Bluejay: A Biological Sequence Browser featuring Knowledge Integration

Gordon, P.M.K., Stromer, J., Turinsky, A.L., Xu, E., Sensen, C.W.

Sun Center of Excellence for Visual Genomics

University of Calgary

Faculty of Medicine

Department of Biochemistry and Molecular Biology

HS1150, 3330 Hospital Drive NW, Calgary, Alberta, T2N 4M1

Canada

☎ (403) 210-9543

Correspondence should be addressed to:

csensen@ucalgary.ca

ABSTRACT

Bluejay is a new browser for biological sequences. It offers users a dynamic and highly customizable display of genomic or proteomic data, with several levels of visual manipulations. It serves as a visual front-end unifying the access to distributed resources, such as bioinformatics data repositories and Web services. Bluejay imports several data types, combining them into a coherent visual model. It can also be integrated with other software tools. The user can interact visually with the whole sequence, functional categories and individual elements. Features of Bluejay include semantic zooming, several types of customizable image granularity, Scalable Vector Graphics (SVG) imagery, and session management. Bluejay was designed for ease-of-use by biologists, and maximum availability. It is available in three versions (http://bluejay.ucalgary.ca/install): 1) a secure applet, especially suitable for new users who do not wish to or cannot install software on their machine, 2) Java Web Start for typical on-line users, and 3) a standalone application for off-line usage. Bluejay was created to interoperate with existing software in a broader bioinformatics context through the extensive use of XML-based standards and protocols.

CATEGORY

- Post-Genomic Management, Integration and Mining
- DNA and Amino acid Sequence Comparison
- Software Application

Keywords: Genome Visualization, Data Integration, Web Services, BioMOBY, XML, SVG

1. INTRODUCTION

The amount of data collected by life scientists is growing at a breathtaking rate. In the last ten years, the number of genomic sequences submitted to GenBank has been doubling faster than every two years. Our ability to extract knowledge from such wealth of data increasingly depends on the access to sophisticated bioinformatics tools. The importance of the underlying gene sequence for all types of bioinformatics analysis cannot be overestimated. A clear picture of genes' functions, locations and interrelations provides a context for the proper analysis of microarrays, protein-protein interactions and all other data derived from genes.

Bioinformatics has some unique challenges. As a rule, life scientists are not well trained in computer technology. Therefore, development of new tools requires balancing power and ease of use. This has given rise to many visual data exploration tools. As the analysis of molecular sequences remains a fundamental problem of bioinformatics, tools for genome function visualization are especially important. There is also a growing appreciation of the complexity of the biological information. System biology is emerging as a fundamental research direction, where the ultimate goal is to combine the knowledge of different biological aspects into a unified model of an organism.

Based on the Web browser interface model, the Bluejay browser is intended to deliver some much-needed innovation in the area of genome function visualization. Our ultimate goal is to provide the life science community with a tool for visual data exploration and integration that is familiar in look-and-feel, while adding functionality that is both more powerful and novel compared to other browsers.

2. MATERIALS AND METHODS

2.1 General Concepts

In this paper we present a biological sequence browser called Bluejay (Browser for Linear Units in Java). Bluejay visualizes sequence data, enables visual data queries, and relates the data to other sources of knowledge. It is designed as Java client-server software that makes substantial use of W3C open standards for XML.

