cMap: the comparative genetic map viewer

Z. Fang [1,*], M. Polacco [1, 2,*], S. Chen [3,†], S. Schroeder [1], D. Hancock [1], H. Sanchez [1] and E. Coe [1, 2]

[1]Department of Agronomy, University of Missouri, Columbia, MO 65211, USA, [2]Plant Genetics Research Unit, Agricultural Research Service, United States Department of Agriculture, Columbia, Missouri and [3]Computer Engineering & Computer Science, University of Missouri, Columbia, MO 65211, USA

ABSTRACT

Summary: cMap, a www comparative genetic map graphical utility, has a search capability and provides comparison of two genetic maps within or between species with dynamic links to data resources and text lists of the shared loci, running in a relational database environment. Currently, maps from three species (maize 'Zea mays L.', rice 'Oryza sativa L.', and sorghum 'Sorghum bicolor L.'), representing over 13 800 distinct loci, are available for comparison at http://www.agron.missouri.edu/cMapDB/cMap.html.
Availability: cMap source code is available without cost on request for non-commercial use.
Contact: fangz@missouri.edu

The cMap prototype was first displayed at the Plant and Animal Genome Meetings in San Diego in January 2001 (http://www.intl-pag.org/pag/9/abstracts/C01_05.html) and was subsequently deployed in Spring 2001. At the time of design, there were few tools available for www genetic map comparisons. Many useful displays have been developed to support an ACEDB data format; these include a comparative map implementation by RiceGenes at Cornell University and the suite of display tools under development at the John Innes Centre (JIC; http://ukcrop.net/software/displays.html), but not yet deployed by that site. The JIC Pairwise Chromosome Map is similar in many respects to cMap, but does not support a search function. It is unclear whether tools integrated with ACEDB data formats are readily portable to a relational database management system. An interesting map comparison utility, prototyped as part of the ISYS bioinformatics tool suite by the National Center for Genome Resources (NCGR; http://www.ncgr.org), is not www based, but does allow dynamic genetic map comparisons of up to three maps and can draw on data from MapMaker outputs (personal communication, Andrew Farmer, NCGR). The National Center for Biotechnology Information (NCBI) Map Viewer (http://www.ncbi.nlm. nih.gov/PMGifs/Genomes/PlantList.html), as currently deployed, presents a searchable comparison of genetic maps for many different species, with links out to GenBank sequences and to genome-specific databases; unlike cMap, the NCBI viewer does not support cross-species comparisons, or the display of comparable intra-species loci when on distinct chromosomes. Genetic map displays are distinct from various comparative genome tools that are based on sequence alignments rather than genetic mapping probes or phenotypic variations. Sequence-based displays include ENSEMBL (http://www.ensembl.org) and the University of California Santa Cruz Genome viewer (http://genome.ucsc.edu). Gramene has employed ENSEMBL to display alignments for cereal grain cDNAs to the sequenced regions of the rice physical map (http://www.gramene.org). Unlike cMap, ENSEMBL does not operate in a relational database environment.

The Genome Information Orderly Tool (GIOT), developed by the Rice Genome Research Program (RGP, Japan; http://rgp.dna.affrc.go.jp) and implemented in 1999, displays an aligned genetic and physical map. It includes many features useful to viewing maps in the 'native' ACEDB (scroll, zoom, link-out to underlying data), providing both general and fine-grained comparative analyses. It has snappy performance over the www. Detailed information about a locus is provided in a pop-up window that has links to external data repositories and screening images. Importantly, the GIOT user interface reads data from a succinct, easily updatable tabular file. By adapting the genetic map display portion of GIOT, we have developed the cMap to run in a relational database environment and have incorporated color coding to distinguish loci associated with the same probe or sequence (in magenta) and loci that are identical (in red). Options include text output and drawing lines for all comparable loci (Figure 1). Maps available for comparison consist

