Plant Molecular Biology **56:** 465–480, 2004.
© 2004 Kluwer Academic Publishers. Printed in the Netherlands.

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

M.C. Sawkins[1,*], A.D. Farmer[2], D. Hoisington[1], J. Sullivan[2], A. Tolopko[2], Z. Jiang[2] and J.-M. Ribaut[1]

[1]*CIMMYT, Apartado Postal 6-641, 06600 Mexico D.F., Mexico (*author for correspondence; e-mail msawkins@cgiar.org);* [2]*NCGR, 2935 Rodeo Park Dr. East, Santa Fe NM, 87505 USA*

Received 28 June 2004; accepted in revised form 16 October 2004

Key words: consensus maps, gene display, integrated bioinformatic tools, marker assisted selection, QTL display.

Abstract

In the past few decades, a wealth of genomic data has been produced in a wide variety of species using a diverse array of functional and molecular marker approaches. In order to unlock the full potential of the information contained in these independent experiments, researchers need efficient and intuitive means to identify common genomic regions and genes involved in the expression of target phenotypic traits across diverse conditions. To address this need, we have developed a Comparative Map and Trait Viewer (CMTV) tool that can be used to construct dynamic aggregations of a variety of types of genomic datasets. By algorithmically determining correspondences between sets of objects on multiple genomic maps, the CMTV can display syntenic regions across taxa, combine maps from separate experiments into a consensus map, or project data from different maps into a common coordinate framework using dynamic coordinate translations between source and target maps. We present a case study that illustrates the utility of the tool for managing large and varied datasets by integrating data collected by CIMMYT in maize drought tolerance research with data from public sources. This example will focus on one of the visualization features for Quantitative Trait Locus (QTL) data, using likelihood ratio (LR) files produced by generic QTL analysis software and displaying the data in a unique visual manner across different combinations of traits, environments and crosses. Once a genomic region of interest has been identified, the CMTV can search and display additional QTLs meeting a particular threshold for that region, or other functional data such as sets of differentially expressed genes located in the region; it thus provides an easily used means for organizing and manipulating data sets that have been dynamically integrated under the focus of the researcher's specific hypothesis.

Introduction

Many researchers in the numerous disciplines of biology find themselves frustrated by the challenges of finding efficient and meaningful ways to manage and explore the data relevant to their research. This is often true even of data that are generated within their own groups, but the task is made still more difficult when expanded in scope to include data made available from other groups, either directly or *via* public databases. A particular challenge is to find ways to integrate and visualize data that are produced by different types of analyses and software tools. This is especially true with respect to genomics, where the amount and diversity of information related to the characterization of molecular markers, quantitative trait loci (QTL), sequence and molecular function is increasing daily. In order to address biological questions more fully and to extract more information from this wealth of data, researchers require tools that will allow them to integrate different datasets in a dynamic, hypothesis-driven fashion and to analyze them within a biologically meaningful framework.

To help address these needs, our research groups in the Applied Biotechnology Center (ABC) at the International Maize and Wheat Improvement Center (CIMMYT) and the National Center for Genome Resources (NCGR) have collaborated on the development of the Comparative Map and Trait Viewer (CMTV), a software component that can help serve as an intuitive and extensible framework for the integration of various kinds of genomic data. Initial work on the tool was begun in 2001, when NCGR and the International Center for Tropical Agriculture (CIAT), CIMMYT, the International Potato Center (CIP) and the International Rice Research Institute (IRRI), centers of the Consultative Group on International Agricultural Research (CGIAR), collaborated in the development of the Comparative Map Viewer (CMV), a prototype of the CMTV. The objective of this collaboration was to develop software components that use the ISYS (Integrated SYStem) integration platform developed by NCGR (Siepel *et al.*, 2001) to access and visualize map data and related information such as germplasm pedigree relationships. The major output of this initial effort was a tool that could display linkage groups or chromosomes from different species and the objects located on them and generate correspondences between the objects on the maps, using an extensible set of algorithms for comparing objects across a set of maps. These correspondences could then be displayed as graphical lines linking corresponding loci between maps in order to illustrate their syntenic relationships; they could also be used to construct a consensus map using these common markers as anchors from which the positions of other markers could be interpolated.

Given the large amount of QTL data generated in-house, CIMMYT and NCGR continued their collaboration in order to add new functionality to the CMV, resulting in the current version of the CMTV tool. In these subsequent phases of development, visualization components were added to the tool enabling a user to display complete QTL results spanning whole linkage groups. At the same time, the use of the inter-map comparisons was considerably extended to enable dynamic alignment of these and other data from multiple maps into a common coordinate framework. This dynamic coordinate transformation applies in different ways to a variety of situations, such as the consensus maps described above and their inputs, or a similar application of the concept to externally curated, marker-rich reference maps that are available in some species. This allows either consensus maps or reference maps to be used as a genetic backbone so that data from diverse experiments (e.g., QTL data from multiple crosses) can be integrated into a common intuitive visual framework. Regions that are consistent (i.e., show similar levels of significance) across individual experimental factors such as cross, trait or environment, or across combinations of these factors, can then be identified and further enriched with functional data such as differential expression levels for genes in the region. The graphical display is ideal for analyzing data from a large number of traits/experiments and for identifying target genomic regions based on particular selection criteria.

We shall focus primarily on the utility of the CMTV in the interpretation of results generated by common QTL analysis software tools, although the tool has been designed quite generally as an extensible framework for visualization of genomic data such as sequence alignments or gene prediction results. Typical of many types of genomic analysis, it can be a daunting task to scan the pages of output generated by common packages to compare QTL locations across different crosses of interest. When the QTL outputs are reported on linkage groups, it can be time consuming to compare outputs across genetic maps to identify common QTL map locations, as maps generally have very few markers in common. When this case arises, one is obliged to undertake a ‘triangulation’ exercise where a reference map, if available, is used as a bridge to compare QTL locations across two genetic maps. This task becomes even more arduous when QTL results are only presented in tables, as is often the case with published research results. These problems illustrate both the basic and common need to translate primary experimental results across contexts with different coordinate systems, as well as the benefits of a flexible, user-driven approach to the problems of data aggregation and customized presentation. With its design around these basic principles, the CMTV has proven useful in aiding the design of the most suitable marker-assisted selection (MAS) strategies and some results from research on drought tolerance in tropical maize conducted at CIMMYT will be used to demonstrate the utility of the tool. By focusing here on the use of the tool to visualize complete QTL results, we also hope to provide a compelling case for the utility of developing more comprehensive approaches to publishing data of this nature, allowing one to identify not only significant or 'hot' regions but also: 1) 'tendency' regions where the likelihood values are high but not significant across all experiments and 2) 'cold' regions that consistently show a low likelihood across experiments. We will also examine its similarities to and differences from other publicly available comparative mapping tools such as BioMercator (Arcade *et al.*, 2004) and cMap (Fang *et al.*, 2003) (http://www.gmod.org/cmap).

Methods

General architecture

Although designed to be usable as a standalone tool, the CMTV was developed to serve as a component of the ISYS software platform for third-party integration developed by NCGR. ISYS offers varied services through *components* – plug-and-play modules that can provide visualization services or access to databases and analytical applications, including web-based data and services. Components within ISYS are loosely coupled and can be unplugged or plugged into a given user's instance of the system without requiring changes to other existing components. This preserves the investment already made in legacy components while leveraging the power of having many tools interoperating in an integrated environment. The two major axes of component integration in ISYS are service brokering and event exchange.

The service brokering mechanisms of ISYS allow components to access functionality provided by other components in two distinct modes. When a component has been specifically designed to make use of a certain type of functionality, this can be achieved via a simple lookup for registered implementations of the desired class of service. The CMTV uses this style of service lookup in a number of places: the set of data sources available to the tool is determined by finding all installed components that implement a map listing and retrieval service; it also uses this method to enumerate algorithmic implementations of services that allow it to establish correspondences between the objects placed on different maps. The main benefit to this approach is that new data sources from simple flatfile parser adapters to database query interfaces and new algorithms for map comparison (e.g., similarity metrics for sequenced marker types) can be added by third party developers to ISYS and will be available to the CMTV without a need for recoding or recompilation. The second style of service brokering is referred to as 'dynamic discovery', whereby arbitrary classes of services are located based on their ability to operate on data selected by the user. For example, if the objects selected on a map have gene symbols associated with them, when dynamic discovery is invoked, they will elicit the offer to lookup their ontological associations if the appropriate component providing access to such information is installed. This technique allows the user to explore the implicit connections between components without any design-time knowledge of specific interfaces, so that new classes of service will often be made available to components that were developed long before they were even conceived.

