NOVA
SCIENCE PUBLISHERS, INC.

WEBGBROWSE 2.1 - A WEB SERVER SUPPORTING MULTIPLE VERSIONS OF THE GENERIC GENOME BROWSER FOR CUSTOMIZABLE GENOME ANNOTATION DISPLAY

RAM PODICHETI
V. KASHI REVANNA
QUNFENG DONG

400 Oser Avenue, Suite 1600
Hauppauge, N. Y. 11788-3619
Phone (631) 231-7269
Fax (631) 231-8175
E-mail: main@novapublishers.com
www.novapublishers.com

ISBN: 978-1-61122-873-1 2011

COMPUTER SCIENCE, TECHNOLOGY AND APPLICATIONS

WEBGBROWSE 2.1 - A WEB SERVER SUPPORTING MULTIPLE VERSIONS OF THE GENERIC GENOME BROWSER FOR CUSTOMIZABLE GENOME ANNOTATION DISPLAY

COMPUTER SCIENCE, TECHNOLOGY AND APPLICATIONS

COMPUTER SCIENCE, TECHNOLOGY AND APPLICATIONS

WEBGBROWSE 2.1 - A WEB SERVER SUPPORTING MULTIPLE VERSIONS OF THE GENERIC GENOME BROWSER FOR CUSTOMIZABLE GENOME ANNOTATION DISPLAY

RAM PODICHETI
V. KASHI REVANNA
AND
QUNFENG DONG

Nova Science Publishers, Inc.
New York

For permission to use material from this book please contact us:
Telephone 631-231-7269; Fax 631-231-8175
Web Site: http://www.novapublishers.com

NOTICE TO THE READER

The Publisher has taken reasonable care in the preparation of this book, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained in this book. The Publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance upon, this material. Any parts of this book based on government reports are so indicated and copyright is claimed for those parts to the extent applicable to compilations of such works.

Independent verification should be sought for any data, advice or recommendations contained in this book. In addition, no responsibility is assumed by the publisher for any injury and/or damage to persons or property arising from any methods, products, instructions, ideas or otherwise contained in this publication.

This publication is designed to provide accurate and authoritative information with regard to the subject matter covered herein. It is sold with the clear understanding that the Publisher is not engaged in rendering legal or any other professional services. If legal or any other expert assistance is required, the services of a competent person should be sought. FROM A DECLARATION OF PARTICIPANTS JOINTLY ADOPTED BY A COMMITTEE OF THE AMERICAN BAR ASSOCIATION AND A COMMITTEE OF PUBLISHERS.

Additional color graphics may be available in the e-book version of this book.

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

ISBN: 978-1-61122-873-1

Published by Nova Science Publishers, Inc. † New York

Contents

PREFACE

Genome browsers are critical bioinformatics tools for biologists to visualize genome annotations and the other sequence features along a reference sequence. GBrowse is one of the most popular genome browsers used by the research community. However, its installation and configuration prove to be difficult for many biologists. We have developed a web server, WebGBrowse, which takes a user-supplied annotation file in GFF3 format, guides users through the configuration of the display of each genomic feature, and allows them to visualize the genome annotation information via the GBrowse software. This chapter describes an upgraded WebGBrowse server, WebGBrowse 2.0, which provides users with a choice to display their genome annotation with different versions of the GBrowse software. The modular design of WebGBrowse 2.0 allows easy integration of future GBrowse upgrades. We have also developed a web-based GFF3 template generator to facilitate the preparation of the required annotation file in the correct format. The entire WebGBrowse 2.0 package is portable and can be freely downloaded and installed locally.

Chapter 1

INTRODUCTION

Each sequenced genome must go through a series of bioinformatics computations to produce biologically meaningful annotation information. Genome annotation data typically include the coordinates of genes, regulatory elements, and repetitive regions with reference to the genome sequence. Genome browsers present an integrated graphical view of all these genomic annotations, which are organized as tracks of information layered along by the reference sequence. Users can navigate through different sections of the genomic sequence by zooming and keyword searching. For many scientists, "Seeing is believing." That is why genome browsers are among the most popular bioinformatics software that are actively used by biologists.

Currently, there exist two types of genome browsers: (i) standalone applications and (ii) web servers. Examples of standalone genome browser systems include GenoViz (http://genoviz.sourceforge.net/) and IGV (http://www.broad.mit.edu/igv). Besides having to install and maintain the software on the user's local computer, the biggest limitation of the standalone genome browsers is their weakness in data sharing. That is, the visualization through the standalone genome browsers is restricted to local users sitting in front of their computers where the software is installed. If different computers must be used (*e.g.*, users are away in the conference or there are collaborators located in different institutions), the software and data file must be copied to other computers for off-site use.

