Genome Analysis

EPIC-CoGe: Managing and Analyzing Genomic Data

Andrew D. L. Nelson¹, Asher K. Haug-Baltzell¹, Sean Davey¹, Brian D. Gregory², and Eric Lyons¹,*

¹BIO5 Institute, School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA
²Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

*To whom correspondence should be addressed.

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract
Summary: The EPIC-CoGe browser is a web-based genome visualization utility that integrates the GMOD JBrowse genome browser with the extensive CoGe genome database (currently containing over 30,000 genomes). In addition, the EPIC-CoGe browser boasts many additional features over basic JBrowse, including enhanced search capability and on-the-fly analyses for comparisons and analyses between all types of functional and diversity genomics data. There is no installation required and data (genome, annotation, functional genomic, and diversity data) can be loaded by following a simple point and click wizard, or using a REST API, making the browser widely accessible and easy to use by researchers of all computational skill levels. In addition, EPIC-CoGe and data tracks are easily embedded in other websites and JBrowse instances.

Availability and Implementation: EPIC-CoGe Browser is freely available for use online through CoGe (https://genomevolution.org). Source code (MIT open source) is available: https://github.com/LyonsLab/coge.

Contact: ericlyons@email.arizona.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Genome visualization is a useful means of inspecting content within the framework of a genome such as gene models, transcriptional information, or SNP data. JBrowse is a powerful browser-based genome viewer that handles large datasets with minimal resource requirements. It supports a wide variety of input files, allowing for fast and seamless data integration. The comparative genomics platform CoGe enables the rapid association of private and public experimental datasets with its large database of genomes (~33,000) to facilitate downstream analyses (Lyons and Freeling 2008). Recognizing the need for an enhanced genome visualization experience when dealing with large and complex datasets, the CoGe Team developed an updated implementation of JBrowse (Buels et al. 2016), which we call the EPIC-CoGe browser. Here we describe the new features associated with the EPIC-CoGe browser that will increase researcher’s ability to visualize genomic information and analyze data associated with those genomes.

2 Features

While genome browsers are incredibly helpful for the dynamic viewing of a genome and its associated annotation data (e.g., genes), most web-based genome browsers center on one or a handful of genomes, and restrict the ability to import data and perform analyses within the genome browser itself. In contrast, EPIC-CoGe allows users to take full advantage of the genomic resources and analysis tools available within the CoGe platform, including adding new data and annotation tracks, and performing comparative analyses across those tracks (Supplementary Fig. S1A, SIB). As described below, EPIC-CoGe includes several enhancements that allow the user to quickly integrate, visualize, search, compare, analyze, and export/share their genomic data.

2.0 Integration with CoGe’s genome management systems

By integrating JBrowse within CoGe, we offer users the ability to rapidly view genomes and genome associated data for all species in the CoGe database. Users can load new genomes and perform genomic/transcriptomic analyses within CoGe using the LoadGenome and LoadExp+ suite of workflows (Grover et al. 2017), and then easily visualize those data within EPIC-CoGe. Comparisons and analyses of data tracks (e.g., methylation overlapping SNPs) can be performed on a mix-
2.1 Search features

EPIC-CoGe offers a number of features for searching through datasets to quickly identify regions of interest. 

**Searching within the track selector:** Tracks can be searched for by name or by type (i.e., show BAM alignment). On the right side of the display, the user can select different coordinates and filters for the track. For instance, the user can search through their experiments for BAM files, or search for a specific BAM file based on its name. Only tracks associated with the same genomic location are found in the track selector, minimizing confusion. (i.e., Homo sapiens RNA-seq data is only associated with the Homo sapiens genome).

**Searching within the main display:** Users can search for features by name using the “Find Features” button from the navigation bar at the top of the track viewer. Depending on the track types present, users can also search tracks within the viewer. For example, SNP tracks can be searched by type or by SNPs that overlap features (Supplemental Fig. S2B).

