

Live Action: Demonstrations of GCP Bioinformatics Products

CMTV: The Comparative Map and Trait Viewer as a Platform for the Visual Integration of Disparate GCP Data

Presented by Andrew Farmer, NCGR

The Comparative Map and Trait Viewer (CMTV) is an open source Java user interface for:

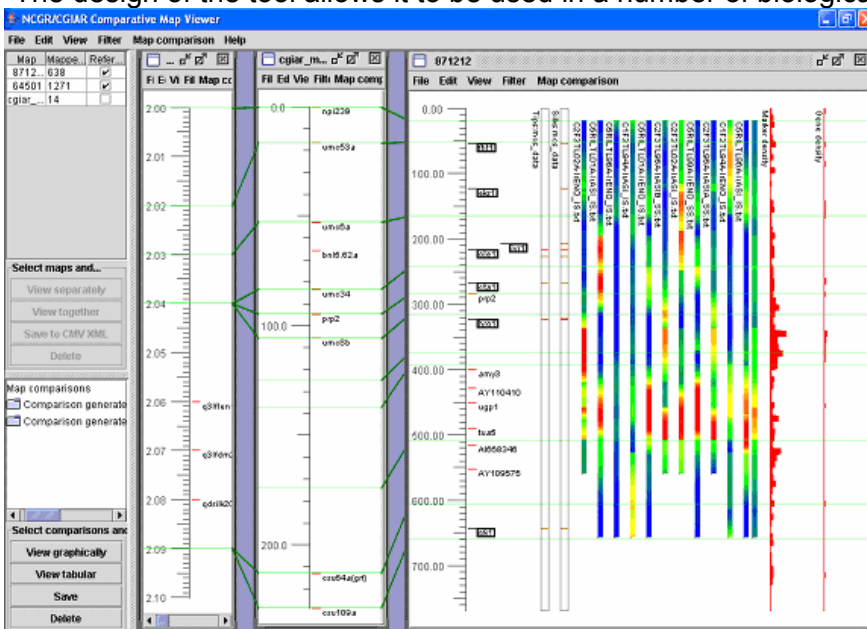
- **Collecting genomic map and map-related data** (e.g. QTLs, gene expression values, sequence data) from disparate sources including local flatfiles, remote public databases, and analysis tools;
- **Utilizing dynamically determined correspondences between different maps** in order to:
 1. construct consensus maps combining results over multiple experiments;
 2. leverage comparative genomic data in syntenic areas of other species, or duplicated regions within the same genome
 3. project data from local and published experiments into a common reference map, allowing a virtual integration of these data into a single synthetic representation
- **Creating dynamic visualizations of the collected data** that may be customized graphically and manipulated interactively (e.g. sorting, grouping, filtering, applying thresholds) to reveal aspects of the data pertinent to the user's current focus and to select subsets of the data for further analysis

CMTV may be used either as a standalone tool or as a plugin component of the open source ISYS framework (Integrated SYStem: <http://sourceforge.net/projects/isys2>) for component-based integration of biological data and tools. In this capacity, it uses the ISYS framework to allow extension of its native functionality through plugins for new data sources, visualization components, and algorithmic implementations of map comparison and consensus generation techniques. In addition, it uses ISYS's "Dynamic Discovery" service brokering and event-based communication to enable runtime interactions with other components that were not specifically designed into the tool.