The other dimension of component integration in ISYS involves dynamic exchange of data *via* events. To date, this has been used mostly with respect to the synchronization of the graphical state of user interfaces; for example, a set of objects selected in the CMTV according to their spatial clustering can trigger the selection of a corresponding set of objects in a gene expression visualization component or an interaction network display, leading to multiple coordinated perspectives of this set of data. The platform allows for arbitrary synchronization of pairs of component instances and components are free to interpret the appropriate response to a given event, so that the coordination may be based on more subtle connections between datasets than simple object identity; for example, selection of a catalyzed reaction in a pathway component might trigger the selection of all loci in the CMTV whose genes are involved in producing the given catalytic function.

Visualization architecture

The visualization framework of the tool has been designed to provide a great deal of flexibility in how the user chooses to have various attributes of the data translated into a graphical context, and to provide clear extensibility mechanisms for developers of new visualization components. This extensibility is important, since it allows the development of new visualization subcomponents (such as the QTL display described below) or map layout mechanisms to take place relatively independently of the high-level features of the tool, enabling a user to choose from a diverse set of data rendering components, possibly developed by different groups.

From the perspective of the end user, the default presentation of any map can be altered by specifying a set of features to include in tracks on the display. Each set is defined by a rule that specifies which of the objects on the map will be included in that track, for example, one track might display only objects of type gene, whereas another might display only objects mapped to the reverse strand of a sequence. The set of ‘mapped object inclusion rules’ can be extended and the logic may be arbitrarily complex, however no mechanism is currently provided for dynamically defining arbitrary new rules. Each of the tracks also has a set of renderers associated with it, allowing the possibility of compound displays for each object in the track; for example, an object can be displayed as a colored bar spanning the coordinates of the location of the object on the map and have a set of labels associated with the bar that display one or more of the attributes of the object (e.g., name, functional class, expression level, etc.).

Another level of graphical customization is supplied by a general mechanism for translating attributes of the underlying map data into attributes of the graphical representation. For example, the user may supply a file that translates mapped object types (e.g., gene, RFLP, SNP) into the color used by a renderer of that object. A more complex example discussed in more detail below is given by user defined mappings from numerical ranges (e.g., significance values, expression levels) and color scales or intensities. The mapping mechanism may be thought of as a sort of data-driven stylesheet mechanism.

Construction of map comparisons

After loading data into the tool by means of one or more of the available data source adapters (e.g., the NCBI Entrez system at http://www.ncbi.nlm.nih.gov/entrez, the MaizeGDB interface at http://www.maizegdb.org/, or local repositories of flatfile data), one is able to construct customized visualizations of the data included on a map by the provider. However, the more interesting aspect of the CMTV framework is its ability to establish meaningful transformations between the coordinates of different maps based on analysis of the corresponding objects between the maps and to use these transformations in order to allow the data from these maps to be visually integrated in a common display framework.

In order to achieve this, the first step is to construct the comparisons between given sets of maps. The CMTV itself defines a simple data model for representing the sets of objects that have been associated across sets of maps, but defers most of the actual construction of these comparisons to external plugin services, in order to allow more sophisticated implementations to be developed by experts. We have developed a small set of simple representative algorithms that are bundled with the package, most that allow one to select particular attribute types (e.g., accession number, marker name, gene function) that have been associated with the mapped objects by the data provider, apply an optional transformation to the character strings (e.g., converting to all upper case or removing non-alphanumeric characters) and then use one of a number of simple algorithms for constructing correspondences between the objects based on basic string matching (e.g., complete matching, substring inclusion or fuzzy matching).

In addition to these fully automated mechanisms, we provide mechanisms for uploading user-defined equivalence relationships among the strings representing the object attributes, as well as the ability to construct a map comparison by manual selection of corresponding objects. Another feature of the tool allows it to utilize information generated by other components concerning pairwise alignment relationships between sequence maps and to interpret these data as correspondences between objects representing the reciprocal similarity relationship between regions on the two sequences.

Once a comparison has been constructed among a set of maps, the user is able to request visualization of the correspondences between maps in the comparison. These are represented as transparent areas overlaid upon and connecting the regions of the two maps where each pair of corresponding objects is located. These connections can appear as lines drawn between points on the maps or as polygonal areas connecting map regions, depending on whether the corresponding mapped objects have point-like, interval-like, or compound locations. In addition, graphical attributes of the transparency regions such as color can be used to represent attributes of the underlying correspondences, such as whether the regions are inverted with respect to one another. It should be noted that the tool allows one to define multiple comparisons involving the same maps using different techniques and to visualize separate comparisons independently, which is useful in comparing the results of different comparison algorithms.

Creation of inter-map coordinate translations

After a comparison has been established between two or more maps using one of the techniques described above, a common coordinate framework for a set of maps can be obtained in two ways: 1) using anchor markers common to all or most of the input maps to construct a consensus map based on the input maps, or 2) using an independently defined reference map on which the data from other maps can be displayed using transformations defined by comparing source maps to the reference map individually.

In the consensus map approach, a set of ‘anchor’ markers must first be selected from the set of mapped object correspondences in a map comparison, to serve as the basis for computing coordinate transformations from the input maps onto the consensus. The selection of a set of anchors from a given comparison may be effected in the tool *via* pluggable algorithmic services; the tool currently only provides one implementation of an automatic anchor selection service, which requires that every anchor is present on all maps and is consistent in order with the other anchors across all the input maps; in the case of ordering inconsistencies across maps, a maximally inclusive set of anchors with consistent order will be automatically selected; other implementations could be developed as ISYS services by third party developers and would be presented as options to the end user without any need for recoding. Regardless of the mechanism used for the selection of the anchors, the user may alter the set of anchors chosen manually before proceeding to the construction of the consensus map; for example, in cases where only a few markers are shared by all input maps, but many are common to a large subset of the maps, the user could add these markers by manual selection through the interface.

The anchor markers are positioned on the consensus map by taking their average cM position across all the input maps where they are located; they can then be used to compute the placement of the non-anchors. This is done for any non-anchor by calculating a ratio between the length of an interval on the input map whose endpoints are defined by anchor markers and, which contains the non-anchor marker and the length of the corresponding interval (as defined by the anchors) on the consensus map. This is shown in Equation 1 below.

$$P_{\mathrm{cm}_y} = P_{\mathrm{cm}_1} + \frac{(P_{\mathrm{m}x_y} - P_{\mathrm{m}x_1})(P_{\mathrm{cm}_2} - P_{\mathrm{cm}_1})}{(P_{\mathrm{m}x_2} - P_{\mathrm{m}x_1})},$$

where, P_{cm_1} and $P_{\mathrm{m}x_1}$ are the locations, in cM, on the consensus map and map x of anchor locus 1; P_{cm_2} and $P_{\mathrm{m}x_2}$ are the locations on the consensus map and map x for anchor locus 2; $P_{\mathrm{m}x_y}$ is the location on map x for locus y and P_{cm_y} is the position of locus y on the consensus map.

In the CMTV display window, the anchor markers on the consensus map are presented in a larger, bold font to differentiate them from the non-anchor markers. When a consensus map is constructed from a map comparison where some corresponding markers are not used as anchors, separate instances of the corresponding markers will be placed on the consensus map, each having its position on the consensus map determined using the ratio appropriate to the map of origin. The user then has the option to merge these corresponding markers (taking the average of their positions on the consensus map) within a threshold distance specified by the user (the default is 10 cM). Note that this is different from the situation where a marker is selected as an anchor but not shared across all maps as described above when the user alters the automatically determined set; in such a case the coordinate transformations to the consensus map from any of the input maps will simply interpolate between the intervals defined by the anchors that are present on that map.

