HymenopteraMine Documentation

Release v1.6

Aditi Tayal, Deepak Unni, Colin Diesh, Chris Elsik, Darren Hagen

Jul 24, 2023

Table of Contents

1 Overview of HymenopteraMine **3**

2 Searching in HymenopteraMine **5**
- 2.1 Quick Search . . . 5
- 2.2 Templates . . . 6
  - 2.2.1 Generate query code . . . 9
  - 2.2.2 Download results . . . 9
  - 2.2.3 Customize output . . . 10
  - 2.2.4 Optional filters . . . 10
- 2.3 QueryBuilder . . . 10
  - 2.3.1 Model browser . . . 11
  - 2.3.2 Examples . . . 12
    - 2.3.2.1 Example 1: Querying for protein coding genes . . . 12
    - 2.3.2.2 Example 2: Querying for protein coding genes on a particular chromosome . . . 14
    - 2.3.2.3 Example 3: Querying for protein coding genes on a particular chromosome and their exons . . . 16
- 2.4 Genomic Regions Search . . . 20

3 Report Pages **23**
- 3.1 Summary . . . 24
- 3.2 Alias and DBxref . . . 24
- 3.3 Transcripts . . . 25
- 3.4 Proteins . . . 26
- 3.5 Function . . . 27
- 3.6 Homology . . . 28
- 3.7 Publications . . . 28
- 3.8 Other . . . 30

4 Lists **31**
- 4.1 Creating Lists . . . 31
- 4.2 Saving Lists . . . 31

5 MyMine **35**

6 API **37**

7 Data Sources **39**

8 BLAST 41

9 How to Cite 43

HymenopteraMine is an integrative resource for genomic data on Hymenoptera, including honeybees, ants, wasps, etc. Powered by InterMine, it provides a user-friendly way to access genomic, proteomic, interaction and literature data. HymenopteraMine is a part of the Hymenoptera Genome Database.

This tutorial is aimed at giving users an introduction to the different parts of HymenopteraMine and how users can make the most of HymenopteraMine.

Main site: http://hymenopteragenome.org/hymenopteramine

HGD YouTube Channel with HymenopteraMine Videos: https://www.youtube.com/channel/UC1NVFd9buEtlbA2mcdq0MXQ

List of available datasets in HymenopteraMine: http://hymenopteragenome.org/hymenopteramine/dataCategories.do

CHAPTER 1

Overview of HymenopteraMine

Below is a brief summary of the layout of HymenopteraMine:

Home – The home page for HymenopteraMine.

MyMine – Once users are logged in, MyMine serves as portal for accessing saved lists and saved templates. Users may also check their account details and manage their account using MyMine.

Templates – List of templates that users may select from based on the nature of their query.

Lists – Allows users to upload lists of genes and perform enrichment analyses. Logged-in users may save their lists for future use.

QueryBuilder – Allows users to build custom queries by browsing the HymenopteraMine data model and customize their results. The queries may be exported to a number of formats including XML.

Regions – Genomic Region Search page where users may enter genomic coordinates and fetch features that fall within the interval. The interval may be extended to increase the range of search.

Data sources – Table of all data sources with their links, date of download, and related publication(s).

Taxonomy - Taxonomic tree for all species in HymenopteraMine (based on the NCBI taxonomy database), given as a reference for homologue queries.

Help- Links to the HymenopteraMine tutorial.

API – Describes the InterMine API that allows users to programmatically access HymenopteraMine.

HGD Blast – Links to a Blast page where users may BLAST their sequence(s) of interest with the Hymenoptera species reference genome, CDS sequences, and protein sequences.

CHAPTER 2

Searching in HymenopteraMine

There are several ways that users may query HymenopteraMine.

2.1 Quick Search

The **Quick Search** enables users to search keywords from any of the datasets on HymenopteraMine. The quick search box is located on the main page and in the upper-right corner of each page.

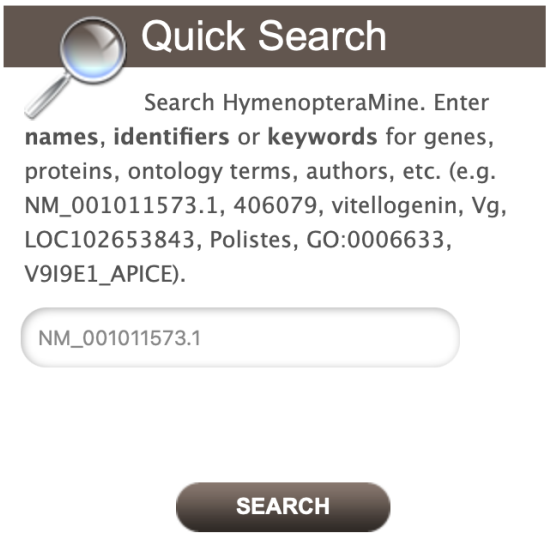


Fig. 1: Quick Search from home page

Users may enter gene names, gene identifiers, or keywords to search for within HymenopteraMine. The wildcard character * may be used to get all results matching the search query.

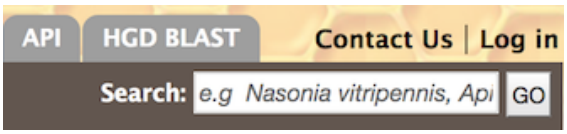


Fig. 2: Quick Search from any page

As an example, to search on the gene **LCCH3**, enter **LCCH3** into the quick search box and click **Search**. The results page displays a summary for the query in a tabular format.

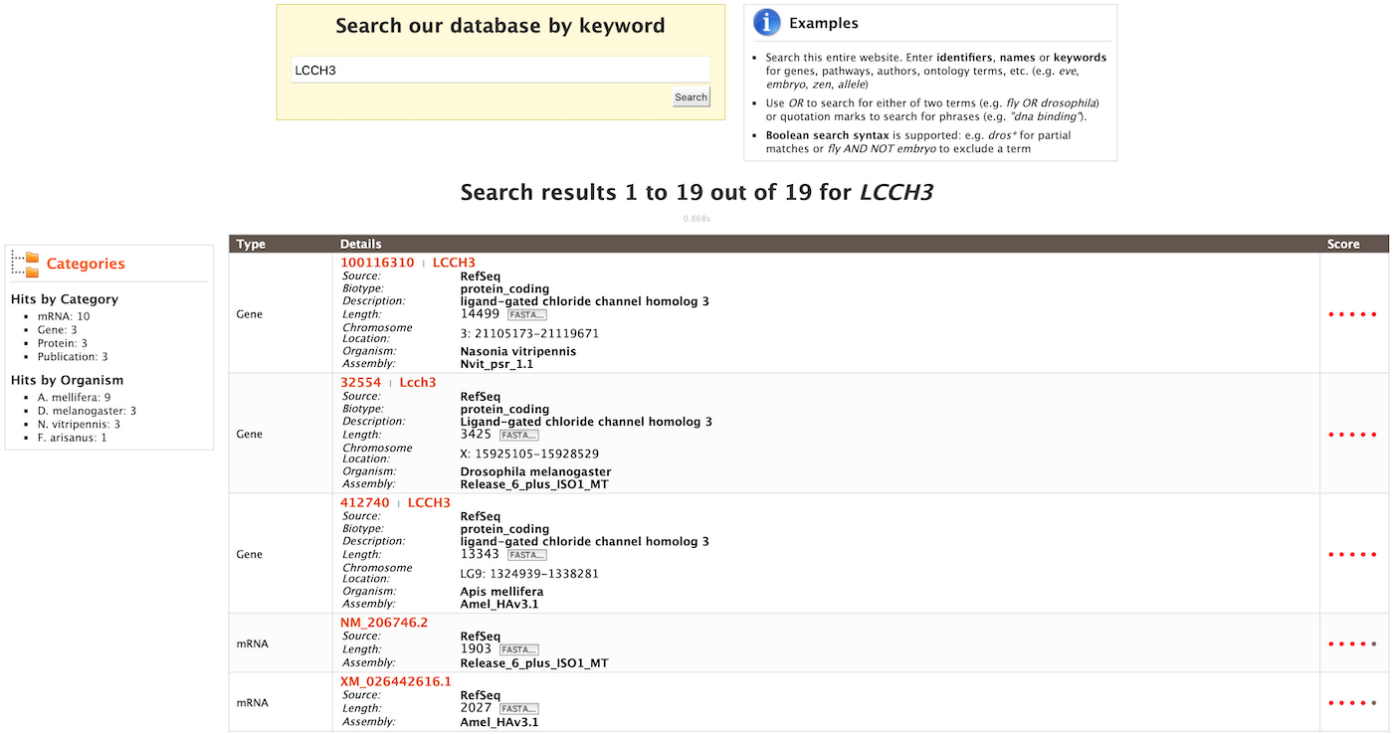


Fig. 3: Example: Search results for LCCH3

The search results may be filtered by **Category** or **Organism**. The score column in the result table indicates the similarity of the query to the result fetched by HymenopteraMine.

The results page may also be converted to a list. To enable this feature, click on **Gene** in **Hits by Category**. Select the gene(s) to be included in the list, and click on the **Create List** button just above the table of search results. Lists are described in more detail in the *Lists* section.

2.2 Templates

Another method of searching HymenopteraMine is through the use of **templates** (predefined queries). Popular templates are displayed on the home page, grouped by category (Genes, Protein, Homology, etc.) The full list of templates may be viewed by clicking the **Templates** menu tab.

As an example, the **GO Term Gene** template queries HymenopteraMine for all genes annotated with a given GO term for a selected organism.

The results page shows all *A. cephalotes* genes having the Gene Ontology term "DNA Binding" (GO:0003677) in their annotation. When logged in, users may create a new list or add these genes to an existing list to perform further

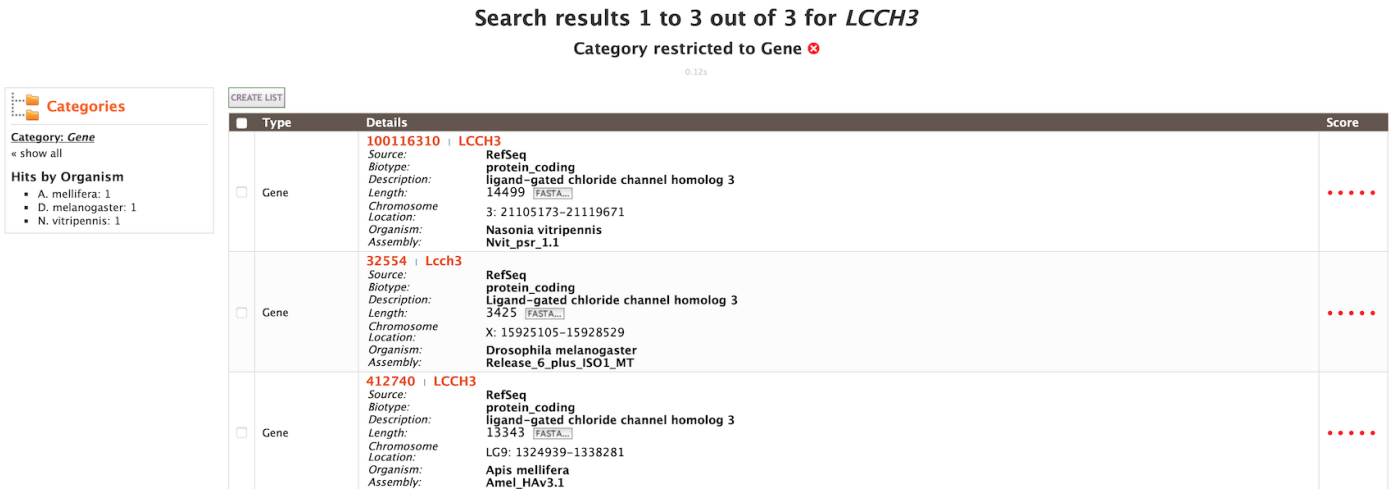


Fig. 4: Example: Search results restricted to Gene category

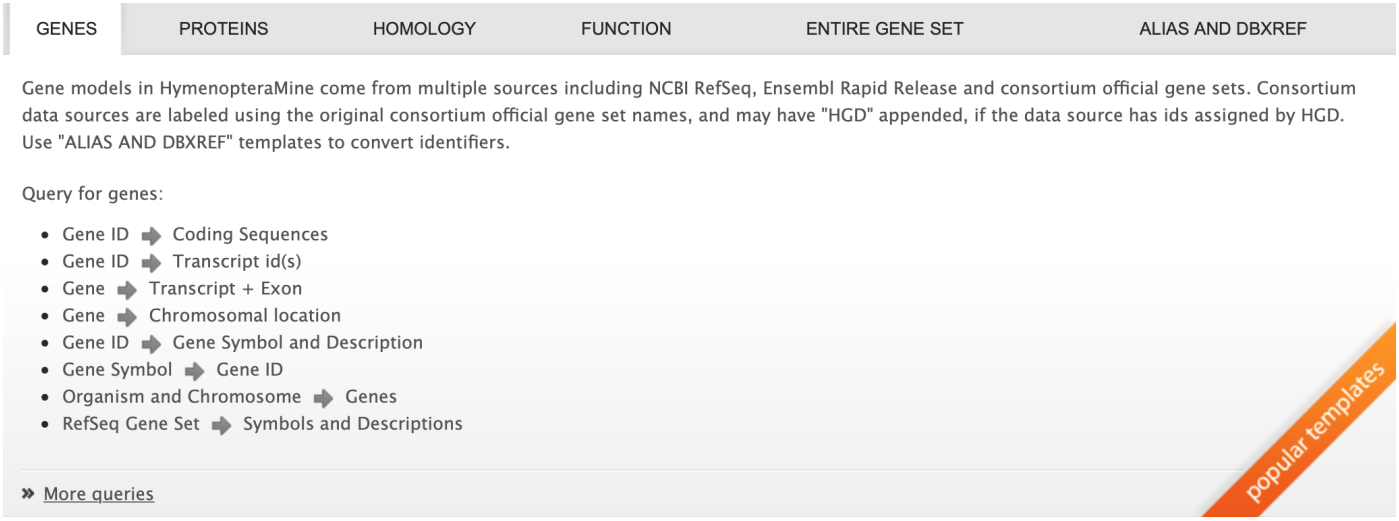


Fig. 5: Popular templates

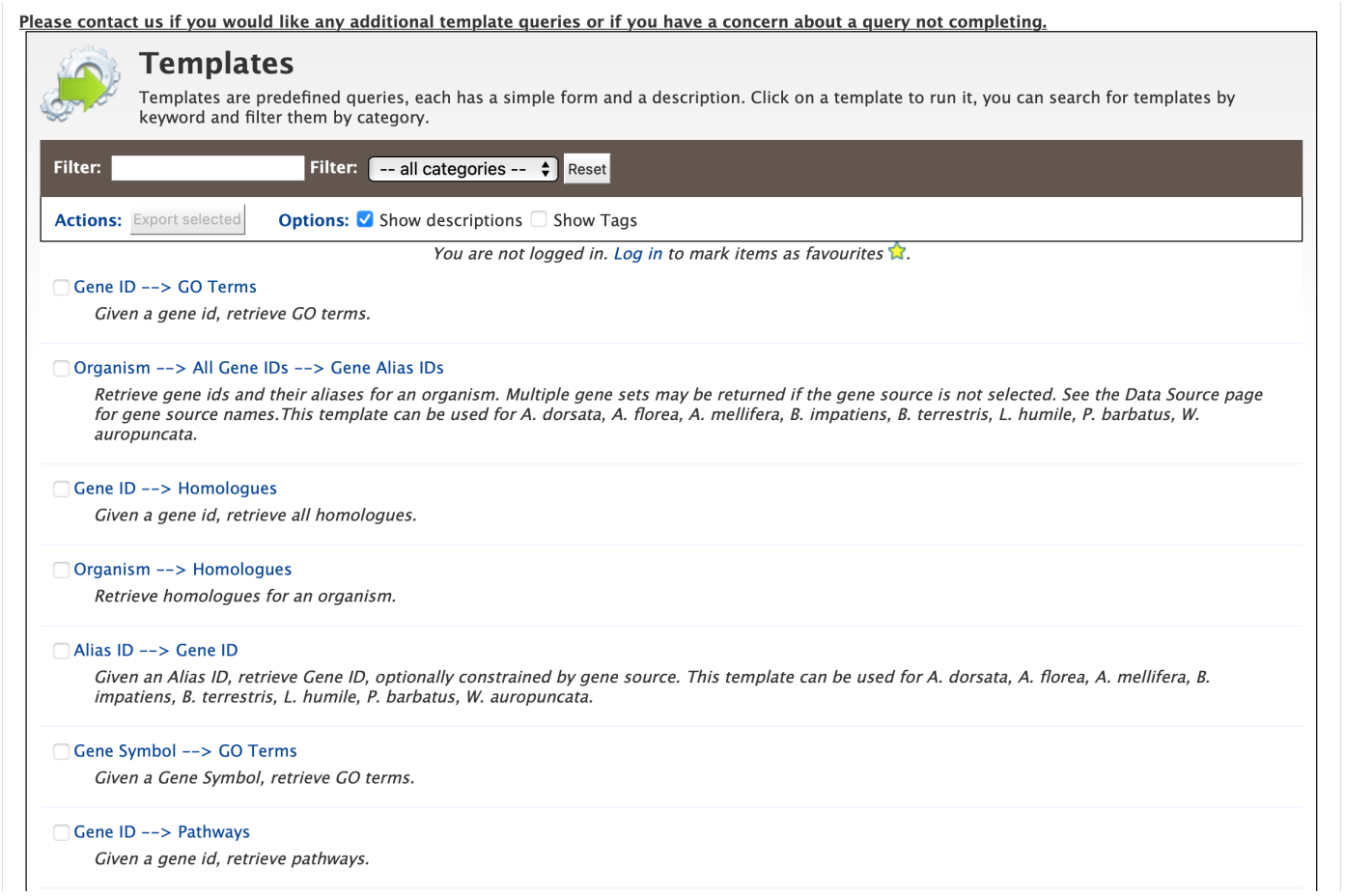


Fig. 6: List of templates on the Templates page

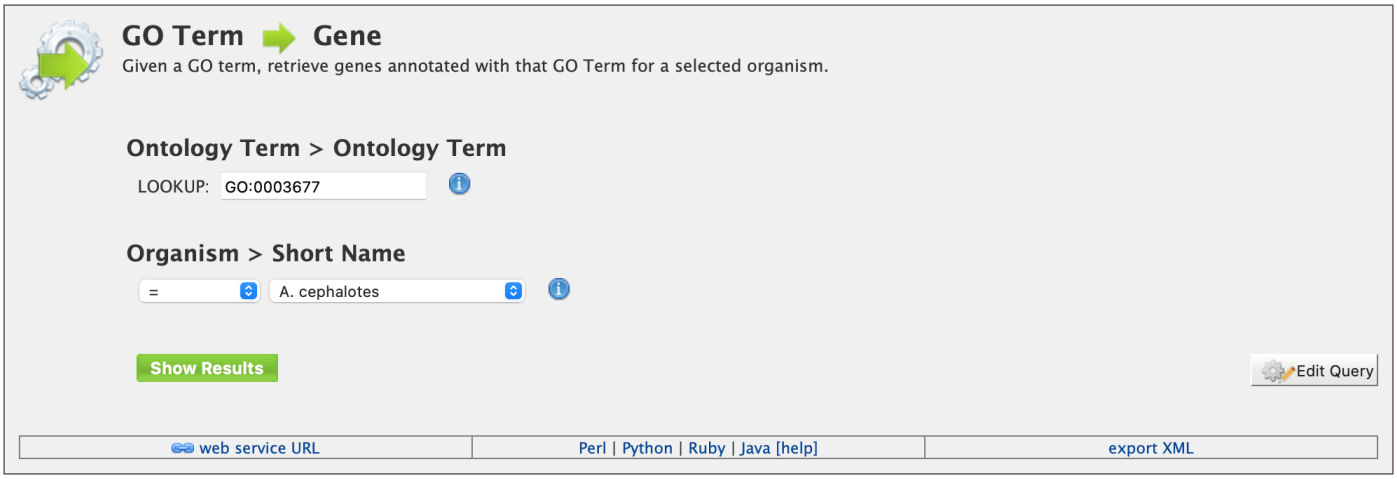


Fig. 7: Example: GO Term Gene template

analyses. Click on the **Save as List** button above the table of results, then choose the column to add to the list. See the *Lists* section for more details on creating and saving lists.

