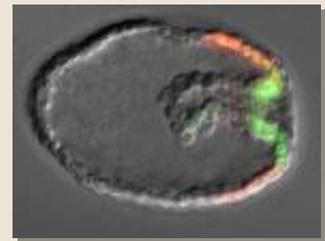


# Echinobase

An Echinoderm Genomic database



<http://Echinobase.org>

We are pleased to announce the public debut of a new web site. Echinobase (<http://Echinobase.org>) was constructed to present the information associated with other echinoderm genomes in addition to *Strongylocentrotus purpuratus* (Sp) our reference species. We have entirely redesigned the web site with new drop down menus and some new pages. We recently have added draft genome assemblies and gene annotations for *Lytechinus variegatus* (Lv) and *Patiria miniata* (Pm). Eventually at least 6 echinoderm genomes from Baylor College of Medicine, Human Genome Sequencing Center (BCM-HGSC) will be presented at Echinobase including the cidaroid sea urchin, *Eucidaris tribuloides*; the sea cucumber, *Parastichopus parvumensis* and the brittle star, *Ophiothrix spiculata*.

The basic structure of Echinobase is similar to that of SpBase (<http://spbase.org>). There is a BLAST sequence search function, a gene search and gene information section, and a genome viewer for each species. Additional information resources like methods pages are smaller for the new species. We encourage you to send us your data for inclusion in the web site.

Echinobase will replace SpBase. The SpBase web site will continue to operate but some of the information will no longer be updated there. We intend to keep SpBase in place to allow outside links to remain intact. Please move your external links to Echinobase.

The new genome sequence assemblies are derived entirely from next-generation sequencing methods at BCM-HGSC. Both the Pm and the Lv assemblies were made from combined Roche 454 and Illumina paired end sequence reads. For Pm the total assembled size is 770.5Mbp with 60,183 scaffolds (N50=52.6Kbp). For Lv the latest assembly produced 835 Mbp in 249,725 scaffolds (N50=42.6Kbp). This latter assembly is of the same scale as the *S. purpuratus* (Sp) version 0.5 used for gene model computation. Additional Pacific Bioscience sequencing is being used to improve the Sp and Lv assemblies. It is expected that Illumina and PacBio sequencing will become the standard method for draft assemblies of these large polymorphic genomes.

Protein coding gene information at Echinobase has expanded considerably. RNA-seq transcriptome information published [here](#) has been integrated into the Sp gene models. Qiang Tu's query tool is available as a quick link from the Sp Echinobase page as before. We are conducting a manual re-annotation of the Sp gene models to remove inconsistencies from the previous version consolidations. These will be posted as they are completed. For the new Lv and Pm assemblies, we have computed gene models using the Maker2 pipeline from GMOD. Although the median gene length was short for these two assemblies, we were able to verify sequences for a majority of transcription factors by reciprocal best BLAST comparisons. As transcripts accumulate and assemblies improve the gene models will also become more representative.

The genome browser GBrowse has been replaced with a JAVA-based software called JBrowse. It is much faster and easier to navigate than the older program. Adding and removing tracks is now a drag-and-drop action and sizing the window is almost instantaneous.

Citation. Since Echinobase/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 #6, 12/16/2013  
Email to: [parulk@caltech.edu](mailto:parulk@caltech.edu)