HymenopteraMine Documentation

Release v1.5

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

Jun 22, 2022

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 . . . 10
  - 2.2.3 Customize output . . . 10
  - 2.2.4 Optional filters . . . 10
- 2.3 QueryBuilder . . . 11
  - 2.3.1 Model browser . . . 11
  - 2.3.2 Examples . . . 13
    - 2.3.2.1 Example 1: Querying for protein coding genes . . . 13
    - 2.3.2.2 Example 2: Querying for protein coding genes on a particular chromosome . . . 15
    - 2.3.2.3 Example 3: Querying for protein coding genes on a particular chromosome and their exons . . . 15
- 2.4 Genomic Regions Search . . . 19

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

4 Lists **31**
- 4.1 Creating Lists . . . 31
- 4.2 Saving Lists . . . 34

5 MyMine **35**

6 API **37**

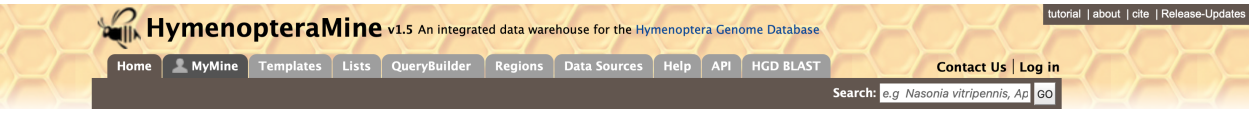
7 Data Sources **39**

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.

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).

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.

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.

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.

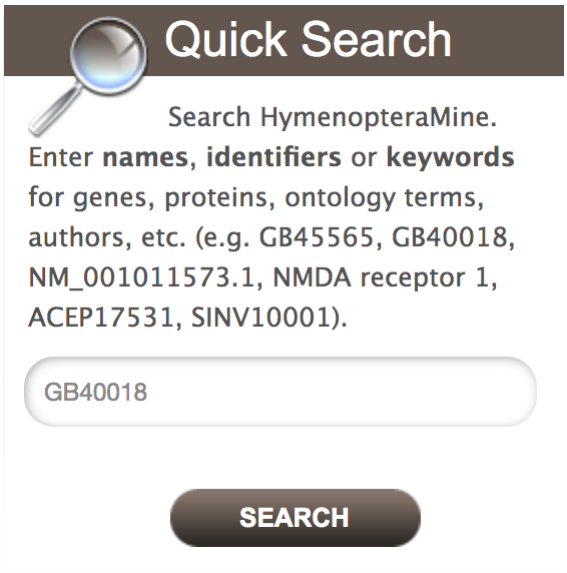


Fig. 1: Quick Search from home page

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.

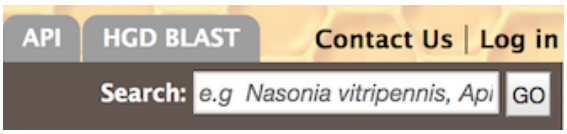


Fig. 2: Quick Search from any page

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.

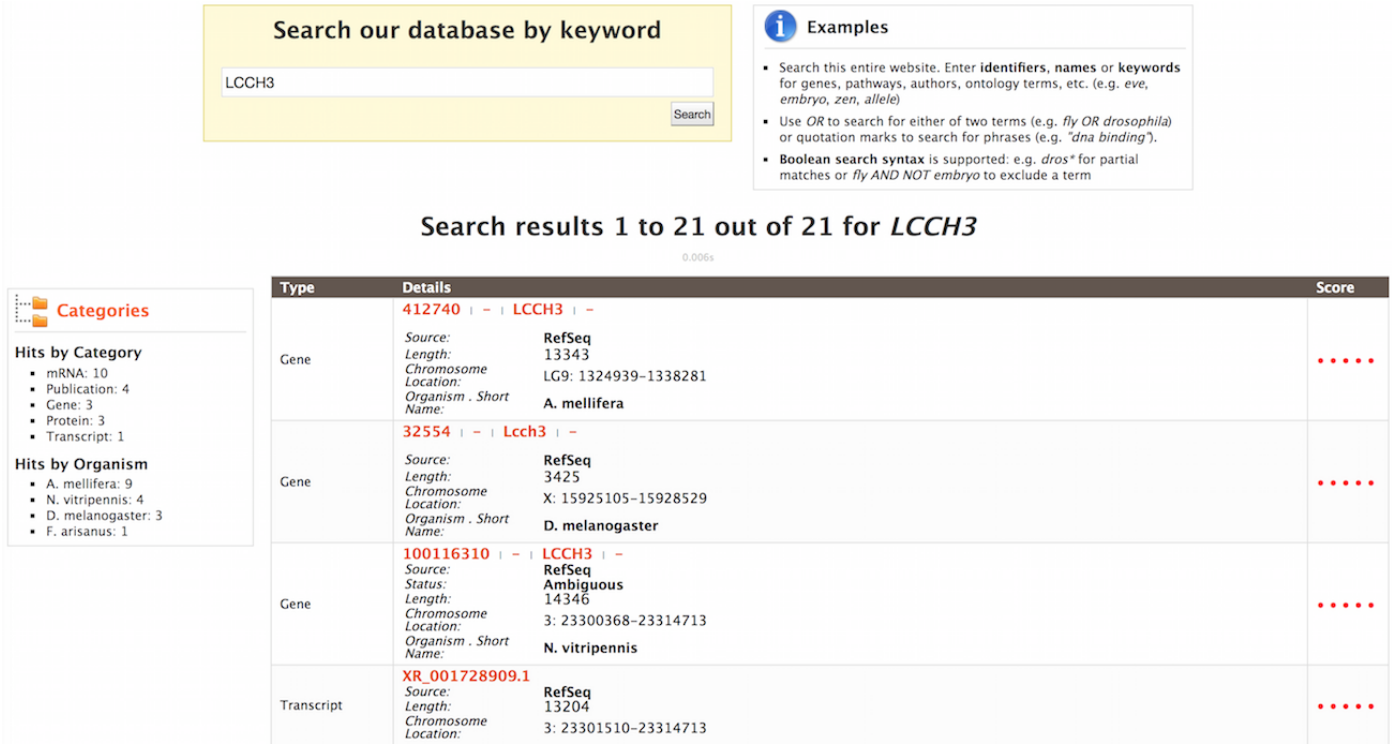


Fig. 3: Example: Search results for LCCH3

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.

The results page shows all the genes having the Gene Ontology term "DNA Binding" in their annotation. When logged