Trail: Query
GO Term → Gene
Given a GO term, retrieve genes annotated with that GO Term for a selected organism.

Manage Columns | Manage Filters | Manage Relationships | Save as List | Generate Python code | Export

Showing 1 to 25 of 1,192 rows

Rows per page: 25 | page 1

GO Annotation Ontology Term Identifier	GO Annotation Ontology Term Name	GO Annotation Ontology Term Description	Code Code	Gene Gene ID	Gene Symbol	Organism Short Name	Data Sets Name	Data Source Name
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	ECO:0007669	105616992	LOC105616992	A. cephalotes	TrEMBL data set	UniProt
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	105616992	LOC105616992	A. cephalotes	HGD GO annotation data set	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	ECO:0007669	105616998	LOC105616998	A. cephalotes	TrEMBL data set	UniProt
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	105616998	LOC105616998	A. cephalotes	HGD GO annotation data set	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	ECO:0007669	105617009	LOC105617009	A. cephalotes	TrEMBL data set	UniProt
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	105617009	LOC105617009	A. cephalotes	HGD GO annotation data set	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	ECO:0007669	105617018	LOC105617018	A. cephalotes	TrEMBL data set	UniProt
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	105617018	LOC105617018	A. cephalotes	HGD GO annotation data set	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	ECO:0007669	105617019	LOC105617019	A. cephalotes	TrEMBL data set	UniProt
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	105617019	LOC105617019	A. cephalotes	HGD GO annotation data set	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	ECO:0007669	105617041	LOC105617041	A. cephalotes	TrEMBL data set	UniProt

Fig. 8: Example: Results after searching on *GO:0003677* (identifier for GO term "DNA binding")

2.2.1 Generate query code

The code for each query may be obtained by clicking on the arrow next to **Generate Python Code** and choosing the desired language from the pull-down menu. The language options are Python, Perl, Java, Ruby, JavaScript, and XML.

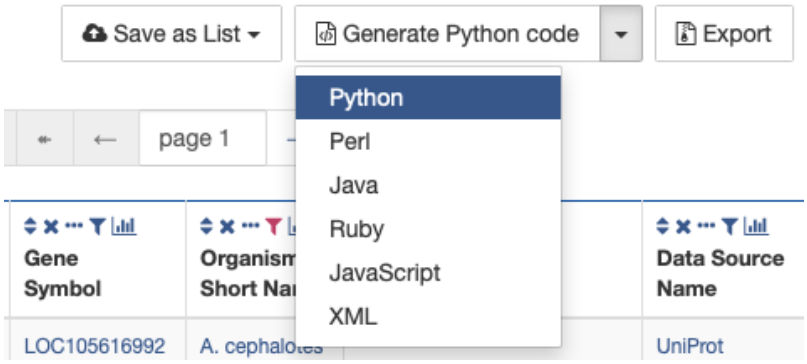


Fig. 9: Generate code options

2.2.2 Download results

The search results may also be downloaded by clicking the **Export** button above the table and choosing the desired format from the pull-down menu to the right of the File name field (blue box in the figure below). Available formats are tab-separated values, comma-separated values, XML, and JSON. When the results contain genomic features, they may also be downloaded in FASTA, GFF3, or BED format. Other options may be specified in the submenu to the left of the download box (orange box in the figure below). By default, all rows and all columns are downloaded, but individual columns may be included or excluded by clicking on the toggles next to the column headers in the **All Columns** submenu. The number of rows and row offset are set in the **All Rows** submenu. Download the results

as a compressed file by choosing GZIP or ZIP format in the **Compression** submenu (default is **No Compression**). Column headers are not added by default but may be included under the **Column Headers** submenu. Finally, the **Preview** submenu displays the first three rows of the file to be downloaded so that the desired format and options may be finalized before beginning the download. When ready, click the **Download file** button to download the results.

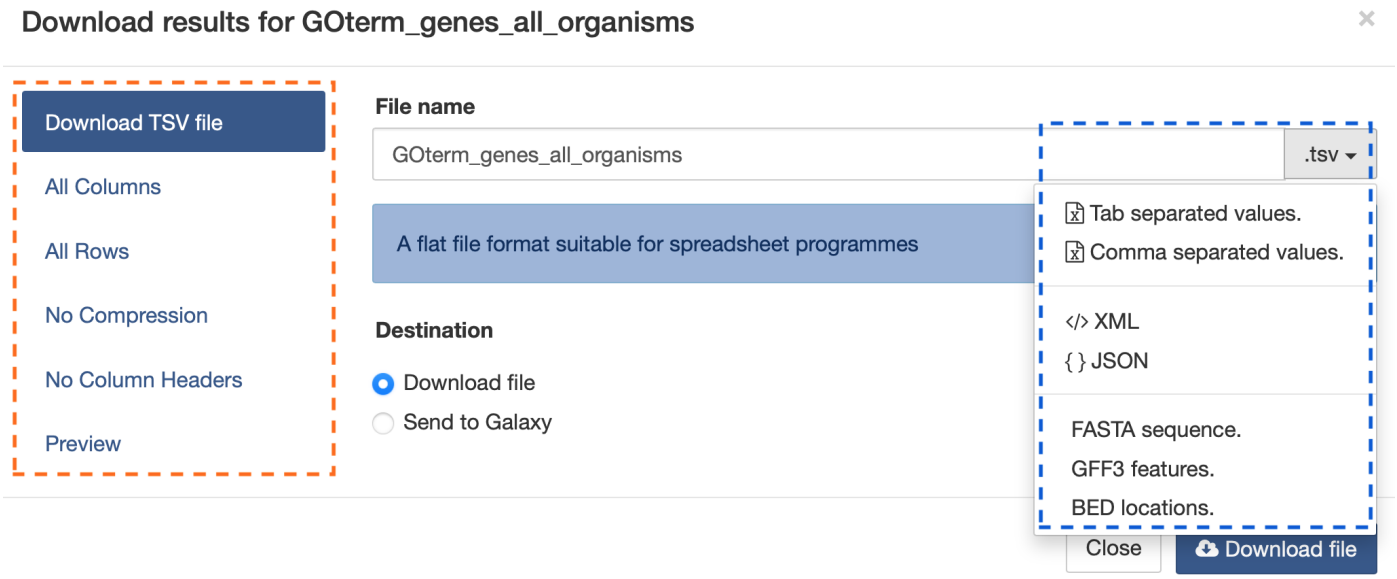


Fig. 10: Download results options

2.2.3 Customize output

Click the **Manage Columns** button to customize the results table layout. Edit or remove active filters by clicking the **Manage Filters** button. Click **Manage Relationships** to specify the entity relationships within the query.

2.2.4 Optional filters

Some templates have optional filters that are disabled by default. For example, the Organism Gene IDs with their Alias IDs template has an additional filter for specifying a gene source (e.g., RefSeq, Ensembl). To enable this filter, click **ON** under **optional** (red box in figure below).

Note: The Query trail link at the top of the page does not work for templates with optional filters. To edit the template query, navigate back to the template page either by clicking on the template name at the top of the query results page or by selecting the template from the Templates tab.

2.3 QueryBuilder

While the templates provided are suitable for many different types of searches, new queries may be built from scratch using the **QueryBuilder**. The possibilities of queries using the QueryBuilder are endless. The output may be formatted exactly as desired, and the query constraints may be chosen to perform complex search operations.

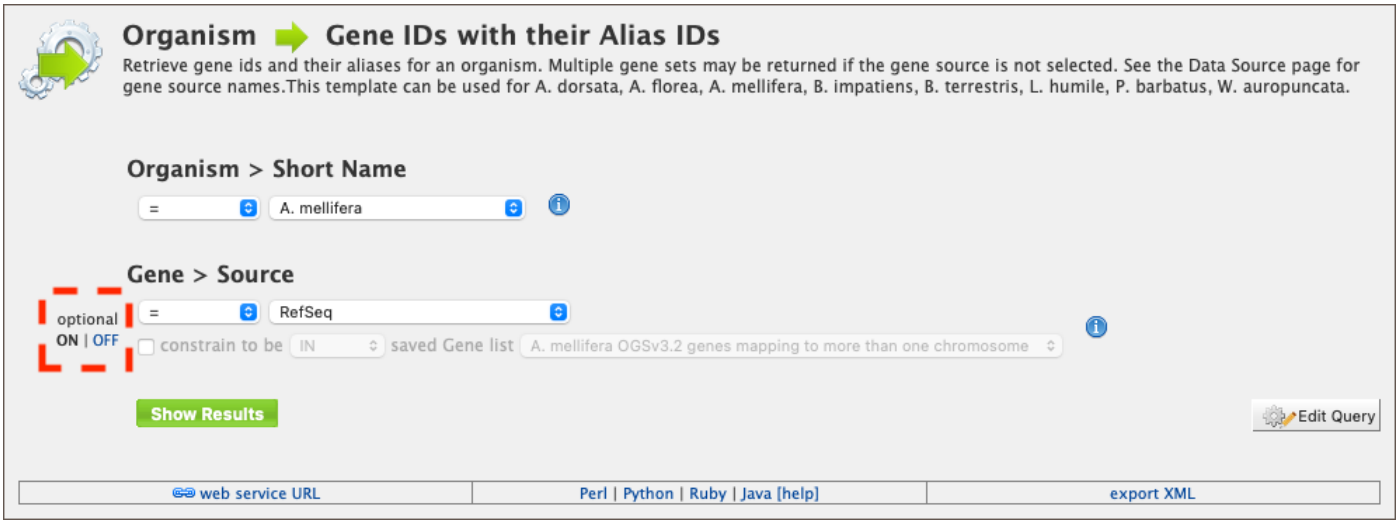
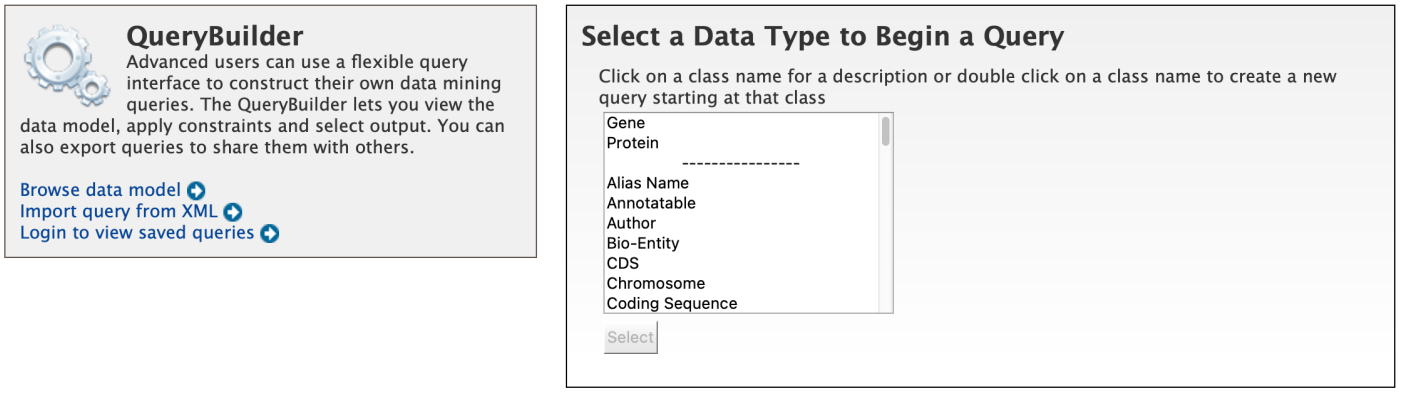


Fig. 11: Example: Organism Gene IDs with their Alias IDs template with source filter enabled



To begin, select a **Data Type**. For example, select **Gene** as a Data Type and click the **Select** button.

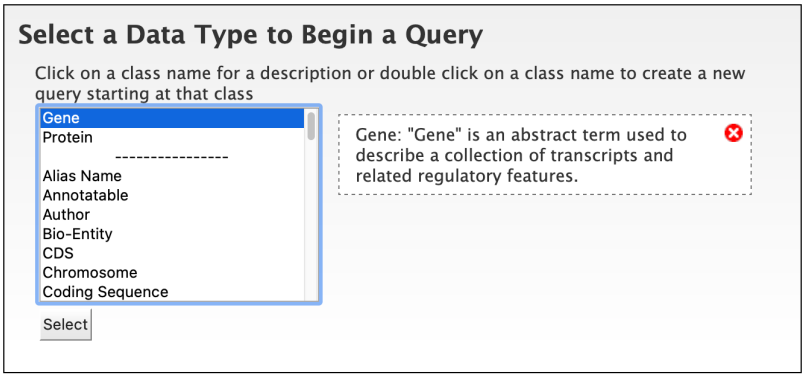


Fig. 12: Example: Gene data type selected in QueryBuilder

2.3.1 Model browser

After choosing a data type, the **Model browser** appears displaying the attributes for the selected feature class.

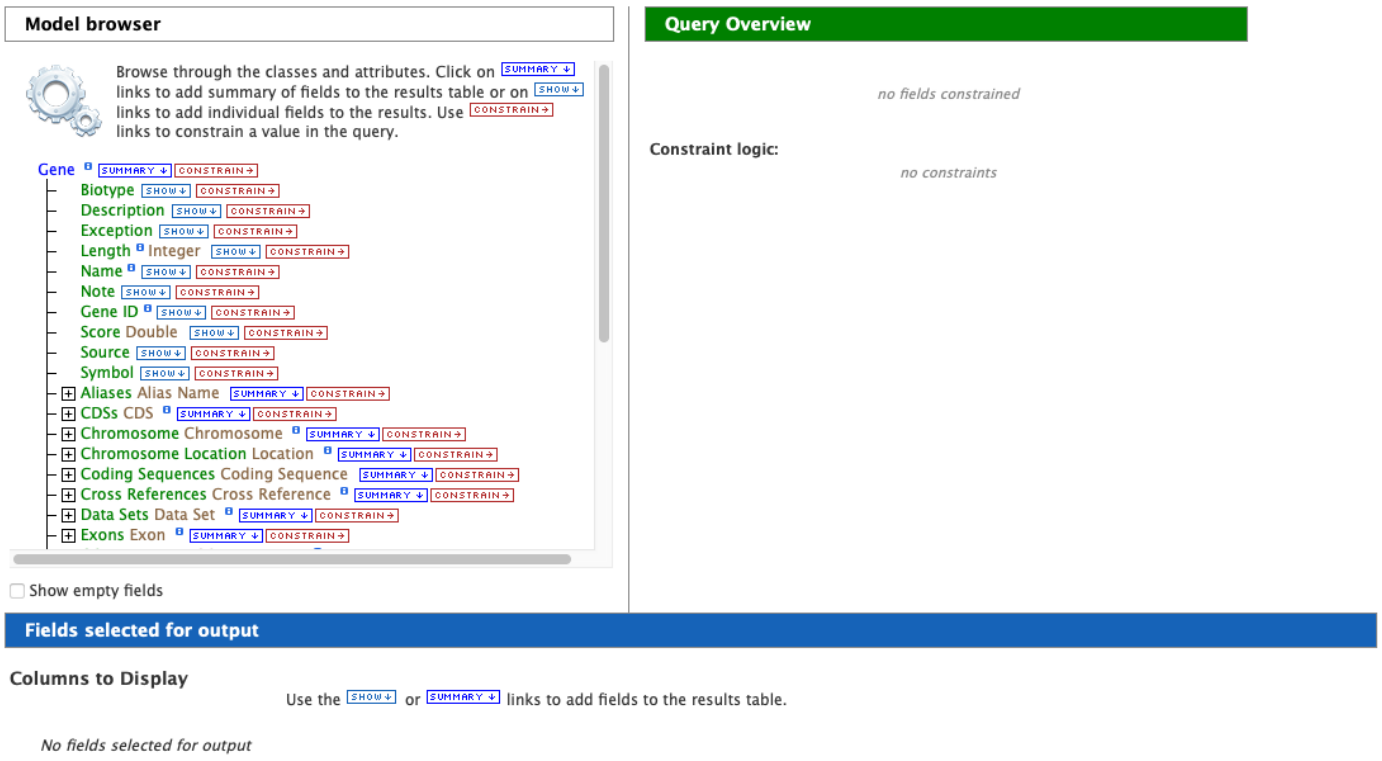


Fig. 13: Model browser with Gene selected as data type

Using the model browser, fields and constraints may be added to the query. Clicking **Show** to the right of an attribute will add that field to the query. Clicking **Constrain** brings up a window with filter options for the attribute selected. The **Query Overview** summarizes the current state of the query; it displays the currently selected fields and constraint logic. The results columns are displayed at the bottom of the page, where they may be rearranged or removed.

