J. Plant Biotechnology(2003) Vol. 5(4). pp. 197~200

Review

Journal of Plant Biotechnology

The *Brassica / Arabidopsis* Comparative Genome Browser A Novel Approach to Genome Browsing

Christopher T. Lewis*, Andrew G. Sharpe, Derek J. Lydiate, Isobel A.P. Parkin

Agriculture and Agri-food Canada, Saskatoon Research Centre, 107 Science Place, Saskatoon, Saskatchewan, S7N 0X2, Canada

Abstract

Scalable Vector Graphics (SVG) has enabled a visually appealing, browser-based application for the display of *Brassica* sequences relative to *Arabidopsis thaliana*, and there are currently more than 70,000 *B. napus* Expressed Sequence Tags (ESTs) displayed. The client side of this browser is based on a Custom Graphical User Interface (CGUI) library which uses SVG, a new web graphics standard, to provide windowing functionality inside the web browser. This windowing functionality, combined with asynchronous data retrieval and client side rendering overcomes two of the key technology imposed drawbacks of current web based browsers: Fixed displays and frequent page reloads. The end result is an intuitive and enjoyable browsing experience. The browser is accessible online from the *Brassica / Arabidopsis* Genomics Initiative (http://brassica.agr.gc.ca). Inquiries about the browser should be directed to LewisCT@agr.gc.ca.

Key words: *Arabidopsis*, *Brassica*, Browser, Comparative, SVG, Visualization

Introduction

The *Brassica / Arabidopsis* Genomics Initiative (BAGI) is a project at the Saskatoon Research Centre of Agriculture and Agri-food Canada, which aims to develop genomic and genetic resources which expedite the characterization of gene function in *Brassica* and *Arabidopsis*. Towards this goal the initiative has developed a number of resources including 3' and 5' sequences for *B. napus* cDNA clones, a *B. napus* BAC library, *B. napus* genetic mapping populations, *Brassica* microsatellite markers, an *Arabidopsis* activation tagged population, and global *Arabidopsis* and *Brassica* microarrays. More information about the initiatives resources can be found online at http://brassica.agr.gc.ca or http://www.brassica.ca.

The current emphasis of the BAGI website is on the *Brassica / Arabidopsis* Comparative Genome Browser (BioViz) (Lewis, 2002), which displays *Brassica* DNA relative to homologous regions of the *Arabidopsis* genome. This imparts contextual information on the *Brassica* sequences through access to the *Arabidopsis* annotation, and provides a loose clustering of *Brassica* multi-gene families relative to their *Arabidopsis* homologue. Currently there are more than 70,000 3' and 5' *B. napus* EST sequences available in the browser, and in the future we plan to include expression information from microarray and Serial Analysis of Gene Expression (SAGE) studies as well as publicly available *Brassica* EST sequences from Genbank and *B. oleracea* genomic sequence from TIGR at http://www.tigr.org.

This manuscript outlines the purpose of the browser, discusses its use of Scalable Vector Graphics (SVG) (Ferraiolo, 2003) and contrasts the functionality offered by an SVG based browser with that of a bitmap based browser. There has been no attempt to provide all the functionality of current genome browsers in BioViz and as such there is no function to function comparison.

Discussion

Description of BioViz

BioViz allows us to efficiently display the alignment of many ESTs over a large region of genomic sequence such as a Bacterial Artificial Chromosome (BAC) (Figure 1). We determined the regions of similarity between our ESTs and the

* Corresponding author, E-mail: LewisCT@agr.gc.ca
Received Sep. 4, 2003; accepted Oct. 20, 2003

