August 2010 Newsletter

http://Spbase.org

Newsletter. This newsletter issue coincides with the posting of a genome sequence update and a series of incremental changes in other parts of the web system. This is Build 5 of SpBase. As before, earlier versions of the data system can be downloaded as compressed archive files.

Genome Assembly. An update of the genome sequence assembly, called version 2.5, is now available. Baylor College of Medicine, Human Genome Sequencing Center (BCM-HGSC) produced reads from the purple sea urchin reference genome using the ABI SOLID sequencing system. This effort is considered an update since the reads were mapped to the existing version 2.1 assembly rather than newly assembling them from a combined pool of reads. The value of these high quality mate pair reads lies in their capacity to join existing scaffolds thus reducing gaps. The total number of scaffolds was reduced by about 25%. At the same time the scaffold length at which 50% of the base pairs were contained in that size or longer increase by 43Kb.

We have removed the version 2.1 genome browser from the web site and added the version 2.5 with mappings to all of the genome features previously represented. In all but the rarest cases the sequence around genes is essentially identical to the previous version. Version 0.5 remains since it is the one from which the current gene models were predicted. To identify the scaffold around your favorite gene sequence (or other feature) copy and paste the gene's nucleotide sequence into the BLAST query, choose version 2.5 genome assembly and the result will indicate the new scaffold. Or you can search in Gbrowse for the gene by name to identify its genomic region.

Gene Annotation. The information about the genes in the sea urchin genome continues to be enriched through searches against a number of sources. We add Pubmed links to gene information pages as new literature comes out. We have tried to resolve overlapping gene annotations by adding comments to all that we have found. An individual page of INTERPROSCAN results from a variety of motif databases has been linked to each gene model. We have added information from the Davidson laboratory Regulatory Gene Collection wherever applicable. This includes primer and amplicon sequences for 508 regulatory genes. We have also added another annotation object: Functional Category. These definitions are derived from the functional categories used in the original annotation publication (Developmental Biology, Volume 300, Number 1, 2006) and extended from motif assignments we have made.

Gene expression data continues to accumulate and we have changed the way we represent it at SpBase. A link on the gene information page takes you to an expression page which shows a spatial expression table and links to several different off-site databases, all of which accumulated the data in different ways. The most recent is the baseline expression for the regulatory genes in the NanoString codeset that Stefan Materna developed.

Website. Links have been added to facilitate movement between pages and gene duplications are cross-indexed in the comments. Biomart has been updated to include functional categories as an attribute.

Citation. Since SpBase is grant supported, we need to justify our usefulness. If you use our resources please cite us using the journal citation on the first page of the web site.

You are receiving this newsletter because you have previously been included in a SUDB mailing. Please let us know if you would like to be removed from our list for these very occasional newsletters. Issue #3, 8/11/2010

Email to: donghe@caltech.edu