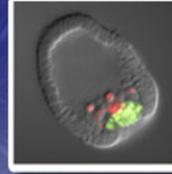# June 2009 Newsletter

Newsletter. We have decided to publish a newsletter with each new version of the SpBase information system and website. We are currently on our third version with the release made public last week. We hope that the changes are definite improvements and that you only notice them when they inadvertently contain errors. Your criticisms and suggestions are most welcome. We will try to explain the changes in this newsletter format each time a new build appears.

History. The principal task of SpBase in the first builds was to transfer the valuable annotation data from the Baylor distributed annotation website including the first 9,941 annotated gene reports. Although reviews of this information were published as Developmental Biology Volume 300, Number 1, the entered data is a unique collection assembled by more than 240 expert annotators. A number of data fields in the gene records remained incomplete and we have worked to fill those whenever possible. In addition SpBase curators have tried to standardize gene names while preserving synonyms used by investigators. In previous builds we also mapped the entire set of predicted gene models to the latest version of the genome assembly, version 2.1. When genes could not be mapped to the new assembly sequence, we included a comment in the gene information page explaining the absence. A download site for sequences and annotation data was initiated and a public bulletin board was installed under the "Community" navigation button. The BLAST databases now include one for the sequences related to our library clones and all of the EST sequences we could find.

Progress. Build 3 of SpBase contains a number of new additions including Gene Ontology terms for the annotated genes, PubMed links, and additional annotations.

We employed Gene Ontology's Standard Operating Procedures to computationally add GO terms to the genes in the Official Gene Set (GLEAN models). The InterPro scan method makes it possible to assign an electronic annotation based on InterPro domains and other criteria. We also manually scanned these annotations to remove redundancies and inconsistencies. A new block on the gene information page contains these assignments.

Starting with the references to annotated genes in the Developmental Biology issue, we have provided PubMed links on gene information pages. We used the SPU numbers to gather these links since there still were a number of gene name discrepancies to reconcile. We are beginning to add PubMed links to papers published since the annotation issue but this will be a slow process with our small staff.

We have continued to manually annotate the Official Gene Set since the website at Baylor made the 5th freeze of the data. The number of annotated genes has risen to 12,648. Only another 10,000 to go!

Feedback. We are acutely aware that the sea urchin genome information contains mistakes and missing data. As quickly as we can we correct the errors that are found or are brought to our attention by you the user. We will try to answer emails or bulletin board questions as quickly as possible. We need your help to clean up the noise that inevitably accumulates in a heap of data of this size. Please refer to the website for email addresses and bulletin board sites.