In contrast, a web-based genome browser easily allows distributed usages for analyzing the same data. The three most popular web-based genome browser systems used by biological research community are the Generic Genome Browser (GBrowse) (Stein *et al.*, 2002), Ensembl's genome browser (Stalker *et al.*, 2004), and the UCSC genome browser (Karolchik *et al.*, 2003). Other web-based genome browsers are also available, *e.g.*, xGDB (Schlueter *et al.*, 2006) and CoGe (Lyons *et al.*, 2008). A web-based genome browser easily allows distributed usages for analyzing the same data. Users do not need to install any software to use them. In principle, most web-based genome browsers can also be installed on the user's local computer. If properly configured, biologists can use such locally installed genome browsers as a web server for data sharing. However, these web-based genome browsers are sophisticated software packages whose installation and configuration instructions are usually intended for professional bioinformaticians. A typical biologist would find it overwhelming to install, configure, and maintain the software. For example, a GBrowse installation has to be preceded by a proper set up of Perl (http://www.perl.org), GD (http://search.cpan.org/dist/GD/), BioPerl (http://www.bioperl.org/), and other dependencies – a non-trivial challenge for biologists who are not computer savvy. In addition, biologists are not always equipped with the adequate computer resources (*e.g.*, the required UNIX-based operating system).

Therefore, in order to allow users to enjoy the functionalities of web-based genome browsers while avoiding the hassle of installation and configuration, we have developed a web server, WebGBrowse (Podicheti *et al.*, 2009), which allows biologists to upload their own genomic annotation data for display. Users can configure the display of each genomic feature visualize their data with an instance of GBrowse pre-installed on a web server. We chose the GBrowse system as the backend workhorse because GBrowse has a large user base due to its installation in many biological research community databases, *i.e.*, existing GBrowse users will find the same navigation and display style of GBrowse in the WebGBrowse server. Users do not need to install any software on their computers to use WebGBrowse, which can be easily accessed by using any standard web browsers (*e.g.*, Firefox, Internet Explorer, etc.).

Since our original WebGBrowse publication (Podicheti *et al.*, 2009), the GBrowse software has gone through significant upgrade from version 1.x to version 2.0. GBrowse 1.x has been widely used in the research community since 2002; GBrowse 2.0 was released in 2009, and represents a significant rewrite of the version 1.x with enhanced user experience by using Ajax programming technology (http:// en. wikipedia. org/ wiki/ Ajax_% 28programming%29) as well as more flexible configuration options for system administrators. Both of these major GBrowse versions will be available simultaneously because version 1.x likely will remain widely used by many major biological databases. In addition, different users may favor different versions of GBrowse due to a desire to continue using GBrowse 1.x for familiarity. In addition, new GBrowse releases may become available, further diversifying the GBrowser instances available for use. Therefore, it is important to provide different versions of GBrowse for users to easily choose and compare. In addition, some users will want to migrate from version 1.x to 2.0 with minimal effort. A naïve solution would be to provide separate servers hosting different GBrowse versions, which not only results in redundant work (*e.g.*, duplication of the web interface) but also causes significant inconvenience for users (*e.g.*, having to upload the same dataset to different servers). Instead, we have developed WebGBrowse 2.0, where multiple versions of GBrowse can be seamlessly integrated. Users only need to upload their dataset once and go through the same configuration process as in the original WebGBrowse sever, then choose to display their genome annotation with either the traditional GBrowse 1.x or GBrowse 2.0. In the following sections, we describe the detailed implementation and usage of WebGBrowse 2.0.

Chapter 2

MATERIALS AND METHODS

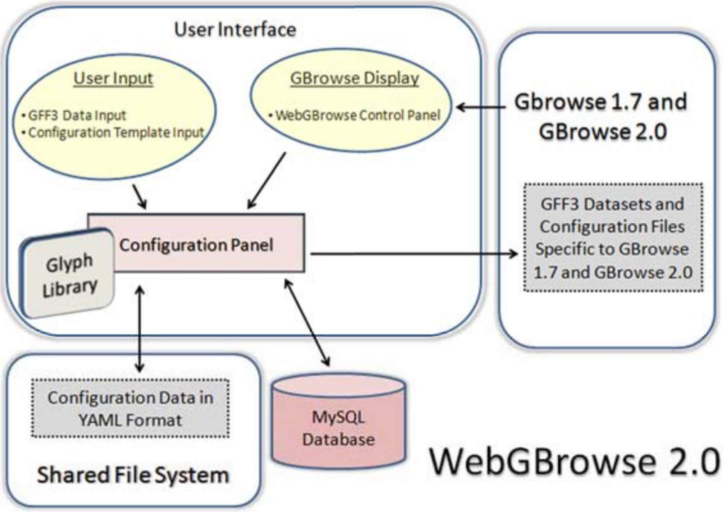
Figure 1 illustrates the architecture of WebGBrowse 2.0. Individual components are described below.