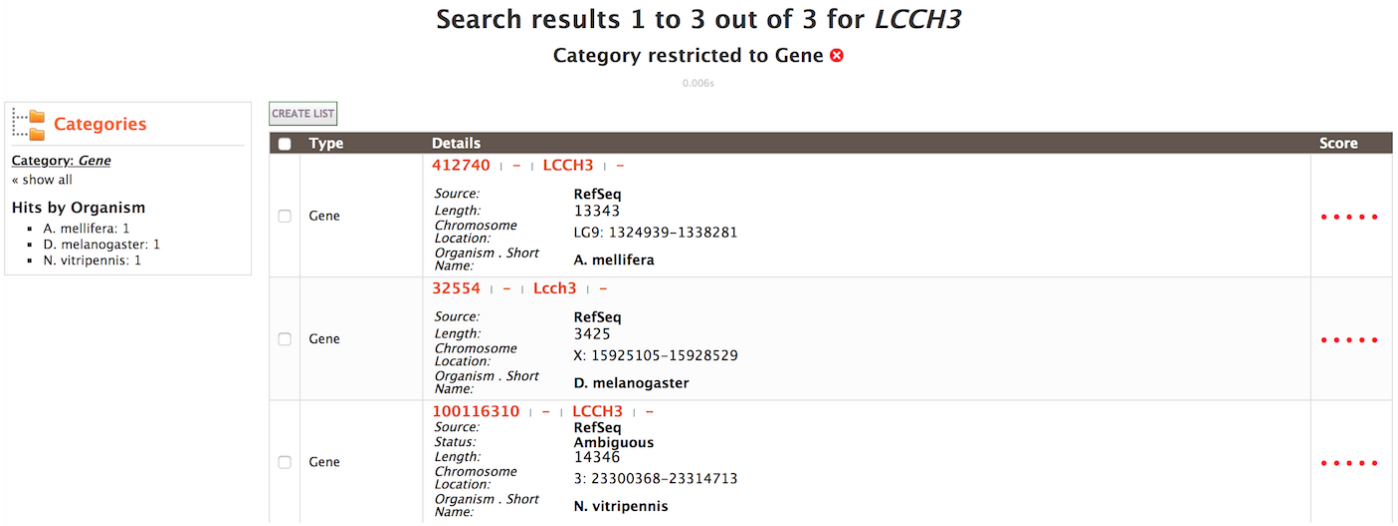


Fig. 4: Example: Search results restricted to Gene category

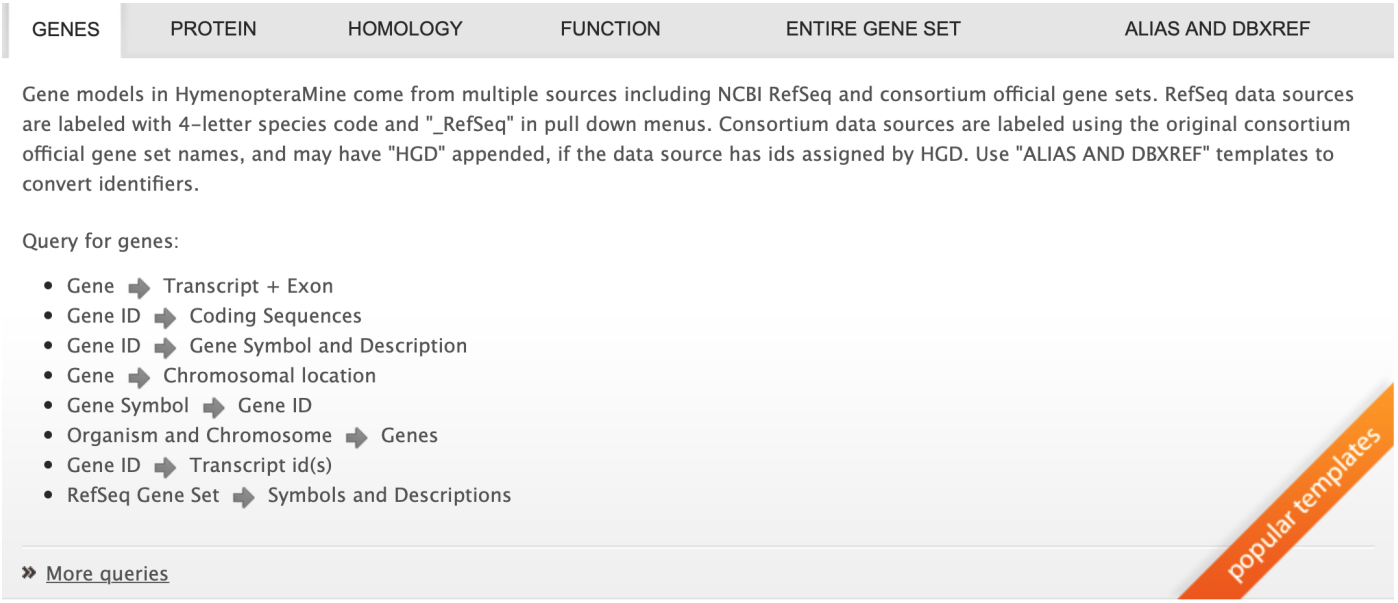


Fig. 5: Popular templates

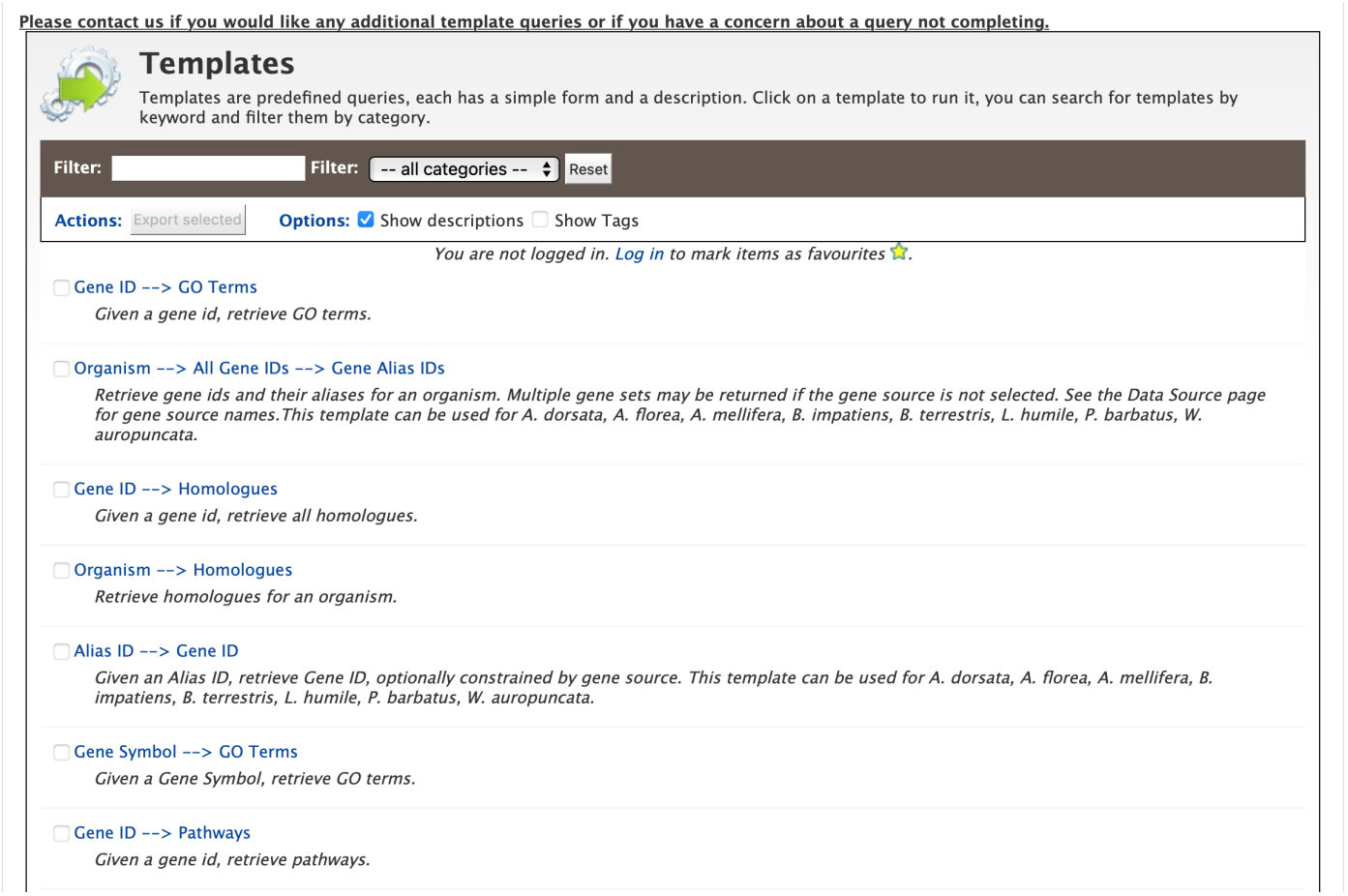


Fig. 6: List of templates on the Templates page

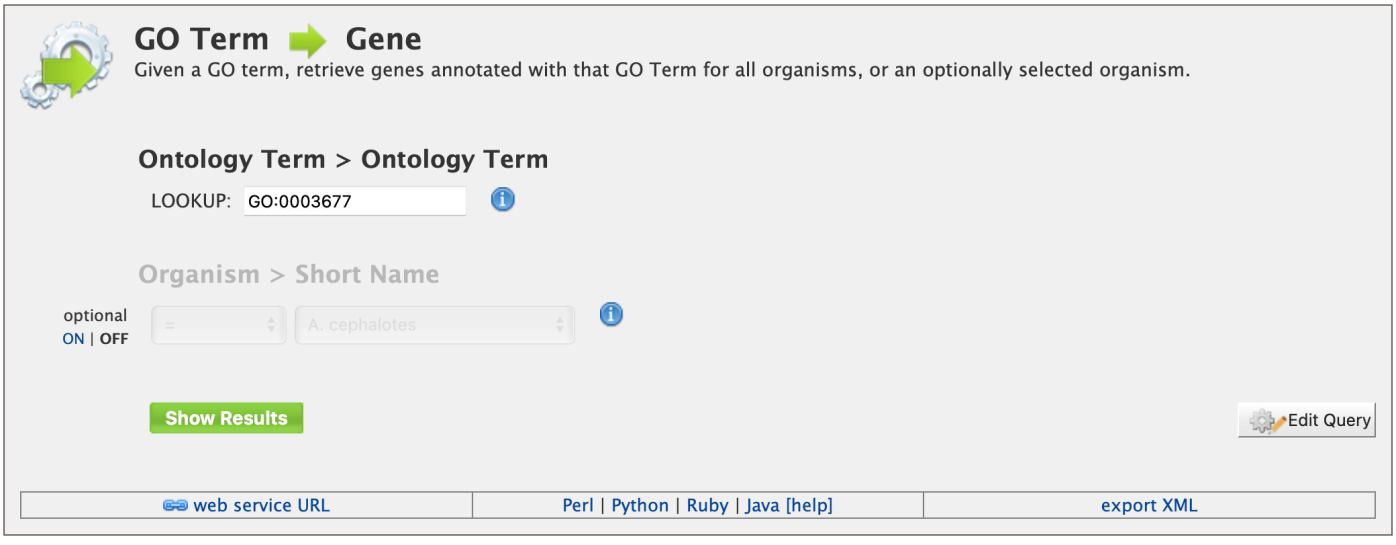


Fig. 7: Example: GO Term Gene template

in, users may create a new list or add these genes to an existing list to perform further 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 all organisms, or an optionally selected organism.
Manage Columns | Manage Filters | Manage Relationships | Save as List | Generate Python code | Export
Showing 1 to 25 of 49,495 rows
Rows per page: 25

Ontology Term Identifier	Ontology Term Name	Ontology Term Description	Code Code	Gene DB identifier	Gene Symbol	Gene 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).	IEA	100104126	Eve	N. vitripennis	GO Annotation from UniProt	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	100107367	LOC100107367	N. vitripennis	GO Annotation from HGD	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	100110062	LOC100110062	N. vitripennis	GO Annotation from UniProt	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	100110095	Scr	N. vitripennis	GO Annotation from HGD	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	100112394	LOC100112394	N. vitripennis	GO Annotation from HGD	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	100113507	LOC100113507	N. vitripennis	GO Annotation from HGD	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	100113526	LOC100113526	N. vitripennis	GO Annotation from HGD	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	ECO:0000501	100113551	LOC100113551	N. vitripennis	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	100113551	LOC100113551	N. vitripennis	GO Annotation from UniProt	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	100113574	LOC100113574	N. vitripennis	GO Annotation from HGD	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	100113609	LOC100113609	N. vitripennis	GO Annotation from HGD	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	100113671	LOC100113671	N. vitripennis	GO Annotation from HGD	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	100113688	LOC100113688	N. vitripennis	GO Annotation from HGD	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	IEA	100113696	Tbp	N. vitripennis	GO Annotation from HGD	HGD
GO:0003677	DNA binding	Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).	ECO:0000501	100113706	LOC100113706	N. vitripennis	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.

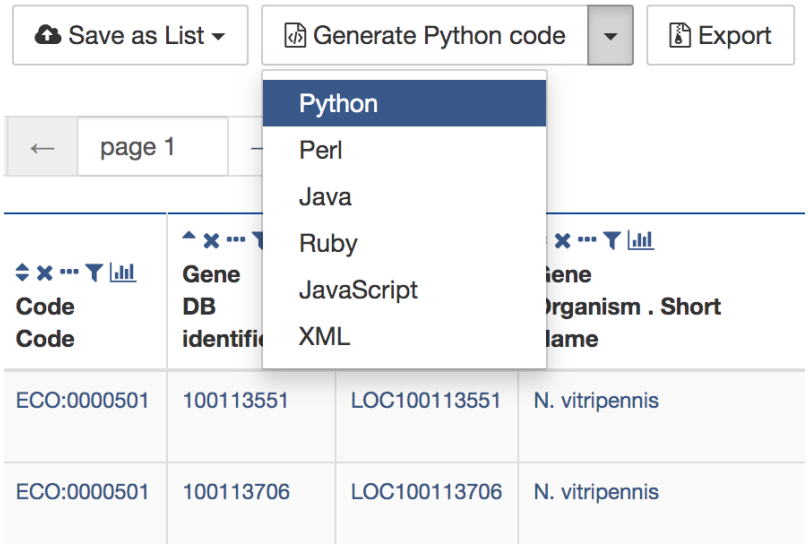


Fig. 9: Generate code options

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.

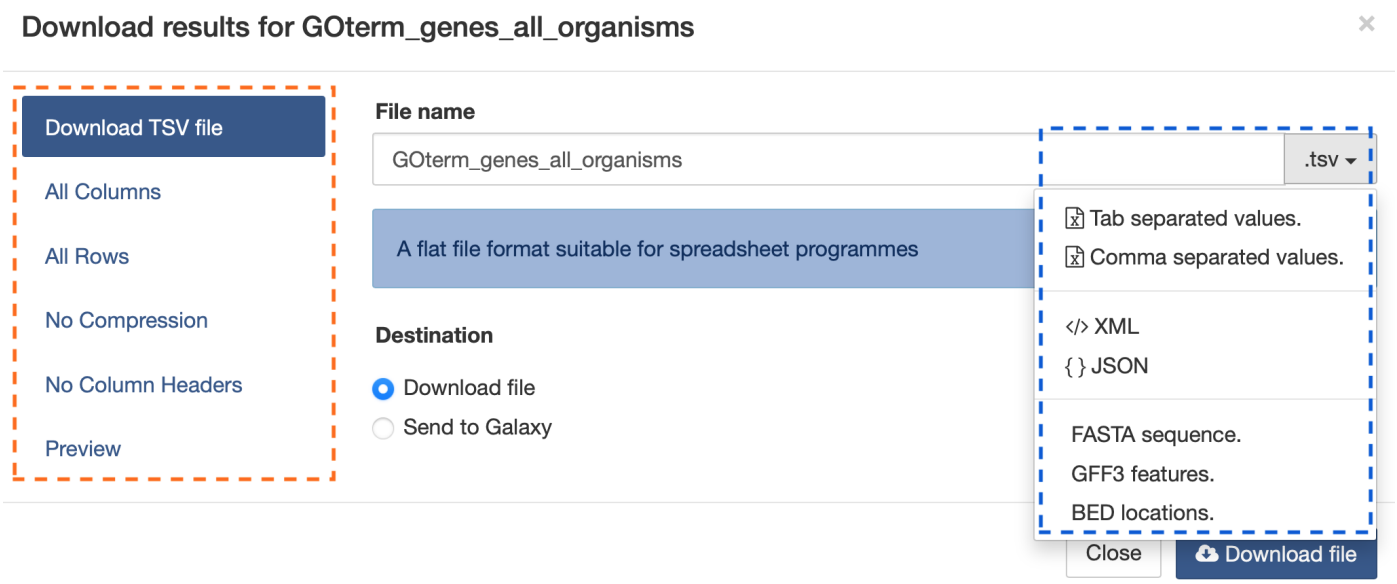


Fig. 10: Download results options

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 GO Term Gene template has an additional filter for specifying an organism. To enable this filter, click **ON** below **optional**.

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.

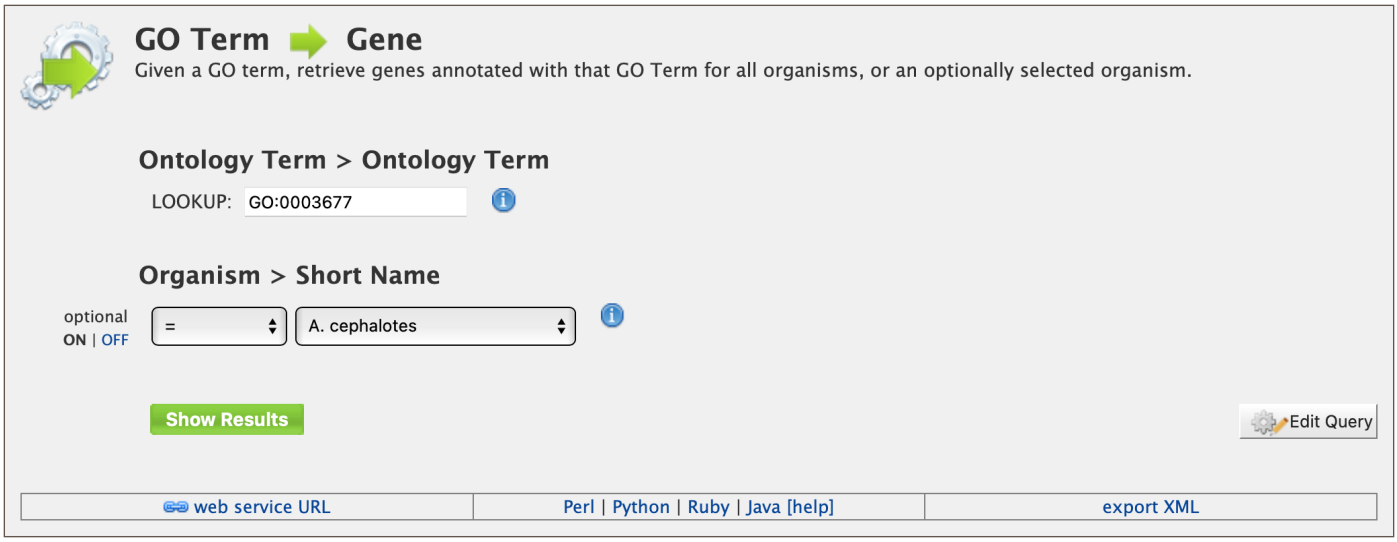
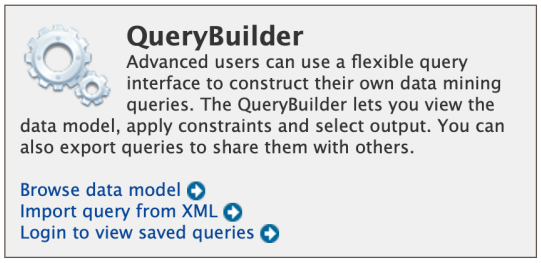
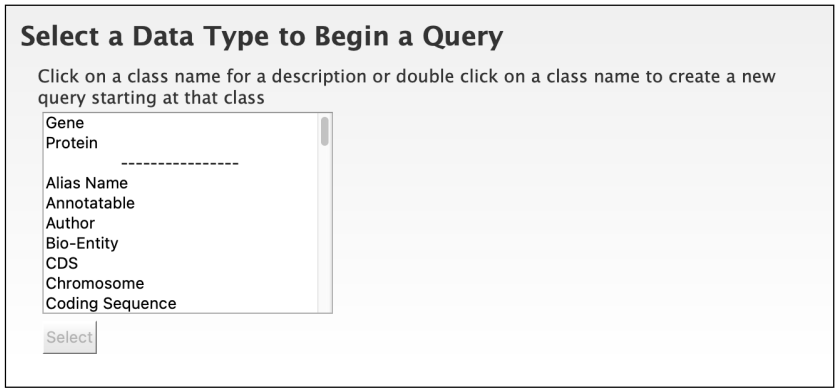


Fig. 11: Example: GO Term Gene template with organism filter enabled

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.





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

2.3.1 Model browser

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

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.