The source code for CMTV is available from: <http://sourceforge.net/projects/cmtv>

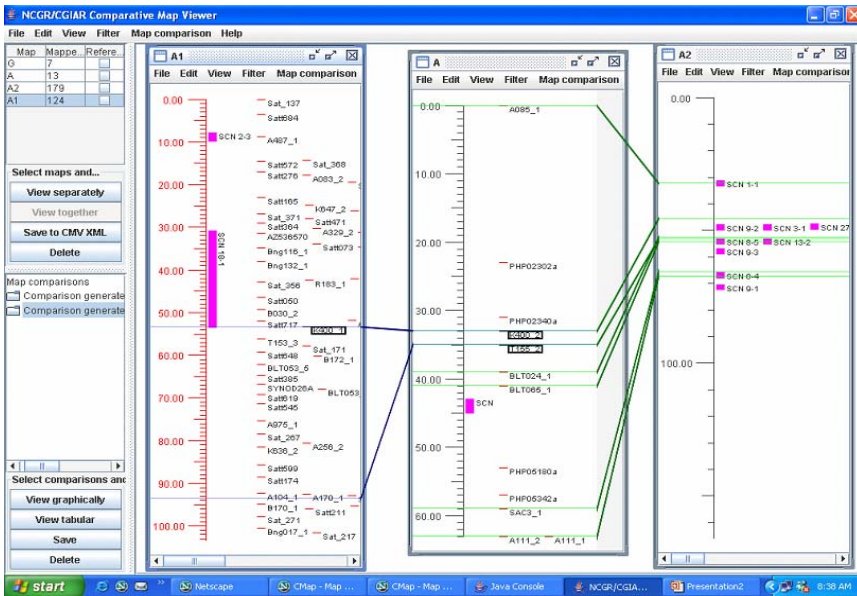
Enabling discovery with CMTV

The design of the tool allows it to be used in a number of biologically relevant contexts. Some examples:



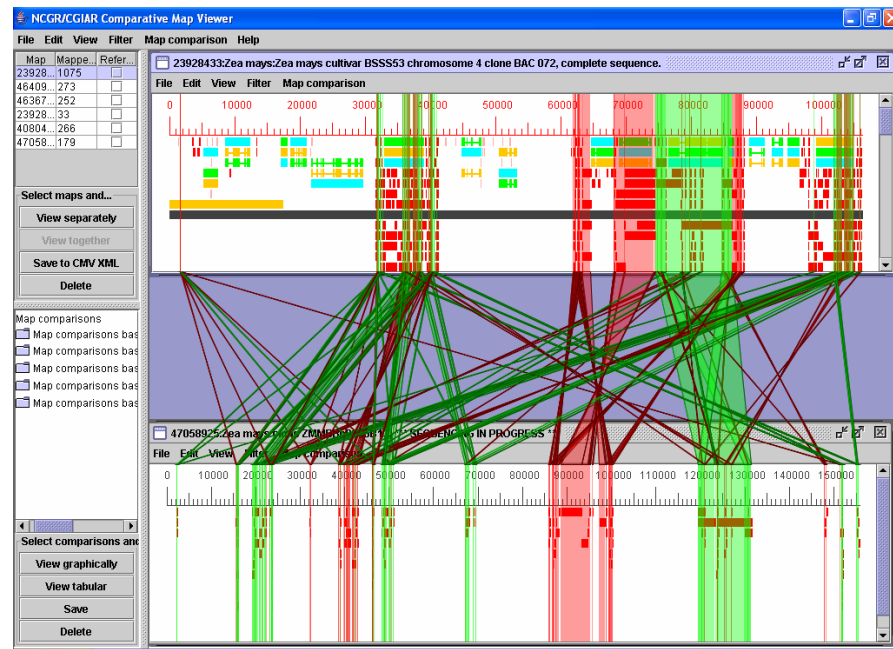
1. Integration of local and public and genetic and functional data

A maize IBM neighbors reference map (right) from <http://www.maizegdb.org> is displayed with the subset of markers linking it to the researcher's local map (center); a number of QTL likelihood distributions for a variety of traits and environments indicating not only regions above the significance threshold, but also regions displaying "tendencies" to explaining the observed variability; a region on the neighbors map on the vicinity of a conserved region of significance has been used to display additional genes on the IBM map in the region, and the "Bins map" (left) has been loaded with data in the region of the corresponding bins. Histograms of marker density and gene density for the complete set of IBM neighbors data have been calculated for the data at hand and are displayed to the far right. Local gene expression data has been loaded for differentially expressed genes in this set. The graphical display is ideal for analyzing large datasets generated from many traits and/or experiments and the identification of consensus regions can be used in the design of marker assisted selection (MAS) experiments.