GBROWSE

The GBrowse software was downloaded from the GMOD web site (http://gmod.org/wiki/Gbrowse). Both version 1.70 (*i.e.*, the latest stable version of 1.x) and version 2.0 of GBrowse are currently installed at the WebGBrowse 2.0 server. To prevent any potential interference of different versions on the same server, they are installed on separate machines with shared physical disk for accessing common data files and the supporting database.

ANNOTATION DATA FILE

Users must supply their genome annotation file in Generic Feature Format version 3 (GFF3) format. GFF3 is a standard format for representing genomic features in a tab-delimited plain text file, which consists of 9 tab-delimited columns that specify each sequence feature (*e.g.*, gene, mRNA, CDS, UTR, or any other sequence characteristic that can be physically mapped to reference genome; specified in column 3) and their corresponding genomic coordinates (start and end positions in column 4 and column 5, respectively). The genomic DNA sequence can also be provided in GFF3 file. The complete format specification for GFF3 is available at http://www.sequenceontology.org/gff3.shtml. Through a provided web form, users can upload their GFF3 files to the WebGBrowser server. Once the GFF3 data file is uploaded, WebGBrowse automatically validates the data file and stores it as a file-based database for the pre-installed GBrowse system.


User Interface
User Input
• GFF3 Data Input
• Configuration Template Input
GBrowse Display
• WebGBrowse Control Panel
Gbrowse 1.7 and GBrowse 2.0
GFF3 Datasets and Configuration Files Specific to GBrowse 1.7 and GBrowse 2.0
Configuration Panel
Glyph Library
Configuration Data in YAML Format
Shared File System
MySQL Database
WebGBrowse 2.0

WebGBrowse interface allows users to input the GFF3 content along with an optional configuration template. A Configuration Panel follows the content validation which provides the interface for individual track configuration. The Configuration Panel is powered by a Glyph Library which provides a configurable set of parameters for each glyph. The configuration panel, as a result of its interaction with the user, outputs the configuration data which is saved into a network file system. This data is finally converted into individual configuration files specific to GBrowse 1.7 or GBrowse 2.0 depending on the user's requirements. The corresponding GBrowse libraries preinstalled on the WebGBrowse server render the GBrowse display. A WebGBrowse Control Panel allows the users to download the resultant configuration information and to navigate back to the configuration panel to continue with further changes to the configuration settings. A MySQL database helps maintain the history of datasets submitted to WebGBrowse associated with each specific email address.

Figure 1. Schematic overview of WebGBrowse 2.0 architecture.

GBROWSE CONFIGURATION FILE

The GBrowse software requires a configuration file for specifying the location of the GFF3 data file as well as other display settings including, for example, feature tracks, shape, font, and color. For every user-uploaded GFF3 data file, WebGBrowse creates a corresponding GBrowse configuration file. Specifically, WebGBrowse identifies a list of genomic features that can be displayed as “tracks” in GBrowse where a track is defined as horizontal display of instances of a particular feature type displayed beneath the reference genomic sequence. For example, predicted genes and proteins can be displayed as two separate tracks in genome browser. The feature list is derived based on the values from the feature type column (*i.e.*, the third column) and the feature source column (*i.e.*, the second column) of the GFF3 data file. Each feature (*e.g.*, gene or quantitative data such as tiling array hybridization intensity values) will be displayed as a different track in GBrowse. Each created configuration file is assigned a unique name based on a time stamp and session ID encrypted by MD5 hash (http://en.wikipedia.org/wiki/MD5). All of the configuration files are stored in the default GBrowse installation directory named *gbrowse.conf*. For GBrowse 2.0, a master configuration file is updated to point to each set of individual GBrowse configuration files.

