

# Using WebGBrowse to Visualize Genome Annotation on GBrowse

Ram Podicheti and Qunfeng Dong<sup>1</sup>

Center for Genomics and Bioinformatics, Indiana University, Bloomington, IN 47405, USA

## INTRODUCTION

Although GBrowse is popular for visualizing genomic features along a reference sequence, its installation and configuration are difficult for many biologists. WebGBrowse is a web server that takes a user-supplied annotation file, guides users to configure the display of each genomic feature, and allows users to visualize the genome annotation with integrated GBrowse software. This protocol guides the user through each step of using WebGBrowse.

## RELATED INFORMATION

An exhaustive and illustrated tutorial is available on the WebGBrowse website at <http://webgbrowse.cgb.indiana.edu/tutorial.html>. A GBrowse user introductory tutorial is available on OpenHelix at <http://www.openhelix.com/gbrowse>. The format specification for Generic Feature Format version 3 (GFF3) is available at <http://www.sequenceontology.org/gff3.shtml>.

## MATERIALS

**CAUTIONS AND RECIPES:** Please see Appendices for appropriate handling of materials marked with <I>, and recipes for reagents marked with <R>.

### Equipment

Genome annotation file in GFF3 format  
Internet-connected computer with web browser

## METHOD

*The procedure below describes how to display a typical genome annotation on GBrowse (Stein et al. 2002) at the WebGBrowse (Podicheti et al. 2009) server. Some nomenclature is defined in Table 1.*

### Uploading the Genome Annotation Data Set

1. Use any modern web browser (e.g., Firefox) to open the URL <http://webgbrowse.cgb.indiana.edu>. You are now at the WebGBrowse Input Form (Fig. 1).
2. Click the link “[Sample GFF3 File]” to download the sample data set, *volvox.gff3* (Fig. 2).  
*This data set presents typical feature types that can be configured to illustrate the default generic display, protein-coding genes, quantitative data display, and so on. We suggest trying the sample data set first before uploading your own GFF3 data set.*

<sup>1</sup>Corresponding author (qunfengd@gmail.com).

**Table 1. Genome browser nomenclature**

Term	Definition
Feature	Sequence characteristic (e.g., gene) that can be physically mapped to reference genome.
Track	Horizontal display of instances of a particular feature type. Tracks are located under the reference genomic sequence. For example, two tracks are used to display gene and mRNA features in Figure 6.
Glyph	Shape of the feature diagram. A WebGBrowse-supported glyph library is available at <a href="http://webgbrowse.cgb.indiana.edu/glyphdoc.html">http://webgbrowse.cgb.indiana.edu/glyphdoc.html</a> .
Configuration	GBrowse display settings (e.g., feature tracks, glyph, font, and color).

3. Click the button “Browse...” in the “GFF3 File” section and upload the sample file (volvox.gff3).
4. (Optional) In the text input field under “Email address,” enter your e-mail address.  
*This allows you to perform the configuration in multiple sessions, have the configuration results sent to your e-mail address, and keep track of all your previous submissions.*
5. Click the button “Submit” to send your data to WebGBrowse. WebGBrowse will validate your input and redirect you to the Configuration Panel.  
*For this example, leave the input field blank for “Configuration File to be used as a template.” Use of this feature is described in Step 18.*

### Working in the Configuration Panel

*The Configuration Panel allows you to add, edit, or delete feature tracks for the GBrowse display of your data set.*

6. Provide a short description for your data set in the “Description” field.
7. From the section “Add New Track” (Fig. 3), select a feature from the Feature menu.  
*The Feature menu lists all the unique features derived from the data set that can be configured into individual GBrowse tracks.*
8. For each selected feature, choose its shape by selecting from the pull-down Glyph menu.  
*A glyph library with a sample image and short description for each selected glyph is also displayed.*

**WebGBrowse Input**

WebGBrowse is designed for configuring GFF3 format annotation files only. For more details on the GFF3 format click [here](#). The input can be either a text file or a .zip or .gz file. While using .zip or .gz, make sure that there is only the gff3 file inside and no sub-directories.

