

IGB (Integrated Genome Browser) and Genoviz: genome and microarray data visualization

John Nicol, Steven Blanchard, Ketan Patel, Ann Loraine
Department of Bioinformatics and Genomics, University of North Carolina at Charlotte

Introduction

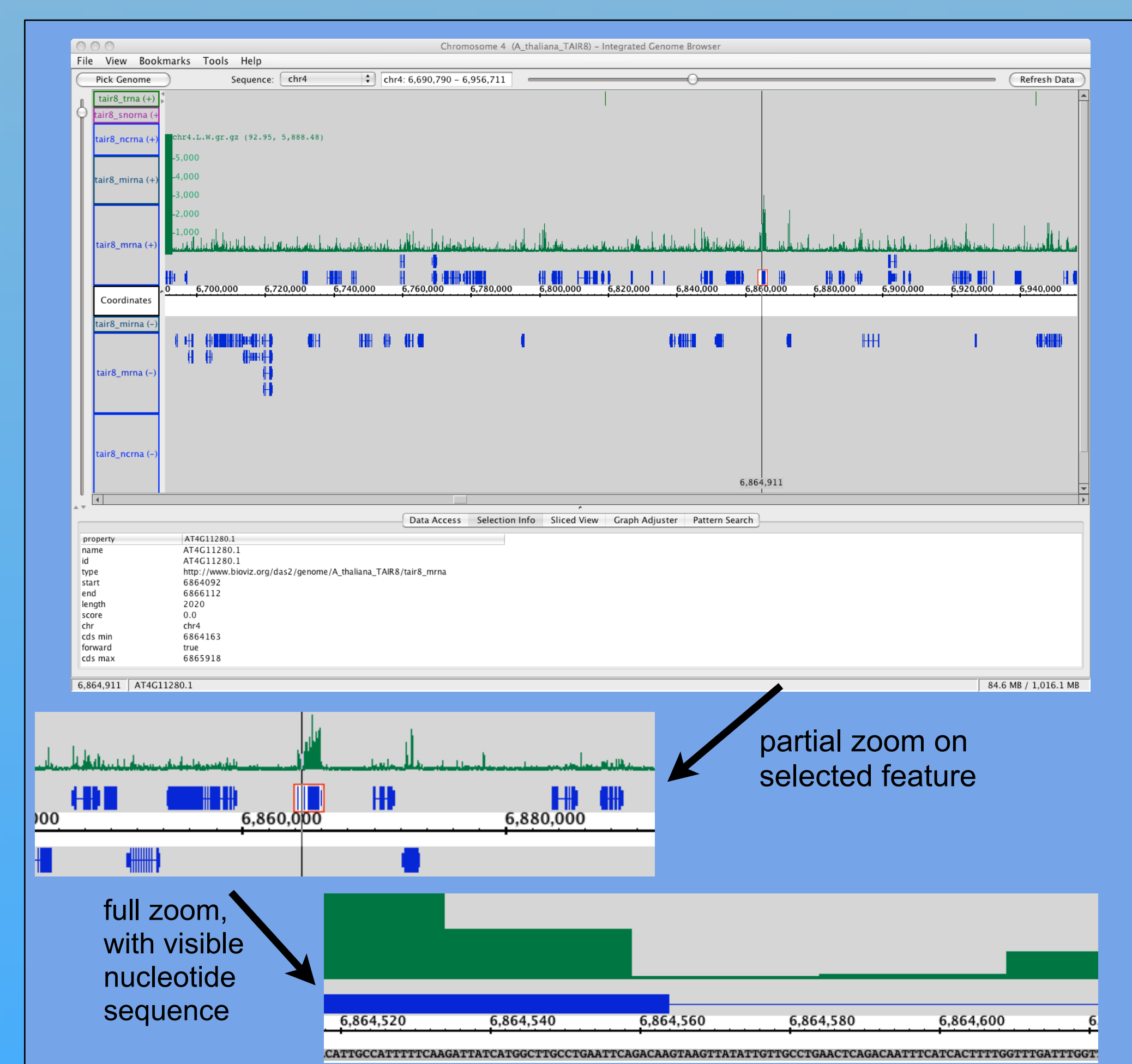
The Integrated Genome Browser (IGB, pronounced 'Ig-Bee') is a desktop program implemented in Java that displays genes, sequence data, and genome-scale data sets in one view. It allows dynamic, real-time zooming and panning through a chromosome or sequence assembly, linking to external databases and servers (including DAS/1 and DAS/2), searching, and many other features and functions. IGB's users include researchers at multiple sites, most of whom use IGB to display and explore genome-scale data sets, especially tiling array data and data from next-generation sequencing experiments. To support Arabidopsis and plant genomics communities, we are developing a DAS/2 site (the ArabiDAS) and a user-friendly Web site where users download IGB and learn about its features.