In the case of non-anchor markers that fall outside the boundaries of the intervals defined by the anchors (i.e., before the first anchor or after the last), the ratio given by the neighboring interval is used. In addition, positions are adjusted so that no markers are given negative coordinates, which can result in the final positions of the anchors having some offset from the average values on the input maps. The transformations established for the non-anchors can be used to translate any coordinate from the input map onto the consensus map, whether or not a non-anchor marker happens to be located at that point; these general coordinate transformations are used by the tool to translate the trait data from the individual crosses onto the consensus map, as will be described below.

If a reference map is used, the coordinate transformations used to place data from the source map onto the reference map are computed automatically from a map comparison between the two. There is no need to choose anchors in this case, since the positions of the markers on the reference are considered fixed and the coordinate translations from a given marker map to the reference map are computed by linear interpolation in the intervals defined by the common markers. This allows the user to create dynamic ‘pseudo’ consensus maps, by loading marker maps onto the reference framework. The CMTV allows a user to denote any map as a reference map and can automatically generate comparisons between the designated reference map and a user-specified set of maps for convenience in utilizing the reference map as a framework for integration of the data from these maps.

Either the consensus map or the reference map approach may be more appropriate depending on the given context and the availability of a reference map for a given genome. Genetic maps suitable for QTL identification are generally not high density, as linked markers closer than 10 cM do not significantly improve the detection of QTLs using typical segregating populations of a few hundred individuals. Therefore, few common markers are generally identified across linkage maps from different crosses developed for QTL identification. This represents a limitation to the first approach, where a reasonable subset of markers should ideally be common to most of the maps in the set. However, this may be the only option when no reference map is available, e.g., for orphan or minor crop taxa. If few common markers have been identified, a possible option would be to add more selected markers to some maps to increase the total number of common markers to all maps. This would require an additional mapping effort, but in general would only necessitate the screening of a subset of each segregating population (e.g., 50 genotypes) and would depend on levels of polymorphism and availability of markers for a given species. The use of a reference map, with a high density of markers is a more attractive option, as the likelihood of finding markers common between a single map and the reference map is high for most of the major crops. Today, a number of high density maps are publicly available and are updated on a regular basis. For example, in maize, the IBM2 neighbors consensus map with approximately 5700 loci, and based on the high resolution IBM (Intermated B73 × Mo17) maps of the Maize Mapping Project (www.maizemap.org) could fulfill the role of a reference map (Polacco *et al.*, 2004). In the current representation, nine maps have been added to IBM2 that share loci, either with IBM or some other map that shares loci with IBM. In addition this map is significantly augmented by the addition of physical map data, which has added approximately another 15 000 additional loci.

The CMTV uses the inter-map coordinate translations as a basis for performing dynamic aggregation of data from multiple independent maps into a common visual framework. For example, the translations based on discrete marker correspondences can be used to adjust the coordinates of continuous LOD distributions for different QTL experiments, allowing them to be aligned together onto a consensus or reference map which makes the comparison more appropriate. The technique of dynamic visual integration may be also be appropriate in some cases such as inter-species comparisons, where constructing consensus maps would be meaningless, but where the drawing of lines between syntenic regions becomes too visually confusing due to the level of genomic rearrangement that may have taken place.

QTL display

Existing mapping tools typically display only the significant QTL regions, primarily because the QTL results stored in databases typically provide summarized data, e.g., position of significant QTL (bin, cM position or bar corresponding to the significant region), the position of flanking markers and/or highest likelihood ratio (LR) value. This is clearly the simplest strategy for the storage of QTL information in a publicly accessible database such as MaizeGDB or Graingenes, as the original source of this information is from publications or personal communications from various research groups. Nevertheless, when the primary data are available (as for locally analyzed results), it can be very informative to compare complete QTL results over an entire linkage group, as this permits one to extract a greater amount of information from the data than simply considering significant regions based on a default cut-off value in the QTL analysis software. Furthermore, the ability to translate the results from multiple experiments across different genetic backgrounds and environmental conditions into a common coordinate framework increases the utility of having the complete QTL results at hand, as patterns of correlation between variation in the significance of a region to a given trait with the variation among different experimental conditions may be suggestive of different mechanisms for the variability in the environmental interactions.

After a transformation has been defined between maps using one of these approaches described above, the tool allows the user to align raw QTL results for any maps whose coordinates can be transformed into the common coordinate framework. To do so, the user queries a repository containing the QTL data and selects the traits to display. Initially, we have implemented a simple flatfile repository for this purpose, although the CMTV is decoupled from the implementation details of its data sources and more sophisticated implementations can be added to the framework. To provide a simple but flexible query capability, selection is based on different metadata fields embedded in the name of the QTL file. A practical example of how fields are currently defined will be presented in the results section. Traits may also be selected on the basis of their meeting a threshold criterion in some common region.

Once target traits have been selected, they are loaded into the tool as numerical distributions samples across the coordinates of the map. (QTL data where numerical distributions are not available can be displayed as regular interval-like objects and can be incorporated with the data where complete results are available.) Using one type of display component for continuous numerical distributions of this sort, CMTV can present the corresponding raw results as colored 'heat' strips along a given linkage group. The color scale represents a log LR value and the user can define mappings between ranges of LRs and color ranges (values within the endpoints of the ranges are interpolated); for example, in the default mapping, colors range from blue (low LR score) to red (high LR score). Another display component is provided that can display this data as a histogram; both styles of display can also be used simultaneously. Various strips can be displayed together on either a linkage group from a single mapping experiment (to compare results across traits and or environments for a given single cross), or on a linkage group from a consensus map or reference map (allowing a comparison of results across traits, experiments and crosses). The same transformations used to convert the positions of the markers are also used to transform the intervals across the QTL data. With the values corresponding to the 'same region' dynamically aligned, a direct visual comparison of LR may be performed by studying the conformity of colors across strips from different maps. The width of the strips can be adjusted making it possible to easily visualize 20–30 QTL profiles at one time on a standard sized computer screen and the user can sort these strips (using the LR score) at any particular point (cM position) in order to group and visualize more easily regions that have similar LR values.

Once a region of interest has been identified across traits/experiments using a set of complete QTL input files that have been selected by the user, other QTL datafiles in the repository can be queried in order to identify additional traits that also have comparable LR values to that of the region selected. In most cases a user would want to apply a high LR threshold to identify traits that also contain regions of significance. This allows the user to dynamically construct sets of QTL data that are relevant to the current line of inquiry, without requiring that the sets be defined *a priori*. These additional traits are aligned in the same manner as the existing traits.

Although the primary use of the QTL strips is to provide a means to visualize genomic regions of interest by taking into account a particular threshold of significance, an effective candidate gene approach requires a finer resolution of the QTL peaks. Therefore within a particular region of interest the CMTV can display a table showing the maximum value of a LOD score and position across a set of traits and environments. This information could also be used as part of the input for statistical analyses such as meta-analysis of QTLs, that will identify the most likely QTL peak/locus by considering QTL data for a target trait across various environments (Goffinet and Gerber, 2000). Although the meta-analysis is not presented here in the context of this work, the statistical algorithms are currently being implemented in the CMTV and the analysis will be available to users in future releases of the software.

Following the selection of interesting genomic regions, it is also possible to query public databases for information at those regions that might have some relevance. This information may include additional QTL that have been identified by other groups and/or the location of candidate genes. For example, in maize, data on the 'bins maps' from MaizeGDB can be downloaded and contain information on different types of objects (e.g., QTLs, genes, markers). A comparison of any marker map to a bins map can be constructed to establish the correspondence between a target region on the map to a set of bins, to display selected objects (based on types) on the bins map in the region.

Visualization of functional genomics data

When a consensus region has been identified, the user may wish to focus in more detail on this region, by incorporating other data and refining and validating the importance of the particular region (i.e., through identification of genes/QTLs/expression data, from both public and local sources). When used as a component of the ISYS system, the CMTV can be synchronized with other components specially designed for displaying gene expression information. However, there are also compelling reasons for achieving a tighter integration of the expression information into the visual framework used to display structural map information. For example, the regions indicated by a set of QTL experiments as being significant for a given trait may contain sets of genetic elements whose differential expression has been assayed in one or more sets of expression experiments. The approach we have taken in order to integrate expression data into the context of the viewer uses a strategy that incorporates three main features:

- using a user-specified set of mapped objects as a query against the expression data in a target dataset; for example, the set of markers in a region indicated as significant by a QTL dataset;
- dynamic translation from the identity of the elements assayed in the expression experiments to their positions on the map from which the query was issued; and
- as much reuse as reasonable of the visualization and manipulation tools developed for QTL data for the resulting map coordinate-based numerical 'distributions' of the expression data.

