BIOINFORMATICS ORIGINAL PAPER
Vol. 21 no. 17 2005, pages 3461–3468
doi:10.1093/bioinformatics/bti555

Genome analysis

SynBrowse: a synteny browser for comparative sequence analysis

Xiaokang Pan[1], Lincoln Stein[3] and Volker Brendel[1,2,*]

[1]Department of Genetics, Development and Cell Biology and [2]Department of Statistics, Iowa State University, 2112 Molecular Biology Building, Ames, IA 50011-3260, USA and [3]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

Received on February 18, 2005; revised on June 15, 2005; accepted on June 25, 2005
Advance Access publication June 30, 2005

ABSTRACT

Motivation: The recent efforts of various sequence projects to sequence deeply into various phylogenies provide great resources for comparative sequence analysis. A generic and portable tool is essential for scientists to visualize and analyze sequence comparisons.

Results: We have developed SynBrowse, a synteny browser for visualizing and analyzing genome alignments both within and between species. It is intended to help scientists study macrosynteny, microsynteny and homologous genes between sequences. It can also aid with the identification of uncharacterized genes, putative regulatory elements and novel structural features of a species. SynBrowse is a GBrowse (the Generic Genome Browser) family software tool that runs on top of the open source BioPerl modules. It consists of two components: a web-based front end and a set of relational database back ends. Each database stores pre-computed alignments from a focus sequence to reference sequences in addition to the genome annotations 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 detail. SynBrowse is portable with simple installation, flexible configuration, convenient data input and easy integration with other components of a model organism system.

Availability: The software is available at http://www.gmod.org

Contact: vbrendel@iastate.edu

INTRODUCTION

Comparative sequence analysis is a powerful approach for decoding genomic information and determining relationships among evolutionarily related species (e.g. Guiliano *et al*., 2002; Bowers *et al*., 2003). The recent efforts of various sequence projects to sequence deeply into various phylogenies, including those of mammals [e.g. human (International Human Genome Sequencing Consortium, 2004), mouse (Mouse Genome Sequencing Consortium, 2002) and rat (Rat Genome Sequencing Project Consortium, 2004)], nematodes [*Caenorhabditis elegans* and *Caenorhabditis briggsae* (Stein *et al*., 2003)], fruitflies [*Drosophila melanogaster* (Adams *et al*., 2000) and *Drosophila pseudoobscura* (Richards *et al*., 2005)], grasses [*Oryza sativa* (Yu *et al*., 2005) and *Zea mays* (http://www.maizegenome.org)] and legumes [*Medicago truncatula* and *Lotus japonicus* (Young *et al*., 2005)], provide great resources for such comparative sequence analysis.

One major focus of comparative sequence analysis is the search for 'synteny', a term that literally means 'same thread', which is used in sequence analysis to mean a set of genes and other features that share the same relative ordering on the chromosomes of two species or between duplicated chromosomes of the same species. Synteny between or among the chromosomes of a species or several species suggests evolutionary relationships of the genomes and can aid in identifying homologous genes and non-coding functional elements such as regulatory elements (Frazer *et al*., 2003). In addition, the identification of syntenic regions can aid in studying chromosome evolution and in identifying genome polyploidization and subsequent diploidization events (Wendel, 2000; Guiliano *et al*., 2002; Bowers *et al*., 2003; Cannon *et al*., 2003).

This type of analysis would be greatly aided by a visualization tool that allows both large- and small-scale synteny to be viewed in the context of the annotations on each of the genomes under study. Presently, there are few such tools that are both easily portable and available under open source terms. Ensembl SyntenyView (Clamp *et al*., 2003) displays conservation of large-scale gene order between two genomes. However, most current web-based genome browsers (Kent *et al*., 2002; Stein *et al*., 2002; Dong *et al*., 2004) allow biologists to visualize and analyze genome information only for a single genome at a time. In these single-genome browsers, cross-species comparison is limited to showing degrees of sequence conservation as a histogram or a 'VISTA Plot' (Frazer *et al*., 2004), and thus these browsers are not conducive to comparison of annotations among multiple genomes.

The Generic Genome Browser, also known as GBrowse (Stein *et al*., 2002), has become very popular in the bioinformatics community due to its portability, simple installation, flexible configuration, convenient data input and easy integration with other components of a model organism system website. Like other web-based genome annotation viewers, GBrowse is able to show regions in the sequence that are conserved in other species, but it is unable to show the relationship among annotated features for more than one genome at a time. In this paper, we describe a companion tool for GBrowse called the SynBrowse, a synteny browser for comparative sequence analysis. It was designed to help visualize cross-species sequence comparisons, such as macrosynteny, microsynteny and

*To whom correspondence should be addressed.

