CMap 1.01: a comparative mapping application for the Internet

Ken Youens-Clark[1],[*], Ben Faga[1], Immanuel V. Yap[2], Lincoln Stein[1] and Doreen Ware[1],[3]

[1]Cold Spring Harbor Laboratories, Cold Spring Harbor, NY 11724, [2]Department of Plant Breeding and Genetics, Cornell University, Ithaca and [3]USDA-ARS NAA Plant, Soil & Nutrition Laboratory Research Unit, Ithaca, NY 14853, USA

ABSTRACT

Summary: CMap is a web-based tool for displaying and comparing maps of any type and from any species. A user can compare an unlimited number of maps, view pair-wise comparisons of known correspondences, and search for maps or for features by name, species, type and accession. CMap is freely available, can run on a variety of database engines and uses only free and open software components.

Availability: http://www.gmod.org/cmap

Contact: kclark@cshl.edu

1 INTRODUCTION

CMap is a generic and extensible comparative map viewer that runs in standard web browsers and aims to assist biological researchers seeking to extrapolate known map data into unknown areas. Comparison of genetic, physical and sequence maps allows researchers to fill in gaps and extend knowledge both within and across species. For example, comparison of map fragments such as FPC contigs to an assembled sequence map or high-quality genetic map can help order, orient and assemble the fragments. Feature order in one map can help aid the selection of additional markers for mapping in another population of interest.

CMap is used by the Gramene project (Liang *et al.*, 2008) to visualize and compare over 200 map sets of various types from 30 plant species; it is also used by many other projects comparing data from plants and animals. Data providers can extensively customize CMap to suit their tastes, configuring the definitions of species, map types, how maps are grouped into sets, how maps are drawn, the types of features displayed, their positions on the maps, how the features are drawn, what correspondences are made between features, how correspondences between maps are aggregated and colored, the evidence codes supporting these correspondences, and more. CMap relies on a relational database and open-source software.

The history of interactive graphical maps includes an early version from AceDB (Stein *et al.*, 1999) which included 'Multi-map' to display comparisons. The similarly named 'cMap' application (Fang *et al.*, 2003) from the MaizeDB was one of the first comparative map viewers to allow cross-species comparisons, but the application appears to be unavailable. National center for genome resources (NCGR) also created the comparative map and trait viewer (CMTV)

(Sawkins *et al.*, 2004) to allow multiple cross-species comparisons. The SOL genomics network (SGN) comparative map (Mueller *et al.*, 2008) viewer is perhaps the closest in features to CMap, but it allows a user to compare only two maps at a time, a limitation not shared by this software. CMap was originally written in 2001 for Gramene, a comparative mapping resource for crop grasses, and has since been contributed to the generic model organism web site (GMOD) project under the GNU Public License. CMap has been downloaded well over 1000 times and adapted by many other groups working on a range of different organisms, such as the legume information system, CottonDB, GrainGenes, the nematode Pristionchus and BeeBase. This article discusses version 1.01 of CMap released on July 1, 2008.

2 METHODS

2.1 Data concepts

The main data components in CMap are species, map types, map sets, maps, features and correspondences. The data administrator decides what constitutes each. Any species or map type is allowed. A map set is simply a collection of maps, and maps are any linear ordered set of features such as a linkage group, chromosome or an FPC contig. A feature is any point or interval positioned on a map such as a genetic marker, an in/del, a centromere, a QTL or a gene. A correspondence can be anything from a shared synonym to a sequence similarity such as a BLAST hit. All data can be loaded from tab-delimited files or manually inserted with tools provided in the CMap distribution.

2.2 Using CMap

To use CMap, the researcher will find there are four main paths for entering the comparative map viewer: the map viewer, the feature search, the correspondence matrix and the map search.

If the user wishes to view a particular map or map set, then the map viewer is a logical point of entry. After clicking the 'Maps' link in the CMap menu, the user is presented with a form to select a species and a map set. The user may choose to view any number of maps in the set to serve as the starting point. Next, the user may select from lists of comparative maps to place on the right and/or left of the reference map(s). The number of correspondences to each comparative map is shown in square brackets to aid the user in deciding which maps to select. After adding comparative maps, the user may choose to add new maps to the right and/or left of the outermost maps; remove, flip or crop maps; change which features and/or labels are displayed; change the size of the image, and more.

If the user has a set of features (markers, QTL, BACs, etc.) he wishes to locate on maps, then he can search for features by clicking on the 'Feature Search' link in the CMap menu. For each feature found, the

---

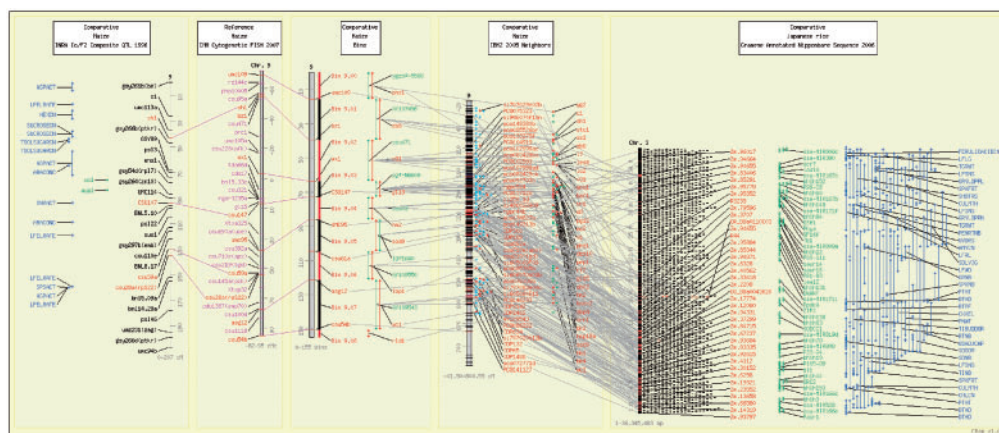[*]To whom correspondence should be addressed.

