Genome analysis

The SGN comparative map viewer

Lukas A. Mueller*, Adri A. Mills, Beth Skwarecki, Robert M. Buels, Naama Menda and Steven D. Tanksley

Department of Plant Breeding and Genetics, Cornell University, Ithaca NY 14853, USA

Received on October 19, 2007; revised on November 17, 2007; accepted on November 27, 2007

Associate Editor: Alex Bateman

ABSTRACT

Motivation: With the rapid accumulation of genetic data for a multitude of different species, the availability of intuitive comparative genomic tools becomes an important requirement for the research community. Here we describe a web-based comparative viewer for mapping data, including genetic, physical and cytological maps, that is part of the SGN website (http://sgn.cornell.edu/) but that can also be installed and adapted for other websites. In addition to viewing and comparing different maps stored in the SGN database, the viewer allows users to upload their own maps and compare them to other maps in the system. The viewer is implemented in object oriented Perl, with a simple extensible interface to write data adapters for other relational database schemas and flat file formats.

Contact: lam87@cornell.edu

1 INTRODUCTION

Genetic maps are an important tool for gaining insights into the genome structure of organisms, facilitating the cloning of novel genes, are the basis of quantitative trait loci (QTLs), and are indispensable in applications such as molecular breeding. By comparing appropriately constructed maps, it is often possible to transfer and relate knowledge, such as positions of QTLs, from one species to another (De Jong et al., 2004) and to shed light on the evolution of a clade of species (Doganlar et al., 2002; Tanksley et al., 1992). With growing sequence databases, particularly EST sequences for many plant species (Van der Hoeven et al., 2002), PCR-based markers have become increasingly popular. Systematic efforts have been undertaken to generate PCR-based comparative genetic maps in several clades of plants, particularly the Solanaceae. CAPS-based (Konieczny and Ausubel, 1993) Conserved Ortholog Set (COS) markers specifically designed to work on orthologs from multiple species have been recently described (Fulton et al., 2002; Wu et al., 2006). COS marker-based genetic maps for different species can readily be compared and the evolutionary history of larger-scale events, such as inversions, exchange of chromosomal segments and other rearrangements, deduced (Wu et al., 2006). The motivation for the comparative viewer presented here was to provide a user-friendly tool to compare the large number of high quality COS comparative maps that are beginning to emerge from different labs for plants in the Euasterid clade on which SGN focuses.

In recent years, a number of comparative map viewers have been created, some web-based and some as stand-alone applications. One of the first web-based programs was a program called cMap running at MaizeDB (Fang et al., 2003), which had the ability to compare two maps side-by-side. More recently, the Gramene site (http://www.gramene.org/) has introduced a powerful comparative mapping program that is also called cMap (Jaiswal et al., 2006) which is used on other sites as well. Genbank’s mapviewer (Wheeler, 2007) also supports comparative functionalities. In addition, stand-alone applications for the display and comparison of genetic maps have been created by a number of authors, such as the CMTV program (Sawkins et al., 2004).

Here we describe a web-based comparative map viewer that has an intuitive user interface, yet provides powerful features and can run off a relational database backend. The program has been specifically designed for the purposes of the SGN website (http://sgn.cornell.edu/; Mueller et al., 2005), a clade-oriented database (COD) for the Solanaceae and related species. Although the viewer program code is integrated into the SGN code base, it has been designed in a modular form, and a stand-alone version is available for download (see ftp://ftp.sgn.cornell.edu/programs/).

2 DESCRIPTION

One of the main design goals of the comparative viewer’s user-interface was to make it easy to explore the data. The user must first select a map from the ‘maps’ menu, and is presented with an overview of the map, showing all chromosomes in the map as small glyphs and displaying a summary paragraph and statistics about the map and markers (Fig. 1A). The overview graph can be used to locate markers by entering marker names in the text field provided and clicking the ‘highlight’ button. The map can be explored in more detail by clicking on one of the chromosome glyphs, which will bring up the selected chromosome as the reference in the chromosome view (Fig. 1B). In this view, the chromosome is displayed with only about a dozen markers displaying their labels to help orient the user on the map (on most maps, there are too many markers to display them all at once). A toolbar is also displayed beneath the map, which allows to manipulate the map images (zooming in and out, enlarging and shrinking the images, and comparing appropriately constructed maps, it is often possible to transfer and relate knowledge, such as positions of QTLs, from one species to another (De Jong et al., 2004) and to shed light on the evolution of a clade of species (Doganlar et al., 2002; Tanksley et al., 1992). With growing sequence databases, particularly EST sequences for many plant species (Van der Hoeven et al., 2002), PCR-based markers have become increasingly popular. Systematic efforts have been undertaken to generate PCR-based comparative genetic maps in several clades of plants, particularly the Solanaceae. CAPS-based (Konieczny and Ausubel, 1993) Conserved Ortholog Set (COS) markers specifically designed to work on orthologs from multiple species have been recently described (Fulton et al., 2002; Wu et al., 2006). COS marker-based genetic maps for different species can readily be compared and the evolutionary history of larger-scale events, such as inversions, exchange of chromosomal segments and other rearrangements, deduced (Wu et al., 2006). The motivation for the comparative viewer presented here was to provide a user-friendly tool to compare the large number of high quality COS comparative maps that are beginning to emerge from different labs for plants in the Euasterid clade on which SGN focuses.