(a)

```
(seq_id source   method     start   end     score  strand  phase  class      targ_id  start    end
 Mt2    coding20 similarity 6663842 6664414 0.277   -       .     Target "Sequence:5" 22047197 22046595
 Mt2    coding70 similarity 6662485 6662697 0.749   -       .     Target "Sequence:5" 22045939 22045724
 Mt2    coding60 similarity 6661977 6662296 0.680   -       .     Target "Sequence:5" 22045057 22044743
 Mt2    coding80 similarity 6661615 6661723 0.847   -       .     Target "Sequence:5" 22044507 22044400
 Mt2    coding80 similarity 6661145 6661381 0.837   -       .     Target "Sequence:5" 22044316 22044080
 Mt2    coding90 similarity 6660986 6661072 0.912   -       .     Target "Sequence:5" 22043977 22043891
 Mt2    coding80 similarity 6660641 6660906 0.874   -       .     Target "Sequence:5" 22043562 22043296
 Mt2    coding70 similarity 6660266 6660518 0.795   -       .     Target "Sequence:5" 22043070 22042910
 Mt2    gene60   similarity 6660266 6664414 0.642   -       .     Target "Sequence:5" 22047197 22042910
```

(b)

```
(seq_id source   method     start    end      score  strand  phase  class       targ_id  start   end
 5      coding20 similarity 22046595 22047197 0.277   -       .     Target "Sequence:Mt2" 6664414 6663842
 5      coding70 similarity 22045724 22045939 0.749   -       .     Target "Sequence:Mt2" 6662697 6662485
 5      coding60 similarity 22044743 22045057 0.680   -       .     Target "Sequence:Mt2" 6662296 6661977
 5      coding80 similarity 22044400 22044507 0.847   -       .     Target "Sequence:Mt2" 6661723 6661615
 5      coding80 similarity 22044080 22044316 0.837   -       .     Target "Sequence:Mt2" 6661381 6661145
 5      coding90 similarity 22043891 22043977 0.912   -       .     Target "Sequence:Mt2" 6661072 6660986
 5      coding80 similarity 22043296 22043562 0.874   -       .     Target "Sequence:Mt2" 6660906 6660641
 5      coding70 similarity 22042910 22043070 0.795   -       .     Target "Sequence:Mt2" 6660518 6660266
 5      gene60   similarity 22042910 22047197 0.642   -       .     Target "Sequence:Mt2" 6664414 6660266
```

Fig. 1. Protein level alignments in the GFF format. In each record line of the GFF file, the method column is 'similarity' and the source column is 'gene' or 'coding' appended with an identity percentage in deciles. (**a**) Alignments from *Medicago* to *Arabidopsis* for *Medicago* database. (**b**) Alignments from *Arabidopsis* to *Medicago* for *Arabidopsis* database. The first nine columns are separated by tabs, while the remaining columns are separated by spaces. The 'gene60' row in (a) refers to gene AC130811_16 on *Medicago* chromosome 2, while the 'gene60' row in (b) refers to gene At5g54250 on *Arabidopsis* chromosome 5 as shown in Figure 5.

homologous genes, and to aid with the identification of novel genes and putative regulatory elements.

METHODS

Generating and formatting alignments

Either nucleotide or protein alignments can be used as the key comparison feature in SynBrowse. Choice of a proper alignment program depends on the application and is up to the user. In our examples, we used the whole genome alignment program BLASTZ (Schwartz *et al*., 2003) to generate nucleotide-level alignments and the spliced alignment program GeneSeqer (Brendel *et al*., 2004) to produce protein to genomic DNA spliced alignments. The alignment data were parsed by scripts developed for supporting SynBrowse into two GFF (General Feature Format) (http://www.synbrowse.org/pub_docs/GFF.txt) files as input into a focus species and its reference species databases, respectively. In each record line of the GFF file, we specified the entry in the method column as 'similarity' and the entry in the source column as 'gene' or 'coding' (for entirely conserved genes or single exons, respectively) appended with a similarity or identity percentage in deciles for protein alignments (Fig. 1) and as 'align' or 'conserved' (for gapped alignments or gap-free alignments, respectively) appended with a similarity or identity percentage in deciles for nucleotide alignments. The value of similarity or identity is shown in the score column.

Identification of synteny blocks

We have developed a program to find synteny blocks between two genomic sequences. This program takes the GFF alignment file of the alignment program output as the input and searches for gene (or alignment) pairs occurring in the same order in both the compared sequences at a distance less than some maximal distance, which can be defined by the user. A set of equal to or larger than a minimum number of such gene pairs is considered a synteny block. This minimum number can be specified by the user, thus allowing flexibility in defining synteny blocks both in terms of alignment quality and extent of colinearity. The derived synteny blocks are also stored in the GFF files. In each file, every record line has the same format as shown in Figure 1, where the method column is 'similarity' but the source column is 'block' for protein alignments or 'region' for nucleotide alignments, appended by an identity percentage in deciles.

Other genome annotation data and databases

All genome annotation data except the alignments described above are in the same format as required by GBrowse. All data are stored in the databases for each species according to the Bio::DB::GFF schema and are uploaded using the BioPerl scripts bulk_load_gff.pl or bp_bulk_load_gff.pl (Stein *et al*., 2002).

Configuring the system

A general configuration file, 'General.conf', is used to manage information applicable to the entire system. In this file, the system administrators are required to set the permutation of comparative species pairs, the tracks of their alignments and other global parameters.

A species-specific configuration file is applied to control the database and the display of the tracks of the comparison features for each species according to the GBrowse configuration format.

Setting up an option to display text-based protein alignments

SynBrowse has a function to generate text-based nucleotide alignments on the fly for the displayed query to reference species alignment blocks. However, it does not produce text-based protein alignments by itself. We designed an option for SynBrowse to extract protein alignments from the file system. This requires the system administrators to store the text-based protein alignments, which are generated by an independent protein alignment program, in a local directory on the server machine. In the configuration file 'General.conf', the administrators need to set the path to the location where the text-based protein alignments are stored.

Software requirements

SynBrowse runs on several software packages. The following software must be installed and configured before running SynBrowse:

(1) GBrowse 1.53 or higher (http://www.gmod.org).

(2) Perl 5.6.0 or higher (http://www.perl.org).

