



Wheat Annotation Working Group Summary of PAG 2007 Meeting and Follow-up Discussion

1. We discussed the need for a standardized, informative name for the BAC clones that are (will) be used in the genome sequencing effort. The proposed format for the clone name is shown below in an example:

TAACSP3BFhA_0001A23 is a specific BAC with the following specifications:

TAA defines the genus/species. Three characters are used since there was concern two would not be enough to clearly define all possible cases.

CSP defines the cultivar. Three characters since we're concerned two won't be enough in future, and to handle cultivars that already have a standard 3 letter designation. (Some have more, but those few would be shortened for the BAC names. Some have less, like Chinese Spring, so they need to be lengthened.) e.g. CSP = Chinese Spring, CNN = Cheyenne.

The third set of 3 characters defines the chromosomal source of the DNA.

Examples are:

3BF = full 3B chromosome

3BL = long arm of 3B

3BS = short arm of 3B

ALL = total genome library

146 = the library of the combined 1D, 4D, and 6D chromosomes. This is an unusual case and not likely to have many other such examples. If necessary, unique 3-character identifiers will be created.

hA defines a restriction enzyme used to make the library, in this case HindIII, with this being the first instance of a HindIII library for this genus, species, cultivar. If a second HindIII library were made from the same source, it would be hB. Lowercase first character to make more readable.



International **W**heat **G**enome **S**equencing Consortium

5207 Wyoming Road

Bethesda MD 20816, USA

Email: info@wheatgenome.org

Website: www.wheatgenome.org

The "_" separates the library name from the specific clone identification within that library. Its main function is to improve readability, instead of the continuous long stream of characters which the eye will tend to blur. Also, BAC libraries and clones within a library are quite different conceptually and in usage.

0001A23 identifies plate number and well position within the plate:
Plate no. 1, position A23 of the indicated library. Note that four digits must be used for the plate number.

Thus, TAACSP3BFhA_0001A23 is clone 0001A23 from the first HindIII library of the full 3B chromosome of Chinese Spring, *Triticum aestivum* ssp. *aestivum*.

In publications, any shorter library names can be used as appropriate for the libraries in that publication, and referenced to the full name in the Abbreviations footnote. Clones may be referred to by clone name only (ex: 0001A23) within a publication and only after being defined - unless fuller names are necessary for clarification.

The registry of BAC library names is at <http://wheat.pw.usda.gov/GG2/BAClibrary.shtml>. It includes specific registries for species abbreviations (TDD vs. TDC, *T. turgidum* ssp *dicoccoides* vs. ssp *dicoccum*) and for germplasm abbreviations.

2. The group agreed to deposit at GrainGenes GFF files of their annotation. Dave Matthews agreed to make this searchable via annotation group and BAC clone.
3. There were two updates to annotation pipelines: Philippe Leroy of INRA on the TriAnnot pipeline. Robin Buell of TIGR
4. The group agreed to include the barley community in annotation efforts, the group's name should be updated to "Triticeae Annotation Group". The list of attendees has been added to the mailing list. We will establish a link to the mailing group on the IWGSC web page.
5. Thomas Wicker and Brent Kronmiller presented data on annotation pipelines. After some more testing it was agreed to incorporate these into the overall annotation pipelines.
6. A proposed annotation nomenclature was made by Francois Sabot. The group preferred the more simplistic, informatic nomenclature. This will be incorporated into the TREP database.