**GFF3 File \***

[Upload GFF3 data file \[Sample GFF3 file\]](#)

WebGBrowse can apply a pre-existing configuration file (generated from WebGBrowse) to your dataset. If you have a configuration file that you wish to begin with instead of starting from scratch, please upload it here. GBrowse configuration files have ".conf" as their file extension.  
Warning: For accurate results, please ensure that the configuration file you are uploading is originally generated using WebGBrowse.

**Configuration File to be used as a template**

[Upload a configuration file \(optional\) \[Sample Conf file\]](#)

Enter your email address if you would like WebGBrowse to send and track your results.

**Email address**

[Enter your email address \(optional\)](#)

**FIGURE 1.** The WebGBrowse Input Form allows users to provide the genome annotation data sets (GFF3 format), configuration templates, and their e-mail address.

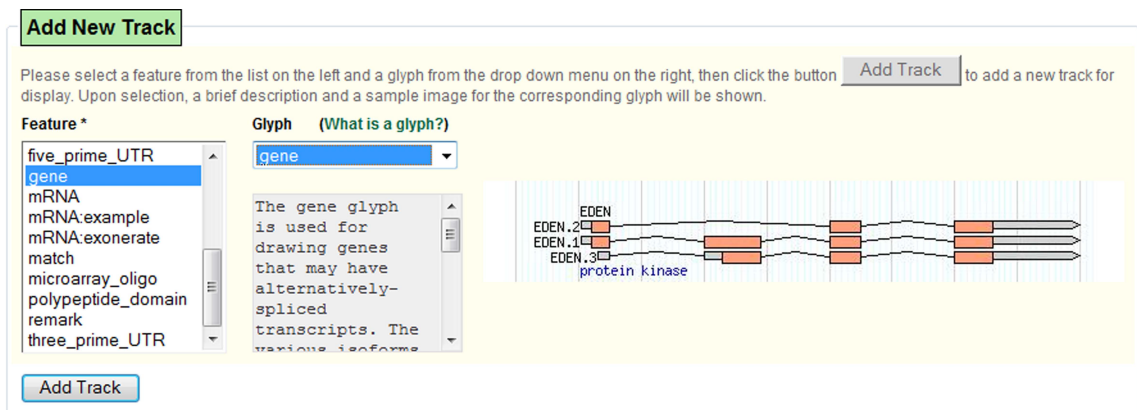
```

##gff-version 3
# file: volvox.gff3 derived from GBrowse Administration Tutorial by Lincoln Stein, 2008

ctgA example contig 1 50000 . . . Name=ctgA
ctgA example remark 1659 1984 . + . Name=f07;Note=This is an example
ctgA example remark 3014 6130 . + . Name=f06;Note=This is another example
ctgA example polypeptide_domain 11911 15561 . + . Name=m11;Note=kinase
ctgA example polypeptide_domain 13801 14007 . - . Name=m05;Note=helix loop helix
ctgA example match 32329 32359 . + . ID=match-seg01;Name=seg01;Note=This is a segment
ctgA example match 26122 26126 . + . ID=match-seg02;Name=seg02
ctgA example match 26497 26869 . + . ID=match-seg02;Name=seg02
ctgA example match 27201 27325 . + . ID=match-seg02;Name=seg02
ctgA example gene 1050 9000 . + . ID=EDEN;Name=EDEN;Note=protein kinase
