

Update to HLA Nomenclature, 2009

The WHO Nomenclature Committee for Factors of the HLA System met during the 15th International Histocompatibility and Immunogenetics Workshop in Buzios, Brazil in September 2008. Discussions took place during this meeting and since on a number of issues relating to HLA Nomenclature.

Current HLA Nomenclature

The convention of using a four-digit code to distinguish HLA alleles that differ in the proteins they encode was introduced in the 1987 Nomenclature Report [1]. Since that time additional digits have been added, and currently an allele name may be composed of four, six or eight digits dependant on its sequence.

The first two digits describe the allele family, which often corresponds to the serological antigen carried by the allotype. The third and fourth digits are assigned in the order in which the sequences have been determined. Alleles whose numbers differ in the first four digits must differ in one or more nucleotide substitutions that change the amino-acid sequence of the encoded protein. Alleles that differ only by synonymous nucleotide substitutions within the coding sequence are distinguished by the use of the fifth and sixth digits. Alleles that only differ by sequence polymorphisms in introns or in the 5' and 3' untranslated regions that flank the exons and introns are distinguished by the use of the seventh and eighth digits.

In 2002 we faced the issue of the A*02 and B*15 allele families having more than 100 alleles. At that time the decision taken was to name further alleles in these families in the rollover allele families A*92 and B*95 respectively. For HLA-DPB1 alleles, it was decided to assign new alleles within the existing system, hence once DPB1*9901 had been assigned, the next allele would be assigned DPB1*0102, followed by DPB1*0203, DPB1*0302 etc.

When these conventions were adopted it was anticipated that the nomenclature system would accommodate all the HLA alleles likely to be sequenced. Unfortunately this is not the case, as the number of alleles for certain genes is fast approaching the maximum possible with the current naming convention.

The following lists the major decisions taken at the by Nomenclature Committee which will be documented more fully in the full report of this meeting, to be published in Tissue Antigens.

1. HLA allele names

With the ever increasing number of HLA alleles described it has been decided to introduce colons (:) into the allele names to act as delimiters of the separate fields. It will be mandatory to include the leading zeros currently included in the alleles, this will help to lessen any confusion in the conversion to the new style of nomenclature, but no further leading zeros will be added to allele names.

Hence

| | | |
|---------------|---------|------------------|
| A*01010101 | becomes | A*01:01:01:01 |
| A*02010102L | becomes | A*02:01:01:02L |
| A*260101 | becomes | A*26:01:01 |
| A*3301 | becomes | A*33:01 |
| B*0808N | becomes | B*08:08N |
| DRB1*01010101 | becomes | DRB1*01:01:01:01 |

For allele families that have more than 100 alleles such as the A*02 and B*15 groups it will be possible to encode these in a single series. Thus the A*92 and B*95 alleles will now be renamed in to the A*02 and B*15 allele series. For example:

| | | |
|--------|---------|--------------|
| A*9201 | becomes | A*02:101 |
| A*9202 | becomes | A*02:102 |
| A*9203 | becomes | A*02:103 etc |

| | | |
|--------|---------|--------------|
| B*9501 | becomes | B*15:101 |
| B*9502 | becomes | B*15:102 |
| B*9503 | becomes | B*15:103 etc |

The names A*02:100 and B*15:100 will not be assigned. In cases of other allele families where the number of alleles reaches 100 these will be numbered sequentially, for example A*24:99 will be followed by A*24:100.

The DPB1 allele names that have been previously assigned names within the existing system will also be renamed, for example:

| | | |
|-----------|---------|-----------------|
| DPB1*0102 | becomes | DPB1*100:01 |
| DPB1*0203 | becomes | DPB1*101:01 |
| DPB1*0302 | becomes | DPB1*102:01 |
| DPB1*0403 | becomes | DPB1*103:01 |
| DPB1*0502 | becomes | DPB1*104:01 etc |

A full listing of old and new HLA allele names will be provided from the IMGT/HLA Database.