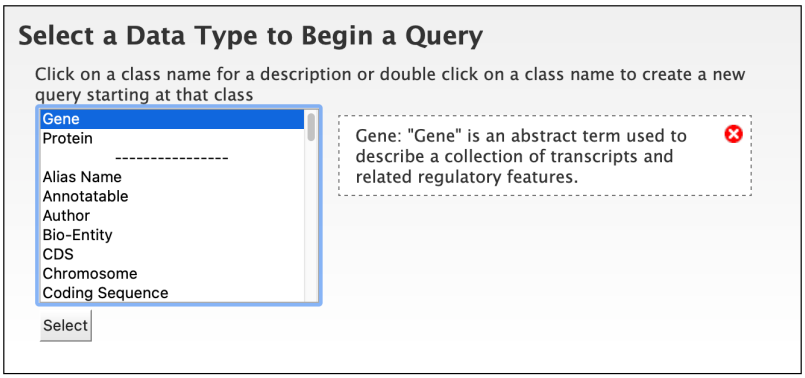


Fig. 12: Example: Gene data type selected in QueryBuilder

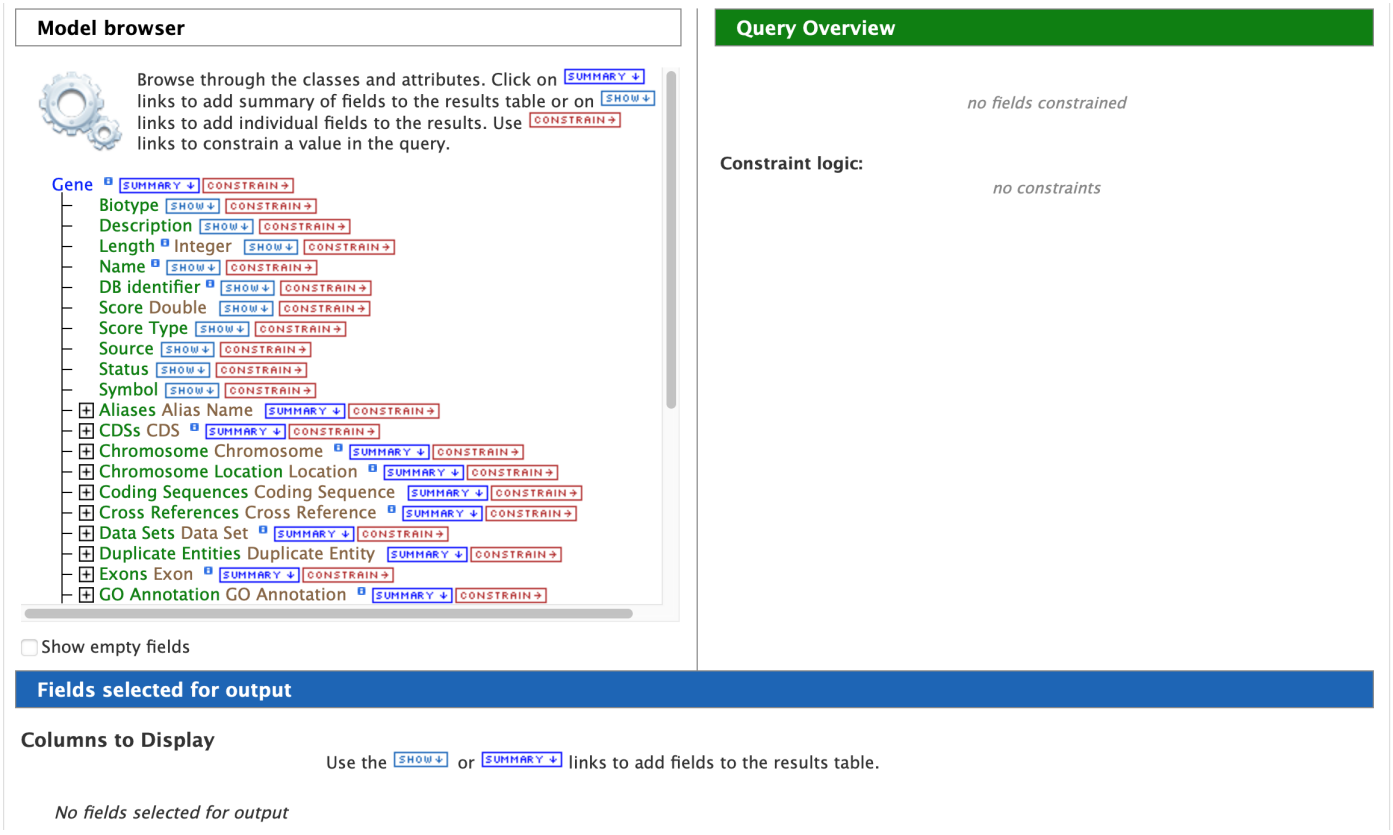


Fig. 13: Model browser with Gene selected as data type

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 **DB Identifier** 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**.

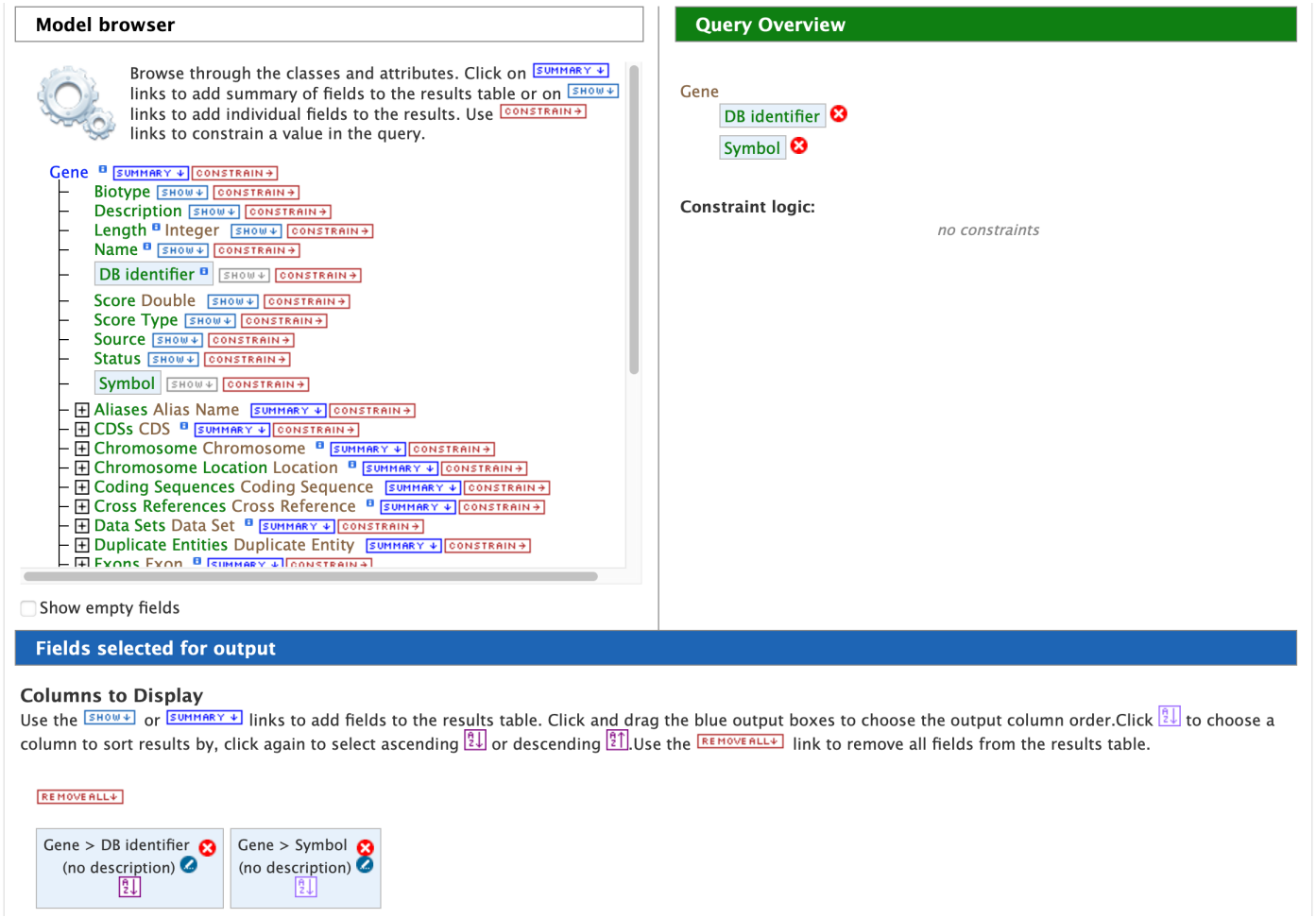


Fig. 14: Step 1: Select fields to be added to the query

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.

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

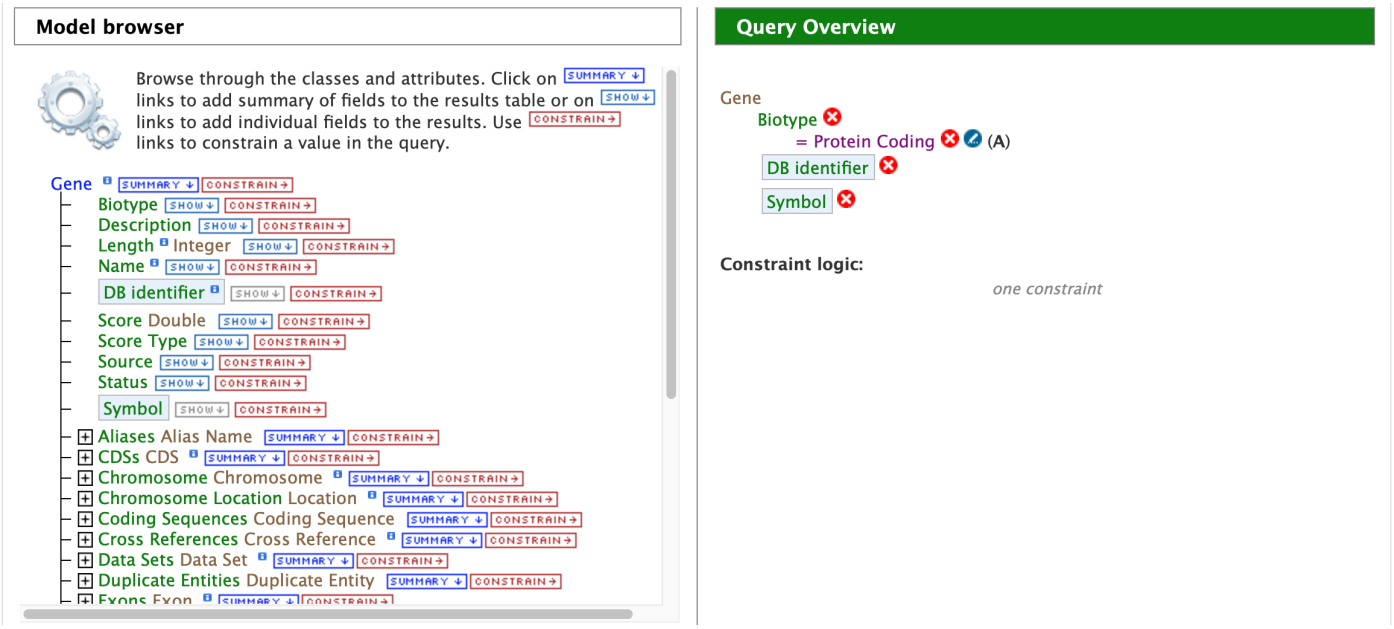
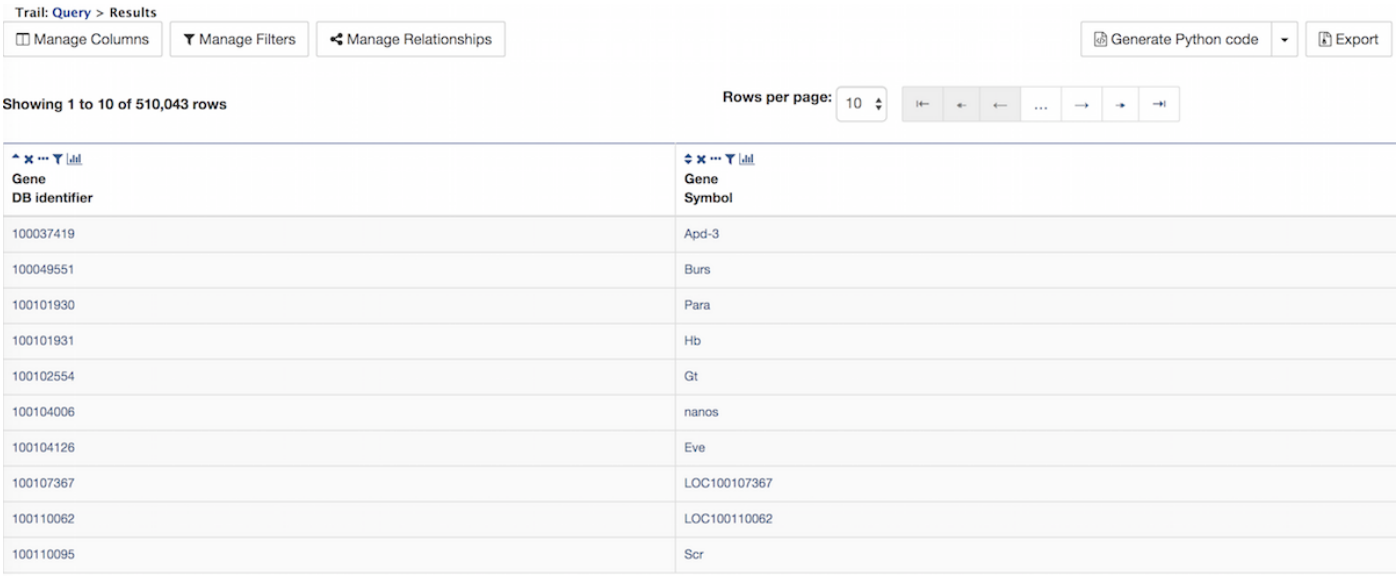


Fig. 15: Step 2: Add a constraint to the query on Biotype



Gene DB identifier	Gene Symbol
100037419	Apd-3
100049551	Burs
100101930	Para
100101931	Hb
100102554	Gt
100104006	nanos
100104126	Eve
100107367	LOC100107367
100110062	LOC100110062
100110095	Scr

Fig. 16: 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.

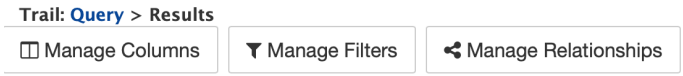


Fig. 17: Click on the query trail to edit the query

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

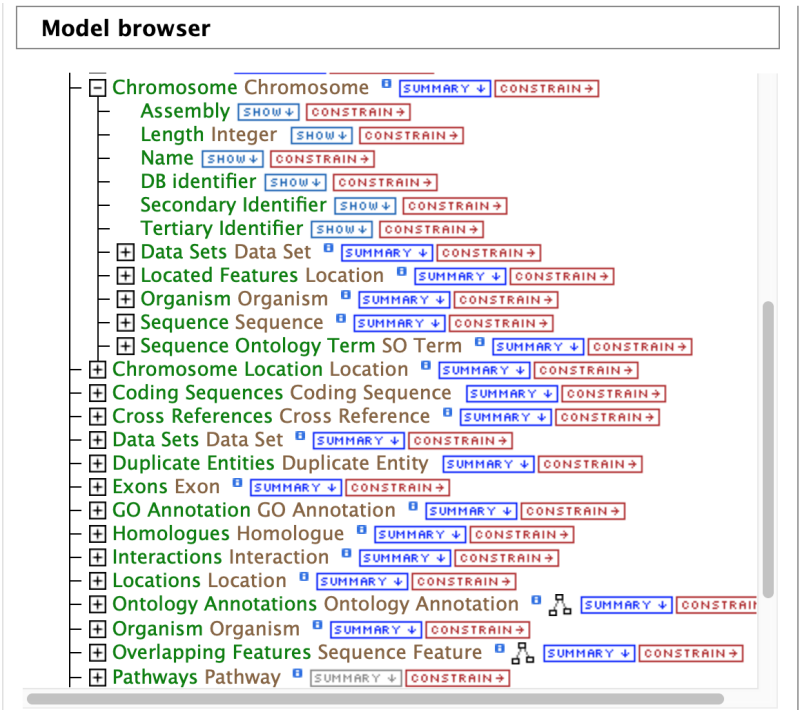


Fig. 18: Step 1: View attributes of the Chromosome feature class

Then click **Constrain** next to the attribute **DB Identifier**. In the pop-up window, enter **NW_003791143.1** into the 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 the DB identifier *NW_003791143.1*.