2.3.2 Examples

The following examples give a more in-depth demonstration on how to use the QueryBuilder. All examples use Gene as the selected data type.

2.3.2.1 Example 1: Querying for protein coding genes

In the Model browser, click **Show** next to **Gene ID** and **Symbol**, which will add these fields to the query. Notice that these two fields appear below Gene in the **Query Overview** section and at the bottom under **Fields selected for output**.

Then click **Constrain** next to **Biotype**. The first drop-down menu defaults to = (equals sign). In the second drop-down menu, select **protein_coding**, then click the **Add to query** button. This adds a constraint to the query to search only for protein coding genes.

Notice that the Query Overview section now shows "Biotype = protein_coding". Also, two types of icons appear next to the attributes. Clicking on the red "X" icon next to an attribute will remove that field or constraint from the query. Clicking on the blue pencil icon next to a constraint brings up the constraint editing window from earlier where changes may be made to the query filters.

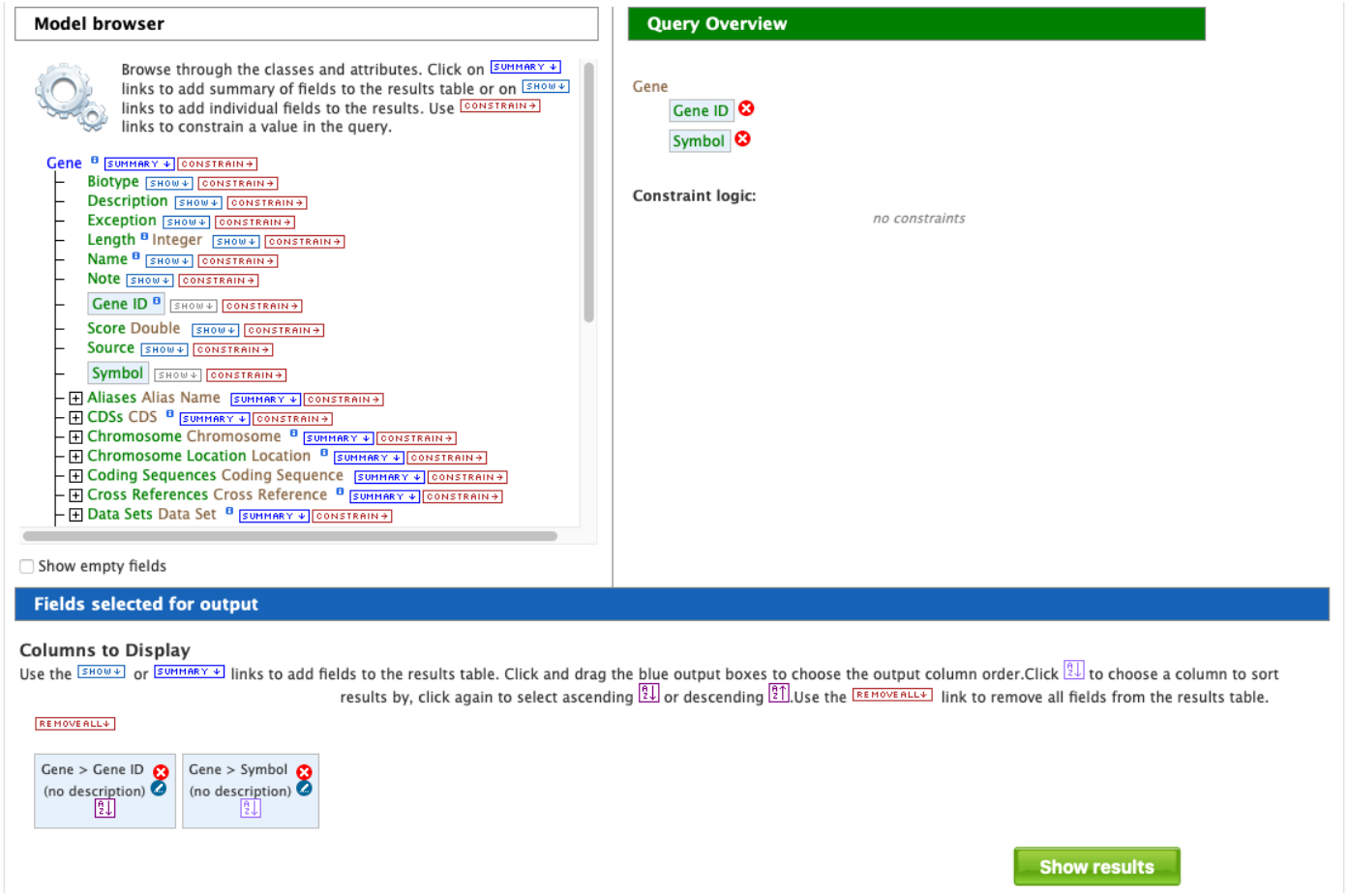


Fig. 14: Step 1: Select fields to be added to the query

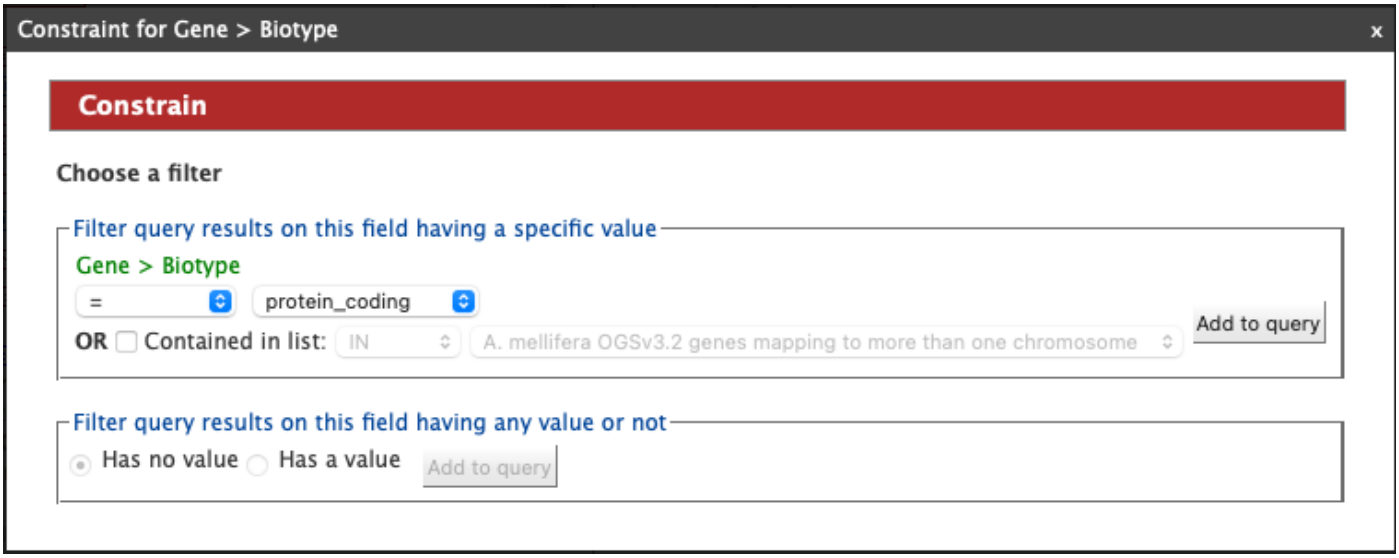


Fig. 15: Step 2: Add a constraint to the query on Biotype

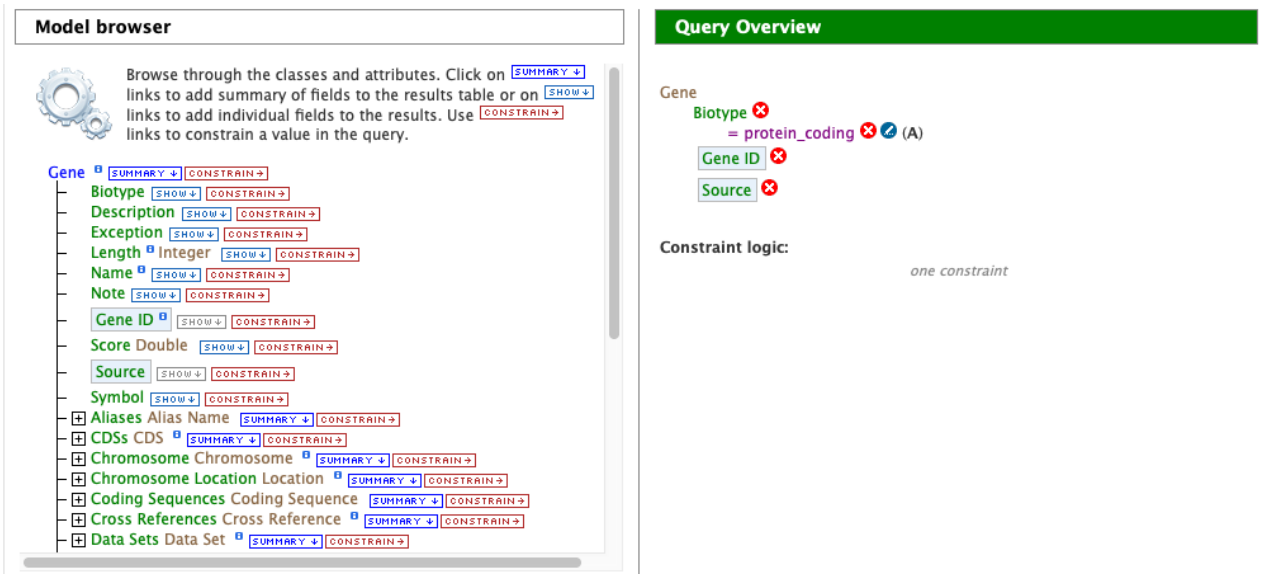
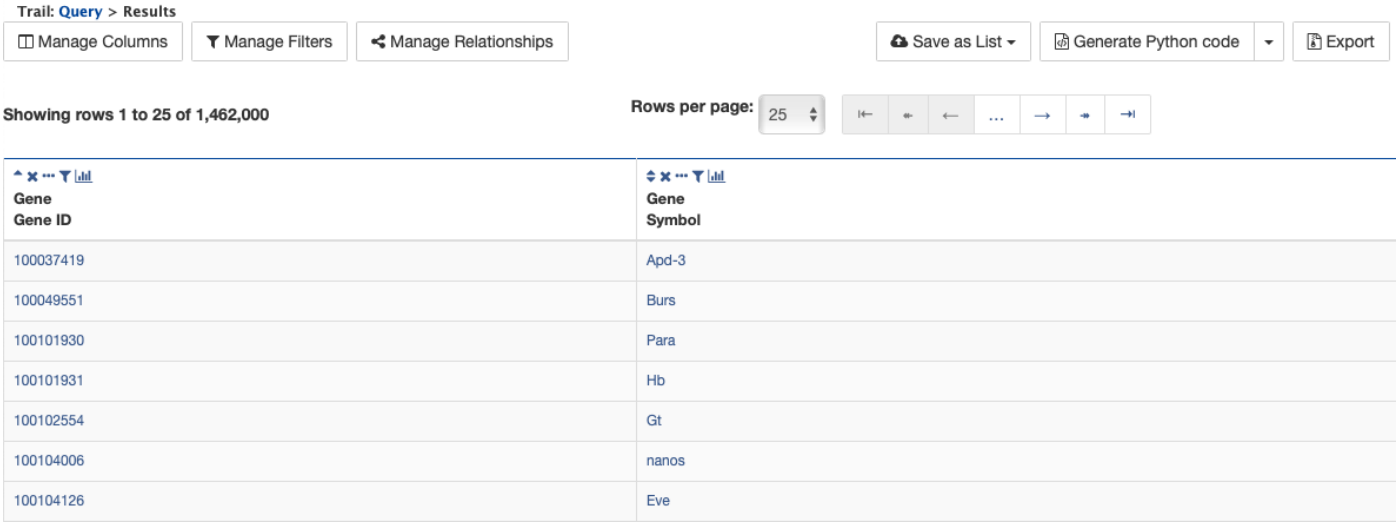


Fig. 16: Constraint added to query

Finally, click on the **Show results** button above the Model browser. The resulting table contains all protein coding genes in the database, with Gene ID and Gene Symbol as the two table columns.



Gene Gene ID	Gene Symbol
100037419	Apd-3
100049551	Burs
100101930	Para
100101931	Hb
100102554	Gt
100104006	nanos
100104126	Eve

Fig. 17: Step 3: Display query results

2.3.2.2 Example 2: Querying for protein coding genes on a particular chromosome

This example will extend the first example to add another constraint to the query.

After running the query in the above example, click on **Query** at the top of the page next to **Trail** to go back to the model browser and edit the query.

In the Model browser, click the + (plus sign) next to the Chromosome feature class to display its attributes.

Then click **Constrain** next to the attribute **Chromosome ID**. In the pop-up window, enter **NW_003791143.1** into the

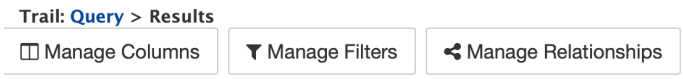


Fig. 18: Click on the query trail to edit the query

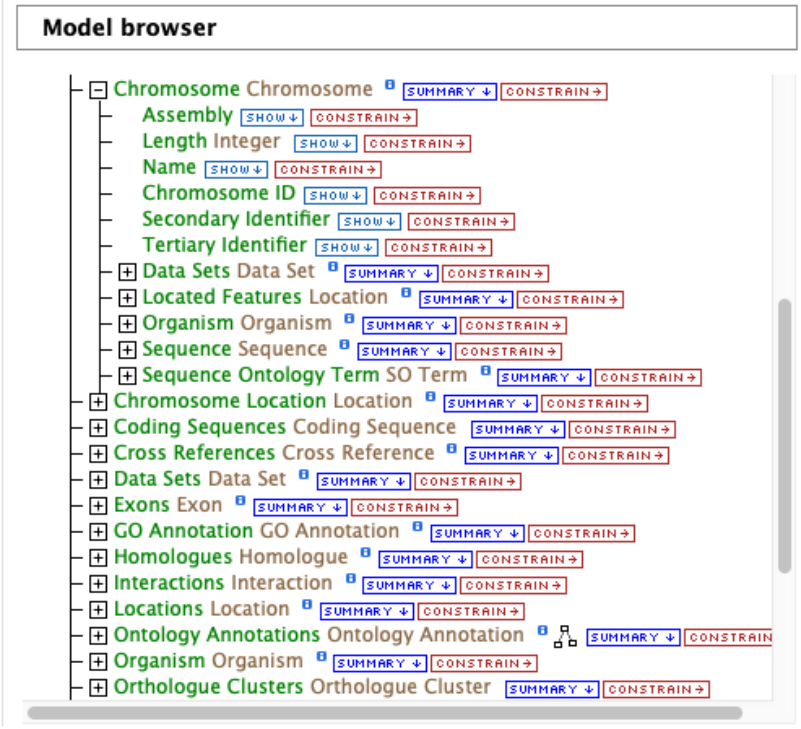


Fig. 19: Step 1: View attributes of the Chromosome feature class

text field, and click **Add to query**. This adds an additional constraint to the query that searches for protein coding genes on a chromosome with ID *NW_003791143.1*.