GLYPH LIBRARY

In order to properly display each feature, the GBrowse system requires that the appearance of each feature (*e.g.*, color, height, shape, font; also see http://gmod.org/ wiki/index.php/CONFIGURE_HOWTO) must be a properly defined “glyph” in the configuration file where glyph refers to the shape of the feature, *e.g.*, using box to represent genes or exons of genes. For inexperienced users, it can be a tedious process to prepare the feature configuration. A key function of WebGBrowse is to provide a user-friendly editor in which a configuration panel lets users choose from more than 40 different glyph styles (*e.g.*, line, box, arrow; also see http://webgbrowse.cgb.indiana.edu/webgbrowse/glyphdoc.html) to properly display the features of interest. Users can also specify the color, font, label and many other characteristics specific to the glyph for the selected genomic feature. Each selected feature can be reviewed for further editing. The above configuration panel is driven by a glyph library that stores various configurable parameters specific to each of the Bio::Graphics::Glyph types in a YAML format file (http://www.yaml.org/), which is used by the HTML::FormEngine Perl library to generate the web forms users interact with.

MODIFICATION OF THE GBROWSE INTERFACE

WebGBrowse preserves any intrinsic functions provided by GBrowse such as the ability to download the sequence or annotation of a displayed genomic region as well as any custom annotation tracks. However, while GBrowse makes all the annotated genomes available in the “Data Source” pull-down menu for visualization, WebGBrowse is designed to limit access to an individual user’s data by disabling the pull-down menu to encapsulate different users. In addition, we have also added a WebGBrowse control panel in the GBrowse display that allows users to return to the track configuration panel if additional editing is desired. It also allows users to download the generated GBrowse configuration file.

A SUPPORTING DATABASE

A MySQL database was implemented to store users’ email addresses and the names of their configuration files. The database is used for associating users with their submitted data. Specifically, when users submit their GFF3 data files, they also have the option of providing their email addresses. If provided, a URL link to the GBrowse display page will be emailed to the user. The email also contains a link to access all the previously submitted datasets from the same user. Providing the email address also allows the user to perform the configuration process in multiple sessions. The progress can be saved at any stage by clicking a "Save Progress" button in the configuration panel and WebGBrowse will send a link to the email address provided. The configuration process can be resumed by clicking on that link at a later time.

GFF3 Template Generator

One of the most common feedbacks we have received from WebGBrowse users is a need for help in preparing the required GFF3 annotation file. Typically, biologists must deal with a large variety of output formats produced by many different bioinformatics programs. For example, there exist more than a dozen of *ab initio* gene prediction programs, each having its own output format. Although some format-conversion tools exist (*e.g.*, the BioPerl script *bp_genBank2gff3.pl* that can convert annotation file in GenBank format to the GFF3 format), specific parsers must be written for many of the existing bioinformatics programs to convert their outputs into GFF3 format. Therefore, for typical biologists who do not have programming skills, they need to seek help from bioinformaticians. Compared with setting up GBrowse, preparing GFF3 data file is a much smaller problem. The involved programming task is not difficult. For example, any bioinformatics students with basic programming skill can usually transform any raw data file into tab-delimited GFF3 files. But most bioinformaticians may be intimidated with GBrowse set-up and maintenance. However, most biologists and many bioinformaticians are not familiar with the GFF3 format specification. Based on the feedback of our users, we have developed a web-based GFF3 template generator, which guides the users step by step through simple web forms on producing a valid sample GFF3 file based on the subset of their genome annotations (*e.g.*, the length of the genome of their interest, the start and end position of some genomic features). The obtained sample GFF3 data file can be used as a template for the programmer to convert the entire dataset.

Chapter 3

NOTES

The following section briefly describes the usage of WebGBrowse 2.0 server. An illustrated tutorial was developed on the WebGBrowse 2.0 website at http://webgbrowse. cgb.indiana.edu/webgbrowse/tutorial.html. Additional description on the usage of the original WebGBrowse server is also available (Podicheti and Dong, 2010). WebGBrowse 2.0 shares many key user interfaces with the original WebGBrowse.

1. Use any web browser (*e.g.*, Firefox) to open the URL http://webgbrowse. cgb.indiana.edu to access the WebGBrowse 2.0 web server (Figure 2A). Click the button "Browse..." in the *GFF3 File* section and upload the genome annotation file. We suggest that users try the sample dataset first, which can be downloaded by clicking the link "[Sample GFF3 File]", which was modified from the GFF3 example provided in the GBrowse installation package. This dataset presents typical feature types that can be configured to illustrate the default generic display, protein-coding genes, quantitative data display, and so on. For large data files, users can choose to upload the compressed *.gz* or *.zip* formats, and WebGBrowse can uncompress such files automatically. Provide an email address in the text input field under "Email address". Although this step is optional, it will allow for configuration in multiple sessions, results to be sent via email, and documentation of previous submissions.