The CMTV currently uses the names associated with the selected mapped objects as a query against a filesystem-based repository of simple delimited gene expression experiment files. These files consist of two or more columns, where the first column represents the identity of the elements assayed and subsequent columns represent the numerical values of some measurement made against the elements (e.g., a ratio of the expression value between treatment and control). The search function will identify any experiments where one or more of the selected mapped objects have been assayed and will return a two-column list representing both the file names and column headers where one or more data points was found; this allows, for example, the grouping of a series of related assays such as a time series into a single file, as well as the ability to select subsets of the data for individual columns in a given experiment series (e.g., a set of knockout mutants, where only a few are of interest in a given context). Each of the items that are selected in this list will have the data retrieved from that column of that file for the given set of mapped elements.

Next, the numerical data from the tabular files will be distributed on the given map with the coordinates of the mapped element to which it corresponds. Basically, the set of numerical values from a single column in a file will be mapped into a single 'discontinuous distribution' where the map coordinates of the assayed element are used as the domain and the measured value of the expression level is the range. These data distribution constructs are then added to the data model of the viewer and the visualization framework automatically invokes whatever renderer has been configured to handle such data. In the current version, the default choice of rendering for this type of data is a color-mapped rendering of the numerical values, similar to that used for the QTL likelihood distributions, but with some specialized logic that allows them to handle the discontinuous nature of this data appropriately.

It should perhaps be noted that the approach taken here is not limited in use to gene expression data, but could be easily reused as is for other sorts of data that can be organized according to this 'tabular' representation. For example, it could be used to display 'graphical genotype' representations for a set of germplasm and their allele values for particular selected markers.

Results

Identifying genetic consensus regions for drought tolerance in maize

One area of research at CIMMYT has been in the dissection of the genetic basis of the response of maize during water-limited conditions at flowering (Ribaut *et al.*, 2002). One approach taken has been the detection of QTL for target traits involved in drought tolerance (Ribaut *et al.*, 1996, 1997). A large QTL dataset has now been generated from several segregating populations. From this, we are identifying regions in the genome that are most consistently involved in the response to drought, which can then be used in a MAS strategy. To do so efficiently requires tools to draw a consensus map in order to compare QTL locations across different experiments and genetic backgrounds.

Consensus map construction

The idea of constructing a consensus map to compare the colocalization of QTL involved in the tolerance of maize to drought was discussed for the first time at CIMMYT about 5 years ago and a set of 50 markers common to the 6 drought maps constructed at CIMMYT were selected to act as anchors before the IBM2 neighbor map was available for use as a general reference map. Now that this is available, it is relatively simple to use it as a common framework for visualization; however, reference maps do not exist for all species and the consensus map functionality of the tool will be relevant for these situations.

Figure 1 illustrates the creation of a consensus map for chromosome 2 of maize. Three individual genetic maps were used to create the consensus map. These three maps represent three CIMMYT maize crosses (comprising different parents and inbreeding levels). The leftmost map is the consensus map produced from the three individual maps and contains 9 anchor (in bold) and 14 non-anchor markers.

QTL alignment

Once a common coordinate system has been established (either by constructing a consensus map or by defining comparisons to a common reference map), one can then overlay and compare several QTL raw results files. The interface allows the user to specify attribute-based filters on the QTL data in order to facilitate the construction of datasets. The attributes currently used to characterize the CIMMYT data are: cross name, inbreeding level, geographical location, cycle of selection, trait and stress intensity. As an example C1F3TL02AGYIS translates as Cross 1 (Ac7643 × Ac7729/TZSRW), F3 families, Tlaltizapan (CIMMYT experimental field station in Mexico), year 2002, cycle A (November–April) and grain yield evaluated under an intermediate water stress level. The CMTV itself is designed to make use of arbitrary sets of attributes used by data provider components with the maps and map objects that it is visualizing.

Example of multiple traits from a single experiment

Figure 2 shows chromosome 2 from a single cross with a number of different traits overlaid. These

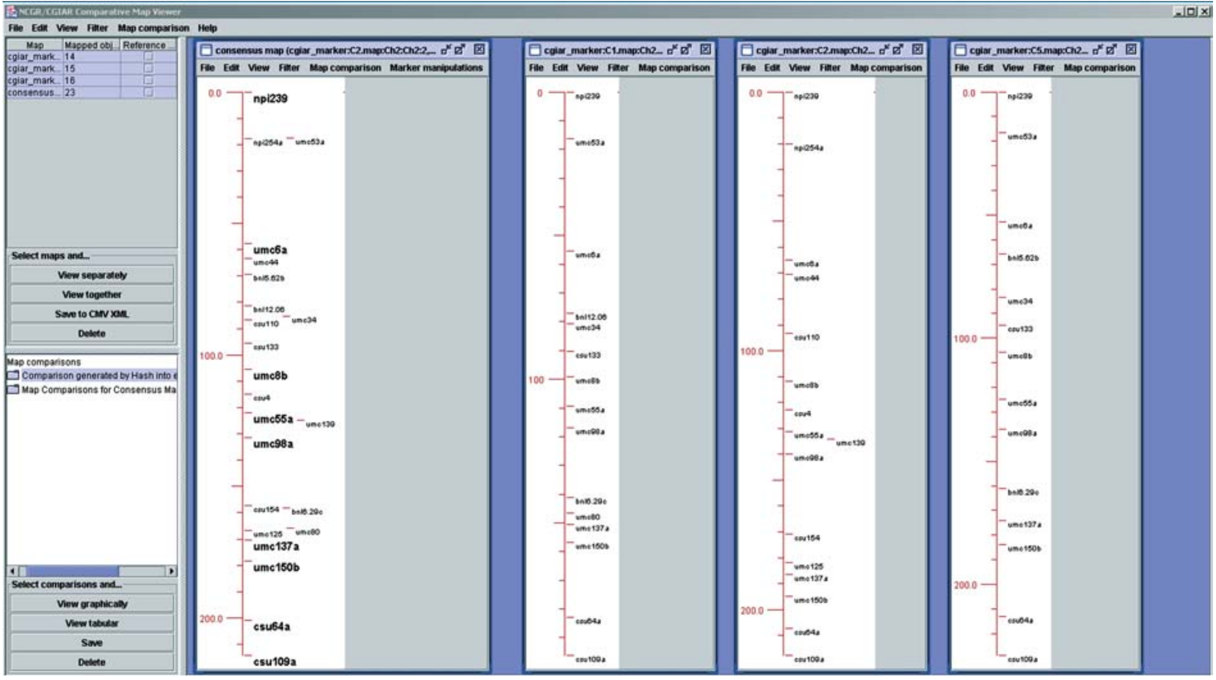
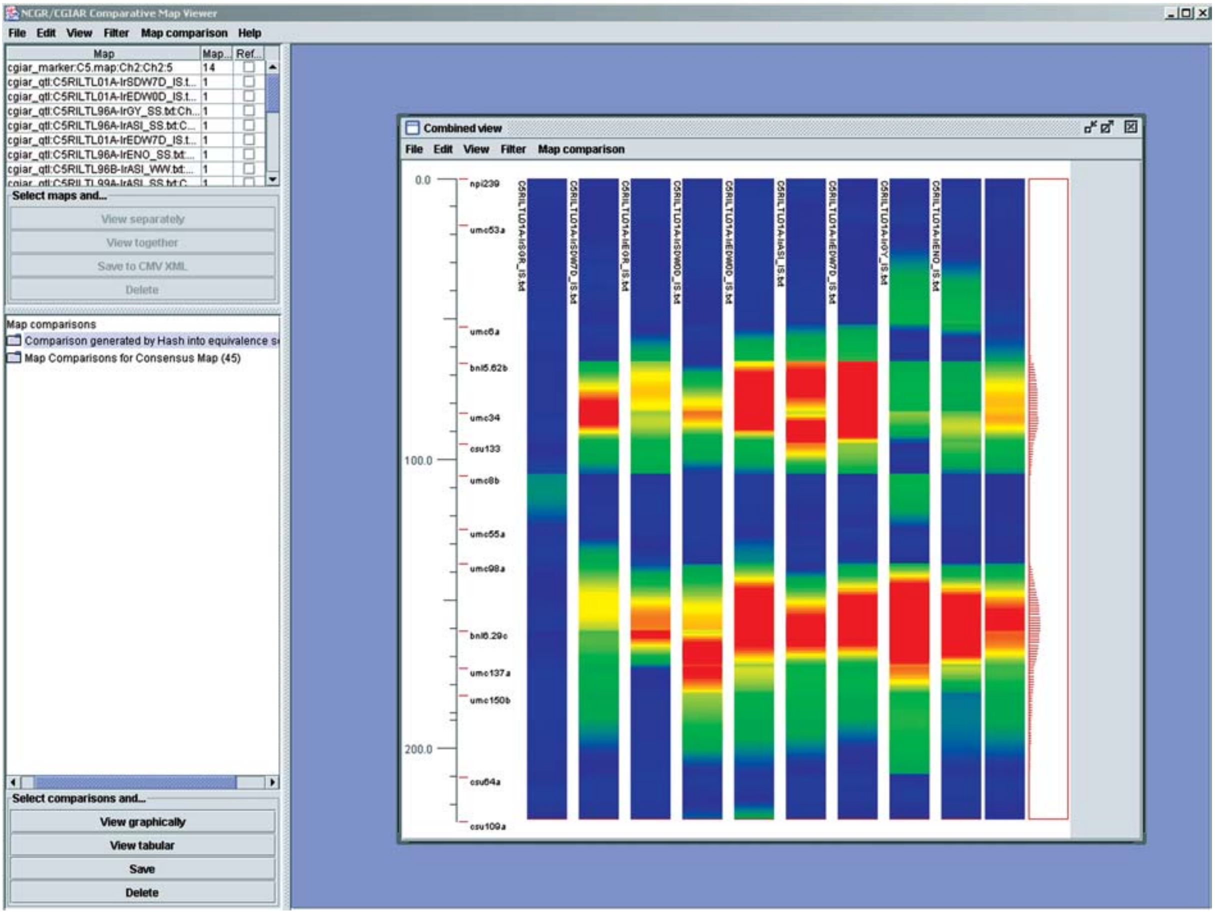

