be sequenced; ladders should contain all common alleles, so that the regular spacing allows for exact typing of the samples. As the automatic determination of allele lengths in capillary electrophoresis depends on many variables, e.g. primers and fluorescent dye used for labeling the primers, allelic ladders must be analyzed under the same conditions as the samples. Allelic ladders can easily be produced by mixing the amplicons of sequenced alleles or by amplifying a mixture of DNA containing the major alleles which have been sequenced (Figures 1, 2).

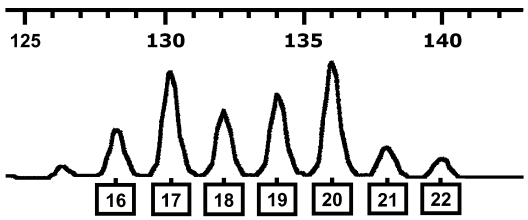
Table 2 Comparison of nomenclatures for alleles of the D6S273 microsatellite

13th IHW#	Proposed nomenclature		NFI[a]		Real length (sequenced)		Oslo protocol, UniSTS: 46793		NCI[b] protocol, UniSTS: 256846	
9053	21	21	16	16	138		133.02		156.56	
9103	18	20	13	15	132	136	127.10	130.94	150.25	154.41
9220	17	20	12	15	130	136	125.07	130.82	148.13	154.46
9370	16	19	11	14	128	134	123.23	129.05	146.10	152.49
9389	19	22	14	17	134	140	128.93	134.84	152.36	158.65

[a] NFI stands for Normalized Fragment Index (used at 13th International Histocompatibility Workshop).
[b] NCI stands for National Cancer Institute, USA.

If allelic ladders are not available, sequenced reference samples encompassing the most common alleles should be taken as standard. These reference samples might also be used in parallel with allelic ladders in order to verify the results. The allelic ladder or the reference samples have to be amplified using the same primers and fluorescent labels and subjected to the same electrophoresis conditions as the samples to be tested.

In rare cases of length polymorphisms due to insertions or deletions outside the repeat region, the alleles are also named according to their fragment size in comparison with the allelic ladder. Automated fragment sizing should have a typing error smaller than half of the distance between common alleles.



Allele designation	Length (bp)	Repeat structure
*16	128	$(GT)_{16}$
*17	130	$(GT)_{17}$
*18	132	$(GT)_{18}$
*19	134	$(GT)_{19}$
*20	136	$(GT)_{20}$
*21	138	$(GT)_{21}$
*22	140	$(GT)_{22}$

Figure 1 D6S273: allelic ladder (calculated length in bp) and sequences. D6S273 is an exact GT repetition located in the 'lymphocyte antigen 6 complex, locus G6D' gene (gene ID LY6G6D N°58530). D6S273 is also known as 142XH6, AFM142xh6, GC378-D6S273, Genome Data Base: 188060, HS142XH6, Z16657, stSG6946. Five different primer pairs have been referenced to Major Histocompatibility Complex database; UniSTS numbers are 46793, 256846, 256847, 464299, and 464300.

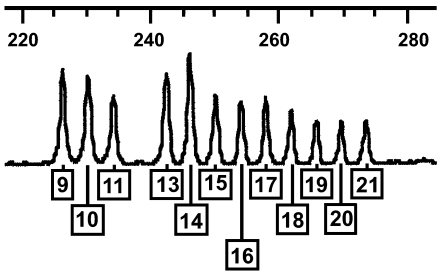
2 – In order to be able to go back to previous data, it is suggested that the fragment sizes should also be reported initially until the proposed nomenclature based on the number of repeats is commonly used. The problem in this respect is that the calculated fragment size is dependent on the primers used and the fluorescent label, as well as the technical system used for electrophoresis. Furthermore, the calculated figures of the length of the amplicon may not be in a linear mathematical relationship with the number of repeats.

It is obvious that the proposed nomenclature is only based on the length of the amplicons; particular cases, e.g. insertions or deletions outside of the repeat region, as well as single nucleotide polymorphisms have to be characterized by sequencing.

Consequences:

- For publications, the group recommends to use the double designation for the microsatellite alleles: the figure based on the number of repeats and the fragment size.
- In order to be able to use the results which have already been published, a dictionary giving the equivalents of the repeat number with various fragment lengths should be established (see e.g. Table 2). This dictionary is similar to the tables which have been used in the first years of the official HLA nomenclature (6, 7). This list must at least include details concerning the primers (UniSTS numbers).
- The editors of scientific journals should insist that articles dealing with HLA microsatellites adhere to the proposed nomenclature.