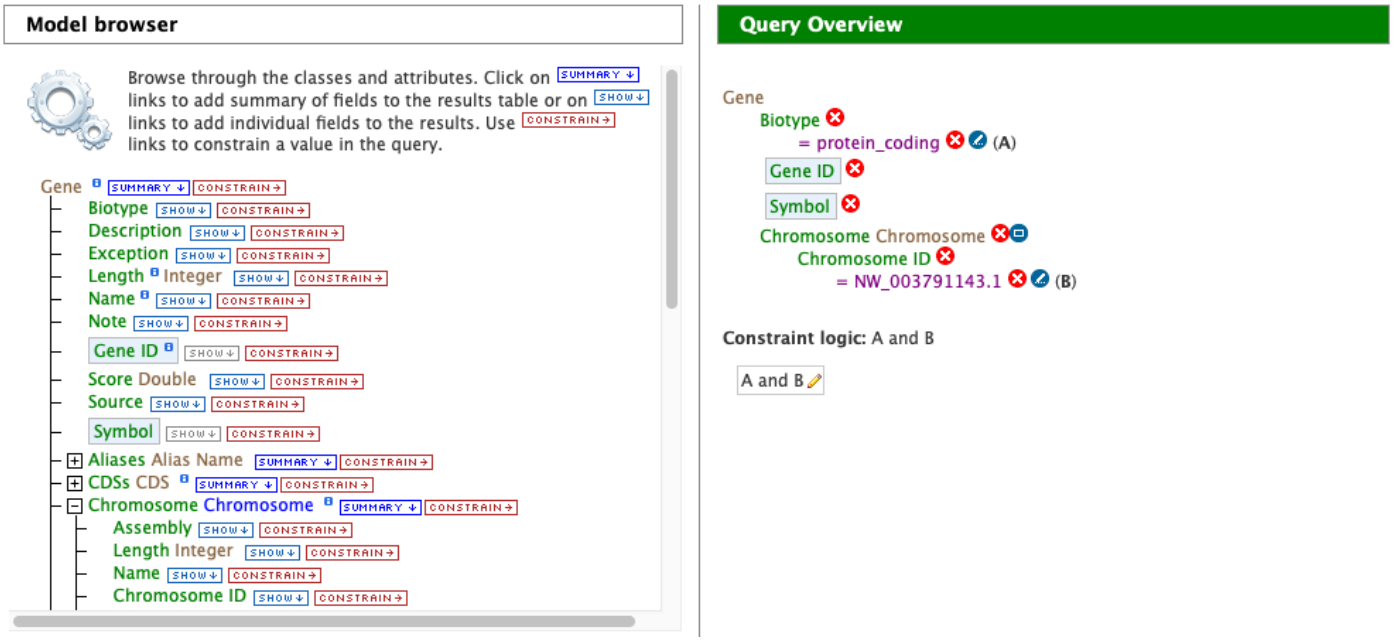
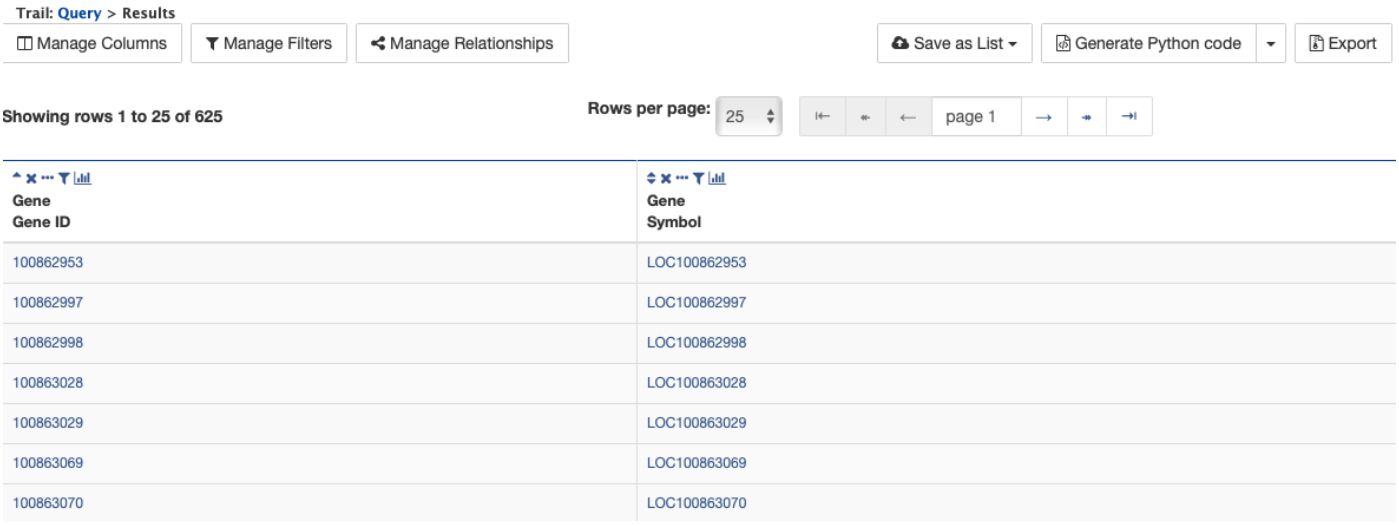


Fig. 20: Step 2: Add a constraint to the Chromosome ID

Click the **Show results** button as before to view the results of the query. The columns are the same as in the first example, but notice that now there are only 625 rows in the table (compared to over 1,400,000 in the first query) due to the additional constraint.



Gene Gene ID	Gene Symbol
100862953	LOC100862953
100862997	LOC100862997
100862998	LOC100862998
100863028	LOC100863028
100863029	LOC100863029
100863069	LOC100863069
100863070	LOC100863070

Fig. 21: Step 3: Display query results

2.3.2.3 Example 3: Querying for protein coding genes on a particular chromosome and their exons

This final example extends the above query to display all exons for each protein coding gene.

As above, click on **Query** at the top of the results page to go back to the model browser and edit the query.

In the Model browser, scroll down to locate the **Exon** feature class, and click the + (plus sign) next to **Exons** to display its attributes. Click **Show** next to **DB Identifier** and **Length**.

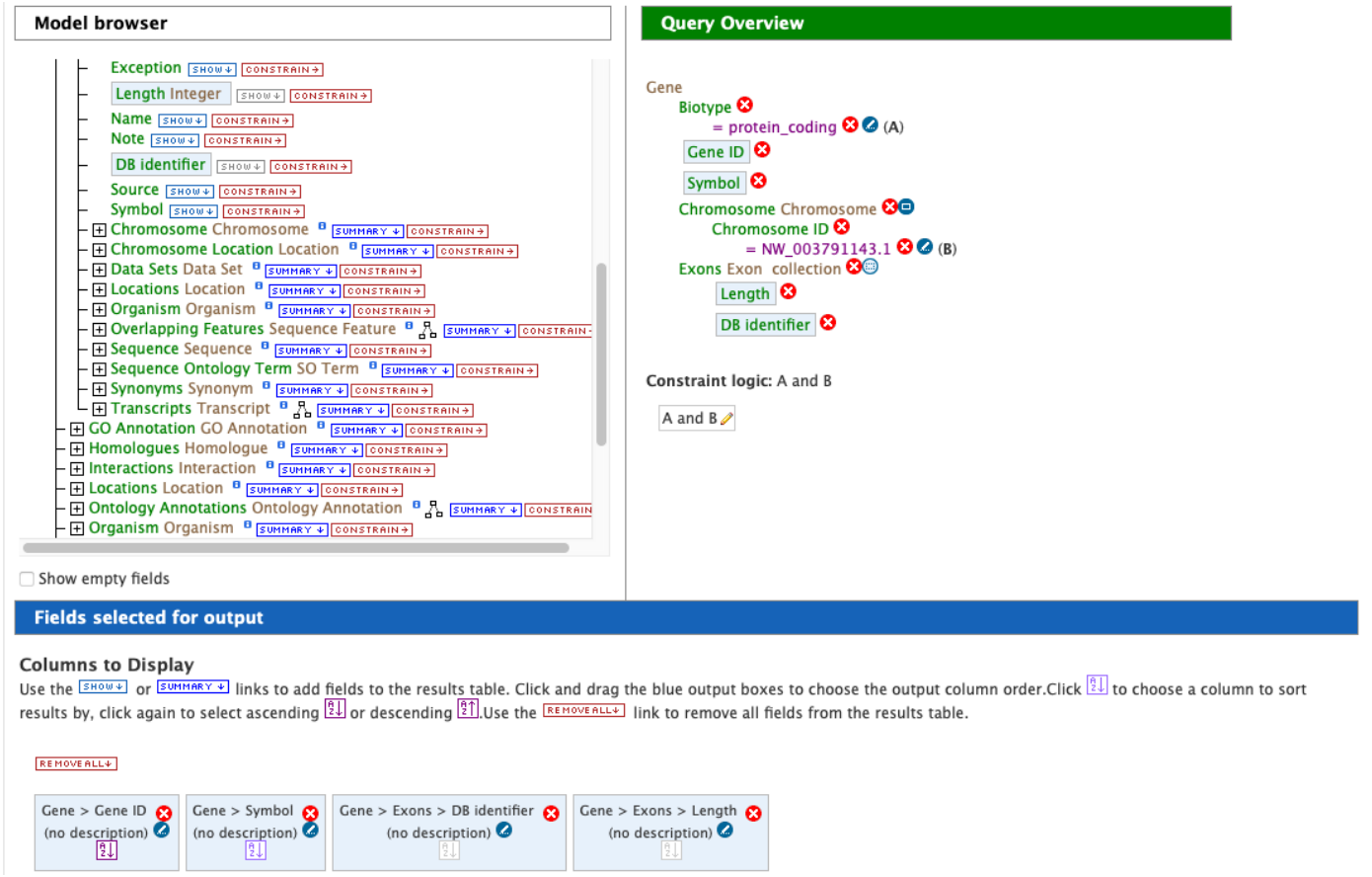


Fig. 22: Step 1: Select Exon fields to be added to the query

The **Query Overview** shows the query in progress, with four fields and two constraints. The additional exon fields have been added to the **Fields selected for output** section. Also notice that a third type of icon, a blue square, appears next to a couple of the attributes. Clicking on a blue square icon next to an attribute brings up a window where the query **Join Style** may be modified. Click on the blue square icon next to **Exon collection** to bring up the Switch Join Style window. The default option is **Show only Genes if they have a Exon**. Change this to **Show all Genes and show Exons if they are present** and click **Add to query**.

Click **Show results** to run the query.

Notice that the results table contains the same rows as in the second example, but now there is a new column, **Gene Exons**. For example, looking at the second row, the Gene with ID 100862997 has 12 exons. Click on the **12 exons** text to expand the table with additional rows containing the DB identifier and Length for each of the 12 exons.

By changing the join style, the exons have been grouped together by gene, making it easier to see how many exons each gene has. By contrast, if the query is run with the default option of "Show only Genes if they have a Exon", the results table adds a new row for each exon.

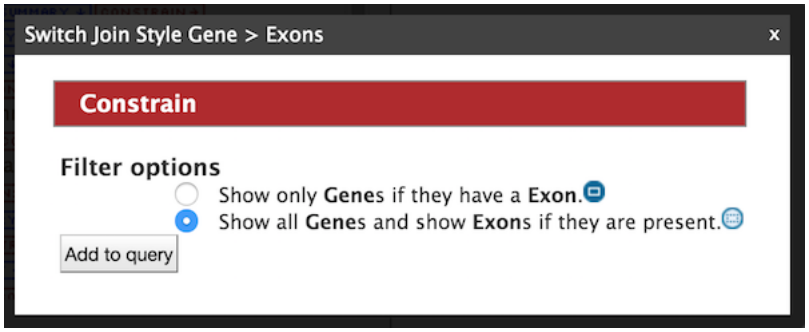


Fig. 23: Step 2: Change the join style

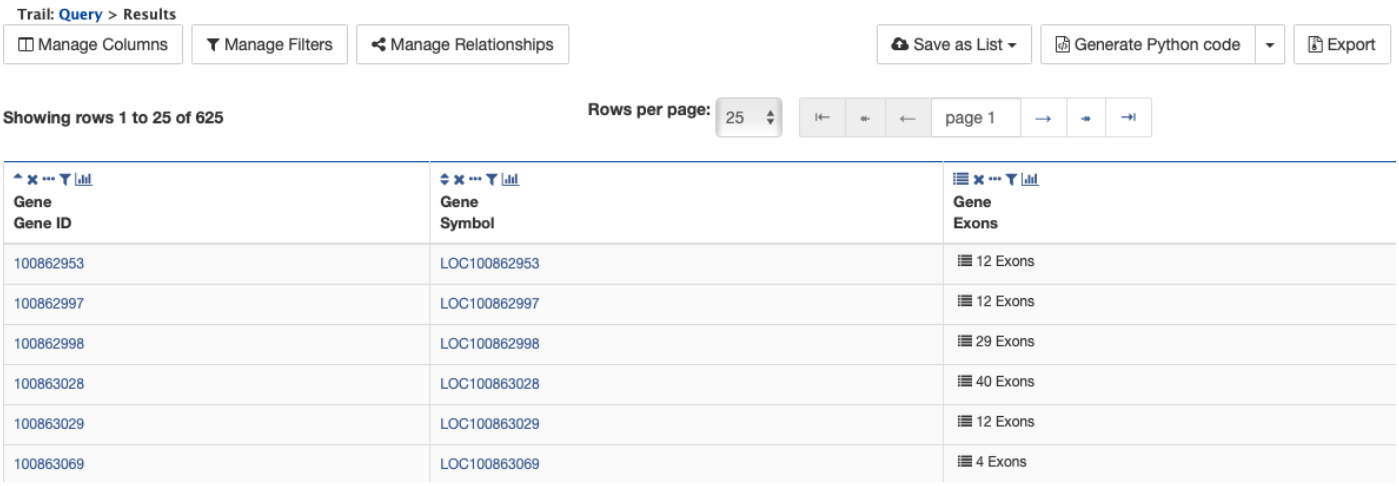


Fig. 24: Step 3: Display query results

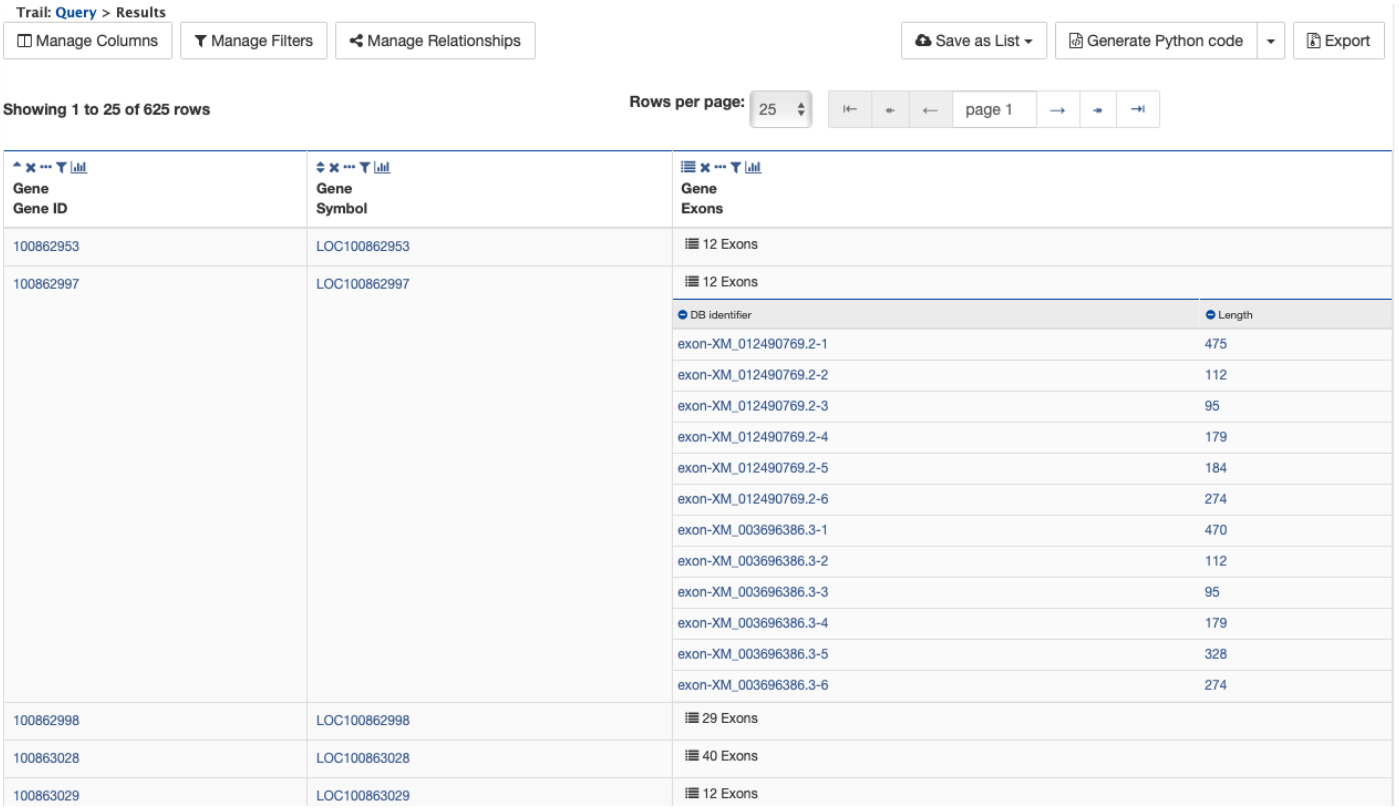


Fig. 25: Step 4: Expand information on exons

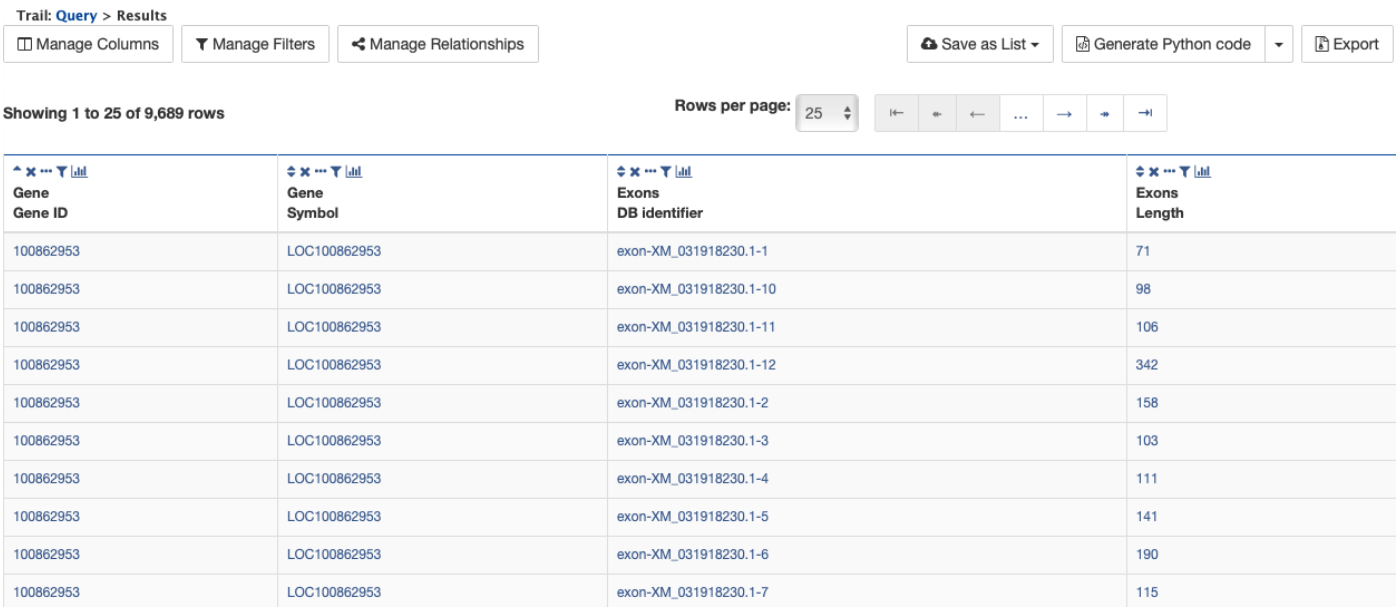


Fig. 26: Same query with default join style for exons

2.4 Genomic Regions Search

The **Genomic Regions Search** is a tool to fetch features that are within a given set of genomic coordinates or are within a given number of bases flanking the coordinates.