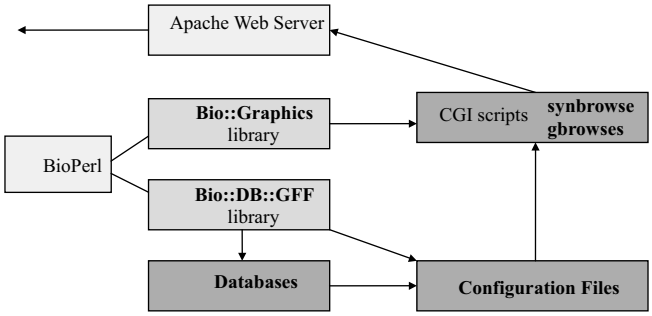

Fig. 2. Organization of SynBrowse software modules. Major and minor modules of our project are highlighted by green and light gray colors, respectively.

(3) BioPerl 1.4.0 or higher (http://www.bioperl.org).

(4) Apache Web Server 1.30 or higher (http://www.apache.org).

Also required are Perl modules GD, DBI, DBD::mysql, Digest::MD5 and Text::Shellwords, all of which are available at http://www.cpan.org

RESULTS

SynBrowse is available for download at http://www.gmod.org. Demonstrations of the software configured for plant cross-species comparisons are available at http://www.synbrowse.org

Software design

SynBrowse is written in the Perl programming language (http://www.perl.org) 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 the text-based administration configuration system of that package. The software runs on any CGI (Common Gateway Interface) compliant web server, but it has only been tested extensively using the Apache web server. The backend can use any of a number of relational databases, including MySQL (http://www.mysql.com), PostgreSQL (http://www.postgresql.org) and Oracle (http://www.oracle.com). As shown in Figure 2, the SynBrowse software architecture is similar to the one used by GBrowse, the main difference being that the entry point for the software is the CGI script 'synbrowse' rather than 'gbrowse.' SynBrowse also extends the BioPerl library in several ways in order to support the display of various alignment types.

Users who are familiar with setting up GBrowse databases will find working with SynBrowse to be similar. To create a SynBrowse website, a researcher will create tab-delimited text files describing the annotations on each of the genomes under study. The user also creates another series of text files describing the alignments among the genomes and then uses the standard GBrowse database load tools such as bp_bulk_load_gff.pl to create the database schema and load data into Bio::DB::GFF databases for each species. After these steps, SynBrowse must be configured using a series of human-readable configuration files. Each species under analysis will have a configuration file that describes its feature types and how they should be displayed (see http://www.synbrowse.org/pub_docs/Arabidopsis.conf for an example). An additional configuration file describes the datasets ('comparison features') that can be used to align the genomes to each other and sets up parameters for the website (see http://www.synbrowse.org/pub_docs/General.conf for an example). This modular design makes the software easy to install and manage and gives the researcher the flexibility of trying out multiple alternative types of alignments, such as the ones based on nucleotide or protein alignments, without rebuilding and reconfiguring the database each time.

The user interface

Once installed and run, SynBrowse provides end users with a web interface to choose cross-species sequence comparisons (Fig. 3). The end user is first prompted to select a 'focus' (or 'query') species, referring to an organism-specific set of sequences to be studied in the context of other 'reference' species. Once a focus species is selected, the interface displays all the other genomes that have been aligned to the focus species along with a list of available feature types or tracks of the focus species in a checkbox group. After a reference species is selected, the interface shows the list of the tracks for the reference species and the search field as well as the selection menus of available comparative alignment types and display levels, from which the user will select. After these preliminaries, the user can search the database for a gene, a BAC, a chromosome or another landmark in order to view patterns of global synteny (Fig. 3), detailed syntenic regions (Fig 4 and 5) or homologous gene pairs (Fig. 6).

The global synteny, detailed syntenic regions or homologous gene pairs found are graphically displayed on both the focus panel (white background) and reference panels (antique white and light yellow background). The comparisons of two sequences are based on an alignment feature created using any nucleotide- or protein-based alignment method. Both ends of a conserved region (a synteny block, a gene, an exon or a DNA alignment) between focus and reference species are connected by colored lines to show the matching regions. These vertical lines are extended through various other features such as predicted genes, expressed sequence tag (EST) spliced alignments, repeats and motifs that have been attached to the focus or reference sequences. The conserved regions between the focus sequence and different reference sequences are shown in different colors based on the level of conservation (Fig. 4).

Because comparisons between related species can result in a very busy display, users can manipulate the display by realigning, zooming in or flipping the region. Clicking on a syntenic block, a gene or another alignment on the focus panel will zoom in to a page showing a detailed view of the clicked object.

The focus panel contains one or more tracks that display the protein or nucleotide alignments used for the genome alignments. For example, line 5 on the white panel in Figure 3 holds a set of protein alignments (synteny blocks) from *Medicago* chromosome 2 to *Arabidopsis* chromosome 5. Clicking one of these blocks will zoom in to a page showing a more detailed microsynteny structure in gene-to-gene comparisons in that block.

Each reference panel also contains an alignment track, which shows the reciprocal alignment from that reference sequence to the focus sequence (e.g. the line Mt2 on the reference panel in Fig. 3). Clicking a nucleotide alignment rectangle on the reference panel will generate text-based nucleotide alignments on the fly, while clicking on the protein alignment rectangles will pop up a window showing the detailed text alignments of the homologous gene represented by the rectangles.