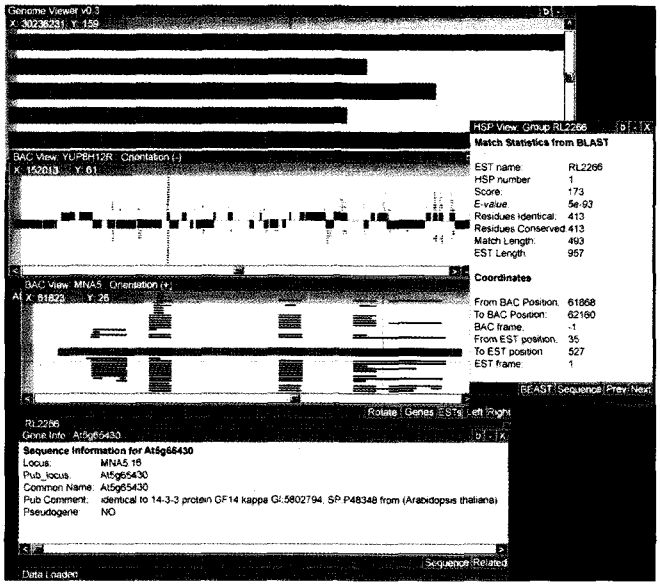

Figure 1. A screen shot of the *Brassica / Arabidopsis* Comparative Genome Browser. This image shows four of the possible views available within the browser. Present is a window showing the 'whole genome view' titled "Genome Viewer v0.3" which presents the five *A. thaliana* pseudo-chromosomes, two windows displaying the 'BAC view' which displays a whole *A. thaliana* BAC that can be scaled and translated to display a sub-region, a 'Gene Info' window which presents information related to the selected gene, and a window showing the 'HSP View' which presents the high scoring sequence pairs related to the selected BLAST hit. In the top right corner of the 'whole genome view' and 'BAC View' windows are boxes indicating the coordinates of the crosshair over the window. The x-coordinate corresponds to the current base pair in the *Arabidopsis* genome or BAC. From both the 'HSP View' and 'Gene Info' window it is possible to access the sequence associate with this feature. From the 'HSP view' it is possible to access the annotation associated with this EST in al MySQL database (*Brassica* EST and Annotation Search Tool or BEAST). From the Gene Info window it is possible to search for related sequences via an on-the-fly BLAST. All features in the 'whole genome view' and the 'BAC View' have a mouse-over event associated with them which displays the feature id in the status bar at the bottom of the window.

BACs in the *A. thaliana* genome using NCBI BLAST (Altschul 1997) and stored hits with an E-value greater than 10e-4 in a MySQL database. BioViz accesses this precomputed data to allow the user to view all ESTs which have similarity to a given BAC, to see how different ESTs align with the predicted *A. thaliana* genes, and to find homologous regions of *Arabidopsis* relative to specific ESTs. The browser is not intended to act as an annotation tool and consequently there is no capacity within the browser to update the underlying dataset. However, the browser could be used to complement an annotation tool by providing a user friendly, publicly accessible (via the web), front end to the underlying database.

There are a number of other genome browsers available (Hubbard 2002; Karolchik 2002; Stein 2002; Couronne 2003) and while they are both useful and informative they have two main technology imposed characteristics which BioViz aims to eliminate through the use of SVG: Static displays and frequent page reloads.

The first limitation inheirent in standard genome browsers, for instance Gbrowse (Stein 2002) and the Ensembl Genome Browser (Hubbard 2002) is that the have a fixed display, which limits the users ability to customize their view of the data. The browsers are based on server generated bitmaps and use an HTML table to provide a pre-defined layout for the interface. Rather than server generated bitmaps and an HTML table, the client side of BioViz is based on an SVG GUI library.

SVG is a relatively new graphics standard based on Extensible Markup Language (XML) (Bray 2000). There are a number of advantages to an XML + SVG based genome bro-wser as compared to the standard bitmap based browser:

1) Because SVG is a vector graphics format you can zoom and pan an SVG image without loosing fidelity (in other words there will be no grainy images). This also means you don't need to wait for the server to generate a new image every time you would like to examine your data from a different perspective.
2) Because SVG is dynamic and interactive, new data can be added to the existing image and the image can be changed in response to user events (i.e. clicking on a gene).
3) Because SVG is an XML based graphics format, developing XML technologies such as Extensible Stylesheet Language Transformations (XSLT) (Clark 1999) can be used to transform XML formatted data into SVG. For instance one might transform MAGE-ML (Spellman 2002) formatted microarray experiments directly into SVG images for display or visual inspection within the browser.