To begin this type of search, click the **Regions** tab on the menu. A form will appear asking for the search parameters (organism, feature types, genomic coordinates, etc.)

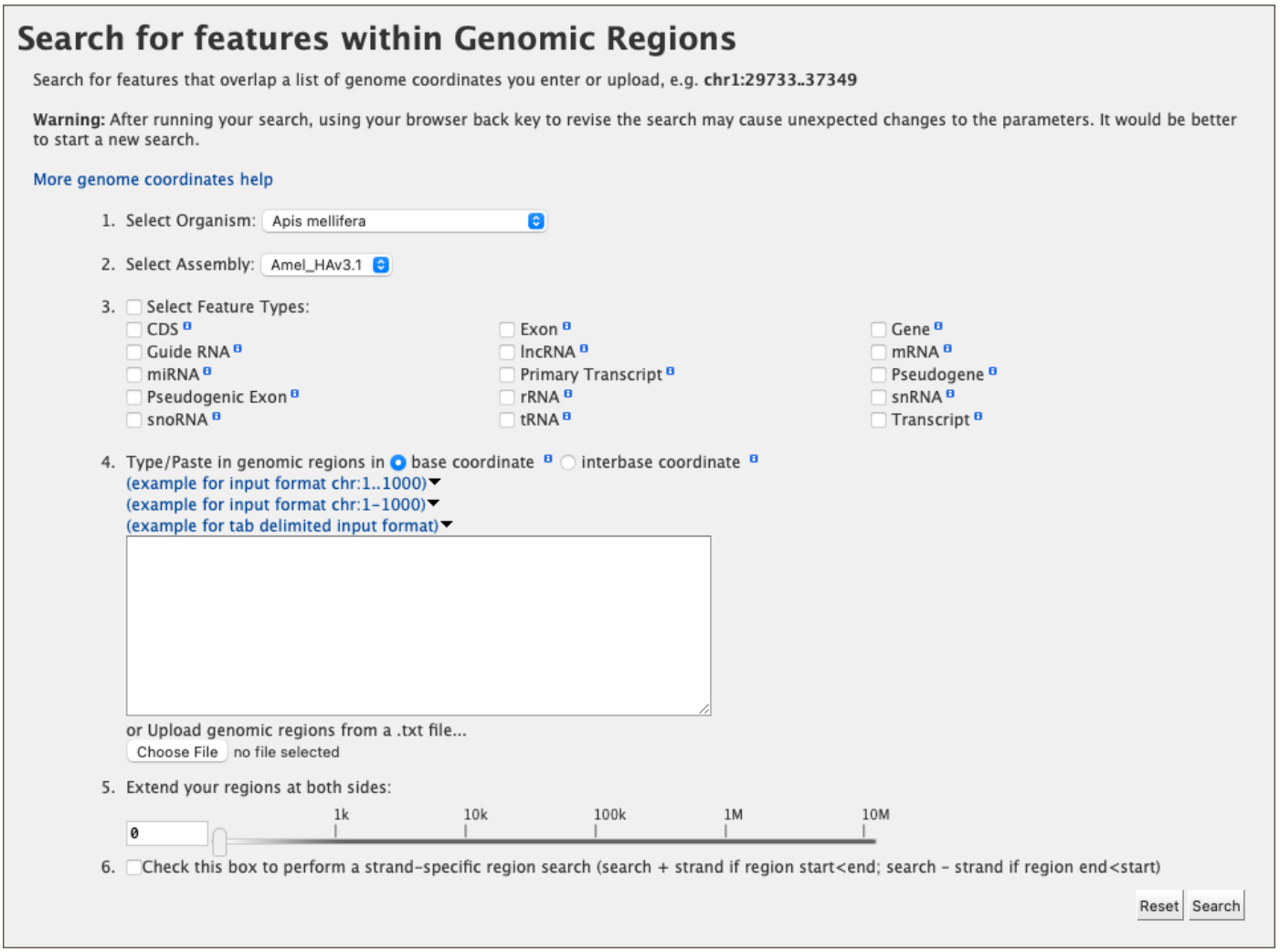


Fig. 27: Genomic regions search form

The coordinates must have one of three formats:

1. chromosome_name:start..end
2. chromosome_name:start-end
3. chromosome_name start end (tab delimited)

Click on the input examples to view a representative set of coordinates in each format. Click the **Genome coordinates help** text near the top of the form for more detailed information on the input format requirements.

The regions may be extended on either side of the genomic coordinates using the slider or the text field next to it. There is also the option to perform a strand-specific region search using the checkbox at the bottom of the form.

As an example, select **Apis mellifera** (default option) from the **Select Organism** drop-down. In this HymenopteraMine release, there is only one assembly per organism, which is selected by default in the **Select Assembly** drop-down. (If multiple assemblies were available, this drop-down could be used to further filter the results by assembly.) Next, check the box next to **Gene** in the **Select Feature Types** section, and enter the following coordinates into the genomic regions search text field:

```
LG5:900000..930000
```

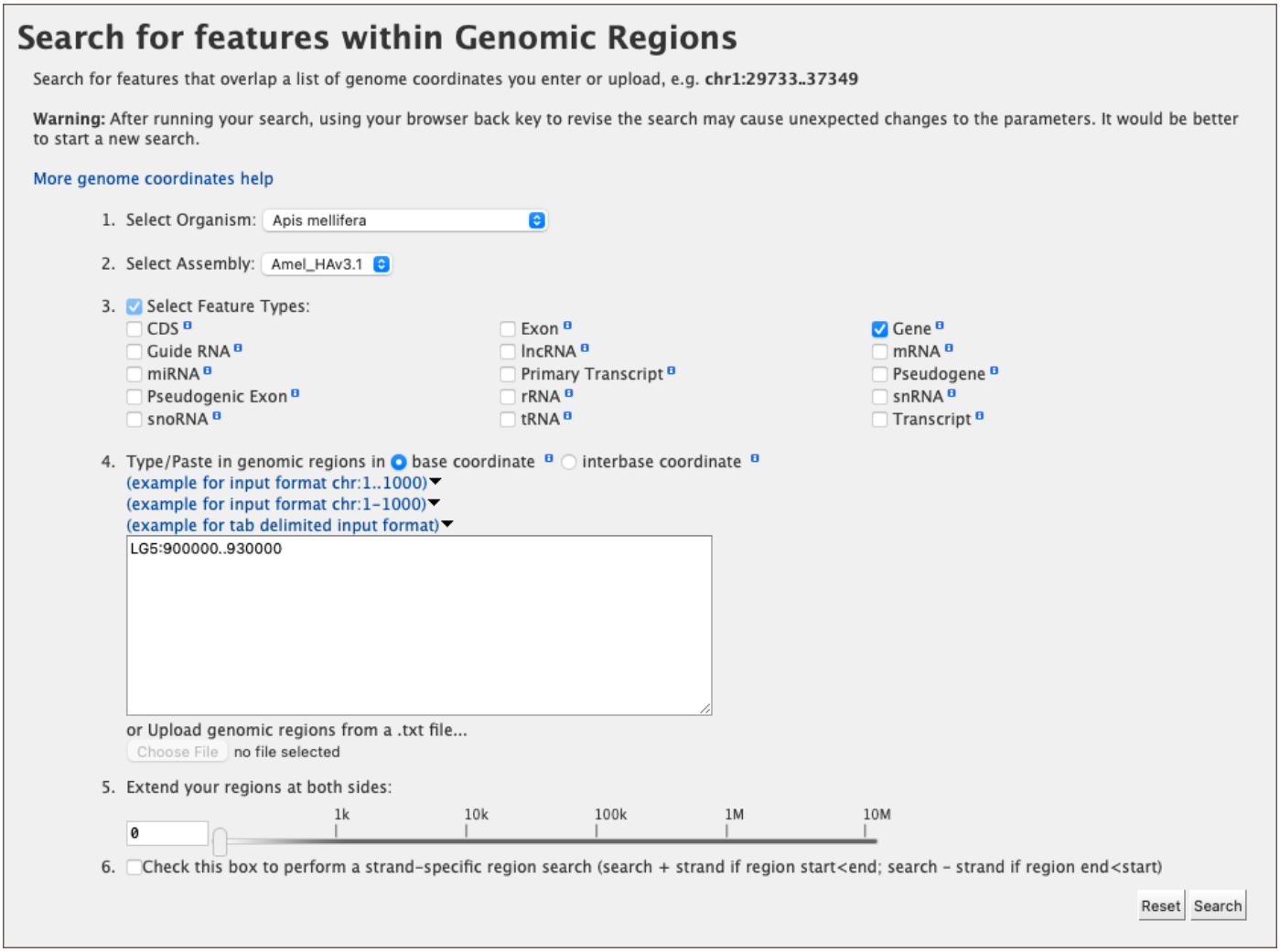


Fig. 28: Example: Genomic regions search on *Apis mellifera*

Then click **Search**. The search results page presents a list of features present within the genomic interval that was entered. In this case, the feature type was limited to Gene. The results may be exported as tab-separated or comma-separated values. If they contain genomic features, there is also the option to saved the results in GFF3 or BED format. The FASTA sequences of the features may also be downloaded.

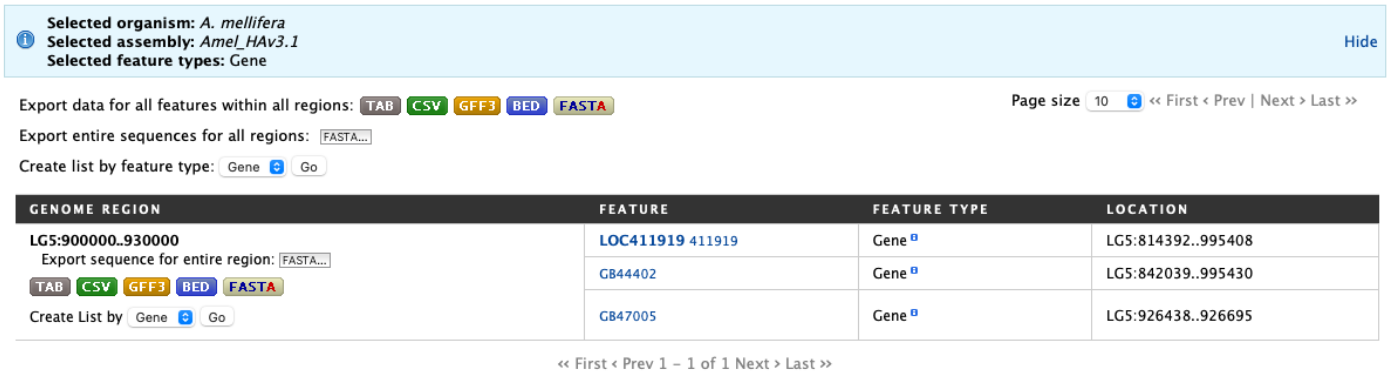


Fig. 29: Example: Genomic regions search results

CHAPTER 3

Report Pages

Every object (e.g., Gene, Protein, Exon) in HymenopteraMine has a report page. The layout of the report page depends on the data available for the object. Report pages may be accessed by clicking on an object name in the results table after running a query.

As an example, on the home page of HymenopteraMine, click on the **Proteins** tab in the **Popular Templates** section. (Refer to the *Templates* section for more details on using templates to search the database.) Click on the **Gene Symbol Proteins** template. In the text field, enter **LCCH3**, and select **A. mellifera** as the Organism. Then click **Show Results**.

Notice that each item in the results table is a hyperlink. Hover over an item to bring up a quick summary window for that item. For example, hover over **LCCH3** to view a summary of the gene with this symbol. The summary contains the gene's biotype, database identifier, description, length, organism, symbol, and source. Similarly, hover over **Q0GQR5_APIME** to view a summary of the protein with this DB identifier.

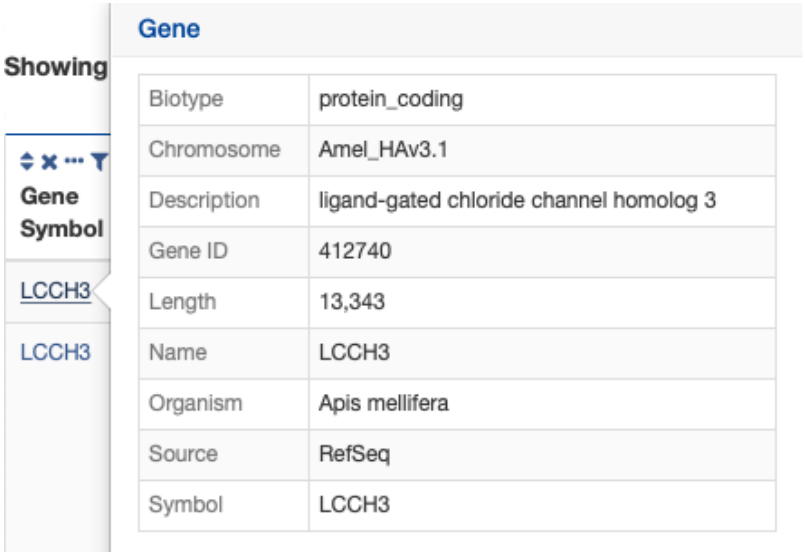


Fig. 1: Example: Summary window for LCCH3

Clicking on an item in the table rather than just hovering over it will bring up its report page. For example, click on LCCH3 in the Gene Symbol column to view its report.

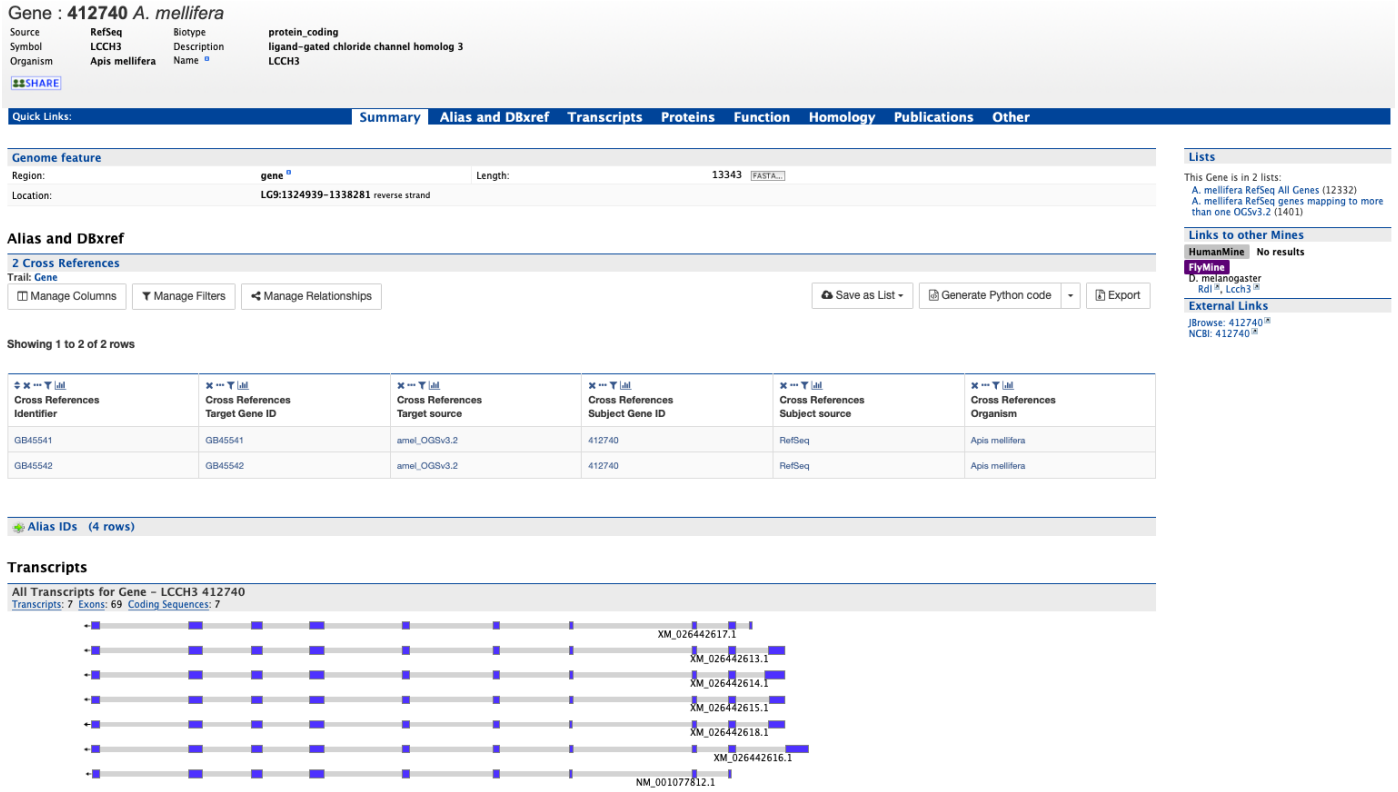


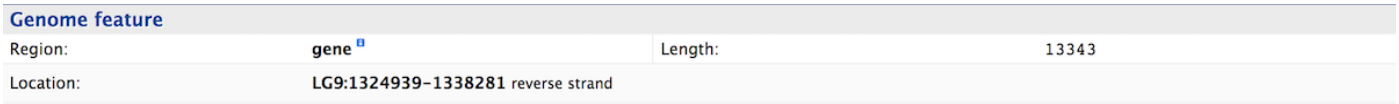
Fig. 2: Report page

The report page provides a complete description for this gene. The header displays the database identifier, followed by the information from the summary window for the gene (organism, symbol, source, etc.) **Biotype** indicates the type of gene; in this case the type is protein coding.