IGB is open source

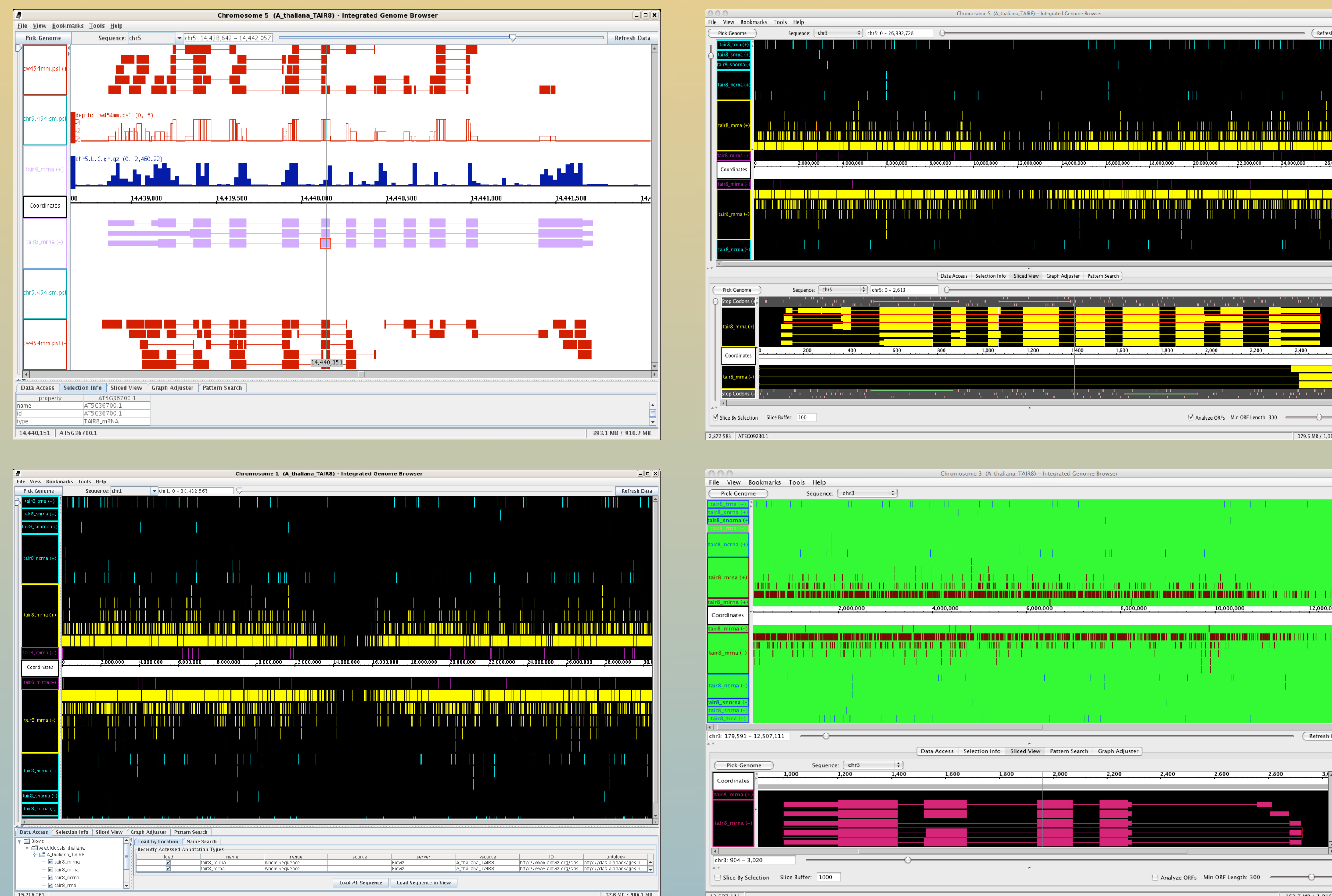
IGB started as a proprietary, internal product at Affymetrix, Inc. In 2004, Affymetrix released IGB and associated libraries as open source software via their Web site. The following spring, it was launched as a Sourceforge project called Genoviz, dedicated to developing visualization applications for genomics and allowing collaboration with developers both inside and outside Affymetrix. The project now includes several part-time developers and one full-time software engineer in our group (JN). In 2008, we received a grant from NSF's Arabidopsis 2010 program to develop IGB and back end data servers for the Arabidopsis research community.