NCGR/CGIAR Comparative Map Viewer
File Edit View Filter Map comparison Help
Map Mapped obj... Reference...
cgiar_mark... 14
cgiar_mark... 15
cgiar_mark... 16
consensus... 23
Select maps and...
View separately
View together
Save to CMV XML
Delete
Map comparisons
Comparison generated by Hash into
Map Comparisons for Consensus Ma
Select comparisons and...
View graphically
View tabular
Save
Delete
consensus map (cgiar_marker:C2.map:Ch2:Ch2:2,...
File Edit View Filter Map comparison Marker manipulations
0.0
npi239
npi254a
umc53a
umc6a
umc44
bnl5.62b
bnl12.06
csu110
umc34
csu133
100.0
umc8b
csu4
umc55a
umc139
umc98a
csu154
bnl6.29c
umc125
umc80
umc137a
umc150b
200.0
csu64a
csu109a
cgiar_marker:C1.map:Ch2...
File Edit View Filter Map comparison
0
100
cgiar_marker:C2.map:Ch2...
cgiar_marker:C5.map:Ch2...

Figure 1. Creation of a consensus chromosome from three individual maps.

traits have been selected for cross C5 (Ac7643 × Ac7729/TZSRW), RIL (recombinant inbred lines), TL (Tlaltizapan), 01A (2001 winter cycle), several target traits and IS (intermediate stress). The nine traits selected are grain yield (GY), ear number (ENO), anthesis silking interval (ASI), ear dry weight at 0 (EDW0D) and 7 days (EDW7D) after pollen shed, silk dry weight at 0 (SDW0D) and 7 days (SDW7D) after pollen shed, ear growth (EGR calculated as EDW7D-EDW0D) and silk growth (SGR calculated as SDW7D-SDW0D). By selecting these traits from a single experiment, the hypothesis that we were trying to test was twofold: 1) that ear and silk DW and growth were genetically correlated when measured at flowering time under water-limited conditions and 2) that traits related to ear and silk development were correlated with three other key traits involved in drought tolerance, GY, ENO and ASI (Ribaut *et al*., 1996, 1997). As ear and silk weight are destructive measurements, the C5 RIL population was planted in four replications under a particular water regime, with two replications used to quantify ear and silk DW and growth and another two replications to quantify yield components, flowering traits and other morphological parameters. Each trait is represented as a single color bar or 'heat strip'. Associated with each heat strip is a text label describing the cross, trait and environment. This is the default view. The user has the option to hide this label, which helps in displaying a larger number of the strips on a single screen.

After selection, the traits are displayed alongside the linkage map of chromosome 2 (Figure 2). A user has the option to sort the strips (using the LR score) at any particular point (cM position) in order to group and visualize more easily regions that have similar LR values. This has been done in Figure 2, where the strips have been sorted at position 150 cM. From left to right in Figure 2 there is progression from low to high LR values.

The final heat strip and corresponding histogram are additional representations that the user has the option of including in the display. Both of these display (in a different manner) the aggregate distribution of the LR data (average values taken at each cM position) found in the individual heat strips, allowing one to condense the data from all the strips. However it should be emphasized that this is simply a way to visualize a large set of data as one strip or histogram and the user should not interpret the underlying data in any statistical manner. The authors are aware that taking an average of a LR score across traits is not



Map	Map...	Ref...
cgiar_marker:C5.map:Ch2:Ch2:5	14	□
cgiar_qtl:C5RILTL01A-IrSDW7D_IS.t...	1	□
cgiar_qtl:C5RILTL01A-IrEDW0D_IS.t...	1	□
cgiar_qtl:C5RILTL96A-IrGY_SS.txt:Ch...	1	□
cgiar_qtl:C5RILTL96A-IrASI_SS.txt:C...	1	□
cgiar_qtl:C5RILTL01A-IrEDW7D_IS.t...	1	□
cgiar_qtl:C5RILTL96A-IrENO_SS.txt:...	1	□
cgiar_qtl:C5RILTL96B-IrASI_WW.txt:...	1	□
cgiar_qtl:C5RILTL99A-IrASI_SS.txt:C...	1	□

Map comparisons
- Comparison generated by Hash into equivalence s
- Map Comparisons for Consensus Map (45)

Figure 2. Complete QTL results for nine morphological traits measured in a single experiment and sorted (low to high LR) at position 150 cM.

statistically correct as the evaluation of LR is not a linear function. This option simply gives the user a ‘quick and dirty’ way to identify hot and cold regions over a number of strips.

From the results displayed in Figure 2 it can be seen that first there is a good correlation across ear and silk weight and growth parameters on chromosome 2, with perhaps the exception of silk growth; and second, ear and silk growth parameters are well correlated with other traits of interest such as GY, ENO and ASI, as demonstrated by the two hot regions identified across traits (between 70 and 100 cM and 140 and 170 cM). The second region between 140 and 170 cM was consistently identified as significant across traits with only SDW7D and SGR having a LR below 12 (equivalent to a LOD of about 3.0). The region located between 70 and 100 cM is also of interest for the nine traits. Four of these have a significant LR value, while the other three show a tendency with LR values approximately 10 (orange) and none below 4. This strongly suggests that within these regions lie genes of interest that are involved in ear and silk growth and other key traits involved in the response of maize to water stress. In addition, three cold regions (all traits with a LR less than 4) can easily be identified on chromosome 2. These are located at the distal parts of the chromosome and the middle (100–130 cM). These results illustrate how the CMTV can be used to identify genomic regions involved in drought tolerance through combining QTL data across target traits measured in a single experiment.