Input sequence data may be encoded in a number of formats. Non-XML data, such as FastA flat files, are converted internally to XML documents. Using XML internally allows us to use existing Java-based freeware for parsing, which can be plugged into Bluejay as a set of separate modules. The logical structure of the document, in the XML form, serves as a template for accessing, displaying and manipulating the underlying data. A proof-of-concept prototype of Bluejay was previously described in [1]. Figure 1 presents the flowchart describing the main data processing steps in Bluejay.
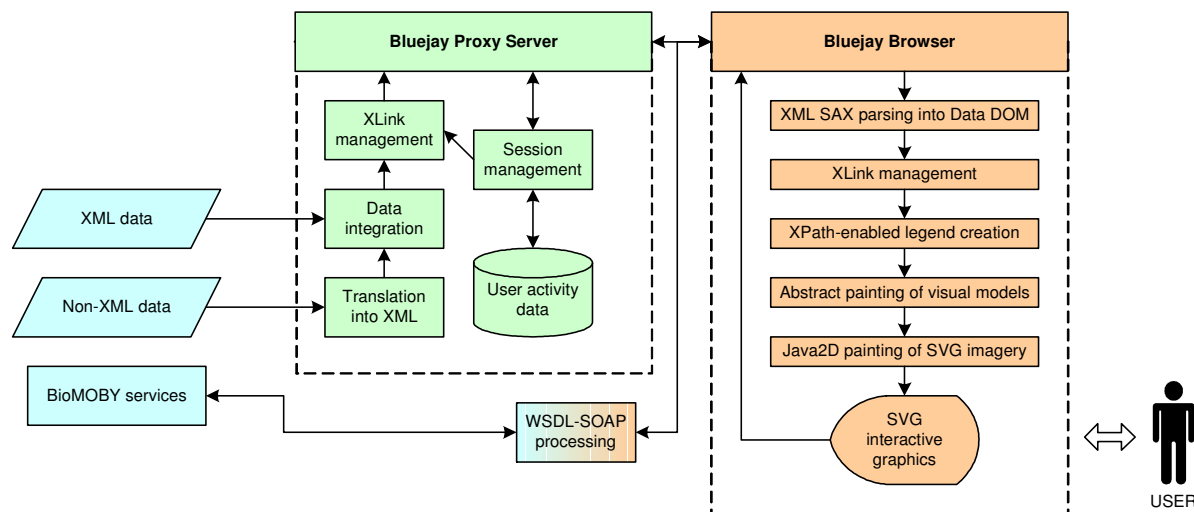

Figure 1. Bluejay components and information flow.

2.2 Proxy Server Side

The proxy server establishes the communication between the Bluejay browser side and the sources of biological content: data, websites and data analysis services. This includes retrieval of remote data requested by the user, data preprocessing, delivery of the data to the Bluejay browser side for visualization, and session management. The need for a proxy originated with applet security restrictions, but has persisted mainly because of applet memory limitations and the potential benefits of centralized knowledge for "intelligently" managing a user's browsing experience. The proxy and client have been designed to not be codependent - if no Internet connection is available, the client application will still run, with a limited loss of functionality.

Data preprocessing may involve several aspects: translation into XML, insertion of XLinks, and personalization. A part of the proxy server is a data translator that translates flat files and other non-XML data into XML before passing the information to the browser. The proxy may also insert additional hyperlinks

into the data, to relate specific sequence elements to external sources of knowledge on-the-fly. Bluejay supports the powerful XLink standard for XML hyperlinks, which allows such advanced features as multiple links originating from the same location, links between third-party data, and persistent link libraries (linkbases). The Bluejay proxy server also performs session management. Users can create and manage projects, logon to a previous work session from a different computer, or switch between several simultaneous work sessions. Future improvement of the session management, such as an upcoming engine for content personalization, will also be handled at the proxy side.

2.3 Client Side

The browser side receives and visualizes the incoming XML data. It also offers the user a range of visual tools to query the data, customize its appearance, and save the results. The client is generally patterned after a Web browser interface, with a location bar, side tabs, and back

and forward buttons to aid in the navigation between XML resources. Non-XML resources that are not translatable by the proxy are delegated for viewing to the user's default Web browser. The client communicates with the proxy over the Web using SOAP [2]. The incoming XML data is parsed into an internal DOM tree [3], where nodes represent nested XML tags. Relevant XLinks, including third-party links, are identified and registered. The visual model is then created as a collection of backbone sequences, such as DNA or protein backbones, with respect to which all other visual elements are placed.

Bluejay painting logically separates high-level placement of elements onto a sequence from low-level rendering of images onto a canvas. High-level painting of an element (e.g. a gene) is performed by specifying its positions on a target sequence (e.g. a DNA molecule). Note that sequence positions are intrinsic to the data and remain constant as the user manipulates the visual model. At a lower level, the Java object representing the target sequence translates positions of sequence features into canvas coordinates and calls the respective Java Graphics2D class. This separation into high-level and low-level painting makes Bluejay architecture more robust and easily extendable to new data types.

The Bluejay graphics is based on the Scalable Vector Graphics (SVG), a cross-disciplinary XML standard for imagery. The Java Graphics2D object for low-level painting is adapted to create SVG tags. The resulting SVG document is an explicit and portable XML representation of the visual model. An SVG-enabled canvas visualizes this model and allows the user to interact with it. Bluejay translates these visual interactions into actions on the original sequence data. To implement SVG, we used the open-source Apache SVG browser Batik [4], plugging it into the Bluejay main browser and connecting it to the data DOM tree. Data access is based on XPointer and XPath standards for DOM [5].