This SVG GUI library, and more generally the dynamic nature of SVG, allowed BioViz to be implemented using a "multi-windowed" format that provides the user with a great deal of flexibility when displaying their data.

The second inherent limitation in the current standard genome browsers is the need for frequent page reloads. For instance, when the user wants to change the current view in either the Ensembl browser or Gbrowse a new image is loaded from the server. Thus when zooming in or out, looking at another region of the genome, or requesting supplementry information a page reload occurs. While functional, the page reload can be distracting and it is unnecessary in an SVG based genome browser.

BioViz was implemented using a client-server methodology with the client side responsible for presentation of the data and the server side responsible for retrieval of the appropriate data at the client's request. Additional information can be requested from the server and because SVG allows a document frag-

ment (or partial image) to be incorporated into the existing image, the data can be added to the view without a page reload - thereby partially removing the need for page reloads.

Furthermore, because an SVG image can be infinitely scaled and translated on the client side without loss of resolution. There is no need for the data to be reloaded each time the user desires a different view of the data - this completely eliminates the need for page reloads. An additional advantage offered by SVG is that requests to the server are asynchronous, which allows the user to continue browsing while the data loads. Asynchronous requests, combined with client side transformations and a flexible display dramatically improve the flow of browsing.

The server side of the browser relies on Perl CGIs to handle client requests, though these could be replaced with Java Servlets or some other mechanism for interfacing with the server. The CGI handles client requests and provides access to the underlying database.

When installing the browser at a new location, the local database administrator (DBA) would have to create CGIs to access their data and return the desired results. These CGIs would implement defined interfaces for message handling and data access. While this makes an "out of the box" installation of the browser impossible, it allows the browser to access an existing database installation, and ensures that the browser is not tied to a particular database managements system or database schema. For instance, the browser could use a database as simple as a list of features and their regions of homology relative to the *Arabidopsis* BACs or it could use an existing MySQL database installation, so long as an appropriate database adapter were created to access the desired data source.

Comparison of BioViz with GBrowse

The Generic Model Organism Database (GMOD), Generic Genome Browser (GBrowse) is based on server generated bitmaps in PNG (Portable Network Graphic) format. The client side of both browsers use JavaScript to provide interactivity, though the use of HTML tables to organize views in GBrowse limits the flexibility of the display. In response to client requests from GBrowse new bitmaps are generated and sent to the client. The resulting page reload forces the users to wait for the new image before they can continue browsing.

The server-side of GBrowse uses CGIs to interact with the system. These CGIs rely on functionality provided by the BioPerl (Stajich 2002) project to fetch the data and render a bitmap representing the data. The use of BioPerl ties GBrowse to Perl which may complicate the use mechanisms other than Perl CGIs for interacting with the server, although the use of BioPerl simplifies data access and display.

The published release of GBrowse uses a MySQL database following either the Bio::DB::GFF or GadFly database schemas, though it is possible to write a new BioPerl database adaptor enabling communication with an existing database installation. This allows an "out of the box" installation of GB rowse, or it can be adapted to work with an existing database installation. Adding features using the Bio::DB::GFF schema requires the DBA to create tab separated feature files in GFF format which can be loaded into the database using the perl scripts provided. A similar mechanism could be employed to add new features to BioViz.

Caveats On Using An SVG Based Browser

As SVG is in its infancy there are a number of technical limitations imposed by its use:

1) There is no native support for SVG in mainstream web-browsers such as Internet Explorer or Netscape Navigator, so SVG must be viewed using a plugin.
2) SVG rendering is not currently hardware accelerated, so translating and scaling the image is processor intensive when displaying large amounts of data. This means that the standard views found in bitmap based browsers may need to be modified to present less data at one time. However this should be possible without diminishing the usefulness of the view.
3) The use of a plugin restricts the browser-OS interaction for certain OS-browser combinations, for instance currently it is only possible to access the clipboard from the SVG plugin using an Internet Explorer/Windows combination.