In recent years, a number of comparative map viewers have been created, some web-based and some as stand-alone applications. One of the first web-based programs was a program called cMap running at MaizeDB (Fang et al., 2003), which had the ability to compare two maps side-by-side. More recently, the Gramene site (http://www.gramene.org/) has introduced a powerful comparative mapping program that is also called cMap (Jaiswal et al., 2006) which is used on other sites as well. Genbank’s mapviewer (Wheeler, 2007) also supports comparative functionalities. In addition, stand-alone applications for the display and comparison of genetic maps have been created by a number of authors, such as the CMTV program (Sawkins et al., 2004).

Here we describe a web-based comparative map viewer that has an intuitive user interface, yet provides powerful features and can run off a relational database backend. The program has been specifically designed for the purposes of the SGN website (http://sgn.cornell.edu/; Mueller et al., 2005), a clade-oriented database (COD) for the Solanaceae and related species. Although the viewer program code is integrated into the SGN code base, it has been designed in a modular form, and a stand-alone version is available for download (see ftp://ftp.sgn.cornell.edu/programs/).

One of the main design goals of the comparative viewer’s user-interface was to make it easy to explore the data. The user must first select a map from the ‘maps’ menu, and is presented with an overview of the map, showing all chromosomes in the map as small glyphs and displaying a summary paragraph and statistics about the map and markers (Fig. 1A). The overview graph can be used to locate markers by entering marker names in the text field provided and clicking the ‘highlight’ button. The map can be explored in more detail by clicking on one of the chromosome glyphs, which will bring up the selected chromosome as the reference in the chromosome view (Fig. 1B). In this view, the chromosome is displayed with only about a dozen markers displaying their labels to help orient the user on the map (on most maps, there are too many markers to display them all at once). A toolbar is also displayed beneath the map, which allows to manipulate the map images (zooming in and out, enlarging and shrinking the images,
jumping to other chromosomes and maps, highlight markers on the map, etc). The user can also click on the chromosome to obtain a zoomed-in view, in which all the markers are displayed, or can choose a comparison map from a pull-down menu in the toolbar. The comparison map menu only displays linkage groups from other maps that share markers with the reference chromosome, ordered by number of markers shared, the linkage group sharing the most markers being displayed on the top. Selecting a comparison map will bring up another map to the right of the reference map, with the common markers of both maps shown on the comparison map, and lines connecting the common markers of the two maps (Fig. 1B).

Entire maps can be compared to each other by clicking on the 'compare entire maps' link in the chromosome view. This will align all linkage groups of a map vertically, and show all the connections between linkage groups of the maps. Up to three maps can be compared in this mode at the same time; the maps to be displayed can be selected from pull-down menus on the top of the page (Fig. 1D).

The map viewer currently supports a number of different types of maps: genetic maps, physical maps, IL maps and cytogenetic maps (Fig. 1C). Different map types can be compared to each other, if appropriate. Maps are versioned, and it is possible to compare different map versions to each other. Currently, 16 maps comprising a total of 197 linkage groups and over 7000 markers are in the SGN database. The SGN mapviewer receives about 12 000 hits every month, about 10% of all the hits of the SGN website.

### 3 METHODS

#### 3.1 Implementation

The program is implemented in object-oriented Perl and consists of about 60 classes with more than 20 000 lines of code including comments. To make the program more portable, the database calls have been factored out into special classes that can easily be implemented for other database schemas or other backend systems such as flat files, while the rest of the viewer code will need no adaptation. An adapter for the popular cMap database is also available (Jaiswal et al., 2006).

#### 3.2 Code availability

The code is freely available on SGN’s FTP site at ftp://sgn.cornell.edu/programs/. The code is released under the liberal MIT license (http://www.opensource.org/licenses/mit-license.php).

### ACKNOWLEDGEMENTS

We would like to thank two anonymous reviewers for their careful reading and suggestions. SGN is supported by the NSF (DBI-0421634) and the USDA #2007-02777.

Conflict of Interest: none declared.

### REFERENCES