ctgA example mRNA 1050 9000 . + . ID=EDEN.1;Parent=EDEN;Name=EDEN.1;Note=Eden splice form 1;Index=1
ctgA example five_prime_UTR 1050 1200 . + . Parent=EDEN.1
ctgA example CDS 1201 1500 . + 0 Parent=EDEN.1
ctgA example CDS 3000 3902 . + 0 Parent=EDEN.1
ctgA example three_prime_UTR 7609 9000 . + . Parent=EDEN.1
ctgA example mRNA 1050 9000 . + . ID=EDEN.2;Parent=EDEN;Name=EDEN.2;Note=Eden splice form 2;Index=1
ctgA example five_prime_UTR 1050 1200 . + . Parent=EDEN.2
ctgA example CDS 1201 1500 . + 0 Parent=EDEN.2
ctgA example CDS 7000 7608 . + 0 Parent=EDEN.2
ctgA example three_prime_UTR 7609 9000 . + . Parent=EDEN.2
ctgA exonerate mRNA 17400 23000 . + . ID=rna-Apple3;Name=Apple3;Note=Predicted
ctgA exonerate UTR 17400 17999 . + . Parent=rna-Apple3
ctgA exonerate UTR 21201 23000 . + . Parent=rna-Apple3
ctgA example BAC 1000 20000 . . . ID=b101.2;Name=b101.2;Note=Fingerprinted BAC with end reads
ctgA example clone_start 1000 1500 . + . Parent=b101.2
ctgA example clone_end 19500 20000 . - . Parent=b101.2
ctgA affy microarray_oligo 1 100 281 . . Name=Expt1
ctgA affy microarray_oligo 101 200 183 . . Name=Expt1
ctgA est EST_match 1050 1500 . + . ID=Match1;Name=agt830.5;Target=agt830.5 1 451
ctgA est EST_match 3000 3202 . + . ID=Match1;Name=agt830.5;Target=agt830.5 452 654
##fasta
>ctgA
cattggtcgggagttgaacaacggcattaggaacacttccgctctctcacttttatacgat
tatgattggttctcttagccttgggttagattggtagtagtagcggcgctaattgctaactg
aattgagaactcgcagcgggggctaggcaaatctgatccagcctgactctctctggaacc
ctgccataaatcaaaagggttagtgcggcacaacgttggaacaacggattagaagacca
acctgaccacaaacgcctaattaaccgggtatctcttcggaaacggcggttctctctctag
atagcgatctgtggtctcaccatgcaatttaaacaggtgagtaagattgctacaatac
gagactagctgtcaccagatgctgttcatctgttggtccttggctcctcgctccggtgatacc