Features - Zooming

IGB supports real-time, dynamic zooming through annotation features associated with a chromosome. Zoomed out, IGB can show an entire chromosome or genome. Zoomed in, IGB shows individual sequence bases.

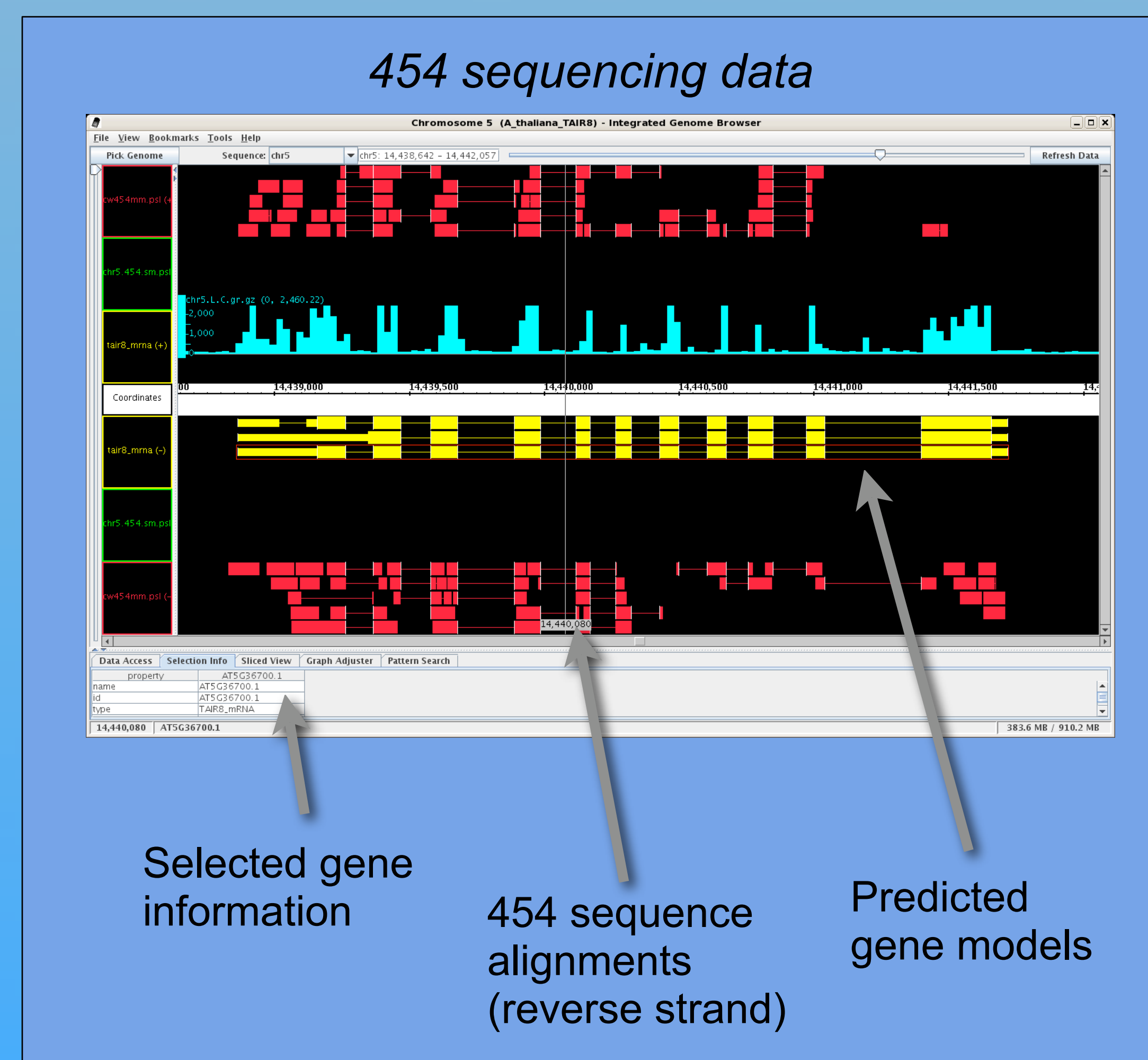


Gallery



Features - Data Access

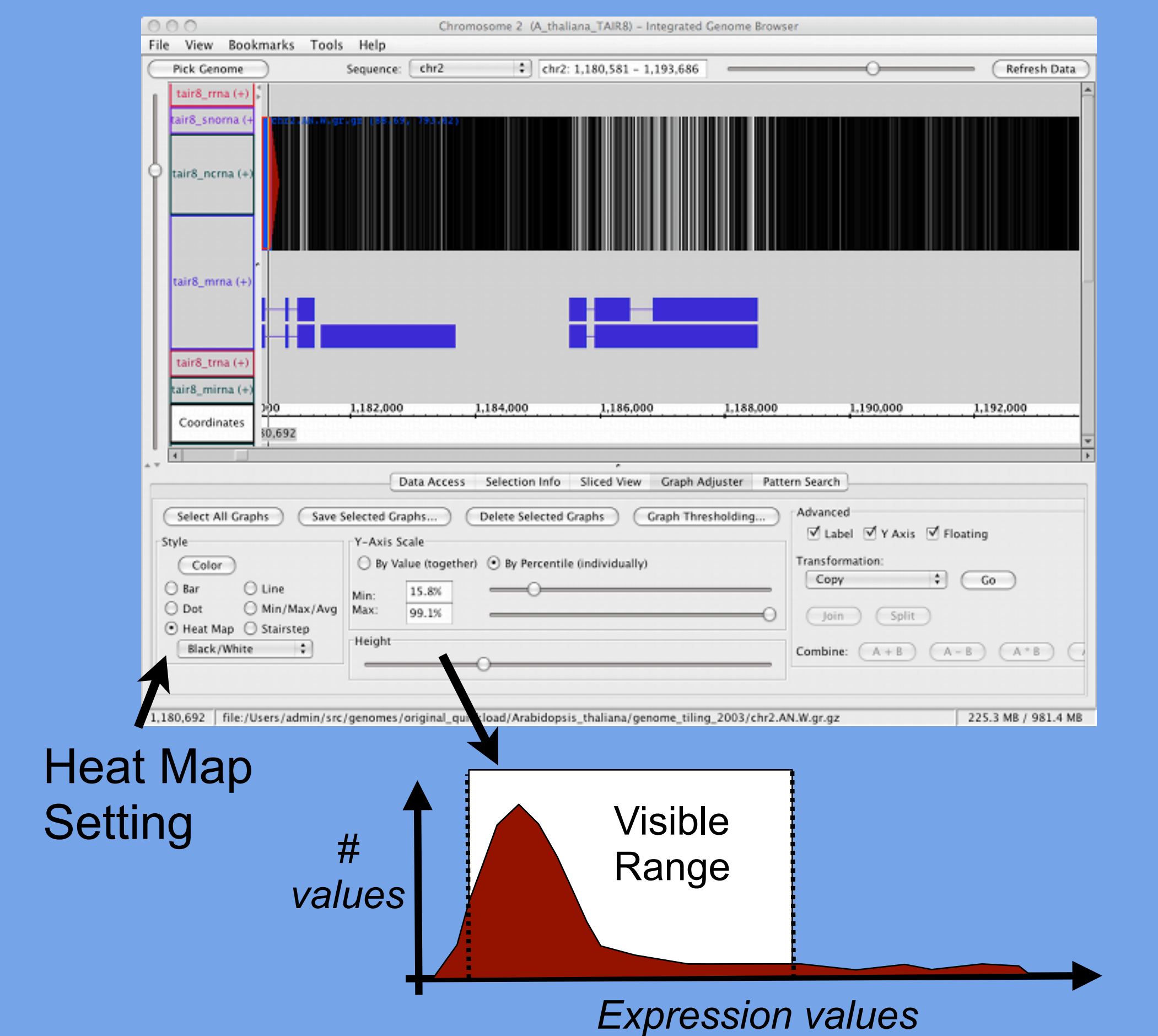
IGB can load data from local files and from Distributed Annotation Servers (DAS). DAS offers simple, Web-based access to data. IGB can read a variety of file formats, including PSL (from the BLAT genome alignment tool), FASTA, and BED format files available from UCSC's Genome Bioinformatics Table Browser. We host a DAS/2 site (the ArabiDAS) that serves Arabidopsis data at <http://bioviz.org/das2/genome>.



