Update to NMDP allele code nomenclature

In April 2010, the WHO Nomenclature Committee for Factors of the HLA System will begin publishing allele names delimited by colon (':') characters to separate the meaningful parts of the name. Full details about this can be found at http://hla.alleles.org/announcement.html.

The allele code system maintained by NMDP will also be modified with this change. This document outlines what changes to the allele code system are required for as smooth a transition as possible. Since the current allele database version is in the 2.x range and the new format will start the 3.x range, we will refer to the "version 2" and "version 3" format in this document.

Representation of alleles and types in version 3

DNA types which use the allele code system will change format to include a colon between allele family and code. For example, DRB1*01AB in version 2 will be DRB1*01:AB in version 3.

Migration of existing codes to version 3 representation

1. The representation of all existing "generic" allele codes (those that consist of only 2-digit members) will be the same as in version 2.

Code	Version 2	Version 3
AB	01/02	01/02
KG	01/09/11	01/09/11

2. The representation of many "allele-specific" codes (those that consist of 4(+1)-digit members) will simply have a colon character added after the second digit of each member (including XX codes).

Code	Version 2	Version 3
AMK	0101/0201	01:01/02:01
CTN	0101/0301N	01:01/03:01N

3. The representation of some allele-specific codes will change. This includes those used for overflow allele types (A*02 in A*92 and B*15 into B*95), those affected by the upcoming renaming of DPB1 alleles and also some types whose alleles have changed names in the past.

Code	Version 2	Version 2 type	Version 3	Version 3 type
CFRG	1501/9501	B*15CFRG	15:01/15:101	B*15:CFRG
CVEG	0201/0224/9201	A*02CVEG	02:01/02:24/02:101	A*02:CVEG
BDVU	0402/0602/5101	DPB1*04BDVU	04:02/51:01/105:01	DPB1*04:BDVU
NUS	0501/0502/0709	Cw*05NUS	05:01/05:09/07:09	C*05:NUS

4. Additionally obsolete allele-specific codes will be removed in version 3. For example, since B*1522 was renamed to B*3543, and FKM is only used at B*35, it is no longer applicable and therefore will be removed. An appropriate <u>one-way</u> v2->v3 mapping table intimated below will be provided as well:

Code	Version 2	Version 2 type	Version 3	Version 3 type
FKM	1522/3517/3532	B*35FKM	17/32/43	B*35:WNF
FAV	0101/0104N/0105N	A*01FAV	01:01/01:04N	A*01:CRY
RMD	0104N/0105N	A*01RMD	01:04N	A*01:04N

The number of such cases is small (currently 233 types). These codes already should not be used anymore. A list of affected types and their mapping has been published on the NMDP Bioinformatics website (http://bioinformatics.nmdp.org/HLA/Ver3_Nomenclature_Allele_Codes/DOCS/ver_2_to_ver_3_mapping_for_removed_codes.xls)

5. Given that the allele naming practice for locus DPB1 differs from that of other loci, allele names for DPB1 overlap those of other loci. This means that some allele-specific codes have been used at other loci besides DPB1. An independent analysis determined that DPB1's name overlap had the only appreciable effect on the uniqueness of existing allele activations. For instance, code CFRG is meaningful at both B and DPB1 because all of the following alleles exist in version 2: {B*1501, B*9501, DPB1*1501, DPB1*9501} So, for codes that are activated at *DPB1 which also conflict with those at other loci,* new codes will be created for use with the DPB1 types.

Code	Version 2	Version 2 type	Version 3	Version 3 type
CFRG	1501/9501	DPB1*15CFRG	FNWN=15:01/95:01	DPB1*15:FNWN
AB	01/02	DPB1*01AB	FNUJ=01:01/100:01	DPB1*01:FNUJ

The number of such cases is small (currently 92 types). In December 2009 NMDP has published the first version the list of types subject to this policy and their mapping on the NMDP Bioinformatics website (http://bioinformatics.nmdp.org/HLA/Ver3_Nomenclature_Allele_Codes/DOCS/ver_2_to_ver_3_DPB1_mapping.xls)

This file will be updated on a regular basis until the end of the transition period (March 31, 2011).

New version 3 requirements

1. Generic codes will need to be able to have members of more than 2 digits.
2. Version 3 introduces notation for "alleles that encode for identical peptide binding domains" ('P') and "alleles that share identical nucleotide sequences for the exons encoding the peptide binding domains" ('G'). These new notations will not be included in allele codes.
3. The NMDP will accept allele names with both 'P' and 'G' designations. If an allele is reported to the NMDP with a 'G', the NMDP will retain that information, but for the purposes of searches and display on search reports will represent the allele as a 'P'. If an allele code request contains a 'P' allele name the designation will be expanded to its corresponding allele string and an allele code will be made which includes the alleles reported in the most recent IMGT database release.

4. During the transition period (April 1, 2010 to March 31, 2011) corresponding allele code lists in version 2 and version 3 will be posted on the NMDP Bioinformatics website with subsequent daily updates. Note: starting on July 1, 2010 there will exist allele codes in version 3 that reference alleles that do not exist in version 2 and therefore the allele codes will not exist in the version 2 list.
5. March 31, 2011 at 2:00 PM CST will be the last day to request the creation of an allele code in the version 2 format. On April 1, 2011, the final "version 2" format allele code list will be published on the NMDP Bioinformatics website. At that point version 2 will be "frozen": the set will not change but will continue to be available for download for use as a reference. After March 31, 2011 all requests for new allele codes will be fulfilled using the version 3 only.
6. As of April 1, 2011 allele codes for alleles in the same family that contain an allele with an expression character will be represented with generic codes in version 3. E.g. the combination A*0311N/A*0312/A*0318 would be represented by a generic code for 11N/12/18 applied at the A*03 family.
7. As of April 1, 2011 allele-specific codes in version 3 introduced before April 1, 2011 which encode alleles of the same family will become generic.

Prepared by Jon Sorbie and Martin Maiers on behalf of the NMDP and the WMDA IT Working Group Subcommittee on Allele Codes (Daniel Baier, Werner Bochtler, Jack Bakker, Steven GE Marsh, Marney Allen).

A .pdf of this document can be found at (*updated 3/17/2010*):
http://bioinformatics.nmdp.org/HLA/Ver3_Nomenclature_Allele_Codes/DOCS/ver3_nomenclature_allele_codes.pdf

version 2 to version 3 mapping table for removed codes (one-way) :
http://bioinformatics.nmdp.org/HLA/Ver3_Nomenclature_Allele_Codes/DOCS/ver_2_to_ver_3_mapping_for_removed_codes.xls

DPB1 version 2 to version 3 mapping table (two-way) (*updated 3/17/2010*):
http://bioinformatics.nmdp.org/HLA/Ver3_Nomenclature_Allele_Codes/DOCS/ver_2_to_ver_3_DPB1_mapping.xls

2010-03-09