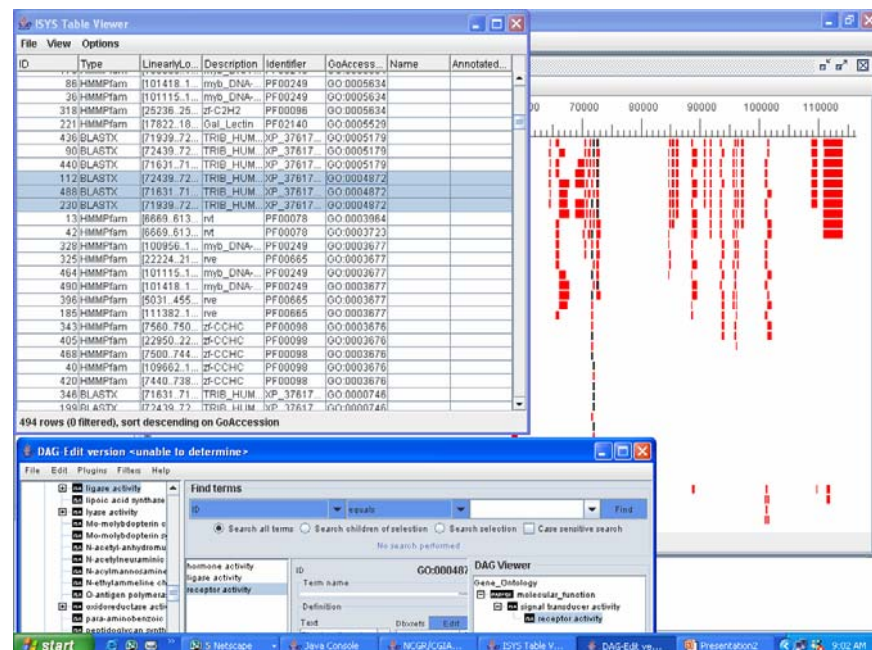
2. Evolutionary fate of duplicated genes

A researcher's local soybean cyst nematode QTL experiment (center) is compared to data from a CMap database maintained at a central public repository (<http://comparative-legumes.org>); the composite maps with which it has markers in common reveal the presence of QTL reported in other experiments for the same trait in the vicinity of corresponding (right) and duplicated regions (left) of the genome. The absence of QTL in the duplicated region is suggestive of loss of function following the duplication event. The public maps have been filtered at the user's discretion to display only QTLs from the same class of traits. Data from other public sources for a model species could be brought into the workspace to further hunt for possible candidate genes of known function in syntenic regions.



3. Interpretation of genome-scale sequence similarity

An annotated genomic clone displays dynamically computed sequence similarity search results against one of the clones matched in a background library. Areas of similarity between the two are displayed on each sequence as red bars, with green links between the two sequences for matches on the same strand and red links for matches on opposite strands. In addition to the clear presence of "noise" due to repetitive elements, one can see an inversion of one of the homologous genes, as well as the conspicuous absence of homology for several of the annotated genes.



4. The map is not the territory: multiple perspectives on the data

Graphical synchronization between components mediated through ISYS allows a division of labor between tools designed for specific purposes. Each of the three tools shown understands a certain aspect of the underlying data set. CMTV displays the positional offsets and groupings of annotation on a genomic clone from a computational pipeline. Table Viewer represents each datum as a row with its properties as columns, and the values rendered as text. An ontology editor (DAG-Edit) displays the Gene Ontology terms associated with the annotations, and their conceptual relationship to other biologically relevant terms. Selection of a term in the hierarchy display causes selection of the associated annotations in the other components; selection of annotations in a region causes selection of the relevant terms in the hierarchy.