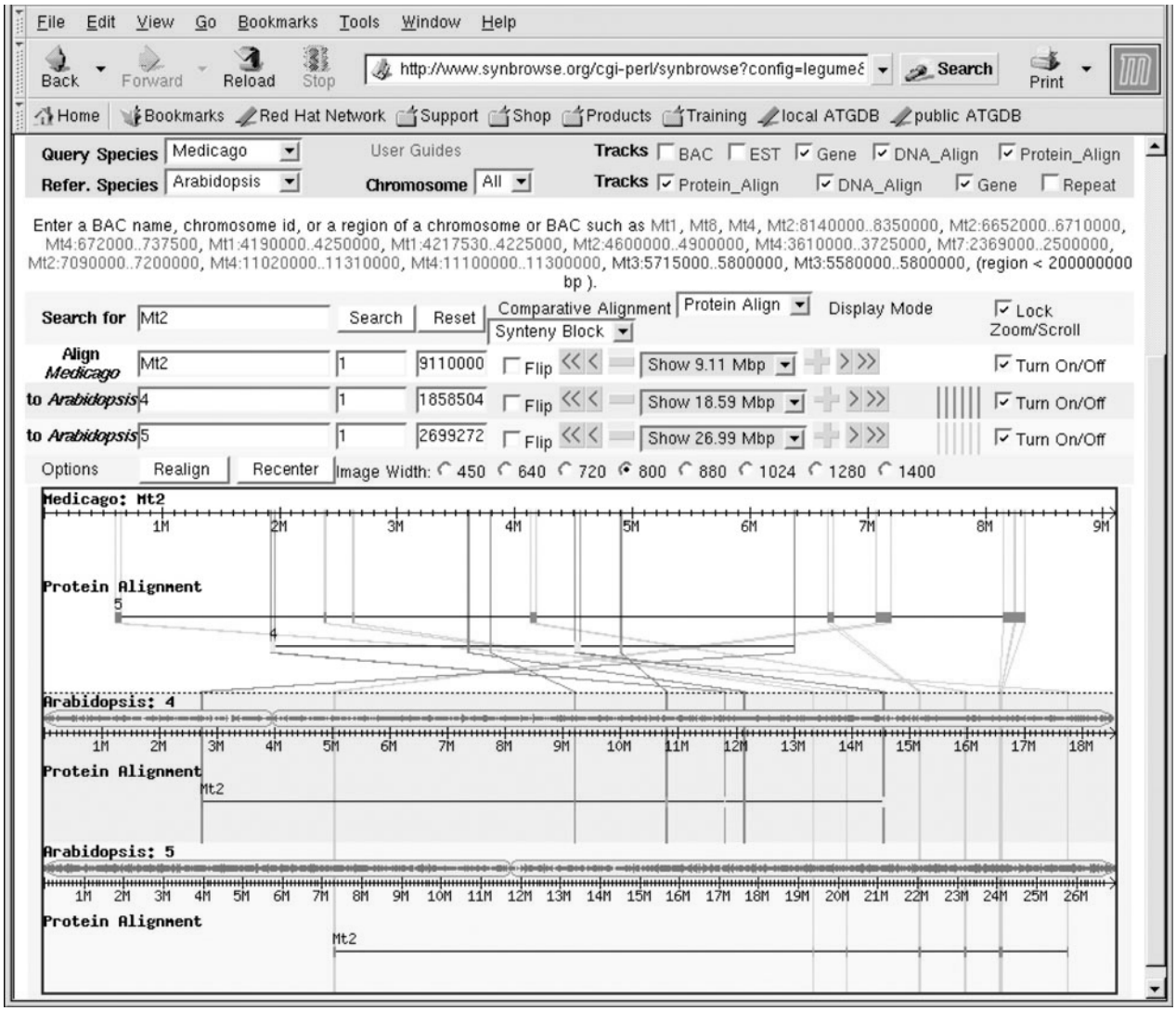

Fig. 3. A screen shot of the web-based user interface and the display of global synteny between *Medicago* chromosome 2 and *Arabidopsis* chromosomes 4 and 5, respectively.

Applications

We use comparisons between *Medicago* and *Arabidopsis* sequences to demonstrate the applications of SynBrowse. *Medicago* bacterial artificial chromosomes (BACs) of phase 2 and 3 were downloaded from GenBank and then assembled based on the overlap tables from the *Medicago* sequencing project at http://www.medicago.org/genome/. Gene structures were predicted using Fgenesh (Salamov and Solovyev, 2000), and tentative gene functions were assigned on the basis of high-scoring hits of the predicted translation products against the NCBI non-redundant protein database using BLASTP. Release 5 of the *Arabidopsis* genome and its annotation data were downloaded from the FTP site of the TIGR *Arabidopsis* genome annotation database (Wortman *et al*., 2003).

