Exploration of Metagenome Assemblies with an Interactive Visualization Tool

Michael Cantor¹, Henrik Nordberg¹, Tatyana Smirnova¹, Evan Andersen¹, Susannah Tringe¹, Matthias Hess², Inna Dubchak ¹,³

¹DOE Joint Genome Institute, Walnut Creek, CA
²Washington State University, Richland, WA
³Lawrence Berkeley National Laboratory, Berkeley, CA

July 2014

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

LBNL-178516

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.
Exploration of metagenome assemblies with an interactive visualization tool

Metagenomics, one of the fastest growing areas of modern genomic science, is the genetic profiling of the entire community of microbial organisms present in an environmental sample.

Elviz is a web-based tool for the interactive exploration of metagenome assemblies. Elviz can be used with publicly available data sets from the Joint Genome Institute or with custom user-loaded assemblies.

Elviz is available at genome.jgi.doe.gov/viz

ABSTRACT

Metagenomics projects, one of the fastest growing areas of modern genomic science, is the genetic profiling of the entire community of microbial organisms present in an environmental sample.

Elviz is a web-based tool for the interactive exploration of metagenome assemblies. Elviz can be used with publicly available data sets from the Joint Genome Institute or with custom user-loaded assemblies.

Elviz is available at genome.jgi.doe.gov/viz

EXPLORING METAGENOME ASSEMBLIES

The Plot
The plot can be zoomed and panned by dragging and scrolling with the mouse. Hovering over each contig will display a small detailed panel. Clicking on a contig brings up an annotation browser (see below).

Interactive Legend
Each contig group can be assigned a color: hidden, shown, or high-lighted.

Search
Contigs can be searched for genes of interest, providing a powerful visual indication of the distribution of function across category in the sample.

Plot Configuration
Each variable in the data can be selected as a plot axis. Range sliders for each variable operate in one of two modes. For axes variables they operate to zoom in/out. For all other variables they act as filters.

RESOLVE
Dense, overlapping plots are typical for metagenomic studies. Elviz offers multiple strategies for visually discriminating features of interest in the data.

Clusters of Flavobacteria are resolved by:
1. Highlighting this group in the Legend.
2. Zooming and panning in the plot
3. Using the Length filter to hide smaller contigs

SEARCH
With Search, the user can identify contigs containing specific genes or functional annotations. The ability to visually search results on the plot immediately illuminates the distribution of function over the different taxa in the sample.

SEARCH
Flavobacteria contigs is now stretched out so that they can be explored individually.

REORIENT
Elviz allows the user to explore relationships among different variables by dynamically assigning them to x-axis, y-axis, and point size, and by binning.

Length is plotted on the x-axis in order to quickly identify the largest contigs in the sample.

DRILL DOWN
Clicking on an individual contig opens an annotation browser allowing the user to examine the contig in detail.

A Flavobacteriaceae contig containing the lipase is examined in detail allowing the user to reach the individual annotation and examine its gene neighborhood.

USER FEATURES

Landing Page
When logged in to the JGI, the home page provides convenient links to recently visited projects. From here the user selects whether to upload their own data or browse JGI datasets.

Upload wizard
To load a project into Elviz, provide a tab or comma delimited file with variable names as the first row. The upload wizard provides a preview of the data file and allows the user to include and exclude columns, as well as to specify initial assignments for x-axis, y-axis, point size and point color.

Implementation

Current advances in web standards have ushered in a new breed of Internet applications that approach the performance and interactivity of desktop tools while maintaining the platform independence and sharing capacity of the web.

Elviz employs WebGL, harnessing the client's graphical hardware (GPU), to render performant interactive displays of tens of thousands of data points. These data are stored on the client (using the HTML5 localStorage API). The client side of the application is written using Angular.js. This combination enables the user to operate Elviz without repeated round-trips to the server for data. The server side of Elviz consists of a RESTful API written in Java to provide data to the application.

When users choose to examine their own assemblies with Elviz (see below) they are given the choice to privately and securely store their uploaded data and project settings on the JGI Elviz server, allowing them to return to it in the future from any computer.

Elviz Data

Using Elviz with JGI data
Investigators can use Elviz to explore either their own metagenomics datasets or datasets created from metagenomics sequencing projects at the JGI. Assembly, annotation, and phylogenetic prediction for JGI projects in Elviz originate from the JGI Integrated Microbial Genome metagenomics pipeline, IMG/M1. All publicly available IMG/M metagenome projects released after Jan 1, 2014 are available for analysis with Elviz.

Using Elviz with user data
Elviz supports exploration of user data in a simple and highly customizable fashion. Users provide a tabular file (tab or comma delimited) in which each row represents a contig and each column defines a variable (e.g. length, GC content). The first row of this file must contain the column headings. The Elviz upload wizard then allows the user to assign columns in their data to the various dimensions of the plot including x-axis, y-axis, point size, and point color. Elviz also provides the capability to define color groups by binning of a user-selected column.

Acknowledgements

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. (DE-AC02-05CH11231).

Contact

jgi-viz@lists.lbl.gov