

Introducing IGB for Arabidopsis: genome & microarray data visualization

Ann Loraine(1), Gregg Helt (2), Ed Erwin(2), Sue Rhee(3) & Eva Huala(3)

(1) Section on Statistical Genetics, University of Alabama at Birmingham; (2) Affymetrix, Emeryville, CA (3) The Arabidopsis Information Resource, The Carnegie Institute, Stanford, CA

Introduction

The Integrated Genome Browser (IGB, pronounced 'Ig-Bee') is a Java software program that displays genes, sequence data, and microarray expression data all in the same view. It allows dynamic, real-time zooming and panning through a chromosome, link-outs to external databases, searching, and many other features and functions.

IGB is open source, free software. IGB was donated to the open source community by Affymetrix in September, 2004 and has since been imported into the Genoviz project at sourceforge.net.

Features

IGB can load data from remote Quickload or Distributed Annotation Server sites or from data files saved on local disk. To make it possible to explore Arabidopsis data using IGB, we set up experimental Quickload sites for Arabidopsis and reconfigured the program to allow link-outs to TAIR Web pages.

Chromosome 2 - a bird's eye view

horizontal zoomer

vertical zoomer

moveable, adjustable data tiers

gene structure annotations

probe set-to-genome alignments

Quickload data checkboxes

genome selection menu

When IGB launches, it accesses its Quickload server and creates a genome selection menu. When a genome is selected, IGB adds Quickload data checkboxes and a chromosome table. Clicking the checkboxes tells IGB to load the data.

probes

probe set design

probe set sequence

Gene model & probe set close-up

For Arabidopsis, available Quickload data include gene structure annotations and probe set-to-genome alignments for Affymetrix ATH1 and AG arrays. We hope to add more soon.

To see a gene model or probe set in detail, click to select it and use the horizontal zoomer to zoom in. To view a TAIR or Affymetrix Web page on a gene model or probe set, click to select it and then right-click. Choosing the **Get Info** option will open a Web page to TAIR or Affymetrix.

Genome Tiling Array Data

The Ecker group provided pre-processed, normalized array data from their 2003 genome tiling array results [1]. We converted the data to IGB's simple graph (.gr) format data files and have made these available for demonstration purposes on the tutorial Web site.

Viewing Genome Tiling Array Data

Genome tiling arrays contain probes spaced at regular intervals throughout the genome. IGB displays probe readings from genome tiling arrays as adjustable, draggable graphs superimposed on the genomic sequence axis. Graph height at any point indicates expression intensity at the corresponding genomic position.

Expression data as genome view graphs

values above VB max

Visible Bounds Max

graph handle

gene model

AT2G18020.1

visible range settings

AT2G18020.1 - ribosomal protein

Expression values captured in a single experiment are usually unevenly distributed and range over several orders of magnitude. This can pose a problem for graph scaling. IGB solves this through its Visible Bounds settings. The Max and Min Visible Bounds set the viewable window for graph values and control how the low & high-intensity readings affect the vertical scale. Probe readings that are below the Min setting are shown at minimum height while probe readings that are above the Max setting are shown at maximum height. Intermediate probe readings vary between these two extremes.

Expression data as genome graphs

anther data

expression signal

2 gene models

root data

AT2G02040.1 - putative peptide transporter, AT2G02050.1 - putative NADH dehydrogenase

anther data

expression signal

ATH1 probes

gene model

AT2G02040.1 - putative cadmium-transporting ATPase

References

[1] Yamada, et al (2003) Empirical analysis of transcriptional activity in the Arabidopsis Genome, Science, v. 302, p 842.