Displaying macrosynteny One of the major applications 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. By default, SynBrowse will display synteny blocks for the whole genome or other large-scale regions. This provides for visualization of the global relationships among the species and acts as the starting point for the navigation to detailed analyses. Figure 3 shows global macrosynteny between *Medicago* chromosome 2 and *Arabidopsis* chromosomes 4 and 5 in the 'Synteny Block' visualization mode. SynBrowse also allows users to view whole-genome synteny or macrosynteny by selecting either the 'Coding_Gene' (coding gene-to-gene comparison) or the 'Coding_Exon' (coding exon-to-exon comparison) as a display mode for the protein comparative alignments. SynBrowse supports an option for the system administrators to set the maximum segment that will be displayed for each display mode in the species-specific configuration files. This prevents freezing of the display in the case that there would be too many data to search in the database for a large genome region. SynBrowse also supports a 'DNA Alignment' (nucleotide alignment comparison region, which may have

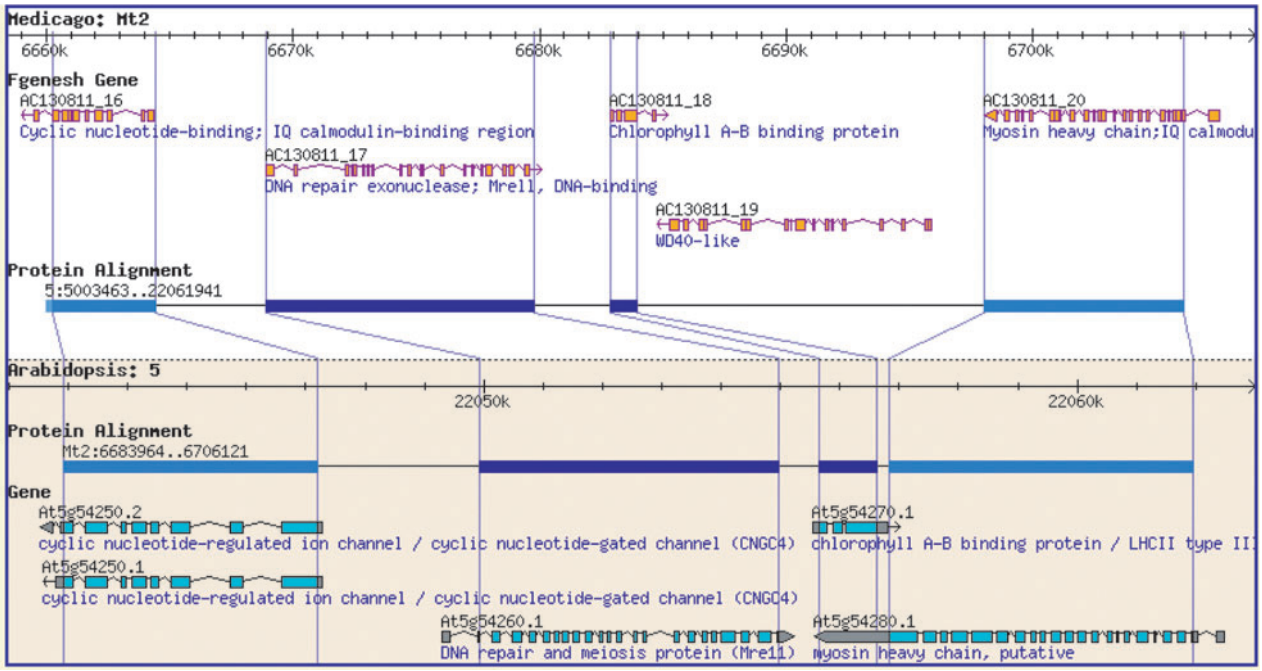

Fig. 4. A microsynteny region between *Medicago* chromosome 2 and *Arabidopsis* chromosome 5. It shows gene-to-gene comparisons using protein alignment as the key comparative alignment type. The protein level alignments are shown by rectangles with different degrees of blue color, where the dark blue represents the highest percent identity (100%), and increasingly lighter blue colors represent lower levels of identity.