Example of multiple traits across experiments

The main objective for constructing a consensus map with the CMTV is to be able to compare QTL results/positions across experiments from different segregating populations in a straightforward manner. In the previous example, a region of potential interest was identified between 140 and 170 cM on chromosome 2 by comparing the position of QTL across a broad set of target traits evaluated under a single environment. The next logical step would be to investigate if a region remains interesting when additional target traits, across several environments and different crosses are added. The objective here is two-fold; first to evaluate the QTL by environment interaction (same cross, different environments) and second, to investigate the 'stability' of QTLs by combining different genetic backgrounds (different crosses). As an example, we chose to display QTL results for ENO (ear number) and ASI (anthesis silking interval) on chromosome 2 of the consensus map; these are two secondary traits highly correlated with grain yield production under water-limited environments and collected over different years and stress levels from Tlaltizapan, Mexico for three crosses (Figure 3). As the QTL data come from different genetic maps, the length of QTL strips may vary if the input maps have different markers as their terminal markers. This is the case with the example in Figure 3. At the end of the chromosome CSU109a is common to all maps and therefore all the strips end at the same position. However, at the top of the chromosome there is no common marker and the QTL strips begin at different positions. On Figure 3, a region of 40 cM can be identified between 140 and 180 cM. Of the traits aligned, 8 are significant for all or part of this region with a LOD higher than 2.5 (red color). For 3 traits a good tendency is observed with a LOD between 1.8 and 2.5 (orange and yellow), while for the two remaining traits no genetic effect has been observed and the LOD is less than 1 (blue and green). Due to the accuracy of QTL data, it is expected that the location of significant regions will vary within our target region from one experiment to another. However, it is clear that this region on chromosome 2 remains important for ENO and ASI across the various experiments. As in the previous example, the importance of this region is further emphasized when additional traits are added to the display in particular those related to carbohydrate and other yield parameters (data not shown). Evidence suggests that carbohydrate regulation plays a key role in plant response to water-limited conditions (Prioul, 1999; Zinselmeier

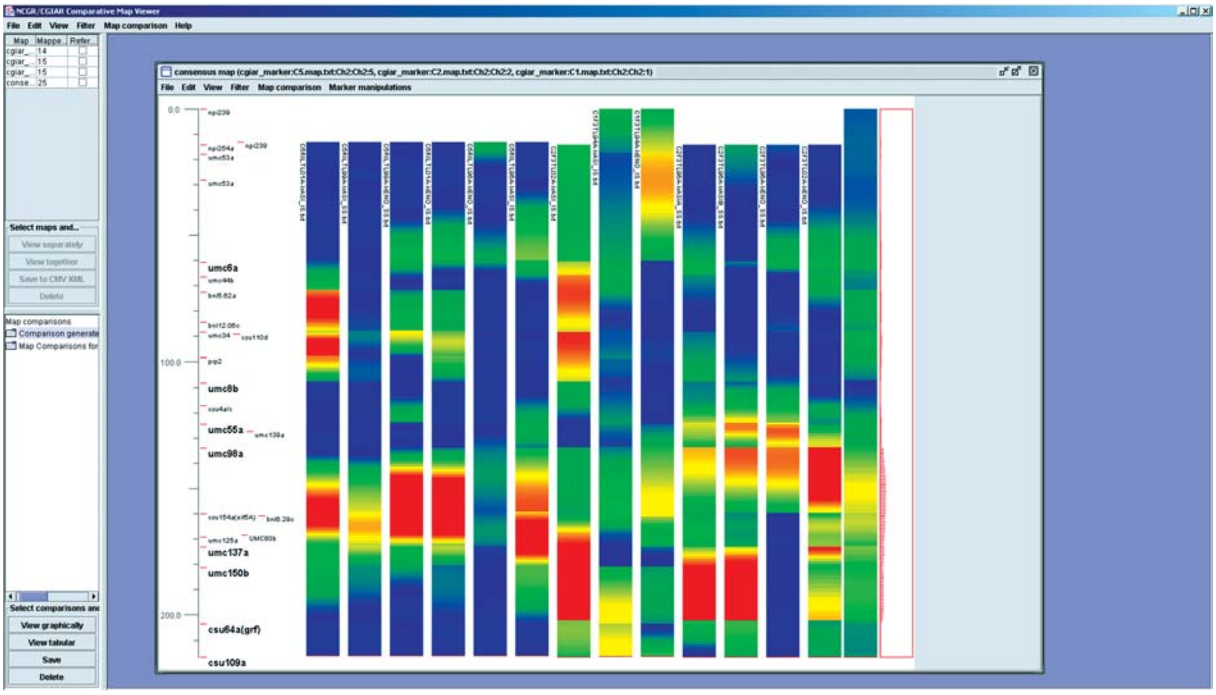


Figure 3. Complete QTL results across multiple experiments and multiple traits for chromosome 2.

et al., 2000; Setter *et al.*, 2001). Furthermore, recent work has begun to reveal the role of sugar sensing and signaling in plants and mechanisms that modulate sugar status and coordinate internal regulators and environmental cues that govern growth and development (Koch, 1996; Smeekens, 2000; Moore *et al.*, 2003). In addition, when the CMTV is used to query external databases other relevant data is found, e.g., genes involved in starch and sucrose metabolism such as *ADP-glucose pyrophosphorylase1* (*agp1*), that maps to bin 2.06 and *UDP-glucose pyrophosphorylase1* (*ugp1*), that maps to bin 2.07; these bins encompass the region identified. These results help underline the stability of this genomic region in maize during the response to water-limited conditions and therefore make this region a very strong candidate for MAS experiment.

Incorporation of functional genomic/expression data and other datatypes

Figures 4A and B illustrate how additional datatypes are integrated into the CMTV. The example chosen again is from chromosome 2. The additional types of data that can be viewed alongside the marker maps and QTL data are described below.

On chromosome 2, changes in expression levels for a number of genes for which map locations are known were studied in detail using RT-PCR at CIMMYT. Studies of differential expression were undertaken on both ear and silk tissue (strips labeled Tips:mcs_data and Silks:mcs_data) (See Figure 4A). Although a small dataset, with a number of the genes falling outside of the two regions highlighted by the QTL results, this is representative of the basic approach for visualization of expression data in the tool.

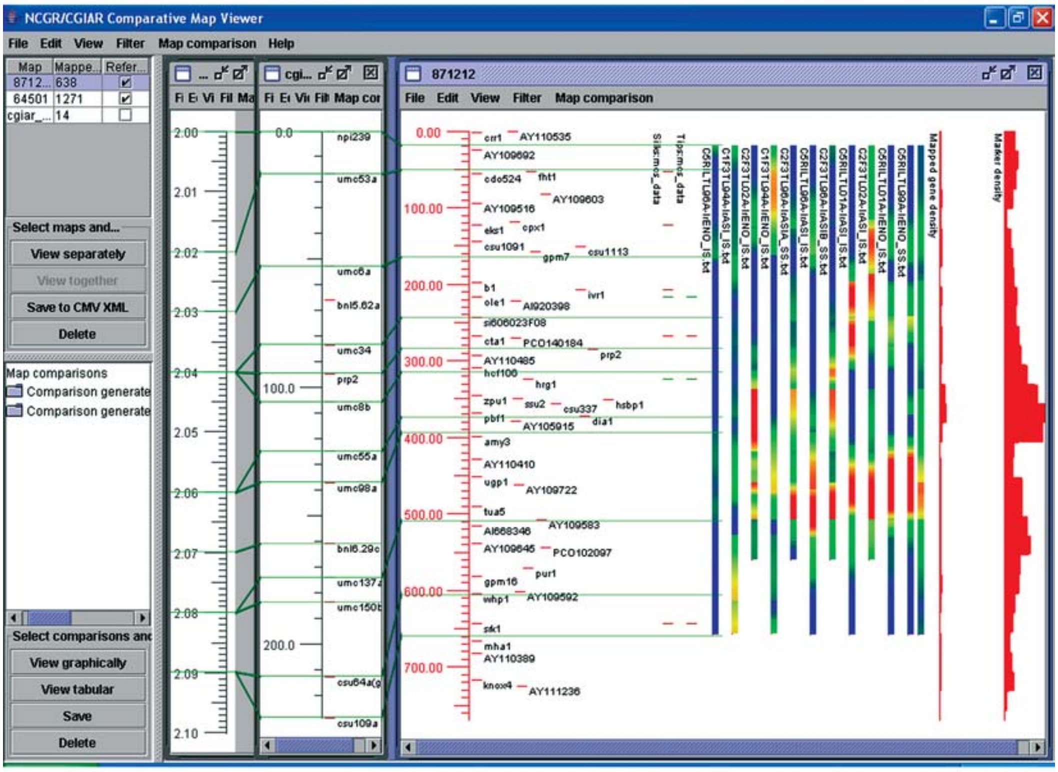
In addition, the CMTV can be used to query publicly available databases to incorporate other genomic data. This is shown in Figure 4a, which presents information on the density of markers/genes along the chromosome, represented as two histograms. A query of the IBM2 neighbors map in MaizeGDB was used to construct the density histograms. The histograms are produced dynamically by specifying the type of data to be considered (e.g., genes, markers) and counting the number of objects on the given map that fall into a sliding window. The first histogram displays the number of markers, while the second histogram is more specific in only displaying genes. More specific information about genes and other QTLs located in a particular region of interest can also be obtained from public databases. In Figure 4b, the region identified earlier corresponds to 2.06–2.08 on the BINs map. The CMTV has then been used to query MaizeGDB to identify other genes and QTL that would add weight to the importance of this region for the trait in question or identify candidate genes of interest. In the example presented, 47 genes have been identified and 80 additional QTL. These include QTL for *q4lfsuc1* (QTL 4th leaf sucrose1), *qSPS2* (QTL Sucrose phosphate synthase2) (Causse *et al.*, 1995) and *qstc25* (QTL Starch concentration 25) (Berke and Rocheford, 1995).