2. Click the button “Submit” to send the uploaded GFF3 data to WebGBrowse and open a “Configuration Panel”, where feature tracks may be added, edited or deleted from the GBrowse display. At the configuration panel, provide a short description for their dataset in the “Description” field. From the section labeled “Add New Track” *(*Figure 2B), select a feature from the “Feature” menu, which lists all the unique features derived from the dataset that can be configured into individual GBrowse tracks. For each selected feature, select its shape from the pull-down menu marked “Glyph”. A glyph library with a sample image and short description for each selected glyph will be displayed.
3. After clicking the button “Add Track”, a floating “Glyph Parameters Form”, where the parameters for the selected glyph are displayed, will appear (Figure 2C). The presented configurable parameter set is specific to the type of glyph chosen. Each parameter field has a brief description explaining the purpose of the parameter. If an email address was provided in step 1, a “Save Progress” button will appear at the top right corner of the configuration panel. Clicking the “Save Progress” feature will cause WebGBrowse to email a link to the address provided where configuration can be resumed at any time. Users can change any default parameter values (*e.g.*, the color of the displayed track). More parameters can be viewed by clicking the link “Advanced Section”. Once finished setting the parameter values, click the button “Save and Continue” to go back to the configuration panel and add all the desired tracks.
4. The configured tracks and their corresponding configuration settings will be listed under the section “Tracks Added” in the configuration panel (Figure 2D). To edit existing track configuration settings, select the track in the section “Tracks Added” and click the button “Edit Track”. Tracks can also be deleted by clicking the button “Delete Track”.
5. After adding and configuring all tracks, select the button “Display in GBrowse 1.70” or “Display in GBrowse 2.0” to visualize the features in via either version of the GBrowse software. A familiar display style of GBrowse will appear (Figure 2E, 2F). Novice GBrowse users can follow the GBrowse

tutorial available at OpenHelix (http://www.openhelix.com/gbrowse).

6. In addition to the conventional GBrowse display, there is a "WebGBrowse Control Panel" displayed at the top of the window. Clicking that button that allows further changes to the configuration file as well as a mechanism to download the generated configuration file. Click the button "Edit Configuration" in the WebGBrowse control panel to return to the configuration panel. To save the configuration to a file, click the button "Download Configuration" in the WebGBrowse control panel. To use the downloaded configuration file as a template while configuring another similar dataset, after performing step 1 with the new dataset, click the "Browse..." button in the "Configuration File to be used as a template" section of the "WebGBrowse Input Form" and upload the configuration file.
7. At the GFF template generator (Figure 3A), first provide a basic description of the reference genome (*i.e.*, name and length) then upload or paste a FASTA-format sequence into the form. Fill out the "feature" table (where each row corresponds to the columns of GFF3 format; Figure 3B) to view or download the generated GFF3 data file (Figure 3C).

Chapter 3

DISCUSSION

Typically, community databases set up web-based genome browsers for their user community. For example, users can browse annotations for fly genomes at FlyBase (Drysdale, 2008), *C. elegans* genomes at WormBase (Harris *et al.*, 2009), vertebrate genomes at Ensembl (Hubbard *et al.*, 2009), and plant genomes at PlantGDB (Dong *et al.*, 2005). In the past, this strategy had worked well for the research community because the numbers of available genome sequences was relatively small. In fact, it used to take a considerable amount of time to get a genome sequenced and the community databases usually partnered with each sequencing consortium to coordinate the display of the sequenced genomes.

Increasingly, however, such a display strategy in centralized community databases is not sufficient to satisfy the diverse needs of researchers. Specifically, with the rapid advent of DNA sequencing technologies, it has become easier for individual biology laboratories to engage in genome sequencing directly. Sequencing targets can be a particular species, strain, or regions of the genome for some individual in a population. This is especially true for member of the microbial research community because the sequencing microbial genomes has become nearly trivial. Due to limited bioinformatics resources, centralized community database will be unlikely to accommodate such diverse genomic data display needs in a timely fashion that meets researchers' demands.