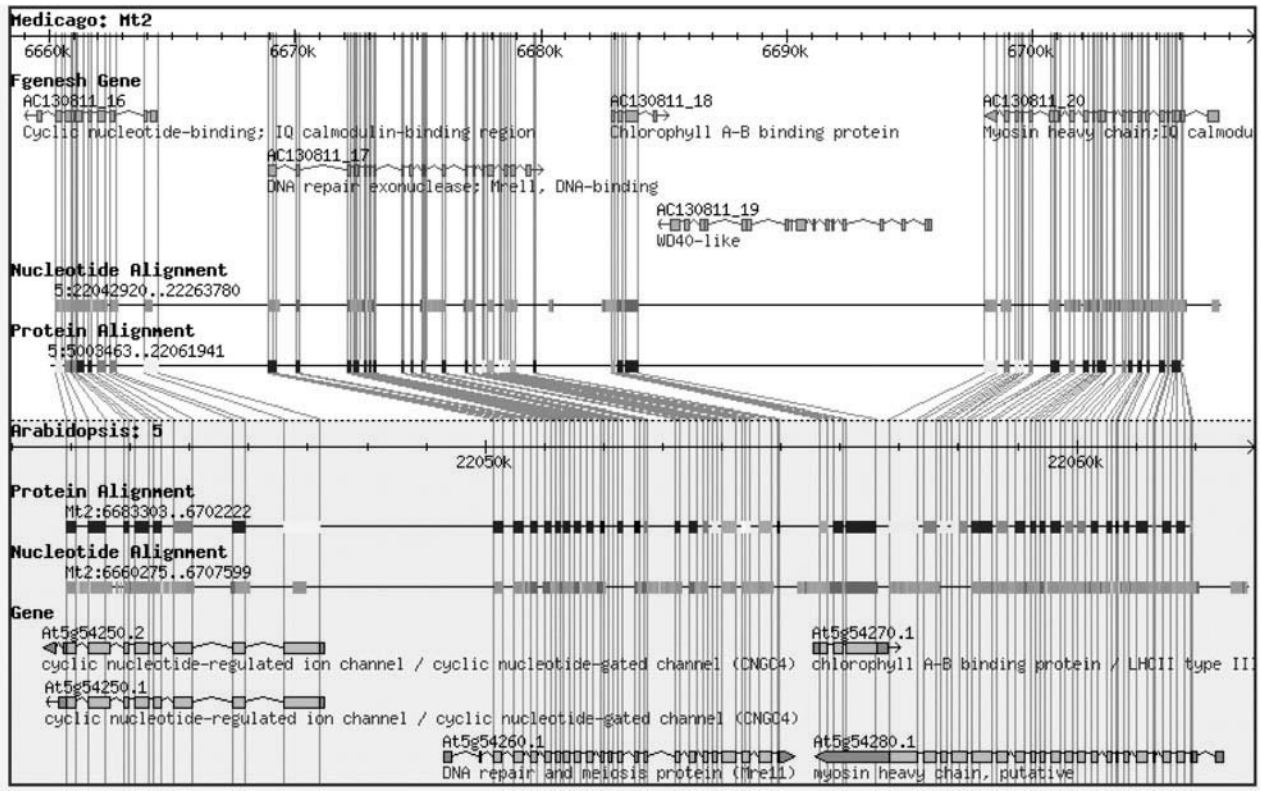

Fig. 5. A microsynteny region between *Medicago* chromosome 2 and *Arabidopsis* chromosome 5. The figure shows exon-to-exon comparisons using protein alignment as the key comparative alignment type.

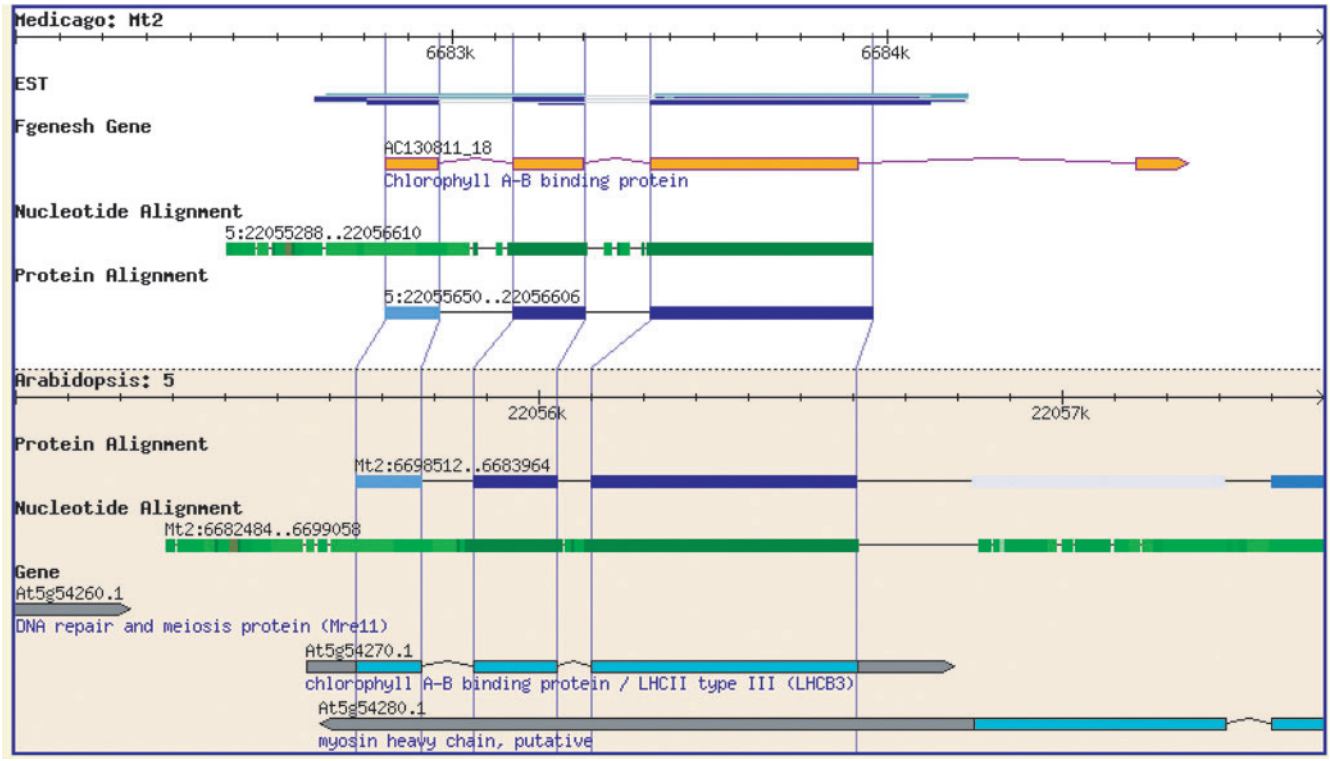

Fig. 6. A putatively homologous gene pair between *Arabidopsis thaliana* and *Medicago truncatula*. Based on the comparison, the Fgenesh prediction for *Medicago* gene AC130811_18 has a likely erroneous extra exon at the C-terminus.

gaps) or 'Conserved Piece' (a small comparison region of successive gap-free pieces of the sequence and pieces of another sequence) mode for those choosing nucleotide alignments for comparison.