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 500,000 in the first query) due to the additional constraint.

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.

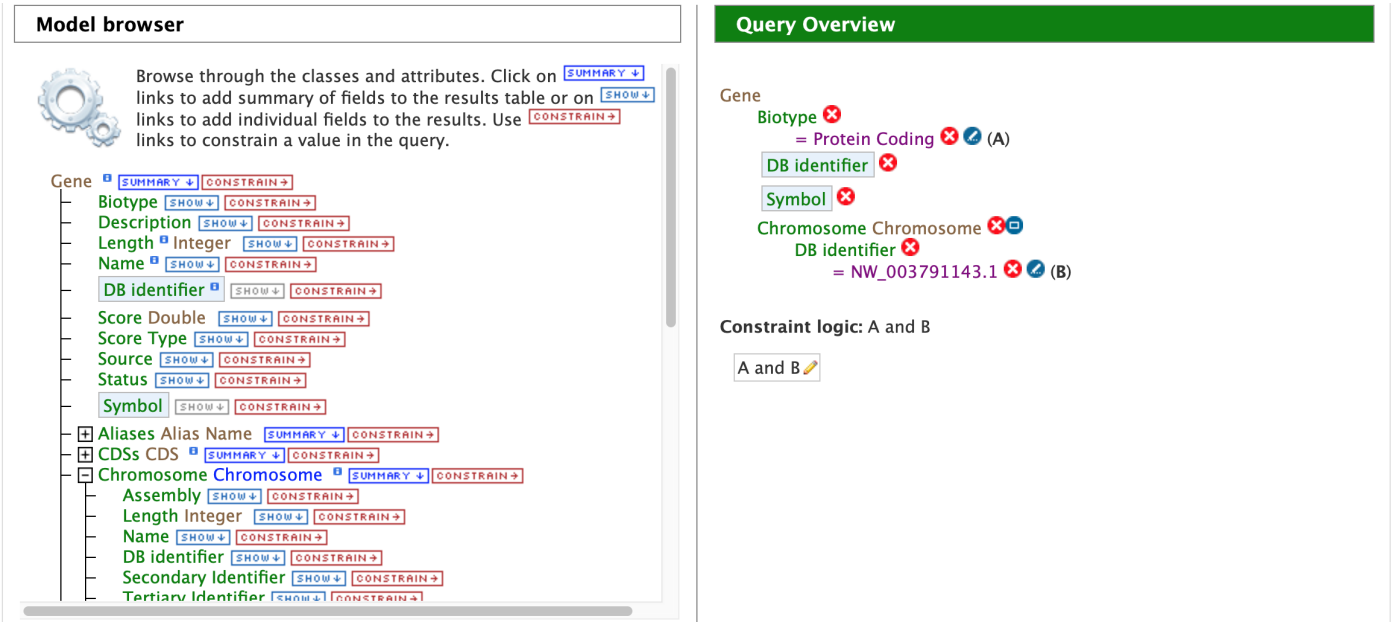


Fig. 19: Step 2: Add a constraint to the Chromosome DB Identifier

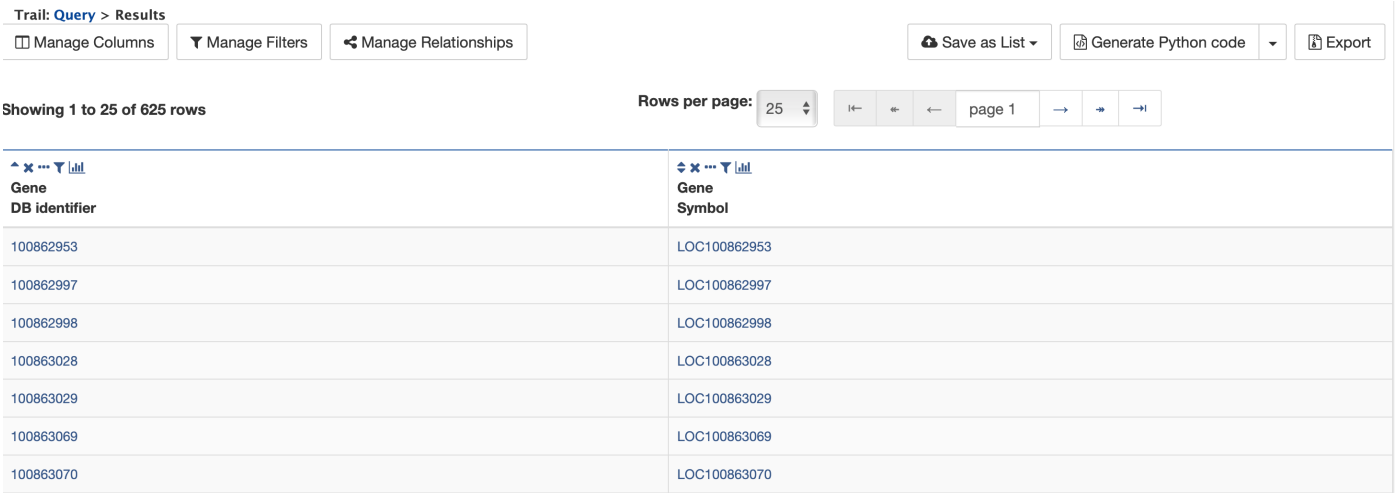


Fig. 20: Step 3: Display query results

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**.

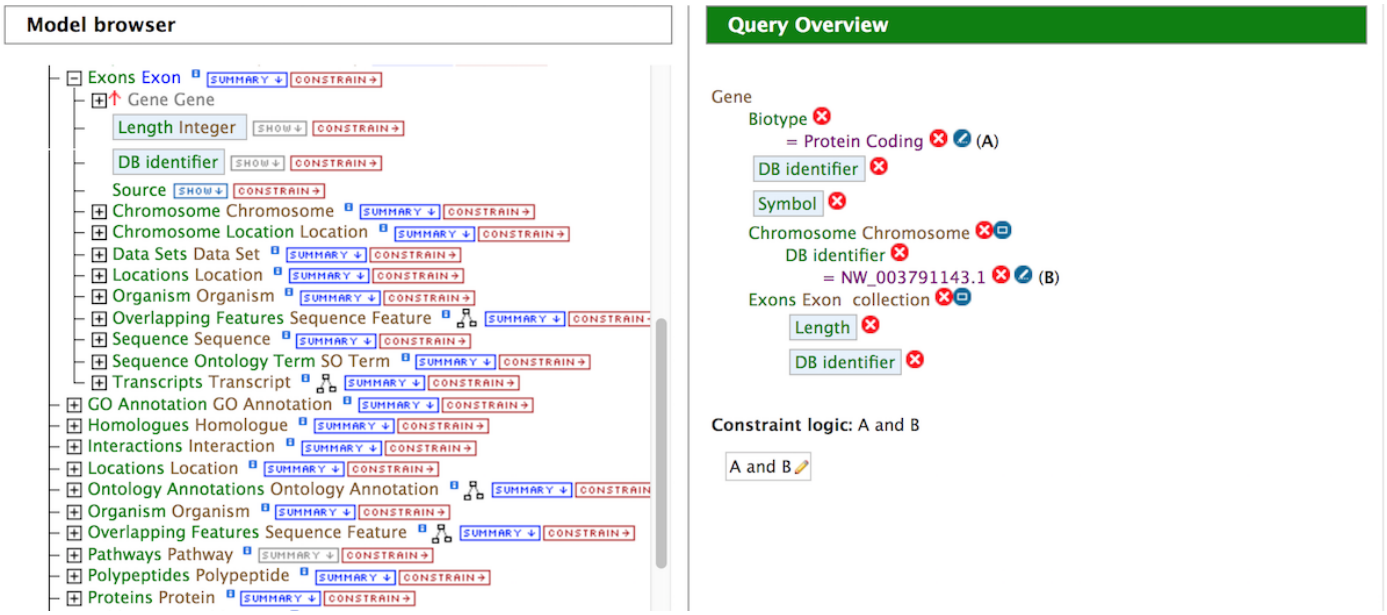


Fig. 21: Step 1: Select Exon fields to be added to the query

The **Query Overview** shows the query in progress, with four fields and two constraints. 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**.

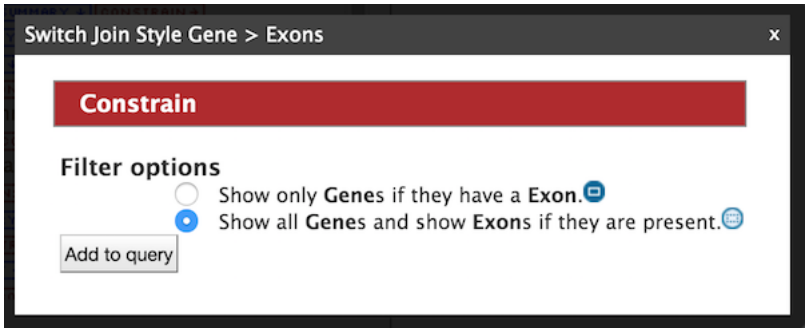


Fig. 22: Step 2: Change the join style

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 DB Identifier 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.