Fig. 1. Many of the key concepts of CMap are shown. Five maps of varying types from QTL to genetic to sequence and from two species are displayed, and more could be added. Map features range from QTLs to genetic markers to bins to genes, and correspondences based on different types of evidence are show in varying colors (http://www.gramene.org/db/cmap).

feature's name, type, species, map set, map, position and aliases are displayed as well as two links for each feature, one that takes the user to the 'map details' page for the map on which the feature occurs with the feature highlighted and another which takes the user to the 'feature details' page showing all the known information on the feature and all of its correspondences.

If the user is interested to know the total number of correspondences from any map set to any other map set, then he can click on the 'Matrix' link in the CMap menu. A table shows a cross-tabulated comparison of all the correspondences among map sets. By clicking on any map set name, the user can limit the matrix to a particular map set. Clicking on the numbers of correspondences takes the user to the map viewer showing the two maps in relation to each other.

The last main method to enter CMap is by choosing the 'Map Search' link in the CMap menu. Here, a user can search for particular maps based on the map name or the number of related maps. The results include the map names, the number of related maps, the number of correspondences and the number of features by type present on the map. Clicking a map's name takes the user to the map viewer page so he may add comparative maps.

In addition to the above four most direct routes to view comparative maps, other entry points are available through the CMap menu. The 'Species' link allows the user to browse the available species and all the map sets from each. From there or from the CMap menu, the user can choose the 'Map Set Info' link to learn more about a particular map set, and from there choose to view one or all of the maps in the map viewer or the matrix. In addition, the user may choose the links for 'Feature Types', 'Map Types' or 'Evidence Types'. Depending on the questions that the user seeks to answer, other routes may prove more revealing. Lastly, a proactive data provider may choose to present pre-formulated views which he knows will be of interest to his community of users by simply copying and pasting a URL into a web page and making a link with text explaining the view or storing this via the 'Saved Links' section of CMap, a function that also allows user to recall previous views.

The data provider has the ability to create multiple CMap databases that are entirely separate from each other. If this has been done, the user can switch among these databases using the 'Data Source' control in the upper-right corner. Users can download data of maps or whole map sets in GFF or CMap's tab-delimited or XML formats. Links to the download page are available on the 'Map Set Info' page and from the map menu buttons in the map viewer. There is also a 'Help' section and a 'Tutorial' to show users how CMap works and its terminology.

2.3 Implementation

CMap is a Perl application that runs on an Apache web server (versions 1 or 2) on Windows and UNIX variants. CMap has a simple relational schema that can be implemented in MySQL, Oracle, PostgreSQL or Sybase. It relies on no proprietary software or SQL extensions, and uses all freely available software. The user interface is a basic HTML/JavaScript page that works with any modern web browser. No registration or permission is required for its use.

Administration of CMap is accomplished through the following three methods: simple text configuration files define system settings such as databases and directories for templates, sessions and temporary files as well as how map, features and correspondences are drawn. The 'cmap_admin.pl' tool is used for importing and exporting data in various text formats (tab, XML, SQL), creating and deleting maps or correspondences, reloading the correspondence matrix or purging CMap's web data cache after loading new data. Lastly, a browser-based administrative tool allows the data provider a basic CRUD (Create/Read/Update/Delete) interface for all the different objects (map sets, maps, features, etc.) in the database.

2.4 Future plans

CMap continues development as a GMOD project. Work is underway on version 2.0 which will feature a major rewrite of the internals, an improved and streamlined database, graphical output in scalable vector graphics, and integration of the Circos circular genome viewer (Krzywinski *et al.*, 2009).

REFERENCES

Fang,Z. *et al.* (2003) cMap: the comparative genetic map viewer. *Bioinformatics*, **19**, 416–417.

Krzywinski,M. *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.* [Epub ahead of print, doi: 10.1101/gr.092759.109, July 24, 2009].

Liang,C. *et al.* (2008) Gramene: a growing plant comparative genomics resource. *Nucleic Acids Res.*, **36**, D947–D953.

Mueller,L. *et al.* (2008) The SGN comparative map viewer. *Bioinformatics*, **24**, 422–423.

Sawkins,M.C. *et al.* (2004) Comparative map and trait viewer (CMTV): an integrated bioinformatic tool to construct consensus maps and compare QTL and functional genomics data across genomes and experiments. *Plant Mol. Biol.*, **56**, 465–480.

Stein,L.D. *et al.* (1999) AceDB: a genome database management system. *Comput. Sci. Engineering*, **1**, 44–52.