The contents of the report page are divided into categories based on the type of information provided.

3.1 Summary

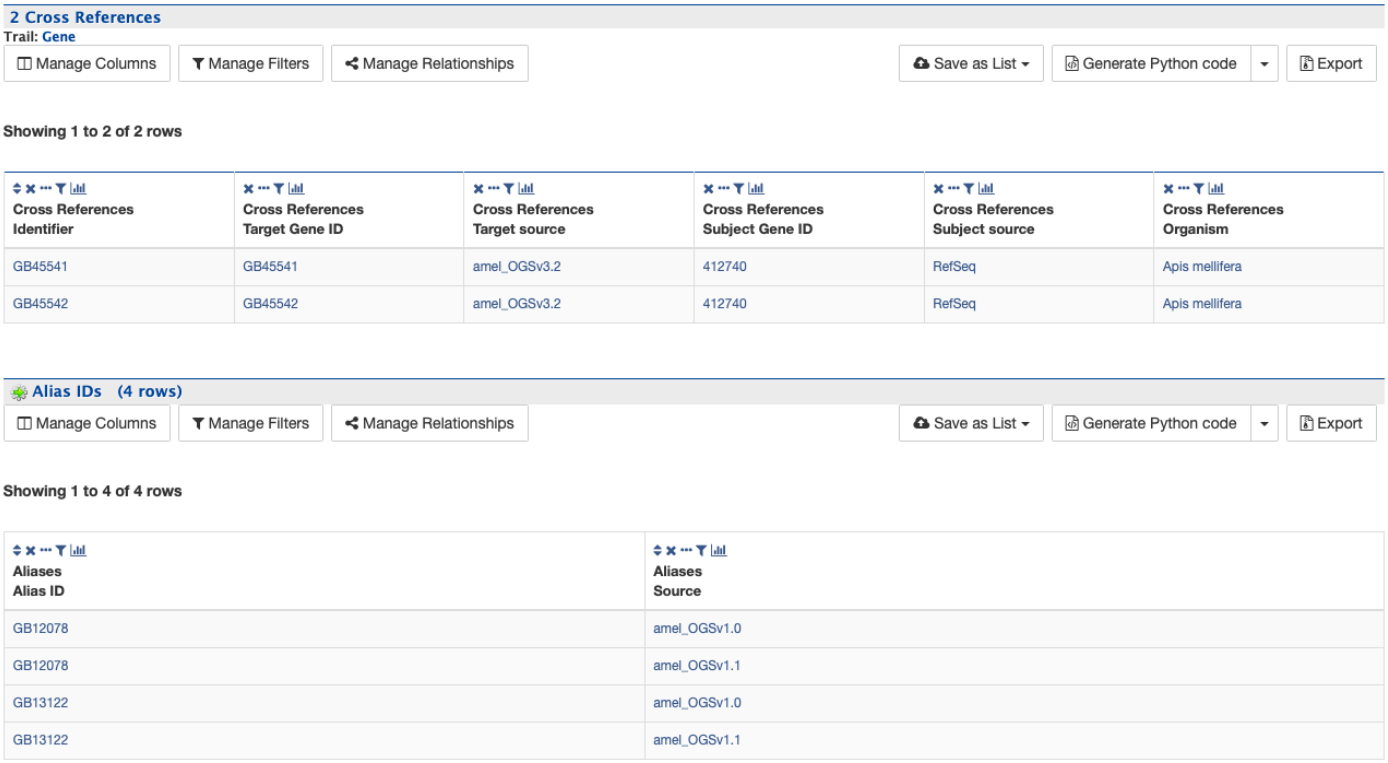
A **Summary** section near the top of the report provides information on the gene such as its length, chromosome location, and strand information.



3.2 Alias and DBxref

The **Alias and DBxref** section displays tables of aliases and database cross references for the gene. In this example, the gene with RefSeq ID 412740 has four aliases and two cross references.

Alias and DBxref



2 Cross References
Trail: Gene

Manage Columns | Manage Filters | Manage Relationships | Save as List | Generate Python code | Export

Showing 1 to 2 of 2 rows

Cross References Identifier	Cross References Target Gene ID	Cross References Target source	Cross References Subject Gene ID	Cross References Subject source	Cross References Organism
GB45541	GB45541	amel_OGSv3.2	412740	RefSeq	Apis mellifera
GB45542	GB45542	amel_OGSv3.2	412740	RefSeq	Apis mellifera

Alias IDs (4 rows)

Manage Columns | Manage Filters | Manage Relationships | Save as List | Generate Python code | Export

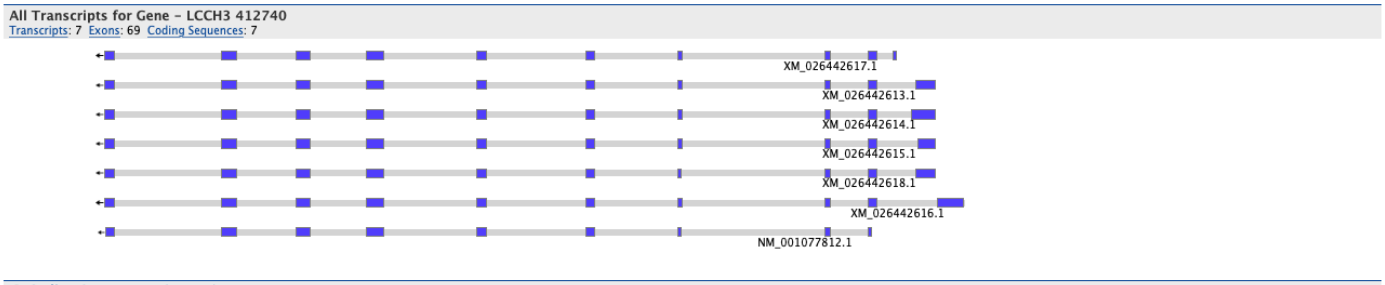
Showing 1 to 4 of 4 rows

Aliases Alias ID	Aliases Source
GB12078	amel_OGSv1.0
GB12078	amel_OGSv1.1
GB13122	amel_OGSv1.0
GB13122	amel_OGSv1.1

3.3 Transcripts

The **Transcripts** section contains information about the gene model, such as transcripts, exons, etc. It includes a diagram visually representing each transcript with its features highlighted (if applicable). In the case of protein coding genes, a table with protein information is also provided.

Transcripts



Coding Sequences (7 rows)

Manage Columns | Manage Filters | Manage Relationships | Save as List | Generate Python code | Export

Showing 1 to 7 of 7 rows

Coding Sequences DB identifier	Coding Sequences Protein Identifier	Coding Sequences Length	Sequence Residues
NM_001077812.1-CDS	NP_001071280.1	1470	ATGCATCACAGGATGTGGTTGCAGCAGATCTTTATCCTGCTGCAGATGATTCATCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCTGATAGATTGGAA AATGTGACGCAAACAATATCACGAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTTCGGCGGAGAACCCCTGTTAGTCGGGATGGATCT...
XM_026442613.1-CDS	XP_026298398.1	1485	ATGCATCACAGGATGTGGTTGCAGCAGATCTTTATCCTGCTGCAGATGATTCATCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCTGATAGATTGGAA AATGTGACGCAAACAATATCACGAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTTCGGCGTTGTTTCATTTCCAGGAGAACCCCTGTT...
XM_026442614.1-CDS	XP_026298399.1	1485	ATGCATCACAGGATGTGGTTGCAGCAGATCTTTATCCTGCTGCAGATGATTCATCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCTGATAGATTGGAA AATGTGACGCAAACAATATCACGAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTTCGGCGTTGTTTCATTTCCAGGAGAACCCCTGTT...
XM_026442615.1-CDS	XP_026298400.1	1485	ATGCATCACAGGATGTGGTTGCAGCAGATCTTTATCCTGCTGCAGATGATTCATCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCTGATAGATTGGAA AATGTGACGCAAACAATATCACGAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTTCGGCGTTGTTTCATTTCCAGGAGAACCCCTGTT...
XM_026442616.1-CDS	XP_026298401.1	1485	ATGCATCACAGGATGTGGTTGCAGCAGATCTTTATCCTGCTGCAGATGATTCATCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCTGATAGATTGGAA AATGTGACGCAAACAATATCACGAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTTCGGCGTTGTTTCATTTCCAGGAGAACCCCTGTT...
XM_026442617.1-CDS	XP_026298402.1	1485	ATGCATCACAGGATGTGGTTGCAGCAGATCTTTATCCTGCTGCAGATGATTCATCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCTGATAGATTGGAA AATGTGACGCAAACAATATCACGAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTTCGGCGTTGTTTCATTTCCAGGAGAACCCCTGTT...
XM_026442618.1-CDS	XP_026298403.1	1470	ATGCATCACAGGATGTGGTTGCAGCAGATCTTTATCCTGCTGCAGATGATTCATCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCTGATAGATTGGAA AATGTGACGCAAACAATATCACGAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTTCGGCGGAGAACCCCTGTTAGTCGGGATGGATCT

Protein Sequences (7 rows)

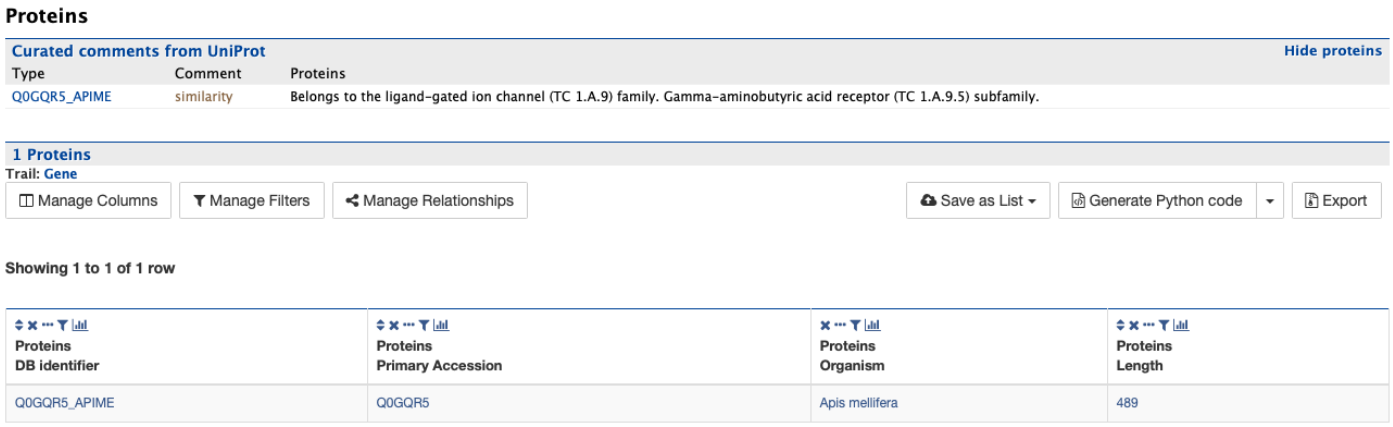
Manage Columns | Manage Filters | Manage Relationships | Save as List | Generate Python code | Export

Showing 1 to 7 of 7 rows

Polypeptides DB identifier	Polypeptides mRNA ID	Polypeptides Length	Sequence Residues
NP_001071280.1	NM_001077812.1	489	MHHRMWLQQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILDGYDIRLRPNFGGEPLLVGMDLTIASFDAISEVNMDYTITMYLNQYWKDERLAFSQEEEVLTLSGDFAEKIWVPDTF FANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPLDSQNCTVEIESYGYTVLDVVMYWKETPVRG...
XP_026298398.1	XM_026442613.1	494	MHHRMWLQQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILDGYDIRLRPNFGVVSFPGEPLLVGMDLTIASFDAISEVNMDYTITMYLNQYWKDERLAFSQEEEVLTLSGDFAEKIW VPDTFFANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPLDSQNCTVEIESYGYTVLDVVMYWKE...
XP_026298399.1	XM_026442614.1	494	MHHRMWLQQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILDGYDIRLRPNFGVVSFPGEPLLVGMDLTIASFDAISEVNMDYTITMYLNQYWKDERLAFSQEEEVLTLSGDFAEKIW VPDTFFANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPLDSQNCTVEIESYGYTVLDVVMYWKE...
XP_026298400.1	XM_026442615.1	494	MHHRMWLQQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILDGYDIRLRPNFGVVSFPGEPLLVGMDLTIASFDAISEVNMDYTITMYLNQYWKDERLAFSQEEEVLTLSGDFAEKIW VPDTFFANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPLDSQNCTVEIESYGYTVLDVVMYWKE...
XP_026298401.1	XM_026442616.1	494	MHHRMWLQQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILDGYDIRLRPNFGVVSFPGEPLLVGMDLTIASFDAISEVNMDYTITMYLNQYWKDERLAFSQEEEVLTLSGDFAEKIW VPDTFFANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPLDSQNCTVEIESYGYTVLDVVMYWKE...
XP_026298402.1	XM_026442617.1	494	MHHRMWLQQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILDGYDIRLRPNFGVVSFPGEPLLVGMDLTIASFDAISEVNMDYTITMYLNQYWKDERLAFSQEEEVLTLSGDFAEKIW VPDTFFANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPLDSQNCTVEIESYGYTVLDVVMYWKE...
XP_026298403.1	XM_026442618.1	489	MHHRMWLQQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILDGYDIRLRPNFGGEPLLVGMDLTIASFDAISEVNMDYTITMYLNQYWKDERLAFSQEEEVLTLSGDFAEKIWVPDTF FANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPLDSQNCTVEIESYGYTVLDVVMYWKETPVRG...

3.4 Proteins

The **Proteins** section provides information about the protein product of gene. The comments section gives a brief description about the protein along with the UniProt accession.

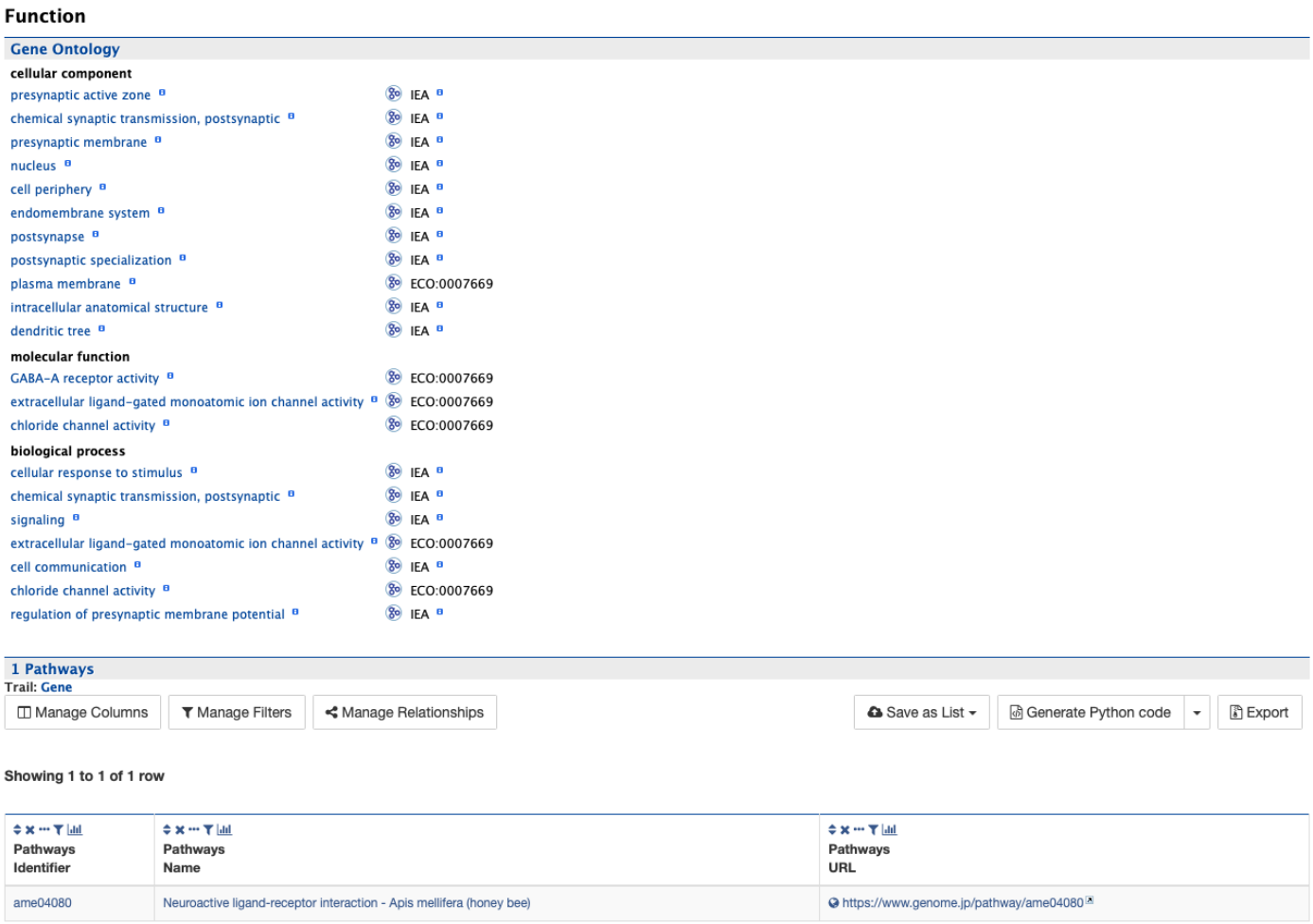


3.5 Function

The **Function** section displays Gene Ontology annotations for a gene. Annotations are divided into three categories:

- Cellular Component
- Molecular function
- Biological process

The GO terms are displayed along with the evidence code indicating how the annotations were derived. If applicable, a table of information on Pathways is also shown.