**Searching a range of values:** Users can quickly identify regions within their experimental track that correspond to a particular value, or range of values. For instance, when examining a track containing expression data, it is simple to identify the transcript with the most read depth by selecting “Search Experiment Data” from the drop-down menu next to the track label (Supplemental Fig. S2C). A pop-up window asks the user if they want to search by maximum value, minimum value, or transcripts that fall within a range of user-selected values. Values can also be selected on a histogram of all values. The results of this search are displayed in a new track.

**Searching data and feature tracks for overlap:** The user can also identify data within a track that overlaps a particular feature based on feature type or feature name. For instance, data within a track that overlaps with a gene feature such as “miRNA” can be quickly identified by clicking on the drop-down arrow next to the track label and selecting “Find Data that Overlaps Features” (Supplemental Fig. S2D). A new track will be generated displaying the search results.

2.2 On-the-fly analyses

The EPIC-CoGe browser allows for multiple dynamic, on-the-fly comparison of datasets.

**Comparisons:** Users can analyze data and feature tracks for either intersection or complement of values with any other track. Analyses are conducted by simply dragging one experimental track on top of another while holding the command/ctrl key. A dialog window then prompts the user about which type of data to identify – unique (non-overlapping) or common (overlapping). A new track displays the results (Supplemental Figs. S3A and S3B).

**Merging tracks:** Multiple tracks can be merged into a new, single track that contains all information from both. The ability to merge tracks can help alleviate “vertical bloat” often encountered with genome browsers, where numerous tracks can become unwieldy. Merged track results are given a name that describes how they were generated (Supplemental Fig. S3C).

**Merging markers:** Datasets can be simplified by merging close markers into a single contiguous marker. By clicking on the drop-down arrow next to the track label, an option to merge markers is given. The user can specify how close two markers should be in order to be merged (i.e., 100 bp is the default). If the user is dissatisfied with the results of a merge, the data can easily be reversed using the revert command within the same window. Simplification can help clarify results from previous searches, especially when viewing a larger genome segment (Supplemental Fig. S3D).

**Links to CoGe’s tools:** When a feature such as a gene is clicked on in EPIC-CoGe, a popup will display information about that feature. This includes metadata about the feature such as its name, annotations, location, length, and GC content. In addition, the popup contains links to other tools in CoGe for additional analyses such as CoGeBlast, SynFind, and FeatView (Supplemental Fig. S4).

2.3 Data export

To facilitate data generation and analysis, the EPIC-CoGe browser includes enhanced save and export data features. Tracks, including search results and results from in-browser analyses can always be saved within CoGe, where they can be stored in notebooks and shared directly with collaborators, all while retaining metadata and provenance. Moreover, because users frequently use their results in downstream analyses, all data tracks can be exported to either the user’s local computer or CyVerse Dataport.

2.4 Portability

One line of HTML in an <IFRAME> allows anyone to embed EPIC-CoGe in another website (Supplemental Fig. S5), or use CoGe’s API to export data tracks to other JBrowse instances. Some advanced analytical features of EPIC-CoGe are available for other JBrowse instances through JBrowse plugins where direct access to the CoGe database are not required; code for those latter features are available through CoGe’s GitHub repository. Together, these features allow researchers to create a genome browser for their genome in minutes, and manage the branding of their genome through their own website. (Supplemental Fig. S6).

3 Availability and requirements

The EPIC-CoGe browser is available through the widely used CoGe web platform (http://www.genomevolution.org; Supplemental Fig. S7). Use requires a modern HTML5 capable browser (ideally Chrome or Firefox). Users can browse any of CoGe’s publically available genomes and experiments without creating a user account, but must register to upload their own data or access restricted data. More detailed use instructions and additional help are available at https://genomevolution.org/zxer.

Acknowledgements

We would like to thank CyVerse (DBI – 1265383) for technical assistance for API development.

Funding

This work has been supported by the National Science Foundation grants IOS – 1339156 and IOS – 1444490.

Conflict of interest: none declared.

References