2.4 Dealing with Large Data Sets

Bluejay employs several techniques to overcome bandwidth, CPU and memory limitations associated with manipulating genome-scale data.

2.4.1. Bandwidth limitations

Two techniques are used to minimize network traffic: compression and skeletonization. The Bluejay client will decompress zipped files in the case that the incoming document was a zipped file or just zip-encoded for transport efficiency by the originating Web server. We

have found that many of the biological XML files achieve five-fold or better compression using the ubiquitous gzip compression utility. Skeletonization involves the transformation of a large document into a pared-down version via an XML transformation stylesheet. The data removed generally contain very detailed information not required for the initial high-level rendering of the document, but when the user "drills-down" into the image (i.e. displays a fraction of the whole molecule in detail). The stubs for these detailed data are represented by embedded XLinks. Skeletonization is an experimental feature in Bluejay, requiring the on-demand brokering of requests for detailed information as a user drills down into a large document such as a eukaryotic chromosome. Together these two techniques can make unwieldy data browsing manageable; for the TIGR *Arabidopsis thaliana* chromosome 2 annotation, the original document size of 30.5MB was reduced to 2.9MB with skeletonization, 3.2MB with compression, and 244KB with both methods used in combination.

2.4.2. CPU limitations

CPU bottlenecks tends to occur during data analysis (such as sequence alignments) and redrawing after layout or position changes. The former is dealt with by farming out analysis to Web services, while the latter is dealt with by using an efficient 'windowing' mechanism. Because the position of features does not change while browsing a document, an index is constructed on document load and constantly reused to only call the paint function for elements which should be visible on the screen. This is important when you consider that a eukaryotic chromosome may have hundreds of thousands of painted elements.

2.4.3 Memory limitations

Applets are restricted to 64MB of memory, making the viewing of very large documents impossible. For bacteria-sized genomes, applet users can take advantage of the "memory-saving" mode of Bluejay, which does not keep the DOM in memory during browsing, but reloads the document from a local disk cache when required. For users with Java 1.3 and above, the Web Start option gives them a one-click launch solution without the memory restriction.

3. RESULTS

3.1 Bluejay gives control to the user

Bluejay can visualize biological sequences ranging from simple bacterial genomes to complex eukaryotic genomes. It currently provides a fully customizable view of the

various sequence features: genes, introns and exons, repeat regions, markers, promoter and terminator sites, G+C% and A+G% content plots, as well as various Gene Ontology classes [7]. Bluejay supports several common data formats: TIGR XML, GenBank XML, FastA, BioML, BSML, Agave XML - which enables it to handle the majority of existing genomic sequence data. Figure 2 shows a sample Bluejay interface.
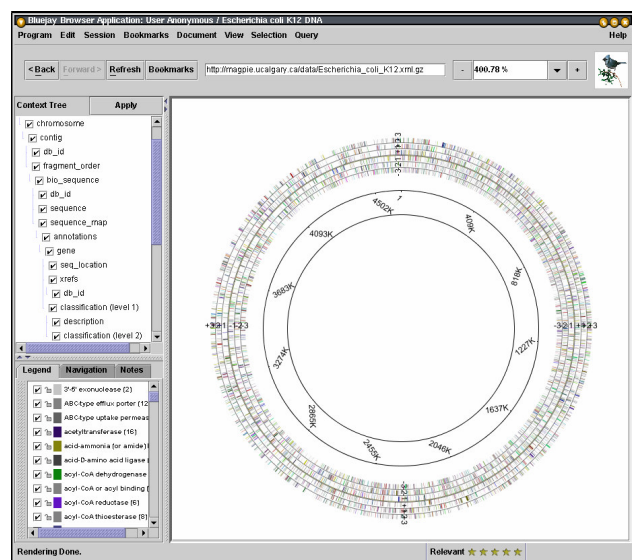

Figure 2. Bluejay displaying the *E. coli* K12 genome

With Bluejay, the user has a wide range of options to customize the visual display, adjust the level of detail, and perform visual operations on the sequence data. The user can also save the customized visual model in a scalable publication-quality format.