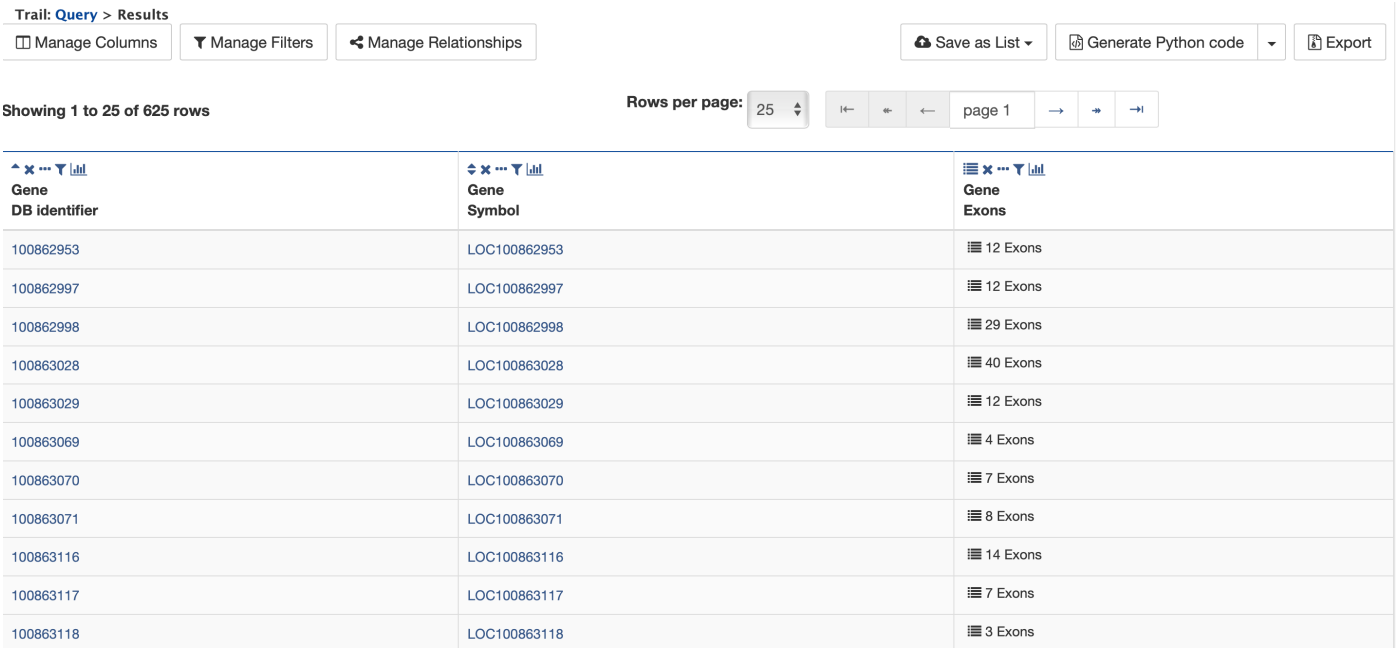


Fig. 23: Step 3: Display query results

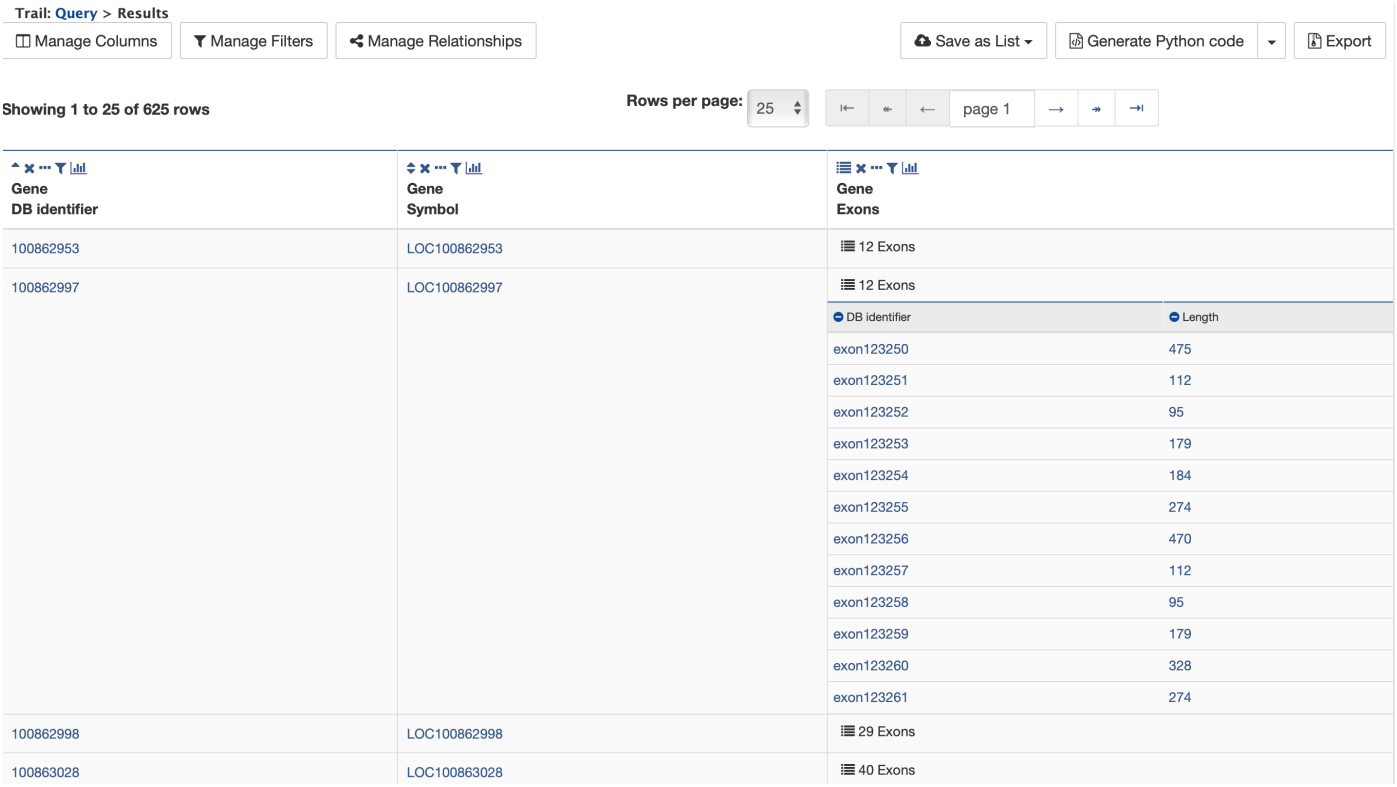
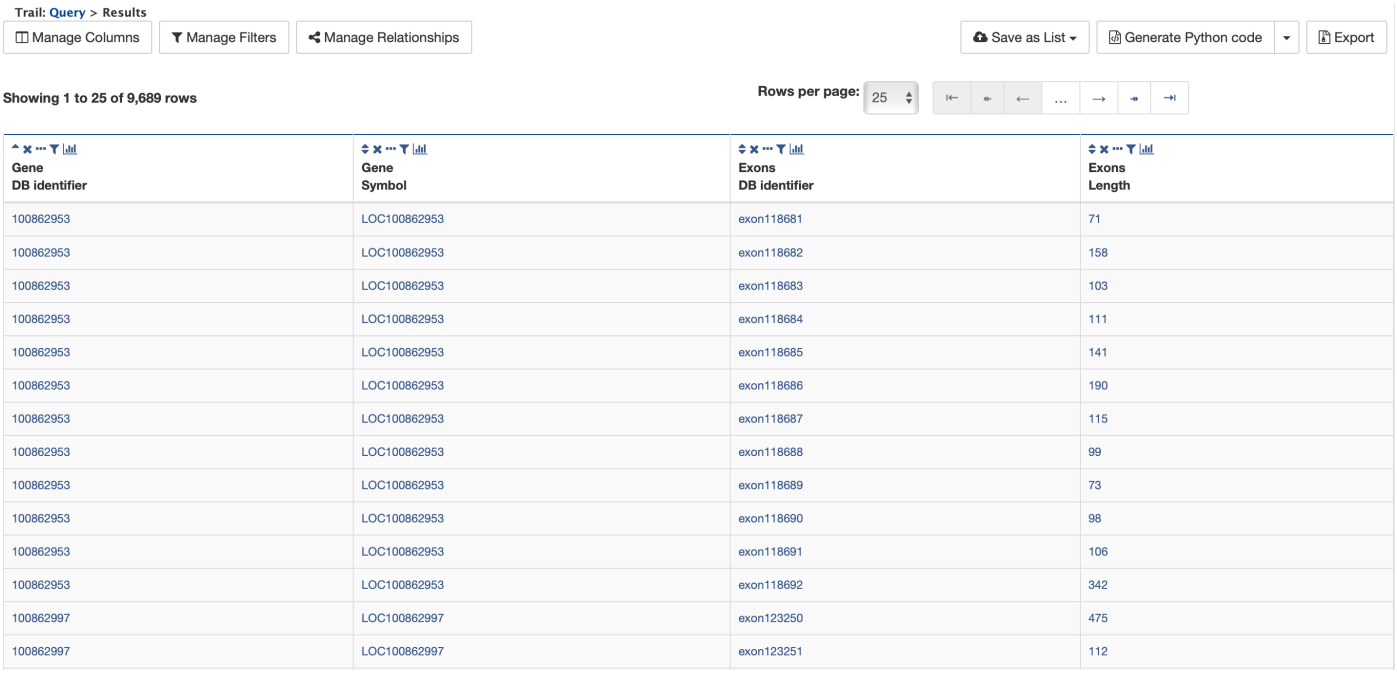


Fig. 24: Step 4: Expand information on exons



Gene DB identifier	Gene Symbol	Exons DB identifier	Exons Length
100862953	LOC100862953	exon118681	71
100862953	LOC100862953	exon118682	158
100862953	LOC100862953	exon118683	103
100862953	LOC100862953	exon118684	111
100862953	LOC100862953	exon118685	141
100862953	LOC100862953	exon118686	190
100862953	LOC100862953	exon118687	115
100862953	LOC100862953	exon118688	99
100862953	LOC100862953	exon118689	73
100862953	LOC100862953	exon118690	98
100862953	LOC100862953	exon118691	106
100862953	LOC100862953	exon118692	342
100862997	LOC100862997	exon123250	475
100862997	LOC100862997	exon123251	112

Fig. 25: 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.)