Visualizing microsynteny Another major application of SynBrowse is the visualization and analysis of microsynteny between sequences from two species. This can be used for detailed gene-to-gene comparisons in a syntenic region as well as exon-to-exon comparisons among orthologous or paralogous genes and thus is valuable for studying gene evolution. In a BAC-size region (100–200 kb), non-comparative alignment features such as gene models, cognate EST spliced alignments, repeats and motifs also appear. Figure 4 shows a conserved microsyntenic region with four gene pairs conserved between *Medicago* chromosome 2 and *Arabidopsis* chromosome 5 after zooming in from Figure 3 in the 'Coding_Gene' comparison mode. In this display, the gene protein level alignments, shown by rectangles with different degrees of blue color, are used as the key comparison feature. The deepest blue represents the highest percent identity (100%), while lighter blue represents less identity. The same gene protein alignments displayed on both the focus and reference panels are connected by two light blue lines at both ends of the alignments. The two connection lines are then extended to both ends of the two conserved genes on *Lotus* chromosome 2 and *Arabidopsis* chromosome 5, respectively, showing an exact match of the conserved genes in length.

Figure 5 shows the microsynteny region with more detailed comparisons after realigning in the 'Coding_Exon' mode. This display uses protein level alignments on the exon scale, with matching segments shown by rectangles with different degrees of blue color, as the key comparison feature. Corresponding ends of each exon pair are connected by light blue lines. The nucleotide level alignments are depicted by rectangles with different degrees of green color. The deepest green represents the highest percent identity (100%), while lighter green represents less identity. These nucleotide alignments are good complementary features showing that non-coding regions, such as some introns and intergenic regions, are also conserved, although the level of their conservation is lower than that of the coding regions.

Identifying genes and motifs SynBrowse can be used to assist in the identification and characterization of genes and motifs of a species by comparing its sequence with a well-annotated reference sequence. For example, *Medicago* and *Arabidopsis* are dicot plant species. The sequencing of *Medicago* is still in progress, while *Arabidopsis* has a complete and well-annotated genome sequence. SynBrowse allows researchers to view the alignment between *Medicago* and *Arabidopsis* and to compare their gene prediction sets. Figure 6 shows the structure of *Medicago* gene AC130811_18, which was originally predicted by the Fgenesh program, in comparison with the orthologous *Arabidopsis* gene At5g54270.1 from the TIGR *Arabidopsis* annotation database. The graphical display shows an extra exon at the right-hand side in the Fgenesh prediction for the gene AC130811_18, which should be corrected by the removal of this exon. Furthermore, the gene product of At5g54270.1 has been putatively identified as

a chlorophyll A–B binding protein/LHCII type III (LHCB3), and AC130811_18 would be predicted to also belong to this family. We should note another benefit of comparative annotation, illustrated using the Figure 5 example. AC130811_18 and AC130811_20 are clearly separated by another gene, whereas their annotated *Arabidopsis* counterparts At5g54270 and At5g54280 appear to overlap. Using the AtGDB display and user annotation tools (Dong *et al*., 2004) we found that this overlap part is an annotation mistake and that the correct 3′ end of At5g54280 is further downstream, based on full-length cDNA evidence (see http://www.plantgdb.org/AtGDB and search for At5g54280).

DISCUSSION

Several comparative sequence visualization tools are publicly available. The VISTA browser (Frazer *et al*., 2004) is a global alignment-based visualization tool used to display the degree of conserved regions between two or more genomes. The PipMaker visualization system (Schwartz *et al*., 2000) is a local alignment-based visualization program that creates static percent identity plots of conserved segments between two or more sequences. Compared with SynBrowse, these two programs are less flexible with respect to input: VISTA relies on AVID nucleotide-level alignments (Bray *et al*., 2003) and PipMaker is built upon BLASTZ (Schwartz *et al*., 2003). K-BROWSER (Chakrabarti and Pachter, 2004) is a VISTA-like browser, which can visualize multiple genomes with several comparison feature tracks at a time. The Ensembl Synteny View tool (Clamp *et al*., 2003) shows syntenic blocks of conserved gene order between two or more genomes. However, it displays synteny only at the levels of macrosynteny and microsynteny and is unable to zoom in to show more detailed conserved gene order at the level of intron–exon structure. The Apollo Synteny Viewer (Lewis *et al*., 2002) is a desktop program, which visualizes synteny relationships at both the macrosynteny and microsynteny level. Artemis Comparison Tool (ACT) (http://www.sanger.ac.uk/Software/ACT/) is another sequence comparison viewer, with the capability to compare several genome regions simultaneously and to conduct microsynteny analysis at the level of intron–exon structure. However, both Apollo and ACT viewers do not run as a web application.

SynBrowse was designed as a generic and portable web-based tool for a comparative sequence analysis, which provides easy installation as a stand-alone product (e.g. http://www.synbrowse.org/cgi-perl/synbrowse?config=legume) as well as potential integration as a component of larger model organism database system, such as Gramene (Ware *et al*., 2002), WormBase (Stein *et al*., 2001) and FlyBase (The FlyBase Consortium, 2002). It uses either protein alignments or nucleotide alignments as the key comparison feature that is complemented by other features. This software lets end users search and browse comparisons between two sequences for syntenic blocks and then zoom in to a set of genes (or exons) or to nucleotide alignments of a syntenic block for more detail. It is useful in the comparative sequence analysis and evolutionary study for visualization and analysis of macrosynteny, microsynteny and homologous genes between two sequences or genomes. It can also aid in the identification of uncharacterized genes, putative regulatory elements and novel structural features in the genomic sequences of different species by comparison with a well-annotated reference sequence, enabling genome curators to refine and edit annotations of species that have incomplete genome annotations. In addition, it can be applied to perform intra-species sequence comparisons and study genome duplications in a species.

