WormBase Newsletter   November 2004

WormBase moves to a three-week release cycle
WormBase has moved to having an interval of three weeks between
version releases, for a trial period until the end of 2004.

Vancouver GFP data set in WormBase
WormBase has added 827 GFP expression patterns generated by the
groups of David Baillie and Don Moerman into the WS134 release
(http://tock.bcgsc.bc.ca/cgi-bin/ce2/eprofile) of the database.
Ninety-one of these experiments have pictures. We will continue to
add new data as and when they are made available.

New   genome-wide   RNAi   data   set   based   on   an   ORFeome   library   in   WormBase
The results from the systematic RNAi screen reported in PMID: 15489339, Rual *et al*., 'Toward improving
Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library' are now available in
WormBase. There are 10951 experiments, 1066 of which show a mutant phenotype. See an example at:
http://www.wormbase.org/db/seq/rnai?name=MV_SV%3Amv_AAB02615;class=RNAi

Keep up-to-date with WormBase!
Are you interested in staying up-to-date with new features and datasets at WormBase? Do you want to
learn about and test experimental features currently under development? Subscribe to our RSS news feeds
which provide concise descriptions of news items at WormBase automatically. You will need an RSS
news reader or aggregator to view the feeds. These include programs like NetNewsWire on the Mac:
(http://ranchero.com/netnewswire/) or SharpReader on Windows (http://www.sharpreader.net/). Browsers
like FireFox (http://www.mozilla.org/products/firefox) also have RSS news reader plug-ins available. We
currently offer two different news feeds. The live site news feed describes new features and datasets as
they become available on the main WormBase server: http://www.wormbase.org/news.rss. A second news
feed on the development server describes new experimental features and datasets currently under testing:
http://dev.wormbase.org/news.rss.

Status of the WormBook project
WormBook, a collection of original reviews on a range of *C. elegans*-related topics, as well as a methods
component (WormMethods), is on target for its late-spring release. The WormBook project is gaining
momentum with approximately 27 of the 131 chapters planned for its first version submitted and currently
being reviewed.

Status of KOG data in WormBase
WormBase has finished importing the conserved eukaryotic orthology groups (KOGs), candidate two-
species orthology groups (TWOGs), and lineage-specific gene expansions (LSEs) from NCBI, so that both
worm and non-worm members are available with up-to-date accession numbers. This data set includes
16,209 up-to-date *C. elegans* proteins, with only 166 remaining obsolete *C. elegans* proteins; in addition,
there are 44,721 up-to-date non-worm proteins. The accession numbers for the non-worm proteins are
derived from Ensembl and Uniprot, with some residual Refseq identifiers where necessary. The proteins
fall into 4,852 KOGs, 225 TWOGs, and 848 LSEs.

Sequencing status of the *C. remanei* and other worm genomes

In the last year NHGRI has funded draft sequencing and assembly of 5 nematode genomes; *C. remanei, C. japonica, CB5161, Pristionchus pacificus*, and *Trichinella spiralis*. All five will be sequenced at the Genome Sequencing Center at Washington University in St. Louis (http://genome.wustl.edu/). An overview of the project and links to the sequencing proposal for the *Caenorhaditis* species is available at http://www.genome.gov/11007952 . The sequencing proposal for the basal nematodes is available at: ftp://genome.wustl.edu/pub/groups/analysis/basal_nematodes/basalnematodes.doc
The whole genome shotgun (WGS) sequencing phase for *C. remanei* was completed early in the summer. The WGS reads and assembly can be searched at http://genome.wustl.edu/blast/client.pl. Fosmid end sequences were recently completed and are in the process of being assembled with the WGS reads. The resulting assembly will then be deposited in GenBank. Two rounds of automated primer-directed sequence improvement will be done to close gaps and improve low-quality regions, followed by a final assembly and submission to GenBank. This is expected to be completed early next year. *Pristionchus* libraries have been made and will be sequenced this month. Preparation of DNA for the other three species is in progress, with sequencing scheduled for later this year and early next year. Questions about the sequencing can be sent to John Spieth (jspieth@watson.wustl.edu).

An update on the *C. elegans* literature search engine, Textpresso
Textpresso, a search engine for *C. elegans* literature which provides full text searching and a novel 'ontology query' function, is expanding to other biological domains. A version of Textpresso for 14,000 yeast papers is being built at the Saccharomyces Genome Database, Stanford, CA. Work will begin at the end of this month to build a version for fly literature at Flybase, Harvard, MA. These new databases will be available to the public in the new year.

Sequencing of Parasitic nematode genomes
The Genome Sequencing Center (GSC) at Washington University has been awarded a 4 year grant from NIH-NIAID to further its analysis of parasitic nematode genomes. Since 1999, the GSC has generated over 245,000 ESTs from 28 nematode species. Included in the new proposal is funding to improve our www.nematode.net database effort, and better integrate it with WormBase and nembase (www.nematodes.org). The purpose is to provide parasitologists and other nematologists with bioinformatics databases and tools that are user-friendly, integrated, and lasting. Project investigators include Richard Wilson (PI), Sandra Clifton, James McCarter, Robert Waterston, and Makedonka Mitreva (Group Leader). For questions, contact mmitreva@watson.wustl.edu.

Latest freeze version of WormBase
WS130 is the fourth and latest freeze version of WormBase. WS130 can be accessed at:
http://ws130.wormbase.org .