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 **A. mellifera** from the **Select Organism** drop-down, 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
```

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.

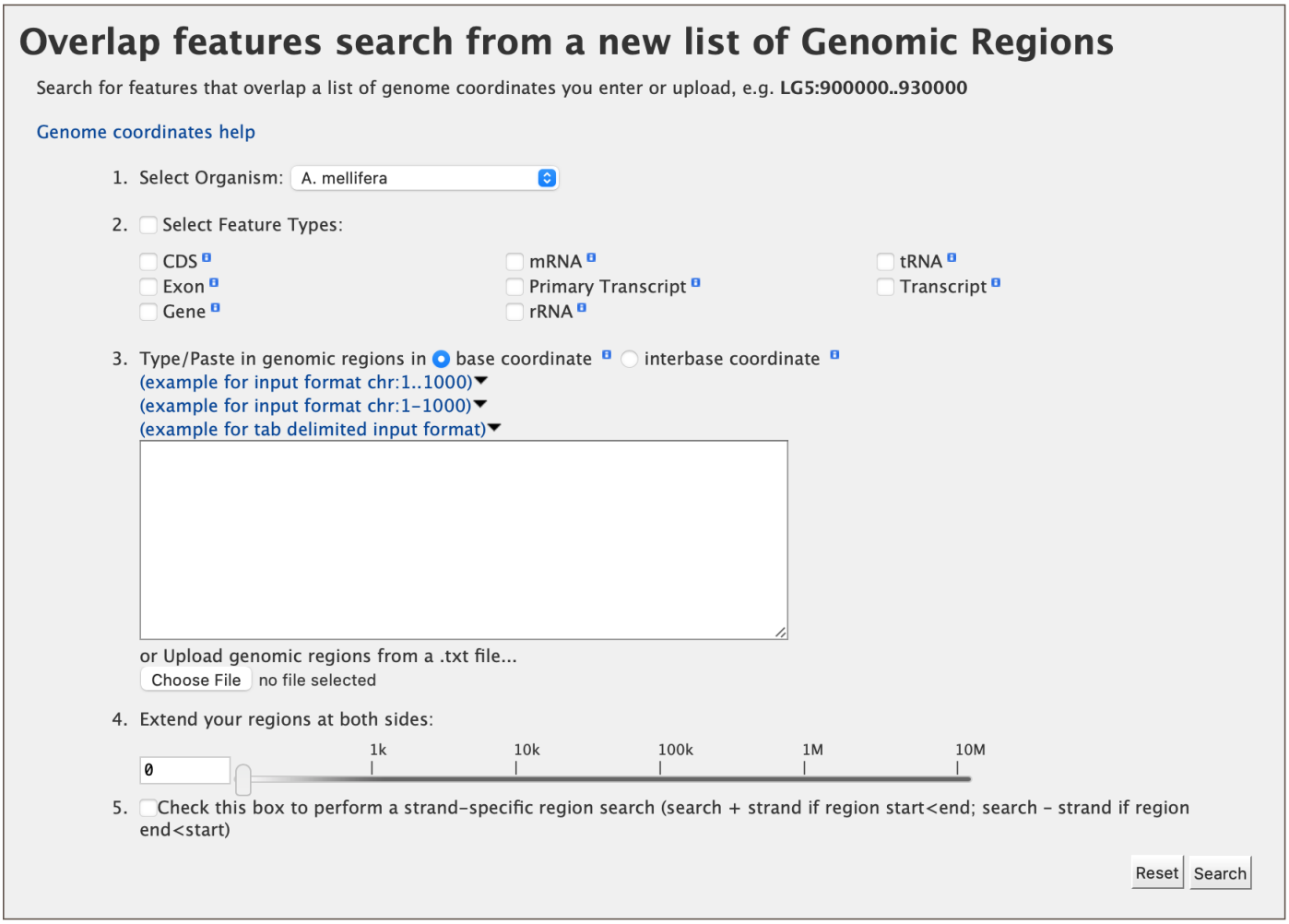


Fig. 26: Genomic regions search form

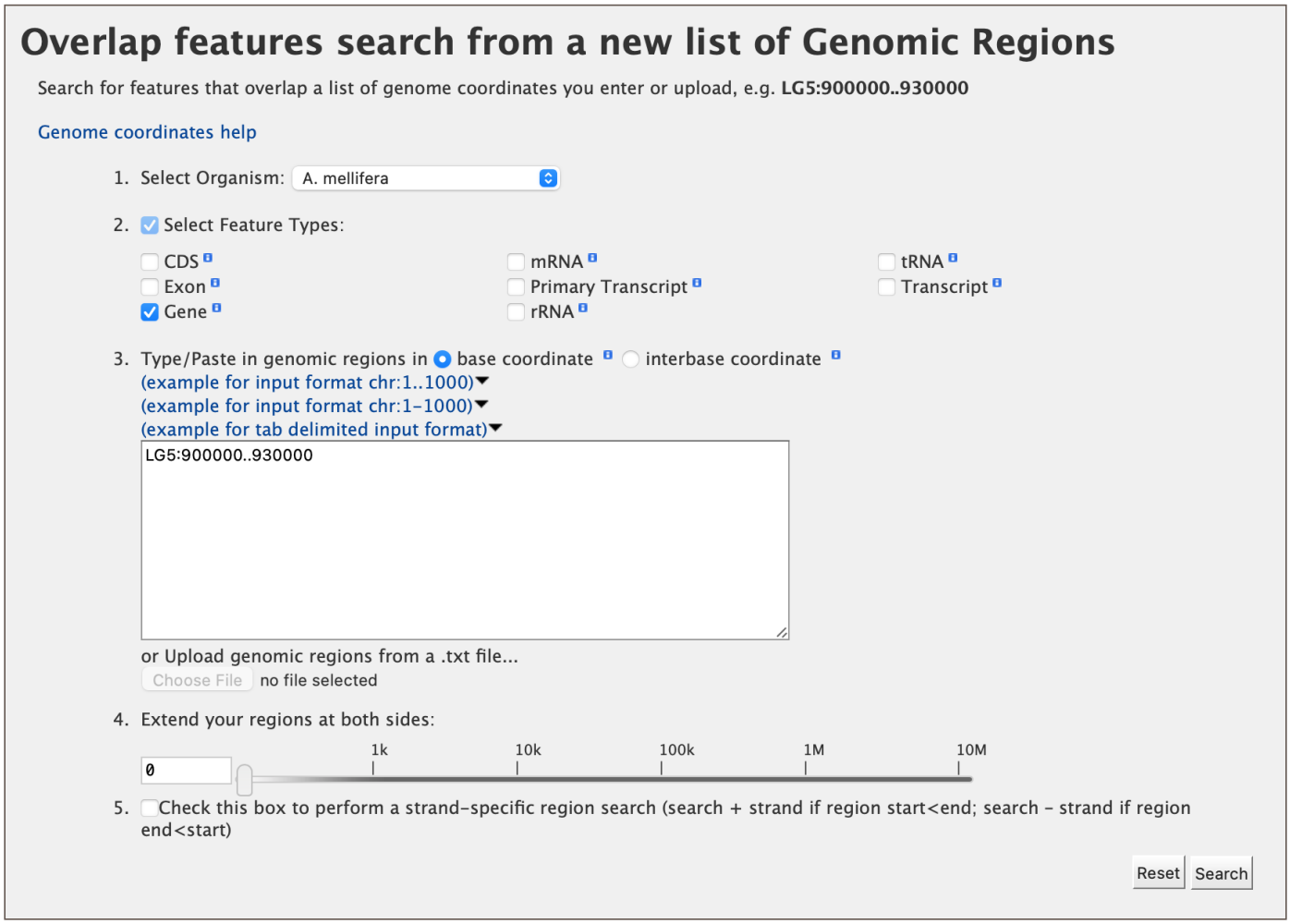


Fig. 27: Example: Genomic regions search on *A. mellifera*

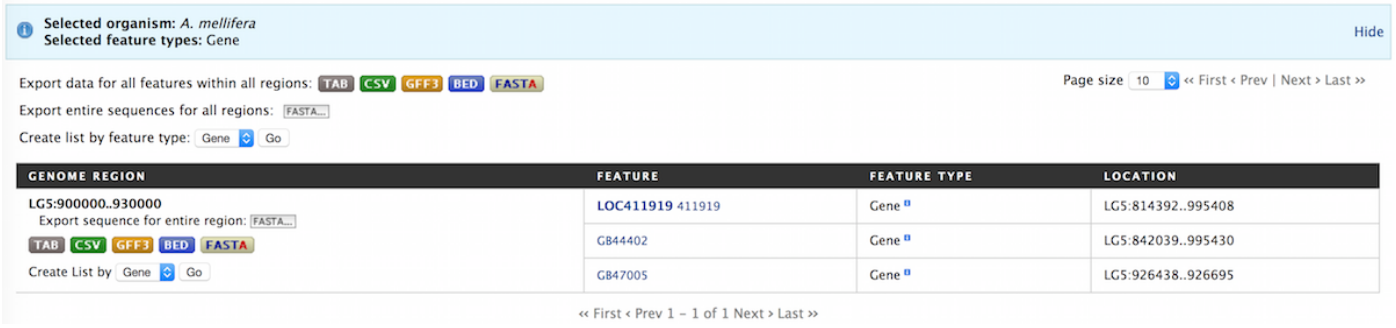


Fig. 28: Example: Genomic regions search results

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 **Protein** 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.

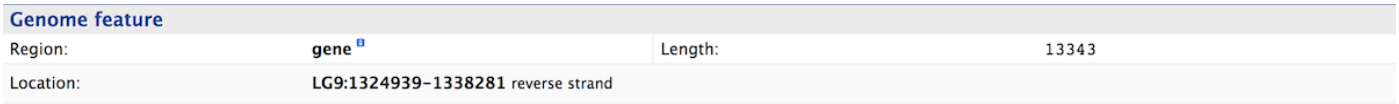
Clicking on an item in the table rather than just hovering over it will bring up its report page. For example, click on LCCH3 in the Gene Symbol column to view its report.

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.



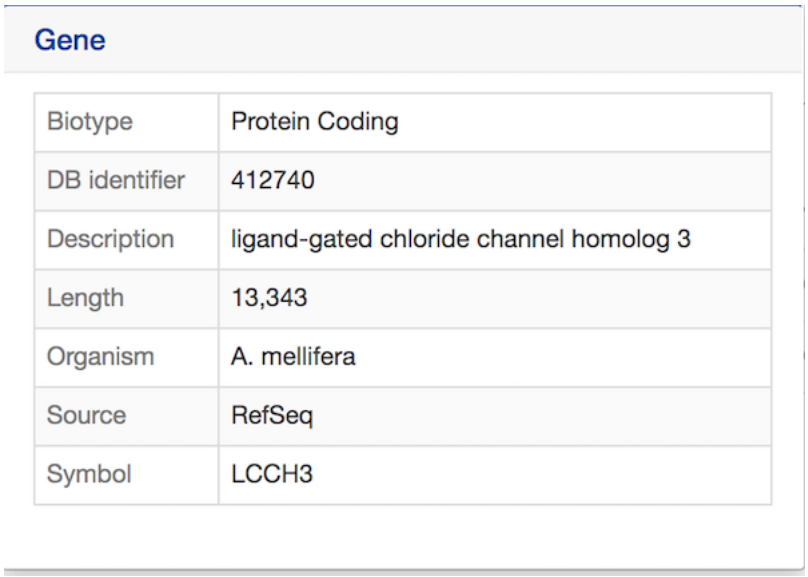


Fig. 1: Example: Summary window for LCCH3

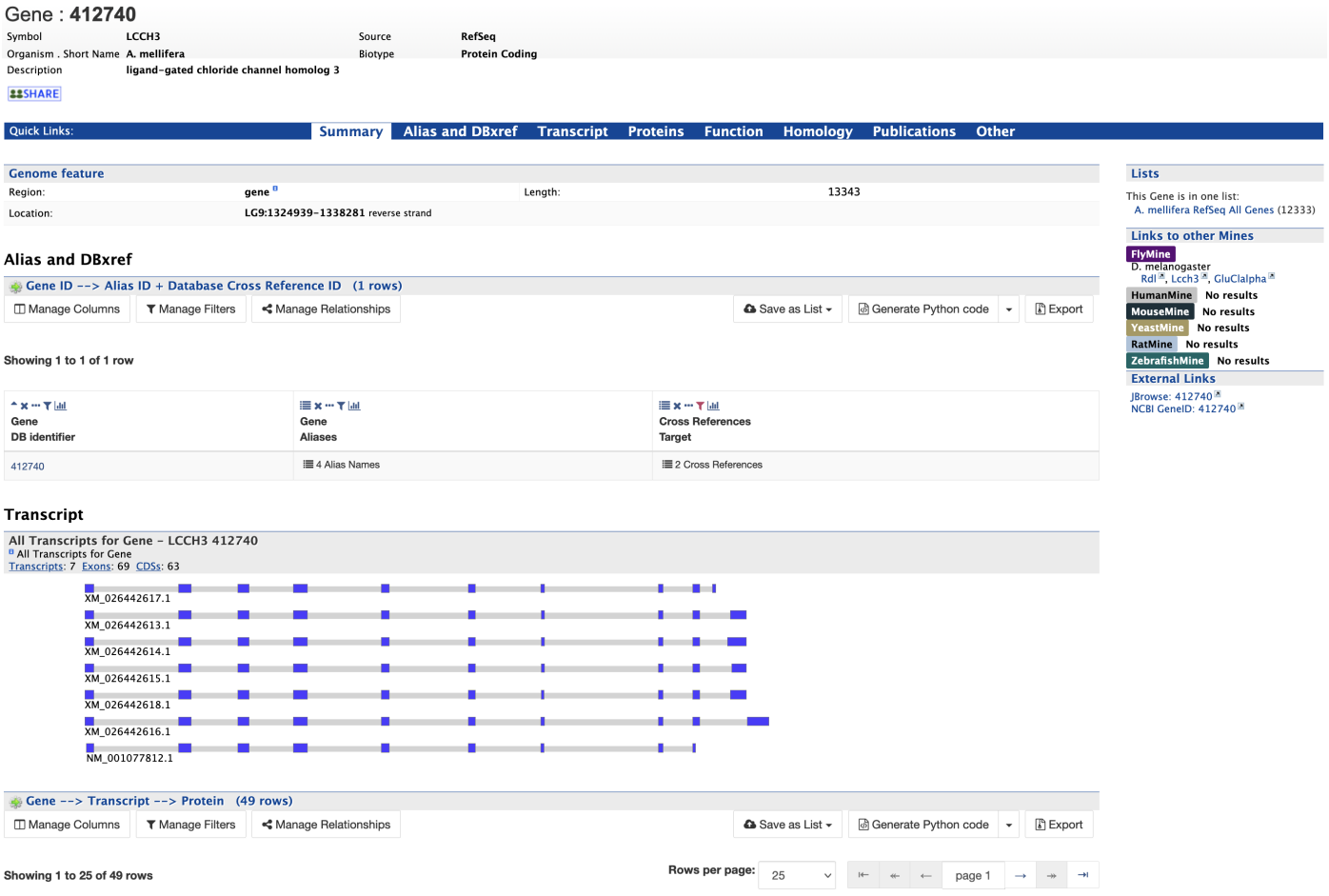


Fig. 2: Report page

3.2 Alias and DBxref

The **Alias and DBxref** section displays a table of aliases and database cross references for the gene. In this example, the gene with DB Identifier 412740 has four aliases and two cross references. Click on the text **4 Alias Names** and **2 Cross References** to expand the table with additional rows containing the ID and Source for each alias and DB Identifier and Source for each cross reference.

Alias and DBxref

Gene ID --> Alias ID + Database Cross Reference ID (1 rows)

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

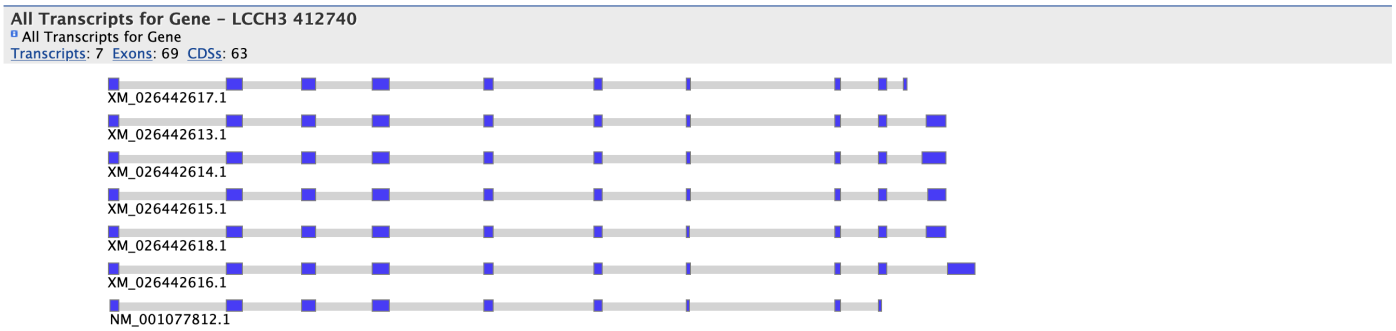
Showing 1 to 1 of 1 row

Gene DB identifier	Gene Aliases		Cross References Target	
412740	4 Alias Names		2 Cross References	
	Alias ID	Alias source	Target > DB identifier	Target > Source
	GB13122	amel_OGSv1.0	GB45541	amel_OGSv3.2
	GB13122	amel_OGSv1.1	GB45542	amel_OGSv3.2
	GB12078	amel_OGSv1.1		
	GB12078	amel_OGSv1.0		

3.3 Transcript

The **Transcript** 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.

Transcript



Gene --> Transcript --> Protein (49 rows)

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

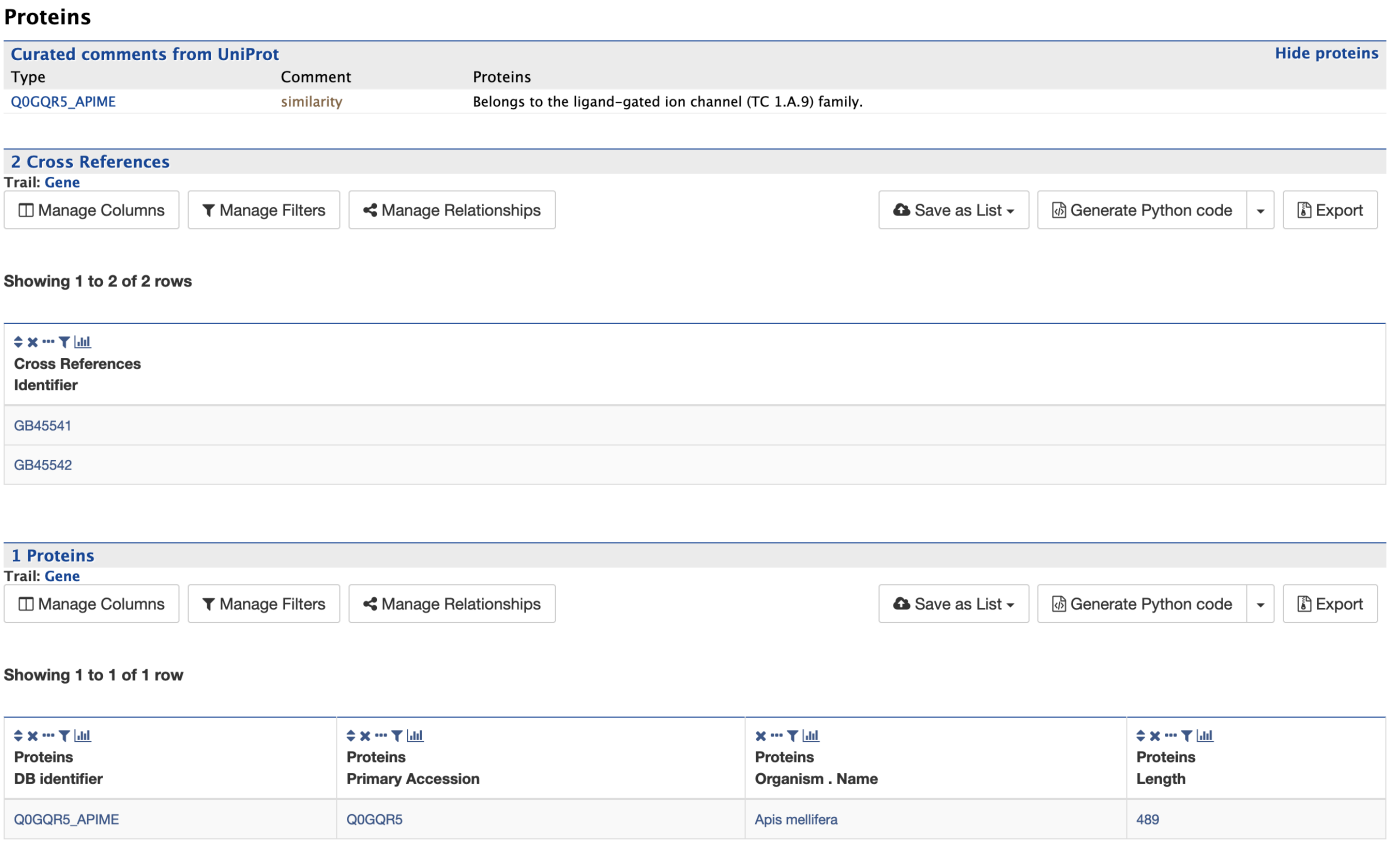
Showing 1 to 25 of 49 rows

Rows per page: 25 | page 1

Polypeptides DB identifier	Transcripts DB identifier
NP_001071280.1	NM_001077812.1
XP_026298398.1	NM_001077812.1
XP_026298399.1	NM_001077812.1
XP_026298400.1	NM_001077812.1
XP_026298401.1	NM_001077812.1

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.

Function

Gene Ontology

cellular component	
cytoplasmic vesicle membrane	IEA
membrane	IEA
transmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential	IEA
integral component of plasma membrane	IEA
postsynapse	IEA
neuron projection	IEA
cytoplasmic vesicle	IEA
chloride channel complex	IEA
GABA-ergic synapse	IEA
chemical synaptic transmission	IEA
cell periphery	IEA
postsynaptic membrane	IEA
synapse	IEA
plasma membrane	ECO:0000501 IEA
GABA-A receptor complex	IEA
integral component of membrane	ECO:0000501 IEA
molecular function	
ion channel activity	IEA
transmembrane signaling receptor activity	IEA
GABA-A receptor activity	ECO:0000501 IEA
transmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential	IEA
extracellular ligand-gated ion channel activity	ECO:0000501 IEA
identical protein binding	IEA
GABA-gated chloride ion channel activity	IEA
biological process	
ion channel activity	IEA
ion transport	IEA
signal transduction	IEA
transmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential	IEA
nervous system process	IEA
signaling	IEA
regulation of membrane potential	IEA
cell communication	IEA
chemical synaptic transmission	IEA
GABA-gated chloride ion channel activity	IEA