As has been stated earlier, the CMTV forms one component of the ISYS system. These components are easily integrated with each other and with web-based resources. Although not presented here, it is entirely feasible for a user to link out to sequence information (such as the molecular databases at NCBI and BLAST services) and tools for viewing pathway information. For example, in Figure 4, the user might want to find sequences for a set of genes in the region of interest, or determine where in the carbohydrate metabolism pathway certain genes of interest are found.

Discussion

A number of other tools similar in some respects to the CMTV tool are publicly available and merit consideration here. cMap (http://www.gmod.org/cmap) is a web-based system for comparative mapping analysis and display that is being developed as part of the Generic Model Organism Database (GMOD) project; this provides an excellent mechanism for publishing sets of maps and predetermined analyses of their correspondences, with many useful navigational features. Other server-side, largely browser-based approaches are taken by tools available at NCBI (http://www.ncbi.nlm.nih.gov/mapview) and MaizeDB (http://www.agron.missouri.edu/cMapDB/cMap.html. The server-side nature of these tools is appropriate for large-scale public resources, but can make it difficult for an end user to integrate local, private data with the public information.

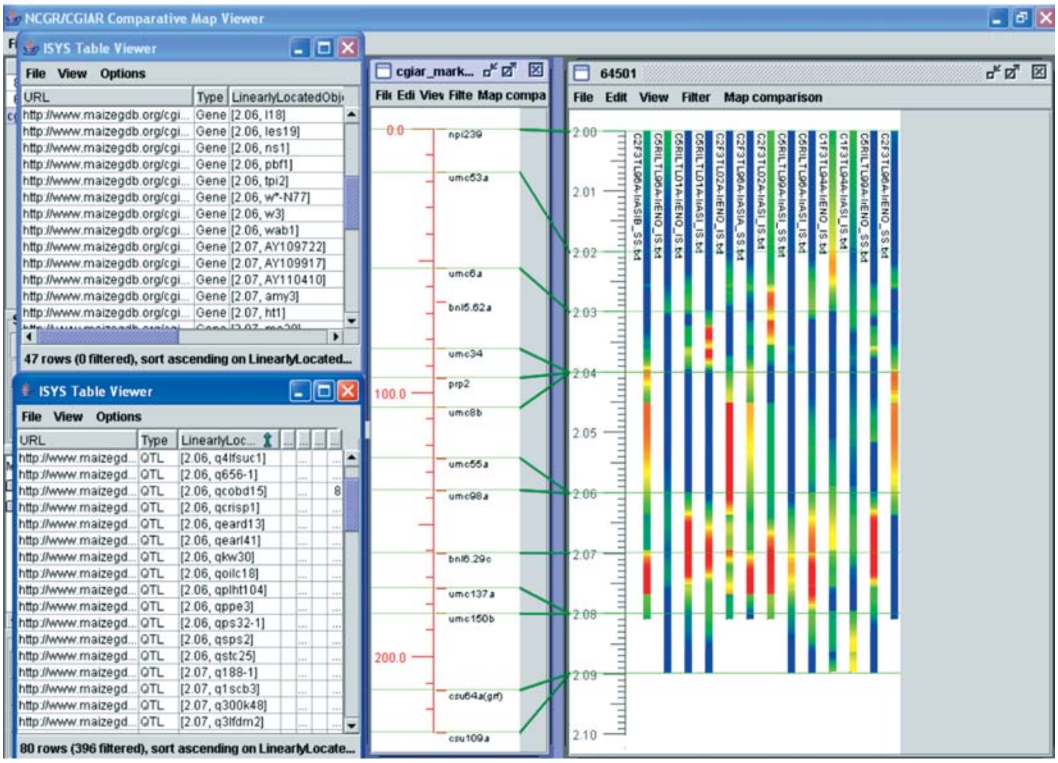
(A)



Map	Mappe...	Refer...
8712...	638	☑
64501	1271	☑
cgiar_...	14	□

Map comparisons
Comparison generate
Comparison generate

(B)



ISYS Table Viewer

URL	Type	LinearlyLocatedObj
http://www.maizegdb.org/cgi...	Gene	[2.06, l18]
http://www.maizegdb.org/cgi...	Gene	[2.06, les19]
http://www.maizegdb.org/cgi...	Gene	[2.06, ns1]
http://www.maizegdb.org/cgi...	Gene	[2.06, pbf1]
http://www.maizegdb.org/cgi...	Gene	[2.06, tpi2]
http://www.maizegdb.org/cgi...	Gene	[2.06, w*-N77]
http://www.maizegdb.org/cgi...	Gene	[2.06, w3]
http://www.maizegdb.org/cgi...	Gene	[2.06, wab1]
http://www.maizegdb.org/cgi...	Gene	[2.07, AY109722]
http://www.maizegdb.org/cgi...	Gene	[2.07, AY109917]
http://www.maizegdb.org/cgi...	Gene	[2.07, AY110410]
http://www.maizegdb.org/cgi...	Gene	[2.07, amy3]
http://www.maizegdb.org/cgi...	Gene	[2.07, ht1]

47 rows (0 filtered), sort ascending on LinearlyLocated...

ISYS Table Viewer

URL	Type	LinearlyLoc...	...	...	...	...
http://www.maizegd...	QTL	[2.06, q4lfsuc1]		...		...
http://www.maizegd...	QTL	[2.06, q656-1]		...		...
http://www.maizegd...	QTL	[2.06, qcobd15]		...		8
http://www.maizegd...	QTL	[2.06, qcrisp1]		...		...
http://www.maizegd...	QTL	[2.06, qeard13]		...		...
http://www.maizegd...	QTL	[2.06, qearl41]		...		...
http://www.maizegd...	QTL	[2.06, qkw30]		...		...
http://www.maizegd...	QTL	[2.06, qoilc18]		...		...
http://www.maizegd...	QTL	[2.06, qplht104]		...		...
http://www.maizegd...	QTL	[2.06, qppe3]		...		...
http://www.maizegd...	QTL	[2.06, qps32-1]		...		...
http://www.maizegd...	QTL	[2.06, qsps2]		...		...
http://www.maizegd...	QTL	[2.06, qstc25]		...		...
http://www.maizegd...	QTL	[2.07, q188-1]		...		...
http://www.maizegd...	QTL	[2.07, q1scb3]		...		...
http://www.maizegd...	QTL	[2.07, q300k48]		...		...
http://www.maizegd...	QTL	[2.07, q3lfdm2]		...		...

80 rows (396 filtered), sort ascending on LinearlyLocate...