*To whom correspondence should be addressed.
† Current address: Computer & Information Science & Engineering, University of Florida, Gainesville, FL 32611, USA
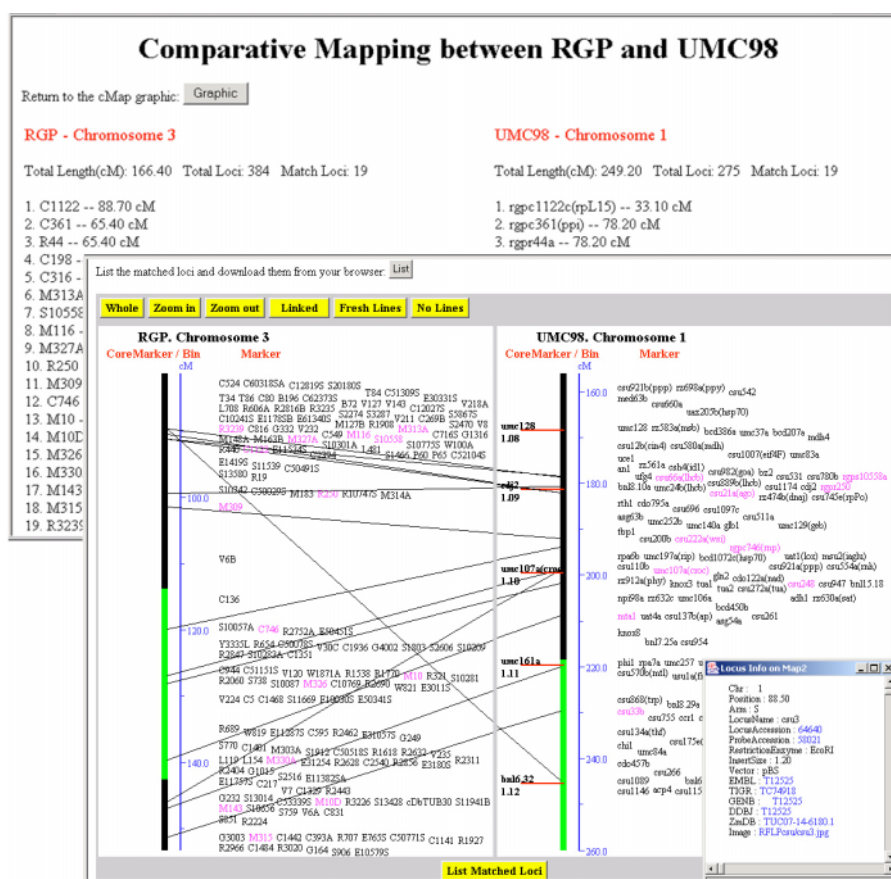
Fig. 1. An inter-species comparison between rice chromosome 3 and maize chromosome 1 using cMap. Top: a text listing of similar loci with their map coordinates on both maps. Bottom: graphic view of the two maps with similar loci in magenta; the pop-up window (right corner at the bottom) presents a few details about locus *csu3* from maize with links to data resources.

of one rice map (3267 loci) from the RGP, one sorghum map (2533 loci) kindly provided by Andrew Paterson and John Bowers of the University of Georgia, and nine maize maps (over 8000 loci), including maps from Ben Burr (BNL 2002; http://burr.bio.bnl.gov); the Maize Mapping Project (IBM; http://www.maizemap.org) and the University of Missouri (IBM, UMC 98, SSR maps); MaizeDB (IBM neighbors); Pioneer Hi-Bred International Inc. (Pioneer99), and Jeanne Romero-Severson of Purdue University (SSR Consensus). All data were semantically checked by MaizeDB (http://www.agron.missouri.edu) before loading into the cMapDB. In the near future, enhancements will include links to TIGR, Gramene, and AGI; anticipated new maps will include maize QTL maps and the 6000 marker rice EST genetic map (RGP).

The cMap has a three-tiered architecture: a relational database cMapDB, the user interface adapted from GIOT described above, and an intermediary data retriever. Entities in the data model represent: (1) core markers and bins; (2) mapped loci, probes, and sequences; and (3) individual maps. Essential attributes in the map entity include linkage group, locus, and position (bin, cM or LOD scores). Error values are not included. All data for two compared maps, except for links which are read from a file, are directly retrieved from the cMapDB and displayed without a pre-calculated table. The intermediary data retriever (a Java socket) communicates between the database and the interface. Switching to Oracle or Microsoft SQL Server, instead of Sybase in the current deployment, would be largely a matter of choosing the appropriate retriever. The cMap configuration also includes a web server (Apache 1.3.3), Perl (5.005.02), and Java (JDK 1.2).

To aid in selection of maps for comparison, the cMap includes a search utility, written in Perl/CGI, that reports all mapped loci and positions relevant to a particular locus, probe, or sequence accession (GenBank/EMBL/DDBJ).

ACKNOWLEDGEMENTS