

## Connecting Integrated Genome Browser to a huge genome database using its open API solves one problem and creates another – and an opportunity.

Ann Loraine<sup>1,2</sup>, Jaya Sravani Sirigineedi<sup>2</sup>, Nowlan Freese<sup>2</sup>

<sup>1</sup> Contact Email: [aloraine@charlotte.edu](mailto:aloraine@charlotte.edu)

<sup>2</sup> University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC 28223

**Project Website:** <https://bioviz.org>, <http://igbquickload-main.bioviz.org/quickload>

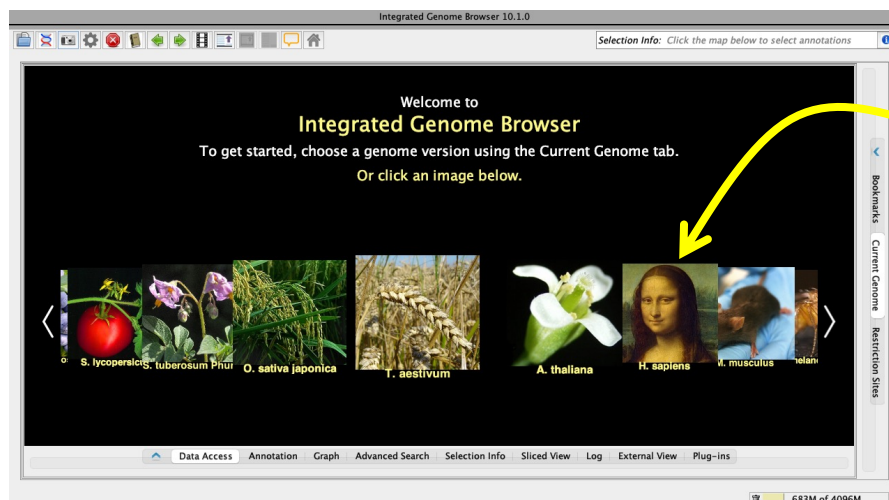
**Source Code:** <https://bitbucket.org/lorainelab/integrated-genome-browser>

**License:** [Common Public License Version 1.0](#)

Integrated Genome Browser (IGB, pronounced “ig-bee”) is a fast, feature-rich, open-source desktop genome browser thousands of researchers have used to explore and analyze genomic data. To support our user audience, we maintain data delivery Web sites called “IGB Quickloads” that supply IGB with 97 reference genome assemblies as of this writing. Of course, IGB can open user-provided genome assembly files, but if a researcher’s desired assembly already exists in an IGB Quickload, they can avoid this inconvenient work. However, we are finding it increasingly difficult to update these Quickload sites as new assemblies are published. Fortunately, many genome database systems now offer robust computational access to their data. By accessing these computational resources, IGB could show new assemblies without our first re-formatting and copying them to a Quickload. To test this idea, we developed a new IGB version that consumes and displays data from one such resource, a JSON-emitting API (application programming interface) from the UCSC Genome Browser system. Now available as an “early access” version at the BioViz.org Web site, this new IGB version can display more than 200 assemblies visible in the UCSC Browser, along with hundreds of experimental data and annotation tracks the API also provides.

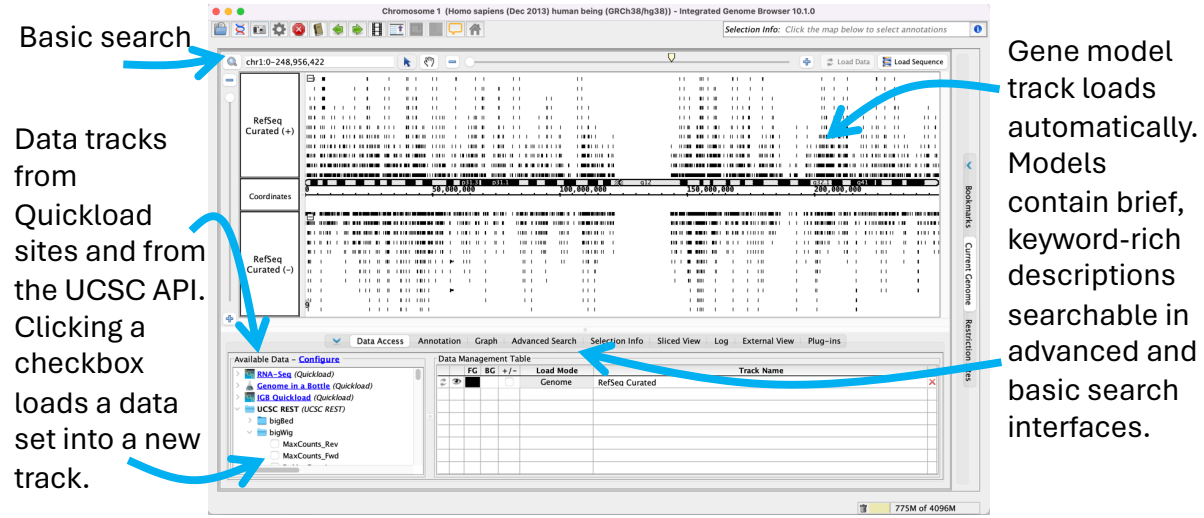
This wealth of annotations and experimental data tracks gives us a new problem to solve, but also an opportunity. The API provides information like track names and data formats, but little about what the data represent. IGB can offer users the option to load these tracks, but without meta-data documenting what they show, how can users understand and interpret what they see? We urge API authors to augment their APIs with richer meta-data about the data sets, such as references to publications, documentation on the data providers’ own web sites, or even ontology terms that formally classify the data sets. With such meta-data, we could create search interfaces for users to find all available data relevant to their scientific interests, reducing their risk of overlooking relevant, previously published work. By providing more complete meta-data, the UCSC team and others like them could make their amazing resources accessible to even more audiences – developers like us and the worldwide communities we serve.

**Figure 1. IGB start screen showing species with available genomes**



Clicking an image opens the newest assembly available via from IGB Quickload for the selected species. Reference gene models annotated with brief functional descriptions also load. Keeping these data up-to-date is hard.

**Figure 2. IGB showing the newest human genome assembly available via Quickload.**



**Figure 3. IGB data chooser interface showing folder contents.**