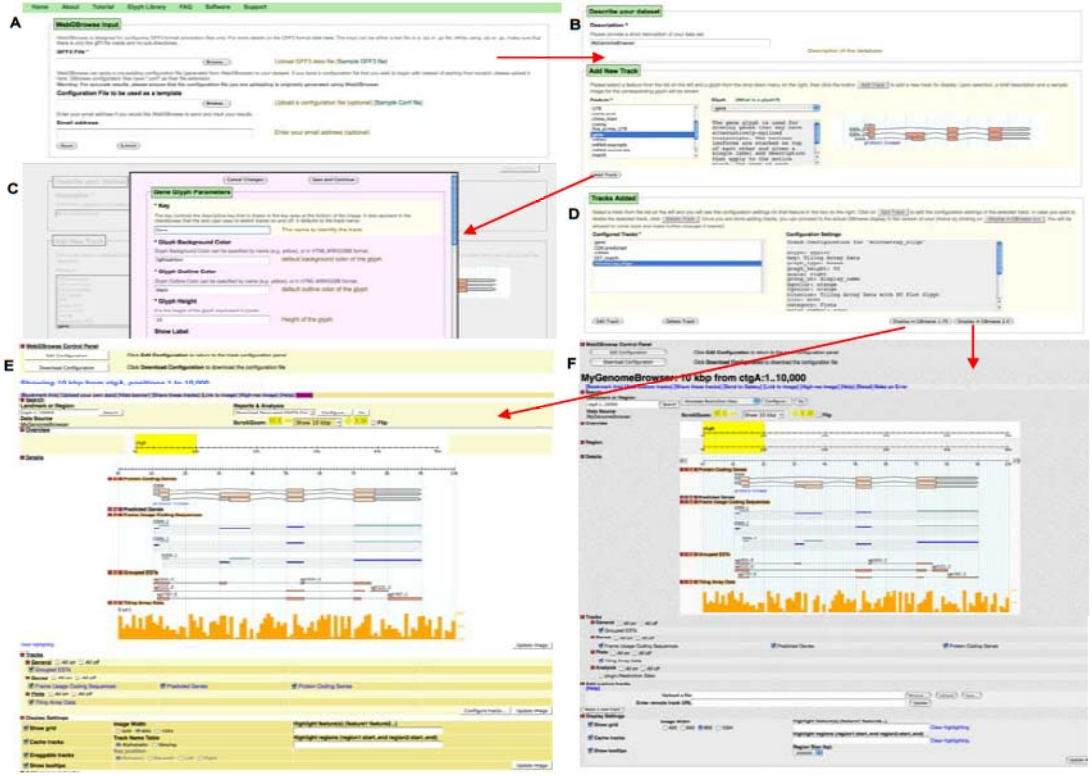
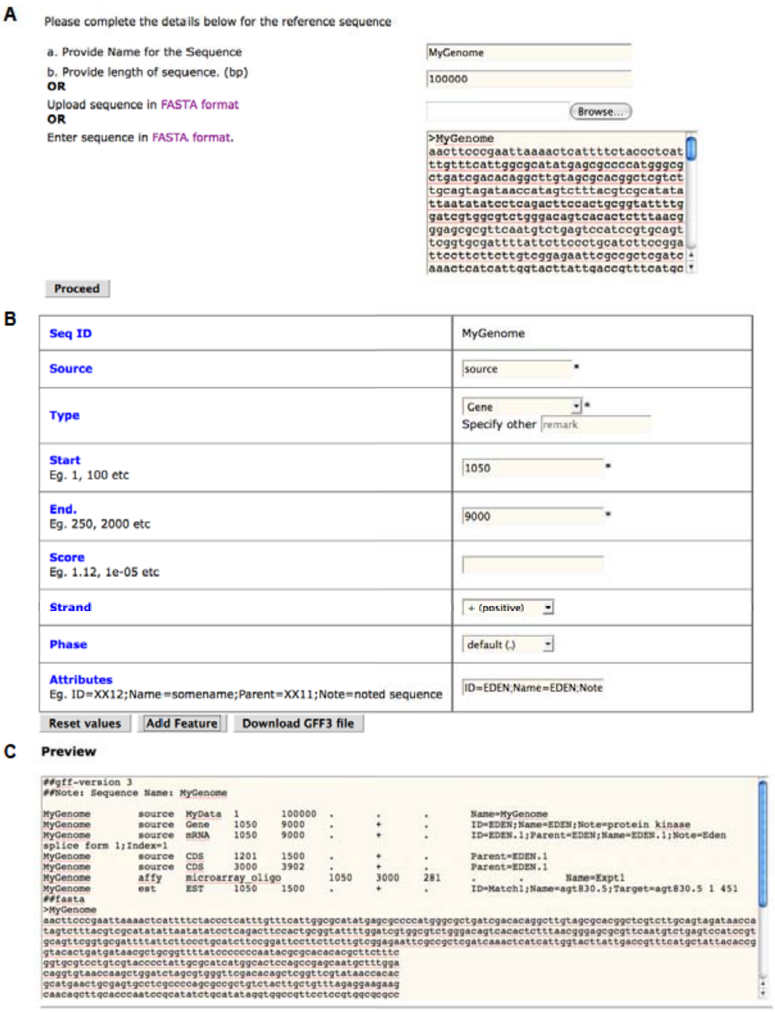


Figure 2. Screenshots of WebGBrowse. See text for details.