2. Naming of HLA-C antigens and alleles

The 'w' will be removed from the HLA-C allele names, but will be retained in the HLA-C antigens names, to avoid confusion with the factors of the complement system and epitopes on the HLA-C molecule often termed C1 and C2 that act as ligands for the Killer-cell Immunoglobulin-like Receptors.

| | | |
|-------------|---------|-------------------|
| Cw*0103 | becomes | C*01:03 |
| Cw*020201 | becomes | C*02:02:01 |
| Cw*07020101 | becomes | C*07:02:01:01 etc |

3. Reporting of ambiguous HLA allele typing

The level of resolution achieved by many of the HLA typing technologies employed today, does not always allow for a single HLA allele to be unambiguously assigned. Often it is only possible to resolve the presence of a number of closely related alleles. This is referred to as an ambiguous 'string' of alleles. In addition typing strategies are frequently aimed at resolving alleles that encode differences within the peptide binding domains, but fail to exclude those that differ elsewhere. For some purposes it is helpful to provide codes that aid the reporting of certain ambiguous alleles 'strings'. The decision was taken to introduce codes to allow for the easy reporting of:

a. HLA alleles that encode for identical peptide binding domains

HLA alleles having nucleotide sequences that encode the same protein sequence for the peptide binding domains (exon 2 and 3 for HLA class I and exon 2 only for HLA class II alleles) will be designated by an upper case 'P' which follows the allele designation of the lowest numbered allele in the group.

For example the string of allele names below share the same a1 and a2 domain protein sequence encoded by exons 2 and 3.

A*02:01:01:01/02:01:01:02L/02:01:01:03/02:01:02/02:01:03/02:01:04/
 02:01:05/02:01:06/02:01:07/02:01:08/02:01:09/02:01:10/02:01:11/02:01:12/
 02:01:13/02:01:14/02:01:15/02:01:17/02:01:18/02:01:19/02:01:21/02:01:22/02:09/02:66/02:
 75/02:89/02:97/02:132/02:134/02:140

This string can be reduced to A*02:01P

b. HLA alleles that share identical nucleotide sequences for the exons encoding the peptide binding domains

HLA alleles that have identical nucleotide sequences across the exons encoding the peptide binding domains (exon 2 and 3 for HLA class I and exon 2 only for HLA class II alleles) will be designated by an upper case 'G' which follows the allele designation of the lowest numbered allele in the group.

For example the string of allele names below have identical exons 2 and 3 nucleotide sequences.

A*02:01:01:01/02:01:01:02L/02:01:01:03/02:01:08/02:01:11/02:01:14/02:01:15/
02:01:21/02:09/02:43N/02:66/02:75/02:83N/02:89/02:97/02:132/02:134/02:140

This string can be reduced to A*02:01:01G

These changes to the HLA Nomenclature will be officially introduced in April 2010. Lists of old and new allele names will be made available through the IMGT/HLA Database (www.ebi.ac.uk/imgt/hla) [2] and be implemented with the April 2010 release of the database.

A full listing of all currently assigned HLA alleles and antigens together with information on the changes documented here will be published in the forthcoming report of the WHO Nomenclature Committee for Factors of the HLA System, 2009 to be published in *Tissue Antigens*.

Prepared by Steven Marsh on behalf of the WHO Nomenclature Committee for Factors of the HLA System.

SGE Marsh, ED Albert, WF Bodmer, RE Bontrop, B Dupont, HA Erlich, M Fernandez-Vina, DE Geraghty, R Holdsworth, CK Hurley, M Lau, KW Lee, B Mach, M Maiers, CH Müller WR Mayr, P Parham, EW Petersdorf, T Sasazuki, JL Strominger, A Svejgaard, PI Terasaki, JM Tiercy and J Trowsdale

References:

1. Marsh SGE, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Geraghty DE, Hansen JA, Hurley CK, Mach B, Mayr WR, Parham P, Petersdorf EW, Sasazuki T, Schreuder GMTh, Strominger JL, Svejgaard A, Terasaki PI, Trowsdale J : Nomenclature for factors of the HLA system, 2004. *Tissue Antigens* (2005) **65** 301-369, *Human Immunology* (2005) **66** 571-636, *International Journal of Immunogenetics* (2005) **32** 107-159
2. Robinson J, Malik A, Parham P, Bodmer JG, Marsh SGE : IMGT/HLA Database - sequence database for the Human Major Histocompatibility Complex. *Tissue Antigens* (2000) **55** 280-287