BioMercator (version 1.1.1) (Arcade *et al.*, 2004), like CMTV, is a Java™-based client side application for visualizing genetic map information that fulfills many of the same basic objectives as the CMTV. Like CMTV, it can be used to compile several maps into a consensus, although it uses an iterative pairwise approach that makes the order that maps are added crucial; also, it uses a similar concept of 'projection' to handle dynamic coordinate system translation. One interesting feature employed by BioMercator but not yet by CMTV is a statistical meta-analysis function (Goffinet and Gerber, 2000) that determines the most probable number of 'real' QTLs from a set of N QTLs detected by independent experiments. The determination of a consensus position for a set of QTL will often reduce the size of the confidence interval around this position, thus reducing the number of genes that may locate to the particular genomic region and facilitating candidate gene identification.

The CMTV is the only freely available molecular marker and QTL display software that displays complete QTL output in an easy and visual manner, thus enabling a user to examine the stability of QTLs across environments and crosses; a key issue in identifying target genomic regions for MAS. Representing output from QTL analyses visually as a colored strip and aligning this to molecular genetic maps is an effective method for identifying stable LR values (high and low) within large QTL datasets. It is also distinguished from other tools in its emphasis on dynamic user-driven customization of data inclusion and presentation. Much of the strength and novelty of the tool is to be able, at a click of a mouse, to display a broad array of attributes on a computer screen, to be able to manipulate this in any manner the user chooses, and to easily link these data to a wealth of additional data stored both 'in-house' and/or in public databases. Although the CMTV was not developed to conduct novel statistical analyses, it has been developed with the idea that it could serve as a front end for specifying the inputs to such analyses and displaying the results, as it currently does for a set of sequence analysis components through its integration into ISYS. Although the example presented here is specifically related to consensus mapping in maize, this tool can be used for comparative and consensus mapping for any species and among genomes of closely related taxa. This tool is already being used to answer research questions at CIMMYT and is furthering our understanding of the genetic basis of drought tolerance at flowering time in maize. It is hoped that in the near future other groups will use the CMTV in addressing their own particular research questions.

The CMTV tool is freely available as part of the ISYS™ platform. To obtain a working copy of the latest version, please contact info@ncgr.org. Use of the tool is actively encouraged and it is hoped that the use of the tool by more researchers will provide feedback on the utility of the tool and will lead to the development of additional functions useful in a variety of research contexts. We are continually enhancing the tool through the addition of new functionality and improvement of existing functions. For example it will be possible in future releases of the CMTV to integrate allelic diversity data obtained from a variety of different types of markers and from a range of genotypes into the tool. The data may be visualized in a similar fashion to the expression data, by representing each genotype as a separate 'strip' where a particular color may represent the allele phenotypic values of genes involved in the expression of target pathways. We would also like to pursue the development of the functionality of the tool with respect to the identification of orthologous regions across different taxa. Although the CMTV can already be used to display syntenic relationships among genomes from closely related species such

◄

Figure 4. (A) Map comparison between maps for chromosome 2. On the far left, the 'bins' map, in the center a marker map for a local experiment and on the far right, the IBM 2 neighbors map (Polacco *et al.*, 2004). This map displays two sets of genes. The first are unselected genes loaded for a significant region of the QTLs. The second, displayed in the discontinuous color strips to the immediate right, are seven selected genes that correspond to gene expression values for two experiments. The color coding used here indicates the level of expression; red represents genes underexpressed in parent Ac7643 relative to parent Ac7729/TZSRW while green represents genes overexpressed in Ac7643 relative to Ac7729/TZSRW. In addition, a set of traits from three different crosses represented by the continuous color strips has been projected onto the map using marker correspondences. Finally, two histograms display the density of markers and genes located on the IBM 2 neighbors map. (B) The same set of QTL results projected onto the bins map, with two tabular views containing the genes and QTLs respectively found in the bins corresponding to the region of most significance for this set of traits.

as maize and sorghum, some of the functions need to be adjusted in order to work more appropriately in situations when subregions of a chromosome in one species translate to subregions on different chromosomes in another species.

Acknowledgements

The authors would like to thank B. Beavis and D. Gessler (NCGR), R. Bruskiewich, G. McLaren and A. Portugal (IRRI), R. Simon and M. Ghislain (CIP), M. Blair and F. Rojas (CIAT) and J. Alarcon (CIMMYT) for their valuable contributions to the development of the CMV. The authors gratefully acknowledge the Rockefeller Foundation and USAID (Comparative Cereals Genomics Initiative) for financial support.

References

Arcade, A., Labourdette, A., Falque, M., Mangin, B., Chardon, F., Charcosset, A. and Joets, J. 2004. BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. Bioinformatics Advance Access published online on April 1, 2004.

Berke, T. and Rocheford, T. 1995. Quantitative trait loci for flowering, plant and ear height and kernel traits in maize. Crop Sci. 35: 1542–1549.

Causse, M., Rocher, J.P., Henry, A.M., Charcosset, A., Prioul, J.L. and de Vienne, D. 1995. Genetic dissection of the relationship between carbon metabolism and early growth in maize, with emphasis on key-enzyme loci. Mol. Breed. 1: 259–272.

Fang, Z., Polacco, M., Chen, S., Schroeder, S., Hancock, D., Sanchez, H. and Coe, E. 2003. cMap: the comparative genetic map viewer. Bioinformatics 19: 416–417.

Goffinet, B. and Gerber, S. 2000. Quantitative trait loci: a meta analysis. Genetics 155: 463–473.

Heisey, P.W. and Edmeades, G.O. 1999. Maize Production in Drought-Stressed Environments: Technical Options and Research Resource Allocation. Part 1 of CIMMYT 1997/98 World Maize Facts and Trends; Maize Production in Drought-Stressed Environments: Technical Options and Research Resource Allocation. CIMMYT, Mexico D.F.

Koch, K.E. 1996. Carbohydrate-modulated gene expression in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47: 509–540.

Moore, B., Zhou, L., Rolland, F., Hall, Q., Cheng, W.H., Liu, Y.X., Hwang, I., Jones, T. and Sheen, J. 2003. Role of the Arabidopsis glucose sensor HXK1 in nutrient, light and hormonal signaling. Sci. Washington 300: 332–336.

Polacco, M.L., Sanchez-Villeda, H. and Coe, E. Jr. 2004. IBM neighbors - mutual enhancement of genetic and physical maps. Maize Genetics Conference Abstracts. 46: P110.

Prioul, J., Pelleschi, S., Sene, M., Theevenot, C., Causse, M., de Vienne, D. and Leonardi, A. 1999. From QTLs for enzyme activity to candidate genes in maize. J. Exp. Bot. 50: 1281–1288.

Ribaut, J.-M., Bänziger, M., Betrán, J., Jiang, C., Edmeades, G.O., Dreher, K. and Hoisington, D. 2002. Use of molecular markers in plant breeding: drought tolerance improvement in tropical maize. In: M.S. Kang (Ed.), Quantitative Genetics, Genomics and Plant Breeding, CABI Publishing, Wallingford, UK, pp. 85–99.

Ribaut, J.-M., Hoisington, D.A., Deutsch, J.A., Jiang, C. and González-de-León, D. 1996. Identification of quantitative trait loci under drought conditions in tropical maize: 1. Flowering parameters and the anthesis-silking interval. Theor Appl Genet 92: 905–914.

Ribaut, J.-M., Jiang, C., González-de-León, D., Edmeades, G.O. and Hoisington, D.A. 1997. Identification of quantitative trait loci under drought conditions in tropical maize: 2. Yield components and marker-assisted selection strategies. Theor Appl Genet 94: 887–896.

Setter, T.L., Flannigan, B.A. and Melkonian, J. 2001. Loss of kernel set due to water deficit and shade in maize: carbohydrate supplies, abscisic acid and cytokinins. Crop Sci. 41: 1530–1540.

Siepel, A., Farmer, A., Tolopko, A., Zhuang, M., Mendes, P., Beavis, W. and Sobral, B. 2001. ISYS: a decentralized, component-based approach to the integration of heterogeneous bioinformatic resources. Bioinformatics 17: 83–94.

Smeekens, S. 2000. Sugar-induced signal transduction in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 51: 49–81.

Zinselmeier, C., Habben, J.E., Westgate, M.E. and Boyer, J.S. 2000. Carbohydrate metabolism in setting and aborting varies. In: M.E. Westgate and K.J. Boote (Eds.), Physiology and Modeling Kernel Set in Maize, Crop Science Society of America, Madison, WI, pp. 25–42.