There are three main ways to customize the level of detail. First, Bluejay provides several view modes, such as a pie-chart summary of sequence features, a graphical view of the sequence, an expanded graphical view with all six open reading frames shown, and a textual view of the sequence as a string of characters. All views are color-coded consistently according to feature types and functional categories as defined in the legend. Secondly, Bluejay provides a semantic zoom, which shows new, finer features of the data at a higher resolution [6]. The zooming range in Bluejay is virtually unlimited. Third, the granularity can be adjusted through the context tree, a visual tool in the main browser window that contains a semantic hierarchy of the XML data elements. Nesting of XML elements often represents a semantic refinement, such as a Gene Ontology hierarchy of terms [7] or additional XML tags with sequence annotation details.

By setting the level of nesting in the context tree, the user can customize the desired level of refinement.

Bluejay allows the user to perform visual data explorations in several scopes. First, sequence-wide operations include switching between linear and circular models of the sequence (important for visualizing bacterial genomes), switching between normal and reverse-complement views of a double-stranded sequence, rotation by a desired angle, and "cutting" at a specified start position, e.g. at the beginning of a gene. Second, an interactive legend enables operations on specific functional categories. For example, the user may click on the legend item "Enzymes" and either ghost out or completely remove all enzyme-producing genes from the visual model of a DNA. By setting the level of details in the context tree, the granularity of the legend can be adjusted, from a longer list of narrow functional categories, e.g. various types of enzymes, to a shorter list of broad categories. Third, the user can access any individual element on the sequence through mouse clicks. For example, by clicking on an image of a gene, the user can access that gene's data and XLinks, and thus be able to request a gene annotation web page or relevant Web-based analysis services (see below).

A key benefit of using Bluejay is the ability to save the image in the SVG format. Unlike pixel-based image format (JPEG, GIF, etc.), SVG allows an image to retain its full quality regardless of the scale. The saved image can then be resized to any resolution, from small images for journal articles to large posters. A wall-size poster image of a complete genome will retain its finest details in SVG format. SVG images can also be edited by any SVG-compliant visual editor (e.g. Corel Draw).

3.2 Bluejay is made easy to use for biologists

The success of bioinformatics software depends critically on its user friendliness. Life scientists are much less likely to use software that requires an above-basic degree of computer expertise. Bluejay provides the ease-of-use through several aspects: ease of access, comfortable interaction, and session management.

Bluejay can be accessed in three versions: as a web applet, one-click Web Start or a downloadable application (http://bluejay.ucalgary.ca/install). For first time users, a web applet is the ideal environment to get familiar with the tool. Bluejay provides a VeriSign-certified secure applet [8], which allows users to save files locally or upload local data into the applet. If a complex task

requires large memory resources, the user may use Web Start or download a platform-independent application version. The only requirement is to have a current version of Java Virtual Machine (JVM) on the user's system, which is commonly available (http://java.sun.com). If Java is not installed, the Bluejay download website provides a link to download both the JVM and Bluejay in a single click. Java enjoys widespread name recognition among biologists, which further reduces the psychological resistance to use Bluejay.

All interaction with the browser is performed through the GUI. No programming, scripting, or command line parameters are required to configure the browser or to analyze the data. This makes work with Bluejay highly intuitive, eliminates any requirements on user expertise, and shortens the learning curve. The Bluejay Help menu contains a step-by-step explanation of the navigation process and main features.

Session management further improves usability. Bluejay can recognize the user and resume a previously interrupted investigation, possibly from a different computer. We shall soon implement a personalization engine to be able to prioritize the content (e.g. data sources and Web services) according to individual preferences. Session management is optional and users may still work with Bluejay anonymously.

3.3 Bluejay is a knowledge integration tool

Meaningful data integration is among the most important technical tasks in the area of bioinformatics. While powerful methods have been developed to analyze various types of data separately, the emerging grand challenge for biology is to combine these analyses into a coherent model of an organism. Bluejay development aims at creating a data exploration tool that can interoperate with other tools and integrate heterogeneous biological information into its visual model.

