

SynBrowse: A Synteny Browser for Comparative Sequence Analysis

Xiaokang Pan¹, Lincoln Stein², and Volker Brendel¹

¹Department of Genetics, Development and Cell Biology, Iowa State University, 2112 MBB, Ames, Iowa 50011;

²Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724

We have developed SynBrowse, a synteny browser for visualizing and analyzing genome alignments both within and between species. It consists of two components: a web-based front end and a set of relational database back ends. Each database stores pre-computed protein and nucleotide alignments from a focus sequence to reference sequences in addition to genome annotations, such as gene models and EST spliced alignments of the focus sequence. The user interface lets end users select a key comparative alignment type and search for syntenic blocks between two sequences and zoom in to view the relationships among the corresponding genome annotations in fine detail.

One of the major applications of SynBrowse is the visualization and analysis of microsynteny between sequences from two or three species. This can be used for doing gene to gene comparisons within a microsyntenic region and thus performing an evolutionary analysis of a gene family. It can also be used for making detailed exon to exon comparisons among orthologous or paralogous genes and is valuable for studying intron gain and loss or other structural changes. Another application of SynBrowse is the display of macrosynteny between a focus sequence and a reference sequence. Macrosynteny between two sequences is usually recognized as a number of synteny blocks. Each block contains two or more homologous gene pairs, and the distances of two nearby homologous gene pairs on both focus and reference sequences are less than a fixed length (e.g., 100 kb). This provides for a visualization of the global relationships among the species. SynBrowse can also be used to assist in the identification of uncharacterized genes, putative regulatory elements, and novel structural features in the genomic sequences of different species by comparison to a well-annotated reference sequence. In addition, it can be applied to perform intra-species sequence comparisons and study genome duplications in a species.

SynBrowse is a GBrowse (Stein et al. 2002) family software tool. It is written in the Perl programming language and depends on the Bioperl library modules Bio::Graphics and Bio::DB::GFF (Stajich et al. 2002). It also uses modules from the GBrowse package, in particular text-based administration configuration system of that package. The software runs on top of any CGI (Common Gateway Interface) compliant web server, but has only been tested extensively using the Apache web server with Unix Linux machines. The backend can use any of a number of relational databases including MySQL, PostgreSQL and Oracle. The SynBrowse software architecture is similar to the one used by GBrowse.

SynBrowse runs on the top of GBrowse, which must be installed and configured before running SynBrowse. Like GBrowse, SynBrowse provides system administrators with simple installation, flexible configuration, convenient data input, and easy integration with other components of a model organism system. This software is available for download at <http://www.gmod.org>. Demonstrations of the software configured for plant cross-species comparisons are available at <http://www.synbrowser.org>.

REFERENCES

- Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gibert, J.G.R., Korf, I., Lapp, H., et al. (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* **12**, 1611-8.
- Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E. Stajich, J.E., Harris, T.W., Arva, A. and Lewis S. (2002) The generic genome browser: A building block for a model organism system database. *Genome Res.* **12**, 1599-1610.