3.6 Homology

The **Homology** section includes information on homologues for the gene.

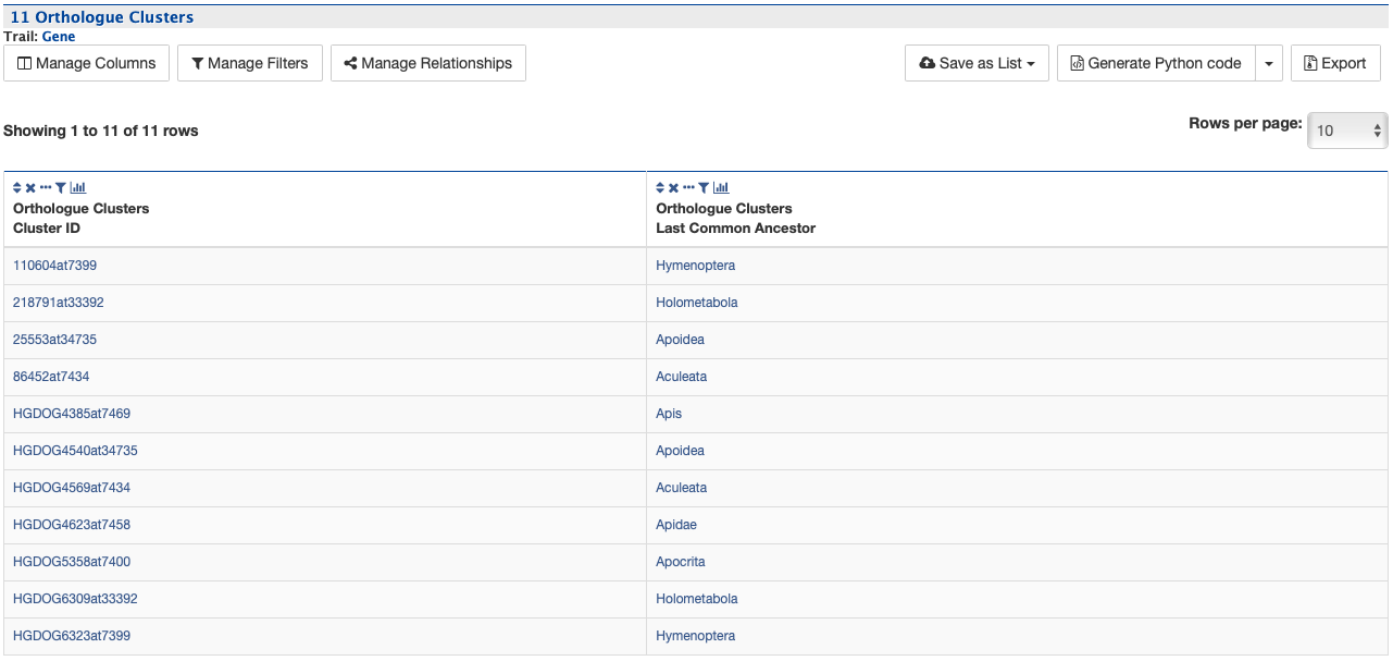
3.7 Publications

The **Publications** section displays a table of publications related to the gene.

Homology

Homologues	
A. cephalotes	LOC105618116 LOC105619338
A. cerana	LOC107999294 LOC107995555
A. colombica	LOC108686942 LOC108687631
A. dorsata	LOC102678684 ENSADOG00000008019 LOC102675465
A. echinatior	LOC105145387 LOC105149353
A. florea	LOC100869944 LOC100871112
A. gifuensis	LOC122856609 LOC122852242
A. haemorrhoa	ENSAHEG00005000570
A. laboriosa	LOC122716650 LOC122714438
A. nigerrimus	BRAKERANGG00000006389
A. nigricornis	ENSANRG00000003003
A. rosae	LOC105691302 LOC105691160
B. bifarius	LOC117211906 LOC117209730
B. breviceps	ENSBBEG00000014451
B. campestris	ENSBCMG00005020967
B. confusus	ENSBCFG00000009981
B. consobrinus	ENSBCSG00000011020
B. cullumanus	ENSBCUG00000014754
B. difficillimus	ENSBDIG00000008503

Fig. 3: Homology section of report page (homologues)



11 Orthologue Clusters

Trail: Gene

Manage Columns | Manage Filters | Manage Relationships | Save as List | Generate Python code | Export

Showing 1 to 11 of 11 rows

Rows per page: 10

Orthologue Clusters Cluster ID	Orthologue Clusters Last Common Ancestor
110604at7399	Hymenoptera
218791at33392	Holometabola
25553at34735	Apoidea
86452at7434	Aculeata
HGDOG4385at7469	Apis
HGDOG4540at34735	Apoidea
HGDOG4569at7434	Aculeata
HGDOG4623at7458	Apidae
HGDOG5358at7400	Apocrita
HGDOG6309at33392	Holometabola
HGDOG6323at7399	Hymenoptera

Fig. 4: Homology section of report page (orthologue clusters)

Publications

Gene ID --> Publication (5 rows)

Manage Columns | Manage Filters | Manage Relationships | Save as List | Generate Python code | Export

Showing 1 to 5 of 5 rows

Gene DB identifier	Publications PubMed ID	Publications First Author	Publications Title	Publications Journal	Publications Volume	Publications Year	Publications Pages
412740	24479613	Elsik Christine G	Finding the missing honey bee genes: lessons learned from a genome upgrade.	BMC Genomics	15	2014	86
412740	19906878	Dupuis Julien Pierre	Homomeric RDL and heteromeric RDL/LCCH3 GABA receptors in the honeybee antennal lobes: two candidates for inhibitory transmission in olfactory processing.	J Neurophysiol	103	2010	458-68
412740	16902773	Jones Andrew K	The cys-loop ligand-gated ion channel superfamily of the honeybee, Apis mellifera.	Invert Neurosci	6	2006	123-32
412740	17065614	Kaplan Noam	ProtoBee: hierarchical classification and annotation of the honey bee proteome.	Genome Res	16	2006	1431-8
412740	17073008	Honeybee Genome Sequencing Consortium	Insights into social insects from the genome of the honeybee Apis mellifera.	Nature	443	2006	931-49

3.8 Other

This last section provides miscellaneous information that doesn't fit into any of the above categories, e.g., data sets including a gene, protein domain regions for a protein, etc.

CHAPTER 4

Lists

4.1 Creating Lists

Users may create and save lists of features, such as gene IDs, transcript IDs, gene symbols, etc. The list tool searches the database for the list items and attempts to convert each identifier to the selected type. Click on the **Lists** tab from the menu to access the full list upload form. A short version of the form is also in the **Quick List** box on the home page.

As an example, enter the following identifiers (comma-separated):

```
GB41586, Sec61Beta, TRAM, Mocs1, mal
```

Leave the **Select Type** as "Gene" and **Organism** drop-down as "Any". Then click **Create List**. A Summary table is displayed with the results of searching for each of the five identifiers in the list.

Next, click **Save a list of 5 Genes**. A **List Analysis** page is presented that contains widgets allowing users to perform analyses on the genes in the list.

The available widgets are Gene Ontology Enrichment, Publication Enrichment, and Pathway Enrichment. Read the **Important Notes for Enrichment Widgets** for special instructions to avoid false positives.

4.2 Saving Lists

Saved lists appear under the **View** tab on the Lists page. For users who are not logged in, lists are saved temporarily; users must log in to save lists permanently. Saved lists may also be accessed from the **MyMine** menu item.

Predefined lists of all genes from different species are also available on the Lists page for all users.

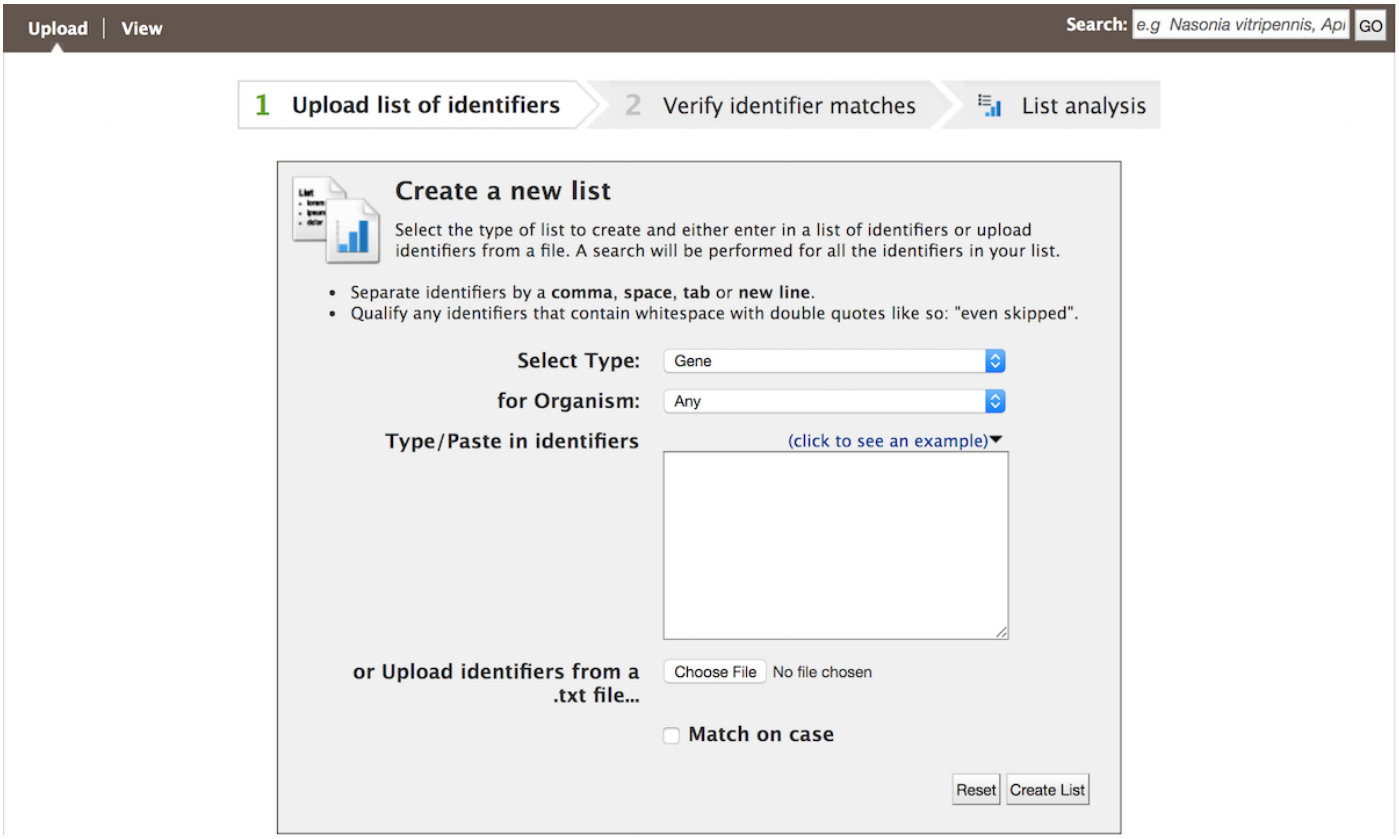
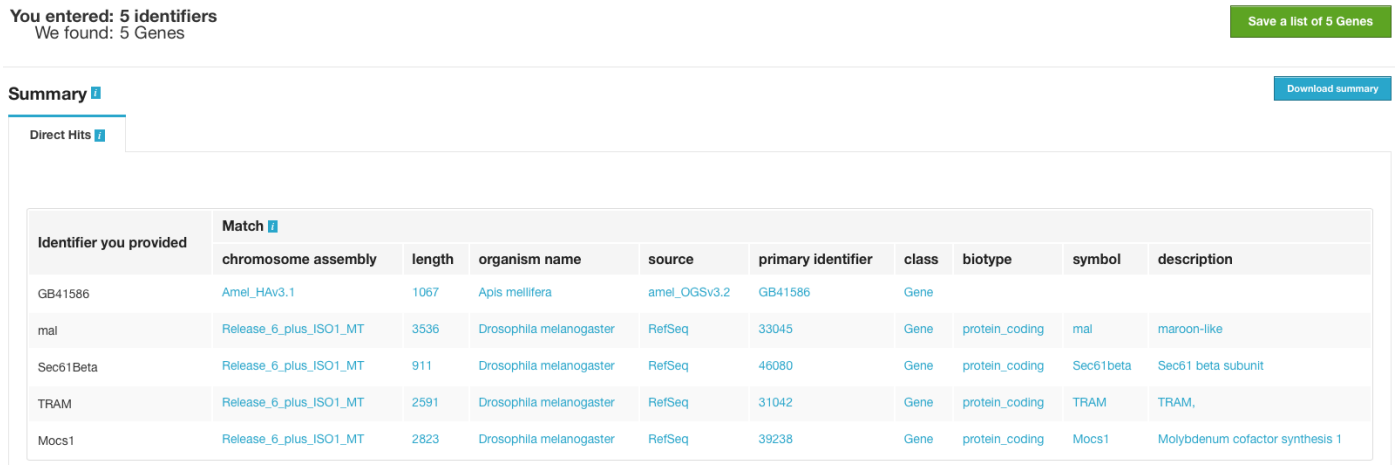


Fig. 1: List upload form



Identifier you provided	Match								
	chromosome assembly	length	organism name	source	primary identifier	class	biotype	symbol	description
GB41586	Amel_HAv3.1	1067	Apis mellifera	amel_OGSv3.2	GB41586	Gene			
mal	Release_6_plus_ISO1_MT	3536	Drosophila melanogaster	RefSeq	33045	Gene	protein_coding	mal	maroon-like
Sec61Beta	Release_6_plus_ISO1_MT	911	Drosophila melanogaster	RefSeq	46080	Gene	protein_coding	Sec61beta	Sec61 beta subunit
TRAM	Release_6_plus_ISO1_MT	2591	Drosophila melanogaster	RefSeq	31042	Gene	protein_coding	TRAM	TRAM,
Mocs1	Release_6_plus_ISO1_MT	2823	Drosophila melanogaster	RefSeq	39238	Gene	protein_coding	Mocs1	Molybdenum cofactor synthesis 1

