



August 2011 Newsletter

<http://Spbase.org>

Newsletter. With this newsletter, we bring you Build 6 of SpBase incorporating a new assembly (Spur3.1) of the purple sea urchin and news about other echinoderm projects. There are also continuing incremental changes in the web system. As always, earlier versions of the data and web system can be downloaded as compressed archive files.

Genome Assembly. The Spur3.1 assembly was accepted into Genbank in July, 2011. This assembly resulted from a rainbow library sequencing strategy wherein four different shotgun libraries with nominal insert sizes of ~300bp, 1k, 3k, 5-6kb were constructed for Illumina sequencing. The assembly is now contained in 32,008 scaffolds and the scaffold size at which one half of the base pairs of the assembly are in larger pieces is 402 Kb. As with version Spur2.6 little change in sequence occurred around annotated genes. We have left Spur0.5 Gbrowse viewers on the website for historical purposes and to allow the migration of your annotations.. We have retired Spur2.6. The blast database for Spur3.1 has been installed as a menu choice on the blast page.

S. purpuratus transcriptome. A nearly complete transcriptome has been assembled by Qiang Tu for the purple sea urchin from 22 individual cDNA samples including 10 embryonic stages, 4 larval stages and 8 adult tissues. The preliminary data is included at SpBase as a Gbrowse track and as gene sequence files. All of the ambiguities have not been removed so only about 13,000 gene sequences are mounted.

Other Transcriptomes. There are now three separate transcriptome databases for other species posted at SpBase. They are derived from Baylor College of Medicine, Human Genome Sequencing Center (BCM-HGSC) submissions in support of additional genome sequencing projects: 1) a transcriptome assembly for embryonic stages of the sea star *Patiria miniata* (Project ID: 52335); 2) a transcriptome assembly for embryonic stages *Lytechinus variegatus* (Project ID: 62467) and 3) a combined assembly of our local Illumina reads with those from BCM-HGSC (Project ID: 62465) for *Eucidaris tribuloides*. These datasets are available to search through a link in the STATIC LINKS sidebar of the home page entitled "Comparative Transcriptomes". In addition to a BLAST database where you can input your own sequences, we also offer a search by name or number of the best SPU gene hit to each transcriptome sequence. These are of course incomplete, we offer them as a tool for gene discovery until more complete and better analyzed datasets become available.

Other Species. A preliminary version (Lvar0.4) of the *Lytechinus variegates* genome assembly has been completed. BCM-HGSC has kindly made it and their transcriptome data available on a BLAST server at: <http://www.hgsc.bcm.tmc.edu/blast.hgsc>. In addition the sequencing center has begun projects for *P. miniata* and *E. tribuloides* genomes. Additional Illumina sequencing for genomes of *S. franciscanus* and *Allocentrotus fragilis* has been completed and are in the assembly phase.

Website. BioMart is temporarily off-line while we install the latest version and correct errors. It will return soon. An expansion of the website into an "Echinobase" is planned to accommodate the new genomes. Stay tuned for new developments. We'll announce them by email and newsletters.

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 #5, 8/10/2011
Email to: parulk@caltech.edu