Please complete the details below for the reference sequence

a. Provide Name for the Sequence: MyGenome

b. Provide length of sequence. (bp): 100000

OR

Upload sequence in FASTA format

OR

Enter sequence in FASTA format.

```
>MyGenome
aacttcccgaattaaaactcattttctaccctcat
ttgtttcattggcgcatatgagcgccccatgggcg
ctgatcgacacaggcttgtagcgcacggctcgtct
tgcagtagataaccatagtctttacgtcgcatata
ttaatatatcctcagacttccactgcggtattttg
gatcgtggcgtctgggacagtcacactctttaacg
ggagcgcgttcaatgtctgagtccatccgtgcagt
tcggtgcgatttattcttccctgcatcttccgga
ttccttcttcttgtcggagaattcgccgctcgatc
aaactcatcattggtacttattgaccgtttcatgc
```

Seq ID	MyGenome
Source	source *
Type	Gene * Specify other remark
Start Eg. 1, 100 etc	1050 *
End. Eg. 250, 2000 etc	9000 *
Score Eg. 1.12, 1e-05 etc	
Strand	+ (positive)
Phase	default (.)
Attributes Eg. ID=XX12;Name=somename;Parent=XX11;Note=noted sequence	ID=EDEN;Name=EDEN;Note

Preview

```
##gff-version 3
##Note: Sequence Name: MyGenome

MyGenome        source  MyData  1       100000  .       .       .       Name=MyGenome
MyGenome        source  Gene    1050    9000    .       +       .       ID=EDEN;Name=EDEN;Note=protein kinase
MyGenome        source  mRNA    1050    9000    .       +       .       ID=EDEN.1;Parent=EDEN;Name=EDEN.1;Note=Eden
splice form 1;Index=1
MyGenome        source  CDS     1201    1500    .       +       .       Parent=EDEN.1
MyGenome        source  CDS     3000    3902    .       +       .       Parent=EDEN.1
MyGenome        affy    microarray_oligo        1050    3000    281     .       .       Name=Expt1
MyGenome        est     EST     1050    1500    .       +       .       ID=Match1;Name=agt830.5;Target=agt830.5 1 451
##fasta
>MyGenome
aacttcccgaattaaaactcattttctaccctcatttgtttcattggcgcatatgagcgccccatgggcgctgatcgacacaggcttgtagcgcacggctcgtcttgcagtagataacca
tagtctttacgtcgcatatattaatatatcctcagacttccactgcggtattttggatcgtggcgtctgggacagtcacactctttaacgggagcgcgttcaatgtctgagtccatccgt
gcagttcggtgcgatttattcttccctgcatcttccggattccttcttcttgtcggagaattcgccgctcgatcaaactcatcattggtacttattgaccgtttcatgctattacaccg
gtacactgatgataacgctgcggttttatccccccaatacgcgcacacacgcttctttc
ggtgcgtcctgtcgtaccctattgcgcatcatggcactccagccgagcaatgctttgga
caggtgtaaccaagctggatctagcgtgggttcgacacagctcggttcgtataaccacac
gcatgaactgcgagtgcctcgccccagcgccgctgtctacttgctgtttagaggaagaag
caacagcttgcacccaatccgcatatctgcatataggtggccgttcctccgtggcgcgcc
```

Figure 3. Screenshots of GFF3 template. See text for details.

Although GBrowse and the UCSC genome browser allow users to add their own tracks, such custom tracks can only be displayed along the reference genomes provided by the centralized community databases. In other words, users still cannot display their own whole genomes of interest. For example, FlyBase currently deploys genome browsers for 12 sequenced species. But if a researcher has just sequenced another

Drosophila species that is not included in FlyBase, he/she cannot add this newly sequenced genome to FlyBase. This is the biggest motivation for developing WebGBrowse: to enable researchers to easily display their own genome sequences and annotation.

Chapter 5

FUTURE WORK

The GFF3 data file can be stored in the GBrowse system either as a file-based or MySQL-based database. We have applied the file-based database mechanism for the simplicity of implementation. In the future, we plan to migrate to a MySQL-based database to provide faster performance for large datasets.

In addition, WebGBrowse is currently developed as a standard web server for manual interaction with biologists. We plan to extend WebGBrowse to utilize Web Services, thus integrating the WebGBrowser system into the emerging network of online bioinformatics resources that facilitate interoperable machine-to-machine interaction among biological databases and other bioinformatics resources. For example, the results of other online gene prediction and alignment tools could be directly sent to WebGBrowse for display.