Conclusions

Using an SVG based genome browser will enable interesting and creative techniques for viewing data relative to a model organism. The multi-windowed display and the asynchronous data retrieval present in the *Brassica / Arabidopsis* Comparative Genome Browser provides a more intuitive and productive browsing environment than standard bitmap based browsers. However, the decision to use an SVG browser will restrict your audience to those willing and able to use a plugin to display SVG (Windows with Internet Explorer, and to a lesser extent Windows or Mac users[1)] using Mozilla or Internet Explorer).

[1)] While there is a beta of the Adobe SVG 3.0 plugin available for Linux, it has been our experience that the graphics do not render well in this version of the plugin. However there is supposed to be a Linux version of the Adobe SVG 6.0 plugin which may resolve the problems in the current version.

Thus the need to reach a broad audience must be balanced with the ability to exploit the novel features of SVG.

Ultimately, to reach the full spectrum of users, the best option might be to provide both interfaces for your data. This could be accomplished using the same server side setup, which returns either SVG document fragments or complete bitmaps. This would allow an SVG based browser such as the *Brassica / Arabidopsis* Comparative Genome Browser to access the full range of functionality present in the standard genome browsers, but provide it with a more flexible interface.

Acknowledgements

This work was partially funded by the Genome Prairie grant 'Functional Genomics of Abiotic Stress of Crops'.

References

Altschul SF, Thomas ML, Alejandro SA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) "Gapped BLAST and PSI-BLAST: A new generation of protein database search programs". Nucleic Acids Res 25: 3389-3402

Bray T, Paoli J, Sperberg-McQueen CM, Maler E (eds) (2000) "Extensible Markup Language (XML) 1.0 (Second Edition)", W3C Recommendation, http://www.w3.org/TR/REC-xml

Clark J (ed) (1999) "XSL Transformations (XSLT)", W3C Recommendation, http://www.w3.org/TR/xslt

Couronne O, Poliakov A, Bray N, Ishkhanov T, Ryaboy D, Rubin E, Pachter L, Dubchak I (2003) Strategies and Tools for Whole-Genome Alignments. Genome Research 13: 73

Ferraiolo J, Fujisawa J, Jackson D (eds) (2003) "Scalable Vector Graphics Specification", W3C Recommendation, http://www.w3.org/TR/2003/REC-SVG11-20030114/

Hubbard T, Barker D, Birney E, Comeron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminiecki L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M (2002) "The Ensembl genome database project", Nucl. Acids Res 30: 38-41

Karolchik D, Kent WJ (2002) "The UCSC Genome Browser", *Current Protocols in Bioinformatics*, Baxevanis, A.D., ed. John Wiley & Sons, Inc

Lewis CT, Karcz S, Sharpe AG, Parkin IAP (2002) "BioViz: Genome Viewer". SVG Open 2002 Conference Proceedings, http://www.svgopen.org

Spellman PT, Miller M, Stewart J, Troup C, Sarkans U, Chervitz S, Bernhart D, Sherlock G, Ball C, Lepage M, Swiatek M, Marks WL, Goncalves J, Markel S, Iordan D, Shojatalab D, Stoeckert CJ Jr, Brazma A (2002) "Design and imple-mentation of microarray gene expression markup language (MAGE-ML)", Genome Biol 3(9): RESEARCH0046

Stajich JE, Block D, Boulez K, Brehner SE, Chervitx SA, Dagdigian C, Fuellen G, Gilbert JG, Korf I, Lapp H, Lehvaslaiho H, Matsalla CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD, Birney E (2002) "The Bioperl toolkit: Perl modules for the life sciences", Genome Res 12: 1611-1618

Stein LD, Mungall C, Shu S, Caudy N, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S (2002) "The generic genome browser: A building block for a model organism system database", Genome Res 12(10): 1599-610