```

**FIGURE 2.** The sample GFF3 file was modified from the GFF3 example provided in the GBrowse installation package. This figure displays partial contents of the sample GFF3 file due to space limitations. GFF3 is a standard format for representing genomic annotation, which consists of nine tab-delimited columns that specify each sequence feature (e.g., gene, mRNA, coding sequence [CDS], untranslated region [UTR], etc., in column 3) and its corresponding genomic coordinates (start and end positions in column 4 and column 5, respectively). The actual DNA sequence can also be provided in a GFF3 file (e.g., the sequence for the genomic contig ctgA). For complete GFF3 specifications, see <http://www.sequenceontology.org/gff3.shtml>.



**FIGURE 3.** Genomic features extracted from the GFF3 file are displayed in the Features box of the Configuration Panel. For each feature, users can select a display style from the Glyph drop-down menu. Each glyph is illustrated with a brief description and an example image. Each Feature-Glyph pair defines a track in the GBrowse display.

9. Click the button "Add Track." This will open the floating Glyph Parameters Form (Fig. 4), where the parameters for the selected glyph are displayed.

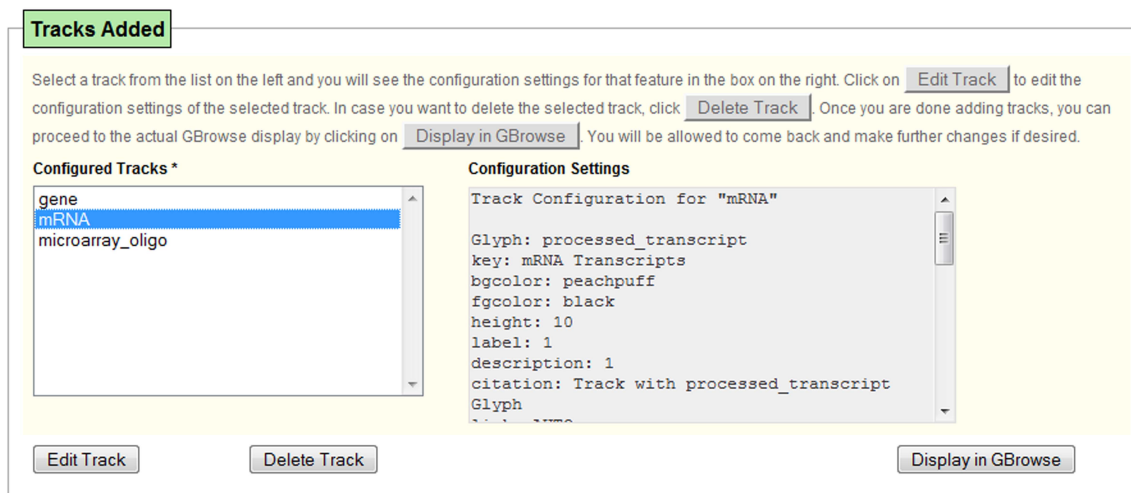
*The presented configurable parameter set is specific to the type of glyph chosen. Each parameter field has a brief description explaining the purpose of the parameter. If you have provided your e-mail address in Step 4, you will see a "Save Progress" button at the top right corner of the configuration panel that lets you save your progress and complete the configuration in a later session. You can save your progress at any stage by clicking on the "Save Progress" button and WebGBrowse will e-mail you a link, which can be later followed to resume your work.*

10. Use the Glyph Parameters Form to change any of the default parameter values (e.g., the color of the displayed track), if desired. More parameters can be viewed by clicking the link "Advanced Section."
11. Once you are done setting the parameter values, click the button "Save and Continue" to go back to the Configuration Panel.  
*The configured tracks and their corresponding configuration settings will be listed under the section "Tracks Added" in the Configuration Panel (Fig. 5).*
12. Repeat Steps 7-11 to add all the desired tracks.
13. To edit any existing track configuration settings, select the track in the section "Tracks Added" and click the button "Edit Track."
14. To delete a track, click the button "Delete Track."
15. After adding and configuring all tracks, click the button "Display in GBrowse" to visualize the features in GBrowse.

*Users will find the same navigation and display style of GBrowse. Novice GBrowse users can refer to the GBrowse tutorial available at OpenHelix (<http://www.openhelix.com/gbrowse>). In addition to the normal GBrowse*

The screenshot shows a web-based configuration form titled "CDS Glyph Parameters". At the top, there are two buttons: "Cancel Changes" and "Save and Continue". The form is organized into several sections, each with a bolded title and a descriptive paragraph. The sections are: 1. "\* Key": A text input field contains "Frame Usage Coding Sequences". Below it is a "Show/Hide Label" link. 2. "\* Glyph Height": A text input field contains "30". Below it is a "Show/Hide Label" link. 3. "Show Label": A checkbox is checked. Below it is a "Show/Hide Label" link. 4. "Show Description": A checkbox is unchecked. Below it is a "Show/Hide Description" link. 5. "Track Citation": A text input field contains "Track with CDS Glyph". Below it is a "Show/Hide Label" link. 6. "Link": A text input field contains "AUTO". Below it is a "Show/Hide Label" link. At the bottom, the text "Track Category" is partially visible. The form has a light pink background and a vertical scrollbar on the right side.

FIGURE 4. The glyph-specific configuration can be edited in the Glyph Parameters Form.



**FIGURE 5.** All the tracks can be reviewed for further editing or deletion before proceeding to the final GBrowse display.

display, there is a WebGBrowse Control Panel (Fig. 6) displayed at the top of the GBrowse that allows you to return to the Configuration Panel to make further changes or to download the configuration file.