Users can access other tools and data resources from Bluejay. This functionality is enabled by an extensive support for XLinks. For example, graphically-rich MAGPIE gene annotation pages [9] (http://magpie.ucalgary.ca) can be launched by clicking on gene representations in Bluejay. MAGPIE currently hosts 90 publicly available genomes and uses powerful sequence analysis accelerators, such as Paracel GeneMatcher2 [10] and TimeLogic DeCypher [11] to produce comprehensive annotations of these genomes. Through XLink insertion and management, Bluejay relates sequence visualization to a wealth of functional annotation information provided by MAGPIE, both visual and textual. XLinks can lead to other desired sources, such as organism-specific databases, NCBI web resources, and PubMed article citations. Bluejay thus serves as a visual front-end to a distributed collection of bioinformatics resources and tools.

Conversely, other bioinformatics tools can launch a Bluejay browser. For example, icon hyperlinks launching a Bluejay applet for each genome are already available from the MAGPIE Web site. They are enabled by a small JavaScript, which simply brings up a Bluejay applet web page, passing it the name of a genome data file associated with the clicked icon. Users of MAGPIE who explore the static sequence annotation pages can thus launch a dynamic browser to explore any of nearly a hundred genomes using the "Bluejay" icon links available on the main genome listings pages.

Figure 3. A portion of the *Sulfolobus solfataricus* genome with oligonucleotides embedded into the visual model. The view also shows repeats (green), and both G+C% (black) and A+G% (red) composition.

Bluejay can combine data from different sources into a meaningful visual model. To date, we have successfully integrated genomic sequence annotations from MAGPIE with oligonucleotide primer computations by Osprey [12]. The combined display (Figure 3) is a helpful visual guide for the production of DNA microarray chips. Our current work is on integrating microarray gene expression data into the Bluejay visualization. As a result, researchers using Bluejay will be able to inspect regions of an organism's genome that are active, and better comprehend genome-wide patterns of protein production and metabolic pathways. Other data types that we plan to integrate in the future include protein modification and protein-protein interaction data.

3.4 Bluejay facilitates biology Web Services

BioMOBY [13] is a protocol consisting of a common XML object ontology for biological entities, and a standardized request/response mechanism for performing analysis of the Web on these objects. Bluejay is the first Java-based BioMOBY client. Bluejay contains semantic mapping between its supported XML formats and the BioMOBY object ontology, allowing the user to directly link from the visually presented data to BioMOBY compliant Web services. Invoking Web-based analysis of the source data simply involves clicking on its Bluejay visual representation, and selecting options from the resulting pop-up menu. Results are displayed in HTML format in a separate window, and can be used to requery BioMOBY, thus forming analysis chains. As BioMOBY services grow in number and diversity we see a clear opportunity to use information gleaned by the proxy to help facilitate service selection.
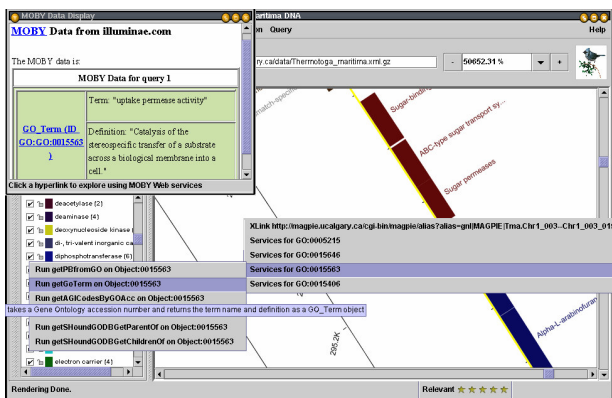

Figure 4. Interface for BioMOBY Web services activated by clicking on the graphical representation of a gene, with service options in cascading popup menus. Menu items have fly-over "tool tips". The results of a service are displayed in a separate window, shown here in the upper left hand corner.

4. DISCUSSION

To meet the growing demand for bioinformatics tools, a number of biological sequence browsers have already been developed. Currently, the most advanced general-purpose browsers include Ensembl Genome Browser [14], NCBI Map Viewer [15], TIGR Comprehensive Microbal Resource [16], and UCSC Genome Browser [17]. These tools deliver their content as dynamically generated web forms and image maps. Other, less well known tools exist that are based on the same principle. The advantage of using these browsers lies in their immediate access to a wealth of information at each facility. In contrast, Bluejay must explicitly relate its data to other knowledge sources through XLinks. However, compared to Bluejay, these browsers have five main limitations.