extracellular ligand-gated ion channel activity	ECO:0000501 IEA
cellular response to stimulus	IEA
ion transmembrane transport	IEA

1 Pathways

Trail: Gene

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

Showing 1 to 1 of 1 row

Pathways Identifier	Pathways Description	Pathways Curated	Pathways Name	Pathways Bsid	Pathways Pathway Map	Pathways Ko Pathway	Pathways Pathway Class	Pathways Short Name
ame04080	NO VALUE	NO VALUE	Neuroactive ligand-receptor interaction - Apis mellifera (honey bee)	NO VALUE	ame04080 Neuroactive ligand-receptor interaction	ko04080	Environmental Information Processing; Signaling molecules and interaction	NO VALUE

3.6 Homology

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

Homology

Homologues	
A. cephalotes	LOC105618116
A. cerana	LOC107995555
A. colombica	LOC108686942 LOC108691567 LOC108687631
A. dorsata	LOC102678684
A. echinatior	LOC105154584 LOC105149353 LOC105145387
A. florea	LOC100871112
A. rosae	LOC105691160 LOC105691302 LOC105690746
B. bifarius	LOC117211906
B. impatiens	BIMP19786 BIMP15224 LOC100740476 BIMP16483
B. terrestris	LOC100644703
B. treatae	LOC117172688
B. vancouverensis nearcticus	LOC117165961
B. vosnesenskii	LOC117239043
C. calcarata	LOC108632131 LOC108629559 LOC108627775
C. cinctus	LOC107262848 LOC107264860 LOC107272752
C. costatus	LOC108772025 LOC108773013 LOC108778372
C. floridanum	LOC106646337 LOC106641367 LOC106642283 LOC106643136 LOC106635744
C. floridanus	LOC105255134 LOC105255944 LOC105248902
C. insularis	LOC118065548
C. obscurior	Cobs_08353 Cobs_14722 Cobs_11662
C. solmsi marchali	LOC105364972 LOC105360539 LOC105362838
D. alloeum	LOC107046645 LOC107043876 LOC107046261 LOC107046593
D. melanogaster	GluClalpha Rdl Lcch3
D. novaeangliae	LOC107194170 LOC107187561 LOC107193885
D. quadriceps	LOC106750929 LOC106749525 LOC106749678
E. mexicana	LOC108554687 LOC108547134 LOC108549167
F. arisanus	LOC105263612 LOC105264309 LOC105267836
F. exsecta	LOC115234995
H. laboriosa	LOC108571111 LOC108570436 LOC108570653
H. saltator	LOC105192239 LOC105182093 LOC105186749
L. albipes	LALB18529 LALB16844 LALB20630 LALB20629
L. humile	LOC105667704 LOC105671992 LOC105678045
M. demolitor	LOC103572802 LOC103568640
M. genalis	LOC117227114
M. pharaonis	LOC105836172 LOC105831134 LOC105839323
M. quadrifasciata	MQUA10111 WN51_08026 MQUA19292 MQUA18245
M. rotundata	LOC100883591 LOC100882308 LOC100883102
N. fulva	LOC114945379
N. lecontei	LOC107218826 LOC107220656 LOC107216469
N. melanderi	LOC116429191
N. vitripennis	LOC100116668 LCCH3 GluCl
O. abietinus	LOC105699045 LOC105699375 LOC105699763
O. bicornis bicornis	LOC114873402
O. biroi	LOC105286156
O. brunneus	LOC116842266
O. lignaria	LOC117609463
P. barbatus	LOC105428153 LOC105423658 LOC105433259
P. canadensis	LOC106790521 LOC106783657 LOC106786215
P. dominula	LOC107069426
P. gracilis	LOC109860178 LOC109862073 LOC109863815
S. invicta	LOC105194445 LOC105199392 LOC105203964
T. cornetzi	LOC108768409
T. curvispinosus	LOC112456292
T. pretiosum	LOC106660520 LOC106652593 LOC106650505
T. septentrionalis	LOC108747081 LOC108749946 LOC108746332
T. zeteki	LOC108724512
V. emeryi	LOC105562681 LOC105566628 LOC105567127
W. auropunctata	LOC105455966 LOC105449299 LOC105448904

Gene ID --> Homologues (338 rows)

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

Showing 1 to 25 of 338 rows

Rows per page: 25 | page 1

Gene DB identifier	Gene Symbol	Gene Organism . Short Name	Homologues Homologue . Primary Identifier	Homologues Homologue . Organism . Short Name	Homologues Type	Homologues Last Common Ancestor
412740	LCCH3	A. mellifera	100116310	N. vitripennis	orthologue	Holometabola
412740	LCCH3	A. mellifera	100116310	N. vitripennis	orthologue	Holometabola