In Figures 1 and 2, we implement the proposed nomenclature as examples. The microsatellite name is the lower D6S#### name: D6S273 in Figure 1, D6S2691 in Figure 2. For information, synonyms are also indicated. Primer pair is referenced using the UniSTS numbers: 46793 in Figure 1, 239082 in Figure 2. For information, alternative primers are also indicated. The equivalence between the two designations of alleles must be documented either in text or in table. In tables, the alleles are indicated both using allele length measured experimentally and a sequence-based repetition figure: the table of Figure 1 shows the equivalence



Allele designation	Length (bp)	Repeat structure
*9	227	$(TTTC)_2$-TTTT-$(TTTC)_6$
*10	231	$(TTTC)_3$-TTTT-$(TTTC)_6$
*11	235	$(TTTC)_3$-TTTT-$(TTTC)_7$
*13	243	$(TTTC)_3$-TTTT-$(TTTC)_7$-CTTC-TTTC
*14	247	$(TTTC)_3$-TTTT-$(TTTC)_8$-CTTC-TTTC
*15	251	$(TTTC)_3$-TTTT-$(TTTC)_9$-CTTC-TTTC
*16	255	$(TTTC)_3$-TTTT-$(TTTC)_{10}$-CTTC-TTTC
*17	259	$(TTTC)_3$-TTTT-$(TTTC)_{11}$-CTTC-TTTC
*18	263	$(TTTC)_3$-TTTT-$(TTTC)_{12}$-CTTC-TTTC
*19	267	$(TTTC)_3$-TTTT-$(TTTC)_{13}$-CTTC-TTTC
*20	271	$(TTTC)_3$-TTTT-$(TTTC)_{14}$-CTTC-TTTC
*21	275	$(TTTC)_3$-TTTT-$(TTTC)_{15}$-CTTC-TTTC

Figure 2 D6S2691 (D6S2939): allelic ladder (calculated length in bp) and sequences. D6S2691 is a compound microsatellite based on TTTC repetition located 41-kb centromeric to the 'chromosome 6 open reading frame 15' (gene ID C6orf15 N°29113). D6S2691 is also known as: BV012631, M6S160, C2_4_4, and D6S2939. Three different primer pairs have been referenced to Major Histocompatibility Complex database; UniSTS numbers are 239082, 256731, and 464234.

between the 128-bp allele name and (GT) 16 allele name of the STR in Figure 1; the table of Figure 2 shows the equivalence between the 227-bp allele name and the (TTTC)2-TTTT-(TTTC)6 allele name. Referring to such alleles in text, we recommend as suggested in Figures 1 and 2 to separate the name of the microsatellite from the allele name by a star ('*'). D6S273*16 and D6S273*128bp, e.g., are equivalent, as well as D6S2691*9 and D6S2691*227bp.

Conflict of Interest Statement

All authors have declared no conflicts of interests.

References

1. Malkki M, Single R, Carrington M et al. MHC microsatellite diversity and linkage disequilibrium among common HLA-A, HLA-B, DRB1 haplotypes: implications for unrelated hematopoietic transplantation and disease association studies. *Tissue Antigens* 2005: **66**: 114–24.
2. Wain HM, Bruford EA, Lovering RC et al. Guidelines for human gene nomenclature. *Genomics* 2002: **79**: 464–70.
3. Gourraud PA, Mano S, Barnetche T, Carrington M, Inoko H, Cambon-Thomsen A. Integration of microsatellite characteristics in the MHC region: a literature and sequence based analysis. *Tissue Antigens* 2004: **64**: 543–55.
4. Marsh SG, Albert ED, Bodmer WF et al. Nomenclature for factors of the HLA system, 2004. *Tissue Antigens* **2005**: 65: 301–69.
5. Bär W, Brinkmann B, Budowle B et al. DNA recommendations – further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. *Forensic Sci Int* 1997: **87**: 181–4.
6. Allen F, Amos DB, Batchelor R et al. Joint report of fourth international histocompatibility workshop. In: Terasaki PI., ed. *Histocompatibility Testing 1970. Copenhagen: Munksgaard, 1970, 17–47.*
7. Allen F, Amos DB, Batchelor R et al. W.H.O. terminology report. In: Dausset J, Colombani J., eds. *Histocompatibility Testing 1972. Copenhagen: Munksgaard, 1973, 3–16.*