Features - Tiling Array Controls

IGB can display numeric data as genome graphs. Genome graphs associate numbers with individual base pair positions. Tiling arrays are DNA microarrays with probes selected from intervals along the genomic sequence.

Tiling array data shown as a heat map.



The Graph Adjuster tab lets users change graph appearance and adjust the range of values displayed in the graph vertical space.

Improvements

Since starting work in September, 2008, JN has made several improvements to IGB and the ArabiDAS, including:

100x speed-up on DAS/2 server startup

The ArabiDAS is implemented as a Java servlet. When it launches, it accesses a feeder directory containing data files, reading the data into memory. By an improved design, we decreased startup time by up to two orders of magnitude.

5x decrease in DAS/2 network traffic

DAS transmits information over http. By implementing data compression we lowered network traffic significantly (resulting in quicker server responses).

Clarifications to DAS/2 specification

The ArabiDAS uses DAS version 2.0, which is under development. We have corrected and clarified specification inconsistencies.

Loading sequence data via DAS/2

The ArabiDAS now allows partial and complete downloads of a chromosome's nucleotide sequence.

Numerous bug fixes in IGB and DAS/2 server

Improvements to IGB testing framework

Updates to Genoviz Documentation

Planned Improvements

Improved data access interface

We plan to re-design IGB's data access interface, making it easier and more intuitive how to activate and control data loading.

Add more content to <http://igb.bioviz.org>

We plan to create user vignettes (like use cases) describing how IGB can solve key visualization problems in genomics, such as interpreting EST alignments.

Acknowledgements

- This work is supported by NSF grant #0820371.
- IGB and Genoviz were generously released as open source by Affymetrix in 2004.
- Sourceforge hosts the Genoviz source code and supports its thriving community of developers.