Data Sources Tab

Introduction

Use the **Data Sources** Preferences tab to

- view, add, or edit data source sites from which IGB can retrieve data.
- load personal genome version and chromosome synonyms files
- change how IGB caches data from remote and local sites

When you choose a species and genome version, IGB downloads meta-data files about available genomes from each site listed in the Data Sources list. If a site contains data for the currently displayed genome, then IGB lists those sites in the **Available Data Sources/Data Sets** section of the **Data Access** tab. And when you load sequence data into IGB by clicking the Load Sequence button, IGB attempts to obtain the sequence data from the data sources in the same order they appear in the Data Sources list.
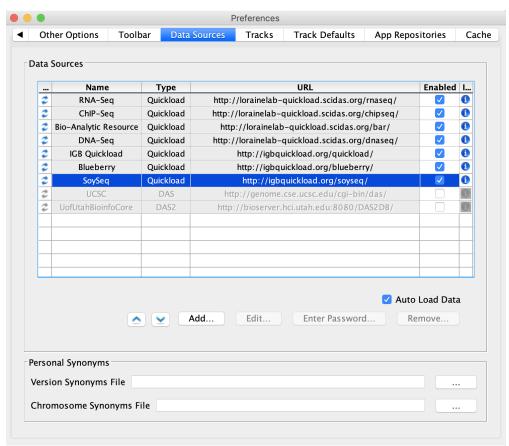
To open the Data Sources Preferences tab

- Select **Data Access > Configure**

or

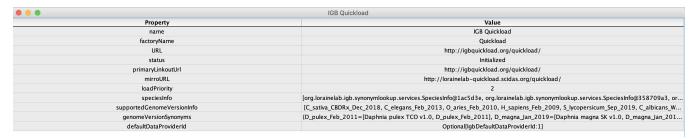- Select **File > Preferences > Data Sources**

Data Sources

Data sources

The **Data sources** table lists all active and inactive data sources that have been added to IGB. Active data sources have a check in the enabled column. When you select a genome version, IGB contacts each active server to determine if the server has any data sets for the selected genome. If a server has data sets for the selected genome, the available data sets will appear as options in the **Available Data Sources/Data Sets** section of the **Data Access** tab.
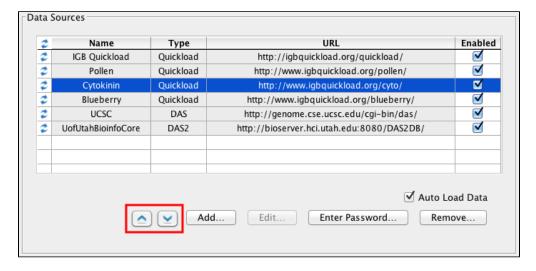
Information About Data source

You can click on the info symbol(i symbol in blue circle) of the Data sources tab to display information about the particular Data source.

Property	Value
	IGB Quickload
name	IGB Quickload
factoryName	Quickload
URL	http://igbquickload.org/quickload/
status	Initialized
primaryLinkoutUrl	http://igbquickload.org/quickload/
mirroURL	http://lorainelab-quickload.scidas.org/quickload/
loadPriority	2
speciesInfo	[org.lorainelab.igb.synonymlookup.services.SpeciesInfo@1ac5d3e, org.lorainelab.igb.synonymlookup.services.SpeciesInfo@358709a3, or...
supportedGenomeVersionInfo	[C_sativa_CBDRx_Dec_2018, C_elegans_Feb_2013, O_aries_Feb_2010, H_sapiens_Feb_2009, S_lycopersicum_Sep_2019, C_albicans_W...
genomeVersionSynonyms	{D_pulex_Feb_2011=[Daphnia pulex TCO v1.0, D_pulex_Feb_2011], D_magna_Jan_2019=[Daphnia magna SK v1.0, D_magna_Jan_201...
defaultDataProviderId	Optional[IgbDefaultDataProviderId:1]

Data sources order

For some genome versions, more than one data source may provide sequence data. When you click the **Load Sequence** button (top right), IGB loads sequence directly from the first server on the list that supports the currently selected genome. Starting with IGB 6.6, you can change the order of data sources listed in the **Data Sources** preferences tab.

To change sequence order, click **Up** and **Down** buttons (red box).

Adding new data sources

Use **Add...** to add a new data source. A new window will open with boxes and a menu for you to specify a new data sources.

Enter user names and passwords

Click **Enter Password...** to enter or change the user name and password IGB uses to access private data sources.

> ⚠ Most data sources don't require authentication. If you select a public (open) data source and click **Enter Password...**, IGB will allow you to enter a user name and password but won't use it when contacting the data source.

Data loading settings

Some servers are configured to automatically load certain foundation data sets, such as the RefSeq gene models, into IGB as soon as the corresponding genome is selected. To block this behavior, uncheck **Auto Load Data**.

Personal synonyms files

Different data sources often refer to the same genome version or chromosome by different names.

For example, hg17, ensembl1834 and H_sapiens_May_2004 all refer to the same genome assembly and sequence. Similarly, some data providers use **ch r1** to refer to chromosome one while others may use **1** to mean the same thing.

IGB provides a system for designating synonyms for sequences and genome versions.

Genome synonyms File

A version synonyms file matches genome versions and allows you to display data from multiple sources in the same window.
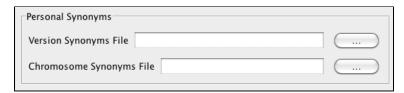
Each time IGB restarts, synonyms will be loaded and merged from all the servers in use as well as from any personal synonyms file that you have specified.

A genome version synonyms file follows a simple tab-delimited format in which synonyms for the same genome appear in the same line, separated by tab characters.

Chromosome Synonyms File

A chromosome synonyms file is tab-delimited. Each line lists possible names for the same sequence.

Each time IGB starts, synonyms will be loaded and merged from all the servers in use as well as from any personal chromosome synonyms file entered here.

Cache Settings and clear the cache

IGB caches data that you have loaded over the web in order to speed the loading process the next time you use it. Files loaded from your local file system will not be cached.

Clicking the **Empty Cache** button forces IGB to clear its local cache of downloaded data sets.

Tip. If you suspect that the cache files have somehow been corrupted, which may happen if a download was halted unexpectedly, unexpected behavior may result. Click **Empty Cache** if data sets you've opened from a QuickLoad site don't appear to have any data or cause IGB to freeze.