SynBrowse integrates with a previously developed software within the Generic Model Organism Database (GMOD) Project (http://www.gmod.org/). It can reuse the existing GBrowse databases and configuration files with the addition of the comparative alignments. So, it is easy to make both the SynBrowse and GBrowse software co-exist in a model organism database system. CMap, another GMOD software component (http://www.gmod.org/cmap/), has also been popular software for genetic and physical mapping and analysis in bioinformatics research community. If CMap is used as the front end for macrosynteny display and then SynBrowse is used to focus on the microsynteny display, this would be a very good combination of the applications for synteny study. SynBrowse is also complementary to Apollo, the GMOD project's genome annotation editor. For a typical model organism database system, we can edit genome annotations with Apollo and then update the SynBrowse databases. In addition, SynBrowse supports Chado, a set of schema modules for building a model organism relational database (http://www.gmod.org/schema/), because SynBrowse databases are created based on the Bio::DB::GFF schema, which was designed to accommodate the integration with GMOD modular schemas.

In the future, we will be seeking to enhance the performance and improve the graphical display of sequence comparison. A priority is to extend the visualization to allow three-way multi-species comparisons so that the focus genome occupies the central panel and the two reference genomes occupy panels above and below the query genome.

We will enhance the system so that end users can upload their own BAC or gene sequences and have the alignments of these sequences compared with the reference sequences generated automatically and then displayed graphically. In addition, we will also integrate genetic map data with SynBrowse, making SynBrowse applicable for both comparative sequence and genetic analyses.

ACKNOWLEDGEMENTS

The authors would like to thank Randy Shoemaker, David Grant and Rex Nelson for suggestions and help in the biological applications of this software during the development period and for reading this manuscript. We also thank three anonymous references for valuable suggestions to improve this manuscript. This work was supported in part by ARS-USDA funds via a subcontract from the National Center for Genome Resources to V.B.

Conflict of Interest: none declared.

REFERENCES

Adams,M.D. *et al*. (2000) The genome sequence of *Drosophila melanogaster*. *Science*, **287**, 2185–2195.

Bowers,J.E. *et al*. (2003) Unravelling angiosperm genome evolution by phylogenetics analysis of chromosomal duplication events. *Nature*, **422**, 433–438.

Bray,N. *et al*. (2003) AVID: a global alignment program for large genomic sequences. *Genome Res.*, **13**, 97–102.

Brendel,V. *et al*. (2004) Gene structure prediction from consensus spliced alignment of multiple ESTs matching the same genomic locus. *Bioinformatics*, **20**, 1157–1169.

Cannon,S.B. *et al*. (2003) Evolution and microsynteny of apyrase gene family in three legume genomes. *Mol. Gen. Genomics*, **270**, 347–361.

Chakrabarti,K. and Pachter,L. (2004) Visualization of multiple genome annotations and alignments with the K-BROWSER. *Genome Res.*, **14**, 716–720.

Clamp,M. *et al*. (2003) Ensembl 2002: accommodating comparative genomics. *Nucleic Acids Res.*, **31**, 38–42.

Dong,Q. *et al*. (2004) PlantGDB, plant genome database and analysis tools. *Nucleic Acids Res.*, **32**, D354–D359.

Frazer,K.A. *et al*. (2003) Cross-species sequence comparisons: a review of methods and available resources. *Genome Res.*, **13**, 1–12.

Frazer,K.A. *et al*. (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res.*, **32**, 273–279.

Guiliano,D.B. *et al*. (2002) Conservation of long-range synteny and microsynteny between the genomes of two distantly related nematodes. *Genome Biol.*, **3**, research0057.

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931–945.

Kent,W.J. *et al*. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.

Lewis,S.E. *et al*. (2002) Apollo: a sequence annotation editor. *Genome Biol.*, **3**, research0082.

Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.

Rat Genome Sequencing Project Consortium (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, **428**, 493–521.

Richards,S. *et al*. (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and *cis*-element evolution. *Genome Res.*, **15**, 1–18.

Salamov,A.A. and Solovyev,V.V. (2000) *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.*, **10**, 516–522.

Schwartz,S. *et al*. (2000) PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res.*, **10**, 577–586.

Schwartz,S. *et al*. (2003) Human–mouse alignments with BLASTZ. *Genome Res.*, **13**, 103–107.

Stajich,J.E. *et al*. (2002) The BioPerl toolkit: Perl modules for the life sciences. *Genome Res.*, **12**, 1611–1618.

Stein,L. *et al*. (2001) WormBase: network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res.*, **29**, 82–86.

Stein,L.D. *et al*. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.

Stein,L. *et al*. (2003) The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol.*, **1**, 166–191.

The FlyBase Consortium (2002) The FlyBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res.*, **30**, 106–108.

Ware,D.H. *et al*. (2002) Gramene: a resource for comparative grass genomics. *Nucleic Acids Res.*, **30**, 103–105.

Wortman,J.R. *et al*. (2003) Annotation of the *Arabidopsis* genome. *Plant Physiol.*, **132**, 461–468.

Wendel,J.F. (2000) Genome evolution in polyploids. *Plant Mol. Biol.*, **42**, 225–249.

Young,N.D. *et al*. (2005) Sequencing the genespaces of *Medicago truncatula and Lotus japonicus*. *Plant Physiol.*, **137**, 1174–1180.

Yu,J. *et al*. (2005) The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol.*, **3**, 266–281.