Fig. 2: Example: Search results for list of five identifiers

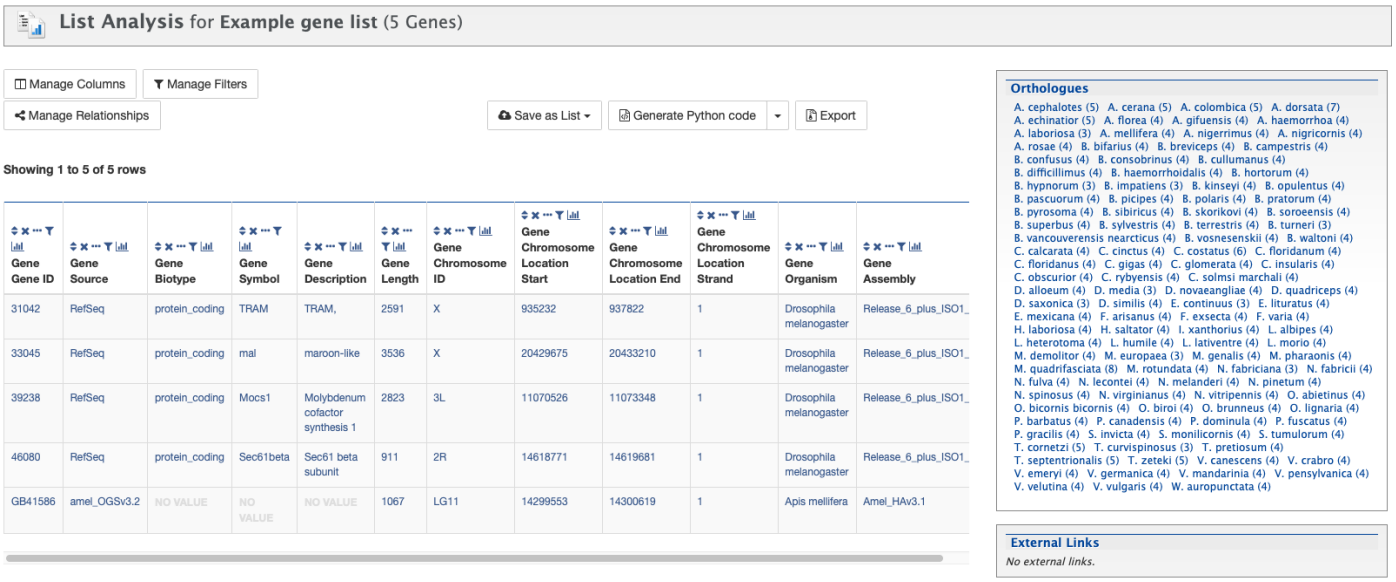


Fig. 3: Example: List analysis for gene list

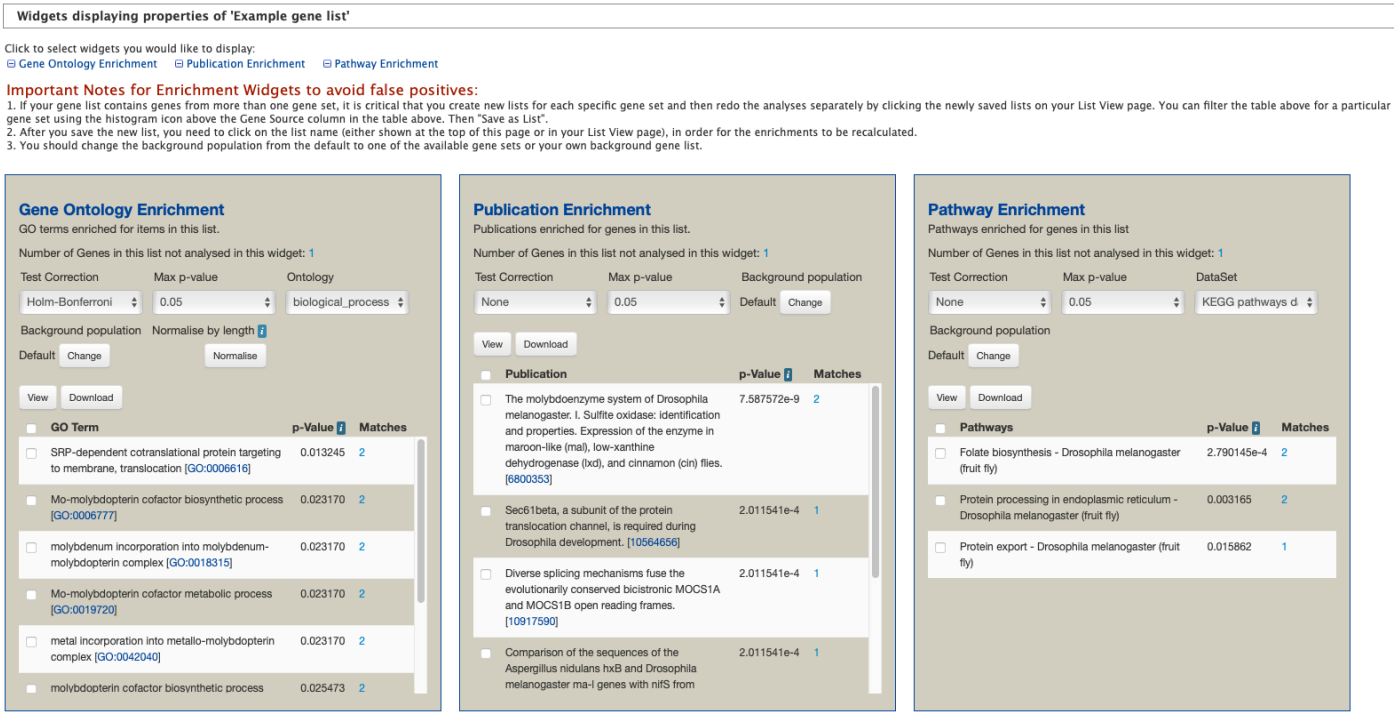


Fig. 4: Example: Available widgets

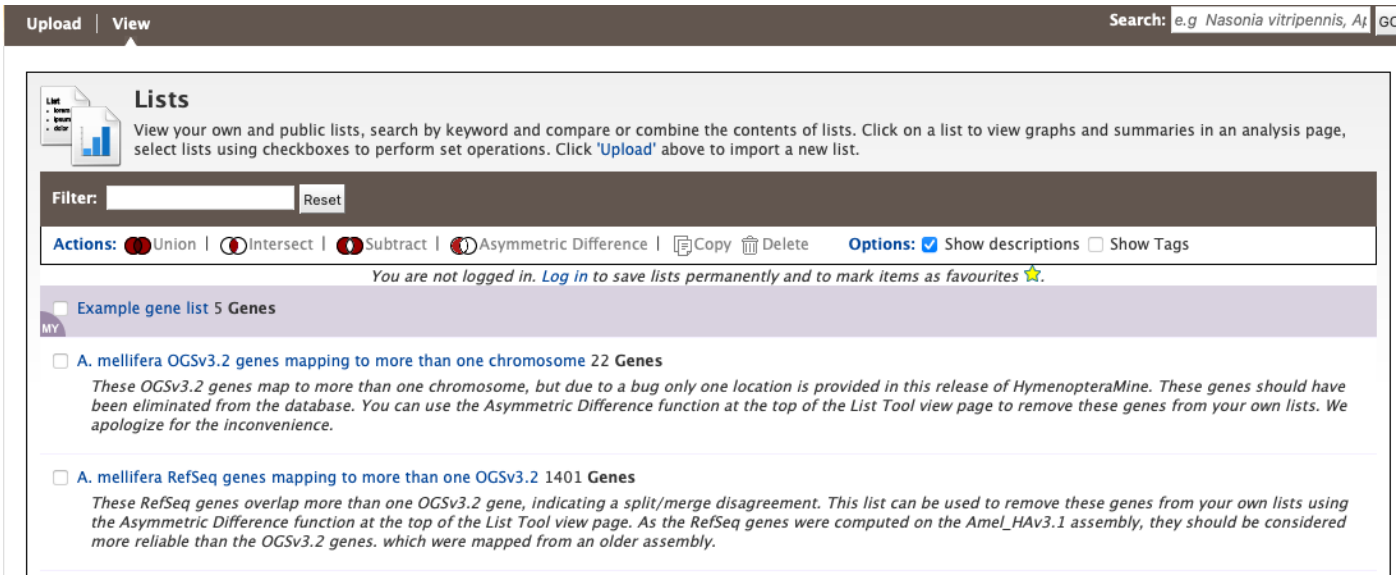


Fig. 5: Saved lists. Lists belonging to user are highlighted.

CHAPTER 5

MyMine

MyMine serves as a portal where logged-in users may manage their lists, queries, templates, and account details.

To access MyMine, click on the **MyMine** menu tab. A submenu appears with six options:

Lists - Saved lists.

History - List of queries recently run.

Queries - List of saved queries.

Templates - Templates created or marked as “Favorite”.

Password - Password reset form.

Account Details - User preferences form.

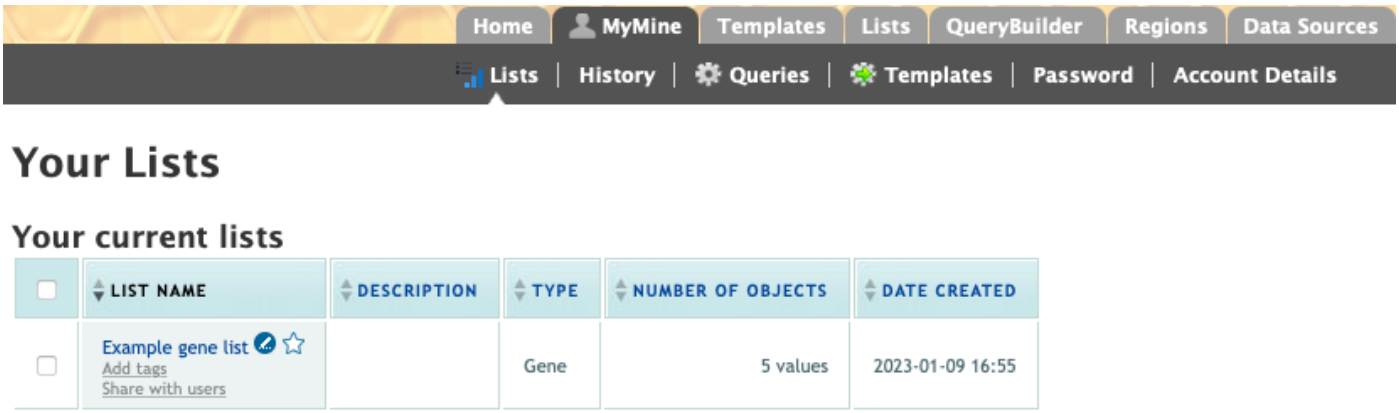
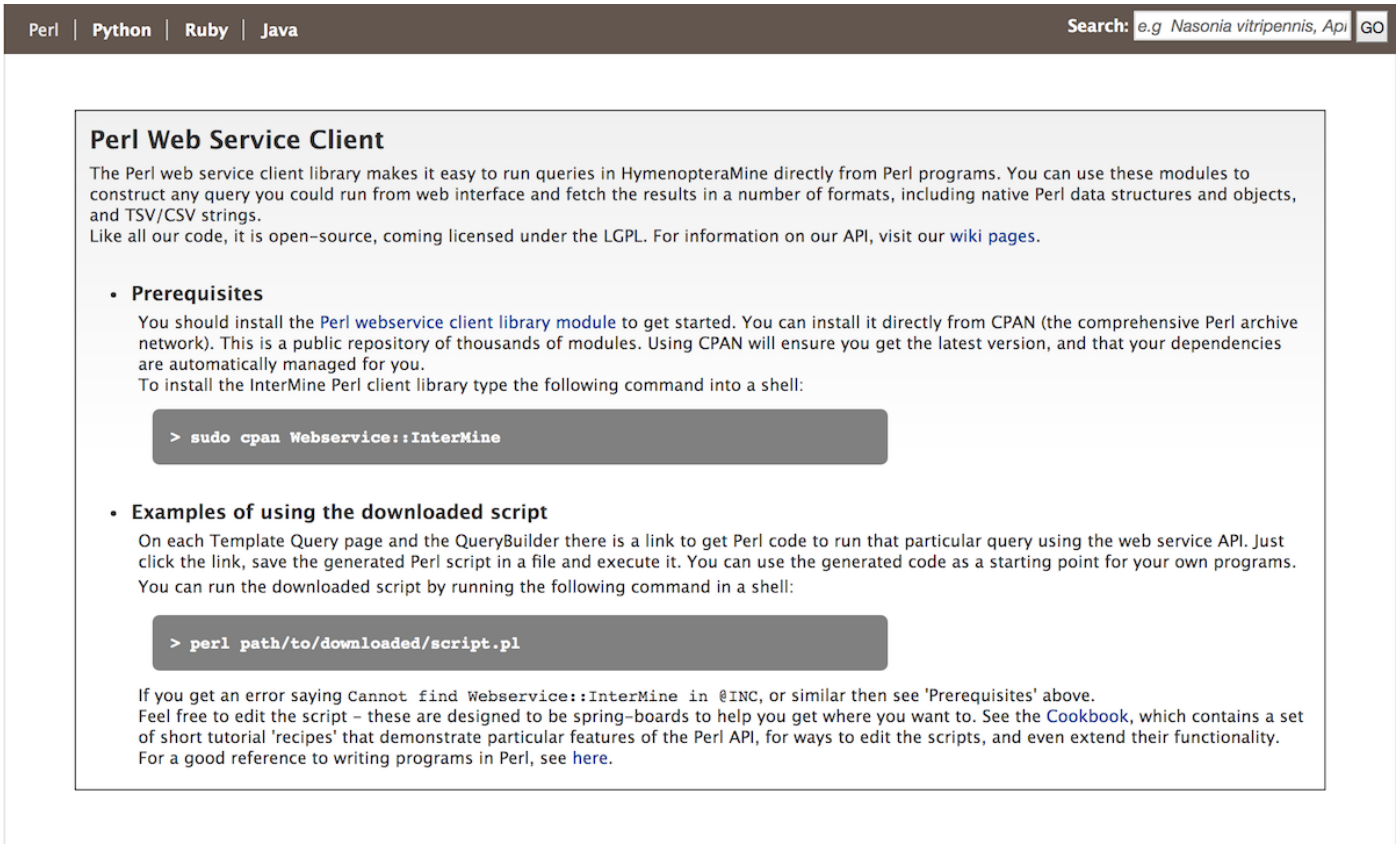


Fig. 1: MyMine

CHAPTER 6

API

An API is available for users who would like to programmatically access HymenopteraMine.



Perl, Python, Ruby, and Java are the languages supported by the InterMine API.

For more detailed information, view the InterMine documentation.

CHAPTER 7

Data Sources

The Data Sources table provides a description of the datasets that are integrated into HymenopteraMine, along with their date of download, version or release, citations wherever applicable, and any additional comments.

Gene Ontology	Gene Ontology Annotations from HGD-GO-Annotation	All species	HGD-GO-Annotation from HGD; 26 Oct 2022	Walsh el al. - PubMed 34747465	HGD GO Annotation Download
	Gene Ontology Annotations from FlyBase	*Drosophila melanogaster*	FlyBase release FB2022_05; 13 Sep 2022	Larkin et al. - PubMed 33219682	FlyBase Download
Pathways	Pathway Information from KEGG	*Acromyrmex echinatior* *Aphidius gifuensis* *Apis laboriosa* *Apis mellifera* *Atta cephalotes* *Bombus bifarius* *Bombus impatiens* *Bombus pyrosoma* *Bombus terrestris* *Bombus vancouverensis nearcticus* *Bombus vosnesenskii* *Camponotus floridanus* *Cephus cinctus* *Ceratina calcarata* *Ceratosolen solmsi marchali* *Colletes gigas* *Cotesia glomerata* *Diachasma alloeum* *Dinoponera quadriceps* *Drosophila melanogaster* *Fopius arisanus* *Formica exsecta* *Harpegnathos saltator* *Leptopilina heterotoma* *Linepithema humile* *Megalopta genalis* *Microplitis demolitor* *Monomorium pharaonis* *Nasonia vitripennis* *Nomia melanderi* *Ooceraea biroi* *Osmia bicornis bicornis* *Pogonomyrmex barbatus* *Polistes canadensis* *Polistes fuscatus* *Pseudomyrmex gracilis* *Solenopsis invicta* *Trichogramma pretiosum* *Vespula pensylvanica* *Vollenhovia emeryi*	KEGG release 104.0; 1 Oct 2022	Kanehisa et al. - PubMed 33125081	KEGG Download
	Pathway information inferred through orthologues from curated human pathways	*Drosophila melanogaster*	Reactome version 80; 5 Apr 2022	Gillespie et al. PubMed 34788843	Reactome Download
Interactions	Interactions from BioGRID	*Apis mellifera* *Drosophila melanogaster*	BioGRID version 4.4.213; 25 Aug 2022	Oughtred et al. - PubMed: 33070389	BioGRID Download
	Interactions from IntAct	*Drosophila melanogaster*	IntAct Release 243; 13 Jul 2022	Del Toro et al. - PubMed 34761267	IntAct Download

Fig. 1: Data sources table

CHAPTER 8

BLAST

Users may perform BLAST queries against the Hymenoptera genomic, CDS, or protein sequences using the BLAST page.

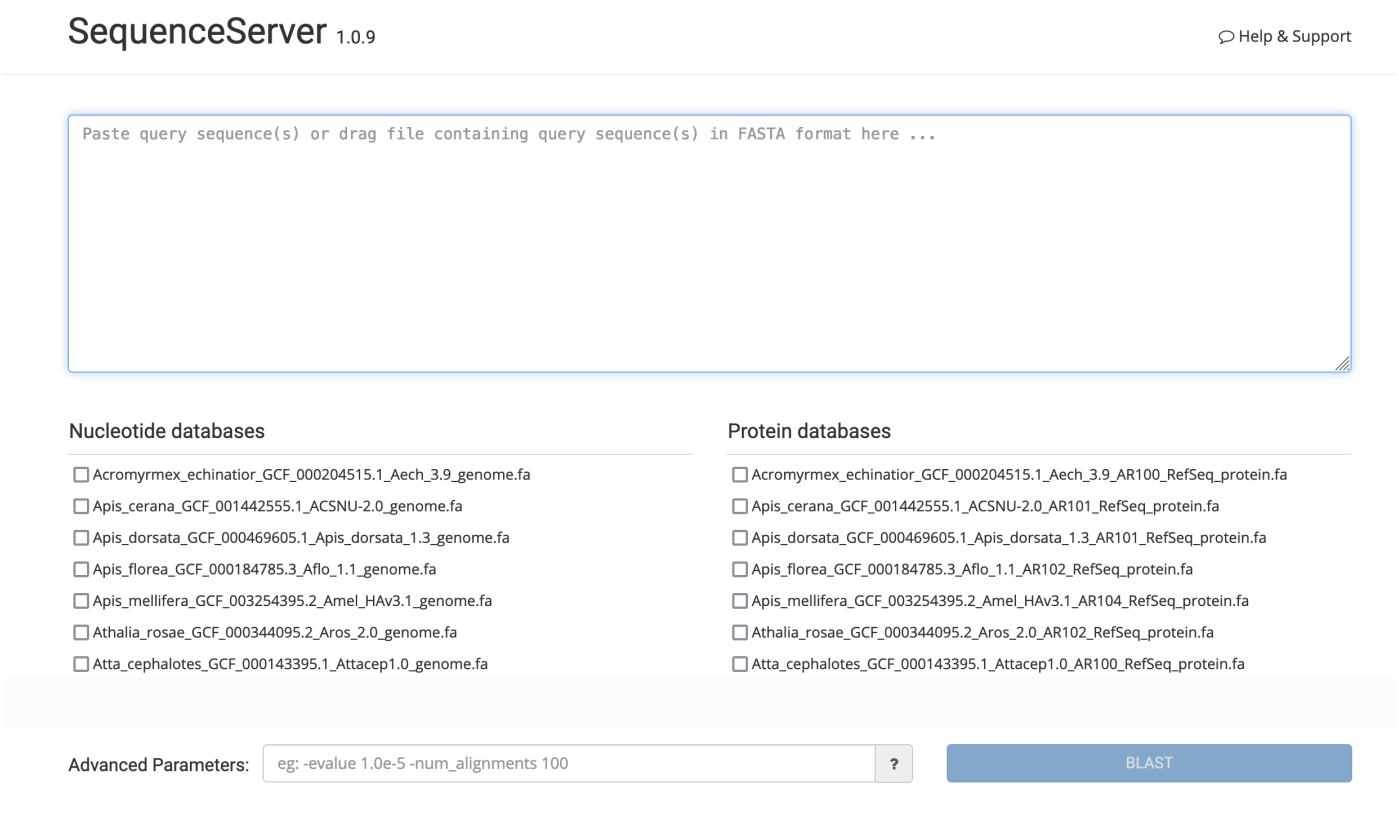


Fig. 1: BLAST search

CHAPTER 9

How to Cite

Visit http://hymenopteragenome.org/data_usage_citing for information on data usage and citing HymenopteraMine.