First, the user has far less control over the sequence appearance. Of the browsers listed above, only the Ensembl browser offers a comparable range of options for customizing the display. Second, the user is forced to communicate with a remote web server for every interaction, which brings up several technical issues such as limitations of bandwidth or availability of a stable Internet connection. Unlike Bluejay, there are no downloadable application versions available. Most of the other major tools were created primarily to showcase the genomic information accumulated at a principal facility (EBI/Sanger, NCBI, TIGR and UCSC). None is, therefore, primarily designed to import additional data for visualization or interoperate with other resources. Fourth, these tools can only export pixel-based images, which, unlike SVG images, lose their quality after resizing. Finally, these browsers do not currently offer any session management or personalization.

Several sequence browsers exist in the form of applets and downloadable applications, e.g. the Vista Genome Browser [18] and the Apollo Genome Annotation Tool [19]. While allowing an easy web access, applets eliminate the need to communicate with a remote web server for each interaction and typically offer more options for manipulating the data. However, unlike the VeriSign-certified Bluejay applet, existing genome browser applets are not secure and, therefore, cannot save or read data files locally. This is a major limitation. Currently, we are not aware of any secure applet tool other than Bluejay for browsing biological sequences.

New genomic browsers with advanced features are being created, such as Gbrowse and Sockeye. Gbrowse [20] is designed as a front end to a Generic Model Organism System Database Project (GMOD). Like Bluejay, it is a

component-based tool that supports BioMOBY services and can be installed as a front end to other data sources. However, Bluejay differs from Gbrowse in several ways. Bluejay is designed to visualize a variety of data types, whereas Gbrowse focuses on comparison of sequence annotations. The Bluejay client itself does not include database components and is purely a front-end visualization tool. The Bluejay interface is not hypertext-based and therefore is not subject to the associated limitations such as static images, browser incompatibilities, and constant round-trips to the remote server for display changes. Bluejay uses components based on interdisciplinary standards that are not a part of Bluejay project, which makes its modular architecture more generic.

Another example of recent advances is the Sockeye browser [21]. Sockeye is the core data viewer of the Bioinformatics of Mammalian Gene Expression project. It is a downloadable Java 3D application that can also be launched using Java Web Start. Key Sockeye features include 3D visualization of genomic sequences and their alignments and support of Ensembl data. However, unlike Bluejay, Sockeye offers considerably less functionality for XML open standards and XML technology, such as XML hyperlinks, XML data manipulation, or SVG graphics. While Sockeye can visualize external data such as GFF or Ensembl data, its interoperation with a wide range of other bioinformatics tools has not been a priority in the Sockeye design.

We further observe that many existing browser applets and applications were created for a specific task. Consequently, they lack the general-purpose nature of Bluejay and its broad range of features. Very few support scalable graphics imagery, the specialized Brassica / Arabidopsis Comparative Genome Viewer [22] and Human Protein Reference Database [23] being notable exceptions. To our knowledge, none of the other existing browsers offers session management.

5.   CONCLUSIONS

Bluejay is a biological sequence browser that reflects the current state-of-the-art in bioinformatics research. It is a platform-independent Java software tool available both as a secure web applet and as a downloadable application. Bluejay provides an intuitive, fully visual exploration of biological sequences, with a wide range of options for customizing the visual model and querying the data. It creates publication-quality scalable images of the data, offers session management options, and is designed to interoperate with other bioinformatics tools.

Given the clear importance of gene function context for all sorts of bioinformatics analysis, we see great potential for sequence visualization and visual analysis. The growing influence of bioinformatics on other areas of biology is a trend that will continue for the forseeable future. This will require new, more powerful, more versatile tools for data analysis and knowledge integration. The dynamic sequence browser presented in this work is a next step in this important direction.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1]   Gordon P, Sensen CW (2000) Bluejay: A Browser for Linear Units in Java. In: Pollard A, Mewhort DJK, Weaver DF (eds.) High Performance Computing Systems & Applications. Kluwer Academic Publishers, 183-194.

[2]   Simple   Object   Access   Protocol   (SOAP), http://www.w3.org/TR/soap

[3]   Document Object Model (DOM), http://www.w3.org/DOM

[4]   Batik SVG Toolkit, http://xml.apache.org/batik

[5]   World Wide Web Consortium, http://www.w3.org.