Chapter 6

AVAILABILITY

The WebGBrowse web server is freely accessible, using a web browser at http://webgbrowse.cgb.indiana.edu. Although typical biologists will not try to install WebGBrowse (they will just use the web server that we have already provided instead), the software, written in Perl, is also available for local installation. WebGBrowse is released the Apache 2.0 open source license so that other developers can further improve the code and contribute to the project and some institutions could limit the need to access to outside web servers.

Chapter 7

CONCLUSION

Instead of relying on centralized community databases to set up user-specific genome browser, biologists need a visualization tool that can be easily used for displaying their own genomes of interest. We have developed WebGBrowse for biologists to simply upload their genome data and display the annotations on integrated GBrowse software at our web server. This allows biologists to enjoy the functionalities of web-based genome browsers while avoiding the installation hassle while also retaining the freedom to configure the display of each genomic feature. Because multiple version of GBrowse system have become available and each has a unique user interface, we have extended the original WebGBrowse to allow researchers to select their favorite version for visualization.

ACKNOWLEDGEMENTS

RP designed and implemented WebGBrowse 2.0. KR implemented the GFF template generator. QD conceived the project and guided the development process, and drafted this chapter. We thank Dr. Carolyn Lawrence for her critical reading of the manuscript. Future correspondences should be sent to Qunfeng.Dong@unt.edu. This work was supported in part by the Indiana METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc.

REFERENCES

Dong, Q., Lawrence, C. J., Schlueter, S. D., Wilkerson, M. D., Kurtz, S., Lushbough, C. & Brendel, V. (2005). Comparative plant genomics resources at PlantGDB, *Plant Physiol*, *139*, 610-618.

Drysdale, R. (2008). FlyBase : a database for the Drosophila research community, *Methods Mol Biol*, *420*, 45-59.

Harris, T. W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W. J., De La Cruz, N., Davis, P., Duesbury, M., Fang, R., Fernandes, J., Han, M., Kishore, R., Lee, R., Muller, H. M., Nakamura, C., Ozersky, P., Petcherski, A., Rangarajan, A., Rogers, A., Schindelman, G., Schwarz, E. M., Tuli, M. A., Van Auken, K., Wang, D., Wang, X., Williams, G., Yook, K., Durbin, R., Stein, L. D., Spieth, J. & Sternberg, P. W. (2009). WormBase: a comprehensive resource for nematode research, *Nucleic Acids Res*.

Hubbard, T. J., Aken, B. L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L., Coates, G., Fairley, S., Fitzgerald, S., Fernandez-Banet, J., Gordon, L., Graf, S., Haider, S., Hammond, M., Holland, R., Howe, K., Jenkinson, A., Johnson, N., Kahari, A., Keefe, D., Keenan, S., Kinsella, R., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Megy, K., Meidl, P., Overduin, B., Parker, A., Pritchard, B., Rios, D., Schuster, M., Slater, G., Smedley, D., Spooner, W., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wilder, S., Zadissa, A., Birney, E., Cunningham, F., Curwen, V., Durbin, R., Fernandez-Suarez, X.M., Herrero, J., Kasprzyk, A., Proctor, G., Smith, J., Searle, S. & Flicek, P. (2009). Ensembl 2009, *Nucleic Acids Res*, *37*, D690-697.

Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., Wang, X., Bowers, J., Paterson, A., Lisch, D. & Freeling, M. (2008). Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids, *Plant Physiol*, *148*, 1772-1781.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Lu, Y. T., Roskin, K. M., Schwartz, M., Sugnet, C. W., Thomas, D. J., Weber, R. J., Haussler, D. & Kent, W. J. (2003). The UCSC Genome Browser Database, *Nucleic Acids Res*, *31*, 51-54.

Podicheti, R. & Dong, Q. (2010). Using WebGBrowse to Visualize Genome Annotation on GBrowse, *Cold Spring Harbor Protoclos*, in press.

Podicheti, R., Gollapudi, R. & Dong, Q. (2009). WebGBrowse--a web server for GBrowse, *Bioinformatics*, *25*, 1550-1551.

Schlueter, S. D., Wilkerson, M. D., Dong, Q. & Brendel, V. (2006). xGDB: open-source computational infrastructure for the integrated evaluation and analysis of genome features, *Genome Biol*, *7*, R111.

Stalker, J., Gibbins, B., Meidl, P., Smith, J., Spooner, W., Hotz, H. R. & Cox, A. V. (2004). The Ensembl Web site: mechanics of a genome browser, *Genome Res*, *14*, 951-955.

Stein, L. D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J. E., Harris, T. W., Arva, A. & Lewis, S. (2002). The generic genome browser: a building block for a model organism system database, *Genome Res*, *12*, 1599-1610.

INDEX