16. To further edit the tracks, click the button “Edit Configuration” in the WebGBrowse Control Panel. This will return you to the Configuration Panel.
17. To save the configuration to a file, click the button “Download Configuration” in the WebGBrowse Control Panel.  
*This is useful for reusing the same configuration setting as a template for a similar data set, as explained in Step 18.*

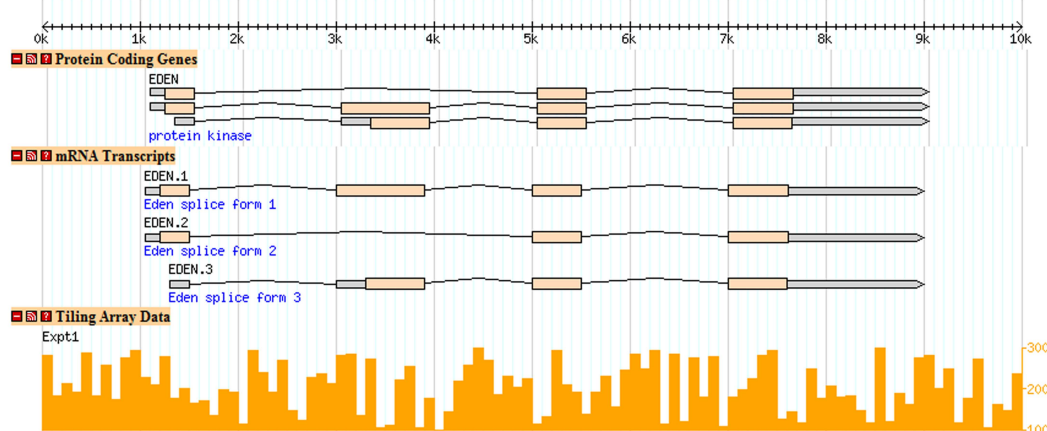
#### WebGBrowse Control Panel

Edit Configuration	Click <b>Edit Configuration</b> to return to the track configuration panel
Download Configuration	Click <b>Download Configuration</b> to download the configuration file

Showing 10 kbp from ctgA, positions 1 to 10,000

[Bookmark this](#) [\[Upload your own data\]](#) [\[Hide banner\]](#) [\[Share these tracks\]](#) [\[Link to Image\]](#) [\[Help\]](#) [\[Reset\]](#)

- [Search](#)
- [Overview](#)
- [Details](#)



**FIGURE 6.** Three tracks (i.e., Gene, mRNA, and tiling array hybridization intensity signals) are displayed in this example. Users can use the embedded WebGBrowse Control Panel to edit or download the GBrowse configuration file.

## Using a Configuration Template

18. To use a pre-existing configuration file as a template to configure another similar data set:
  - i. After performing Step 3 with the new data set, click the “Browse...” button in the “Configuration File to be used as a template” section of the WebGBrowse Input Form (Fig. 1).
  - ii. Select a previously downloaded configuration file (see Step 17).
  - iii. Continue at Step 4.

## DISCUSSION

GBrowse is a sophisticated software package that is typically installed and configured by professional bioinformaticians. WebGBrowse enables biologists to simply upload their own genomic data for visualization within a preinstalled GBrowse system. Through embedded web forms, users can easily configure the display of each genomic feature. For institutions that are interested in installing WebGBrowse locally, instructions are available at <http://webgbrowse.cgb.indiana.edu/software.html>. WebGBrowse supports the latest GBrowse production version 1.69. Currently, a completely new version of GBrowse, 2.0, is being actively developed for improved user experience. Once GBrowse 2.0 is stable for production use, we will also incorporate it into WebGBrowse. Additionally, to assist biologists who are not comfortable with the GFF3 format specifications, we are developing a web-based GFF3 template generator to facilitate the preparation of the required GFF3 input file.

## REFERENCES

- Podicheti R, Gollapudi R, Dong Q. 2009. WebGBrowse—a Web server for GBrowse. *Bioinformatics* **25**: 1550–1551.
- Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, et al. 2002. The generic genome browser: A building block for a model organism system database. *Genome Res* **12**: 1599–1610.