[6]   Bederson B, Hollan J (1994) Pad++: A zoomable graphical interface for exploring alternate interface physics. Proc. ACM User Interface Software & Technology, 17-24.

[7]   Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V,

de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R; Gene Ontology Consortium (2004) The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32 Database issue: D258-61.

[8] VeriSign Inc., http://www.verisign.com

[9] Gaasterland T, Sensen CW (1996) MAGPIE: Using multiple tools for automated genome interpretation. Trends in Genetics, 12:76-78.

[10] Paracel GeneMatcher2, http://www.paracel.com/.

[11] TimeLogic DeCypher, http://www.timelogic.com/.

[12] Gordon P, Sensen CW (2004) Osprey: a comprehensive tool for the design of oligonucleotides for DNA sequencing and microarrays. Submitted.

[13] Wilkinson MD, Links M (2002) BioMOBY: an open source biological web services proposal. Brief Bioinform. Dec 3(4):331-41.

[14] Birney E, Andrews D, Bevan P, Caccamo M, Cameron G, Chen Y, Clarke L, Coates G, Cox T, Cuff J, Curwen V, Cutts T, Down T, Durbin R, Eyras E, Fernandez-Suarez XM, Gane P, Gibbins B, Gilbert J, Hammond M, Hotz H, Iyer V, Kahari A, Jekosch K, Kasprzyk A, Keefe D, Keenan S, Lehvaslaiho H, McVicker G, Melsopp C, Meidl P, Mongin E, Pettett R, Potter S, Proctor G, Rae M, Searle S, Slater G, Smedley D, Smith J, Spooner W, Stabenau A, Stalker J, Storey R, Ureta-Vidal A, Woodwark C, Clamp M, Hubbard T. (2004) Ensembl 2004. Nucleic Acids Res. Jan 1;32 Database issue:D468-70.

[15] Feolo M, Helmberg W, Sherry S, Maglott DR. (2000) NCBI genetic resources supporting immunogenetic research. Rev Immunogenet. 2(4):461-7.

[16] Peterson JD, Umayam LA, Dickinson T, Hickey EK, White O. (2001) The Comprehensive Microbial Resource. Nucleic Acids Res. 29(1):123-5.

[17] Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D, Kent WJ; University of California Santa Cruz. (2003) The UCSC Genome Browser Database. Nucleic Acids Res. 31(1):51-4.

[18] 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 Res. 13(1):73-80.

[19] Lewis SE, Searle SM, Harris N, Gibson M, Lyer V, Richter J, Wiel C, Bayraktaroglir L, Birney E, Crosby MA, Kaminker JS, Matthews BB, Prochnik SE, Smithy CD, Tupy JL, Rubin GM, Misra S, Mungall CJ, Clamp ME (2002) Apollo: a sequence annotation editor. Genome Biol. 3(12):RESEARCH0082. Epub 2002 Dec 23.

[20] Stein LD, Mungall C, Shu S, Caudy M, 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.

[21] Astakhova T, Bilenky M, Birney E, Fu T, Hassel M, Melsopp C, Rak M, Robertson AG, Siddiqui A, Sleumer M, Jones SJM (2004) Sockeye: A 3D Environment for Comparative Genomics. Accepted Nov. 2003. See Montgomery S.B., Astakhova T., Bilenky M., Birney E., Fu T., Hassel M., Melsopp C., Rak M., Robertson A.G., Sleumer M., Siddiqui A.S., and Jones S.J.M., Sockeye: A 3D Environment for Comparative Genomics. Genome Research, 2004, 14[5]:956-962..

[22] Lewis CT, Sharpe AG, Lydiate DJ, Parkin IAP (2003) The Brassica / Arabidopsis Comparative Genome Browser: A novel approach to genome browsing. J. Plant Technology, 5(4):197-200.

[23] Peri S, Navarro JD, Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B, Gandhi TK, Chandrika KN, Deshpande N, Suresh S, Rashmi BP, Shanker K, Padma N, Niranjan V, Harsha HC, Talreja N, Vrushabendra BM, Ramya MA, Yatish AJ, Joy M, Shivashankar HN, Kavitha MP, Menezes M, Choudhury DR, Ghosh N, Saravana R, Chandran S, Mohan S, Jonnalagadda CK, Prasad CK, Kumar-Sinha C, Deshpande KS, Pandey A (2004) Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res. Jan 1;32 Database issue: D497-501.