412740	LCCH3	A. mellifera	100116310	N. vitripennis	orthologue	Hymenoptera
412740	LCCH3	A. mellifera	100116310	N. vitripennis	orthologue	Hymenoptera
412740	LCCH3	A. mellifera	100116668	N. vitripennis	orthologue	Holometabola
412740	LCCH3	A. mellifera	100116853	N. vitripennis	orthologue	Holometabola
412740	LCCH3	A. mellifera	100644703	B. terrestris	orthologue	Aculeata

3.7 Publications

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

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:

1. Chromosome Distribution
2. Gene Ontology Enrichment
3. Protein Domain Enrichment
4. Publication Enrichment
5. Pathway Enrichment
6. Orthologues

The selection of widgets provided on the List Analysis page depend on the contents of the list. In the above example, three widgets appear: Gene Ontology Enrichment, Publication Enrichment, and Pathway Enrichment.

Read the **Important Notes for Enrichment Widgets** for special instructions to avoid false positives.

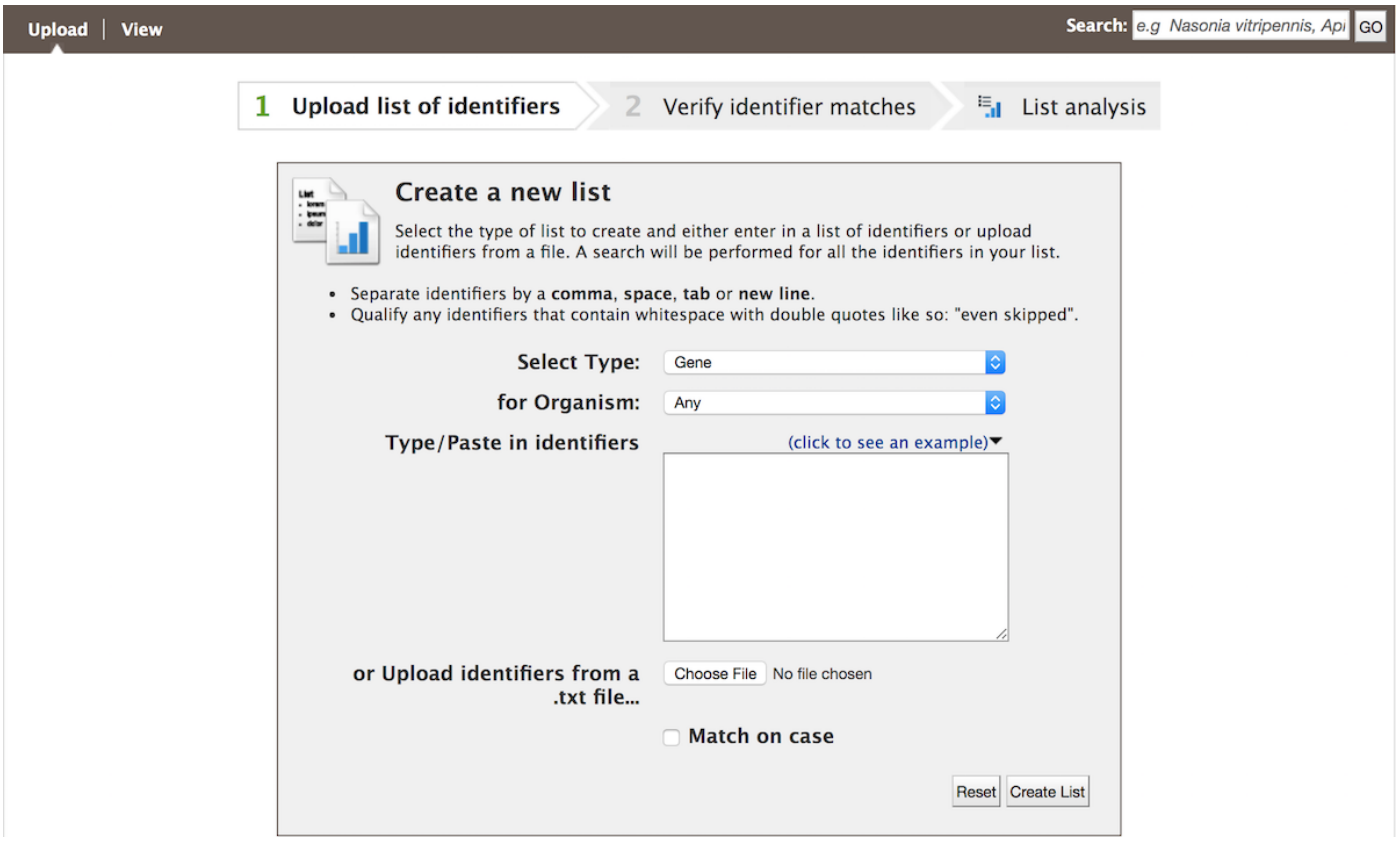
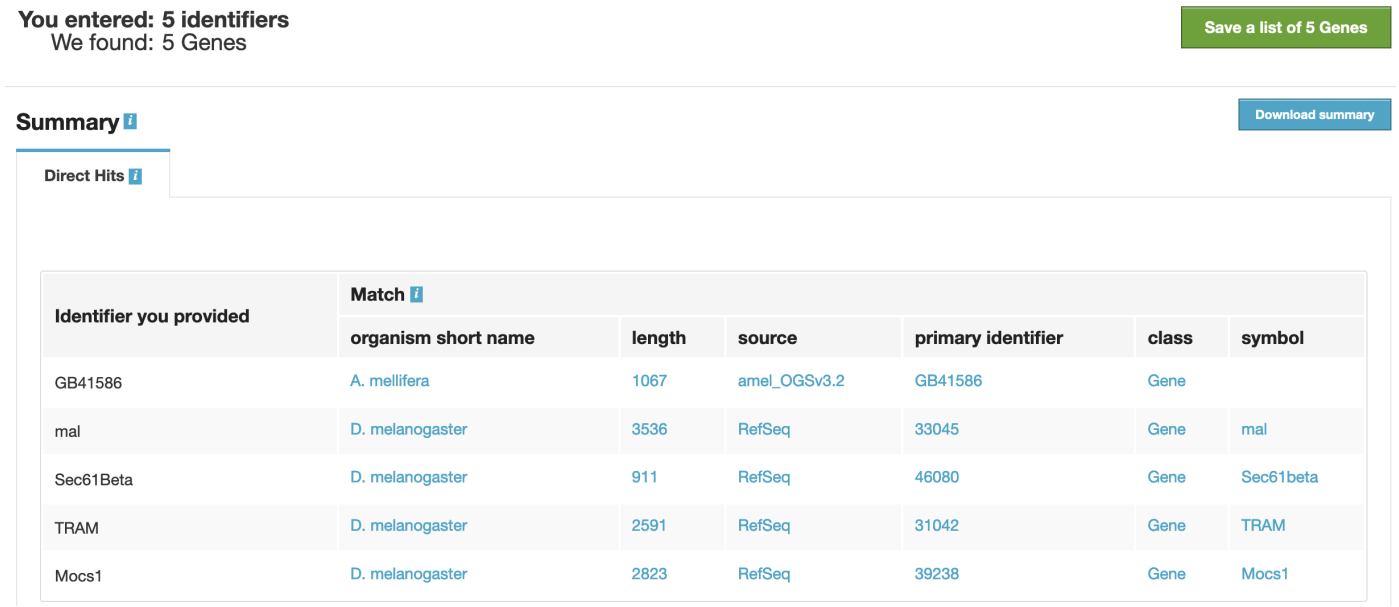


Fig. 1: List upload form



Identifier you provided	Match					
	organism short name	length	source	primary identifier	class	symbol
GB41586	A. mellifera	1067	amel_OGSv3.2	GB41586	Gene	
mal	D. melanogaster	3536	RefSeq	33045	Gene	mal
Sec61Beta	D. melanogaster	911	RefSeq	46080	Gene	Sec61beta
TRAM	D. melanogaster	2591	RefSeq	31042	Gene	TRAM
Mocs1	D. melanogaster	2823	RefSeq	39238	Gene	Mocs1

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

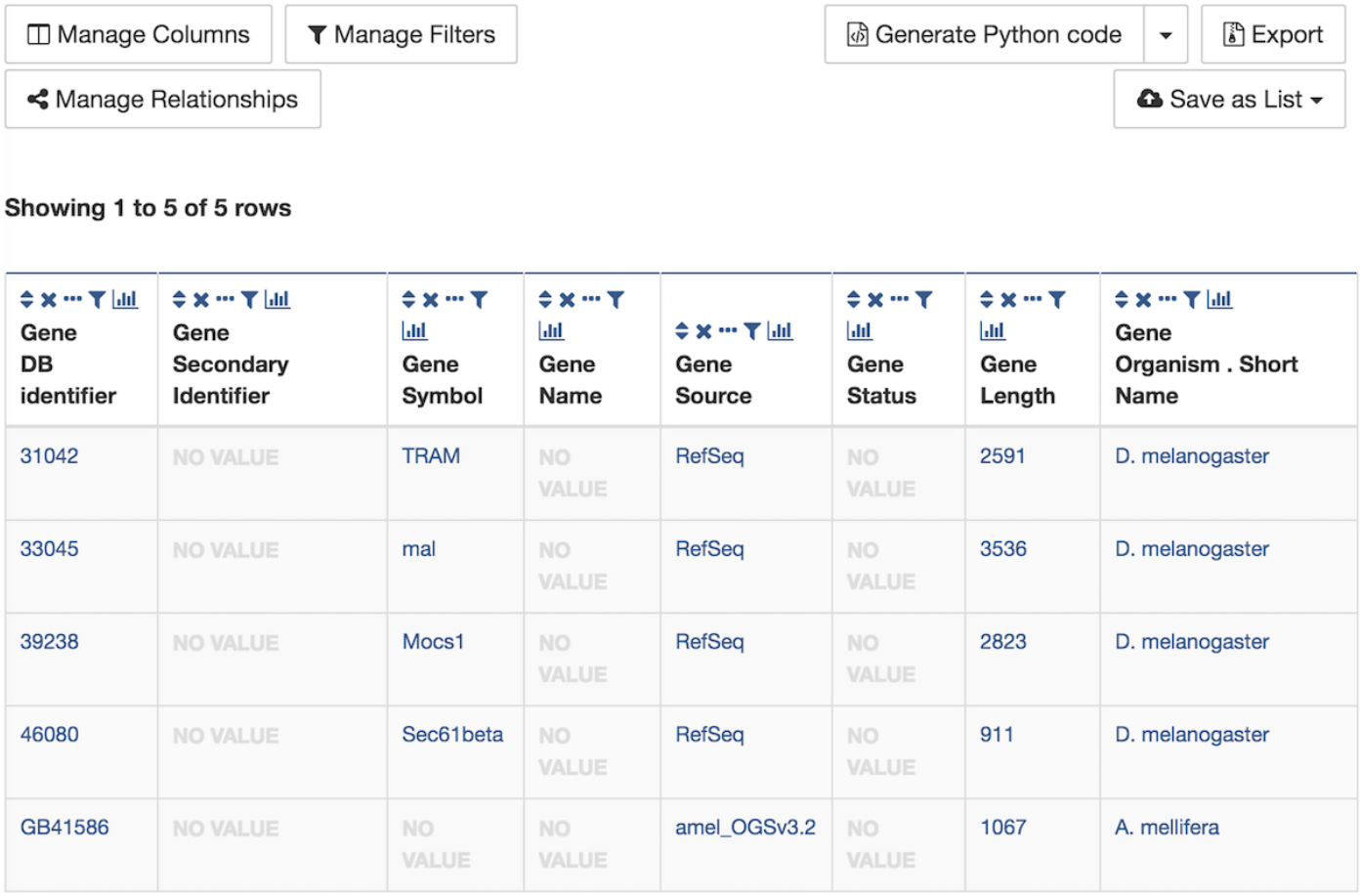


Fig. 3: Example: List analysis for gene list

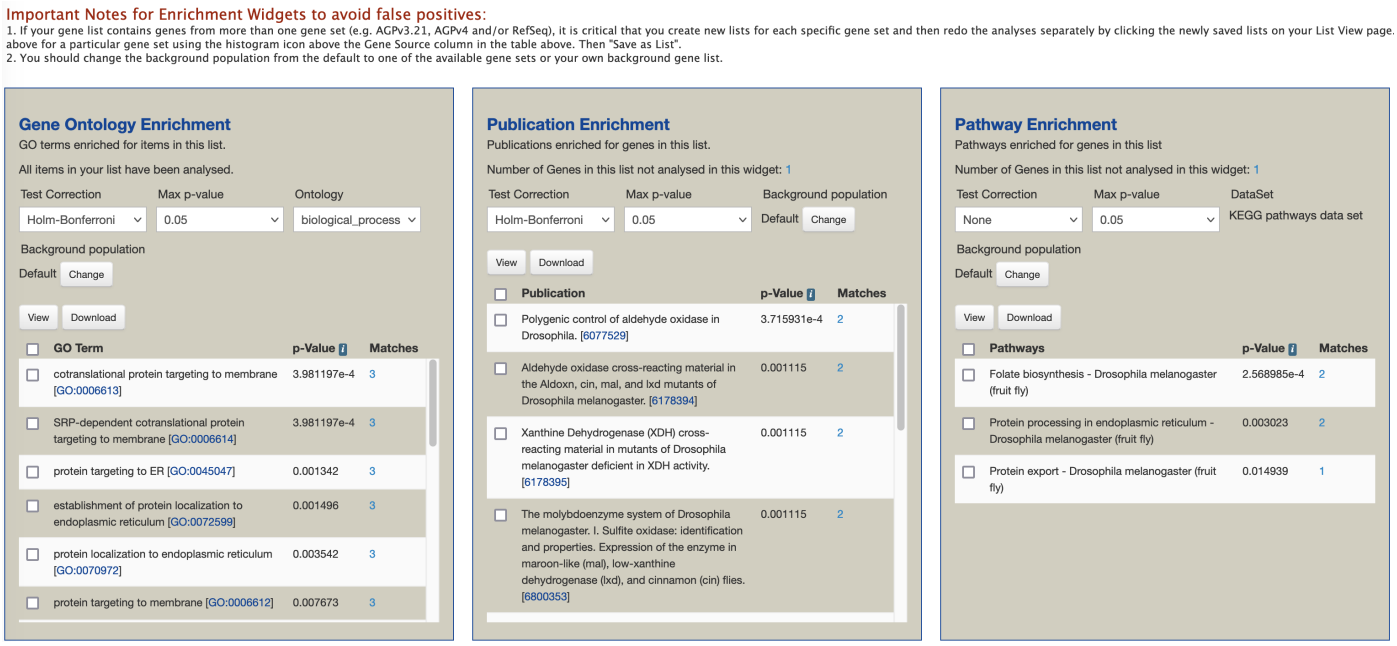


Fig. 4: Example: Available widgets

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.

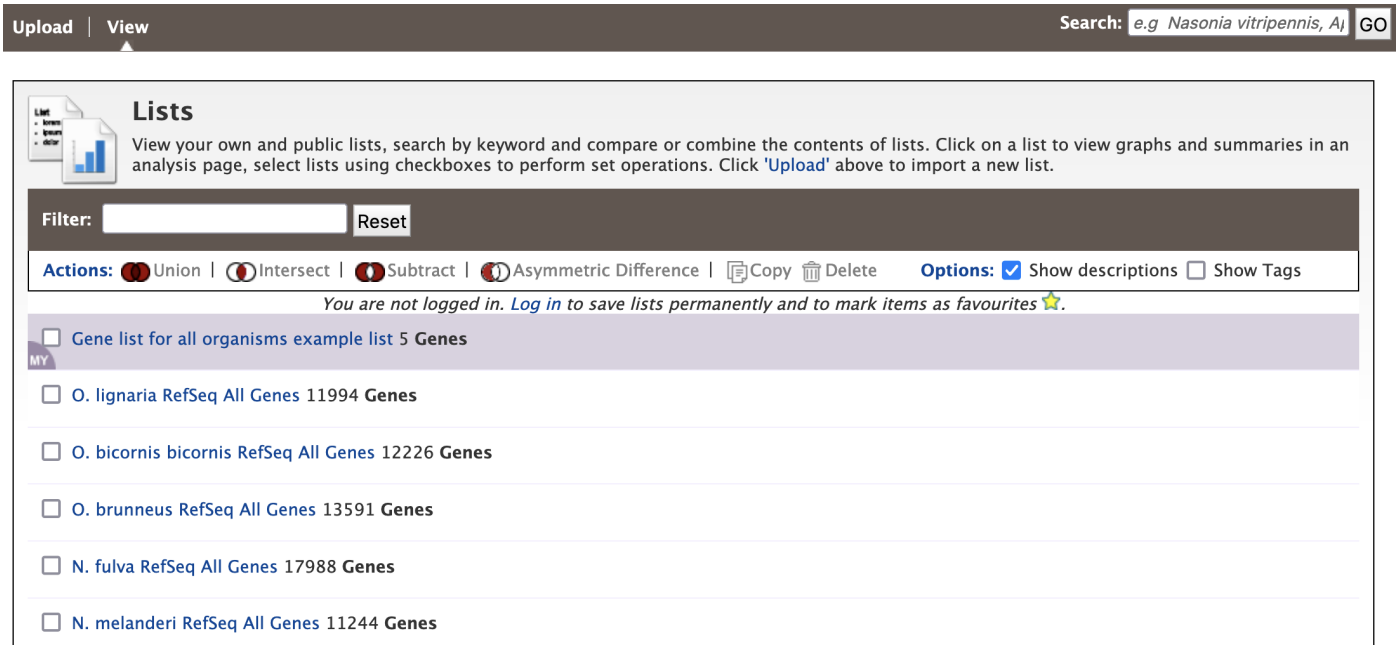


Fig. 5: Saved lists. Lists belonging to user are highlighted.

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.

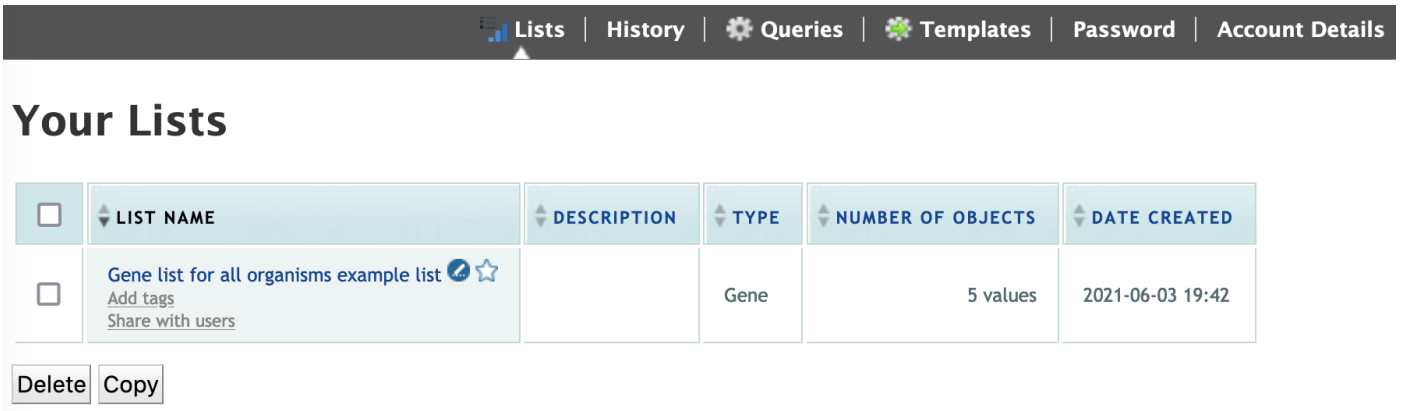
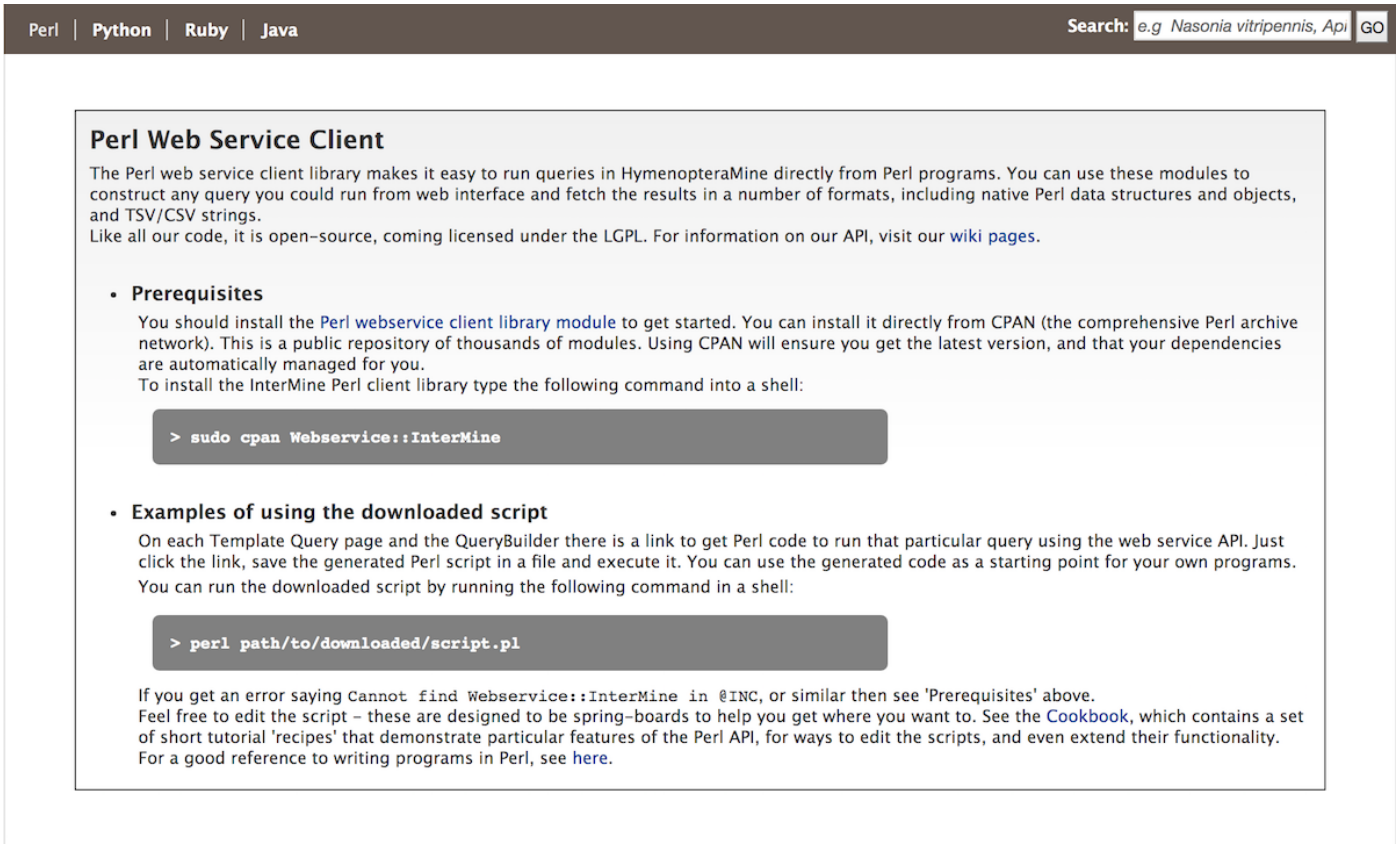


Fig. 1: MyMine

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.

Data Category	Data	Organism	Source	PubMed	Link
	Consortium Official Gene Set	Acromyrmex echinatior	aech_OGSv3.8	Nygaard et al. - PubMed: 21719571	HGD OGS GFF3 Download
		Apis mellifera	amel_OGSv3.2	Elsik et al. - PubMed: 24479613	HGD OGS GFF3 Download
		Atta cephalotes	acep_OGSv1.2	Suen et al. - PubMed: 21347285	HGD OGS GFF3 Download
		Bombus impatiens	bimp_OGSv1.0	Sadd et al. - PubMed: 25908251	HGD OGS GFF3 Download
		Camponotus floridanus	cflo_OGSv3.3	Bonasio et al. - PubMed: 20798317	HGD OGS GFF3 Download
		Cardiocondyla obscurior	cobs_OGSv1.4	Schrader et al. - PubMed: 25510865	HGD OGS GFF3 Download
		Dufourea novaeangliae	Dufourea_novaeangliae_v1.1	Kapheim et al. - PubMed: 25977371	HGD OGS GFF3 Download
		Eufriesea mexicana	Eufriesea_mexicana_v1.1	Kapheim et al. - PubMed: 25977371	HGD OGS GFF3 Download
		Habropoda laboriosa	Habropoda_laboriosa_v1.2	Kapheim et al. - PubMed: 25977371	HGD OGS GFF3 Download
		Harpegnathos saltator	hsal_OGSv3.3	Bonasio et al. - PubMed: 20798317	HGD OGS GFF3 Download
		Lasioglossum albipes	lalb_OGSv5.42	Kocher et al. - PubMed: 24359881	HGD OGS GFF3 Download
		Linepithema humile	lhum_OGSv1.2	Smith et al. - PubMed: 21282631	HGD OGS GFF3 Download
		Megachile rotundata	Megachile_rotundata_v1.1	Kapheim et al. - PubMed: 25977371	HGD OGS GFF3 Download
		Melipona quadrifasciata	Melipona_quadrifasciata_v1.1	Kapheim et al. - PubMed: 25977371	HGD OGS GFF3 Download
		Nasonia vitripennis	nvit_OGSv1.2	Werren JH et al. - PubMed: 20075255	HGD OGS GFF3 Download
			Nvit_EviGene	Werren JH et al. - PubMed: 20075255	HGD OGS GFF3 Download
		Ooceraea biroi	armyant.OGS.V1.8.6	Oxley PR et al. - PubMed: 24508170	HGD OGS GFF3 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.

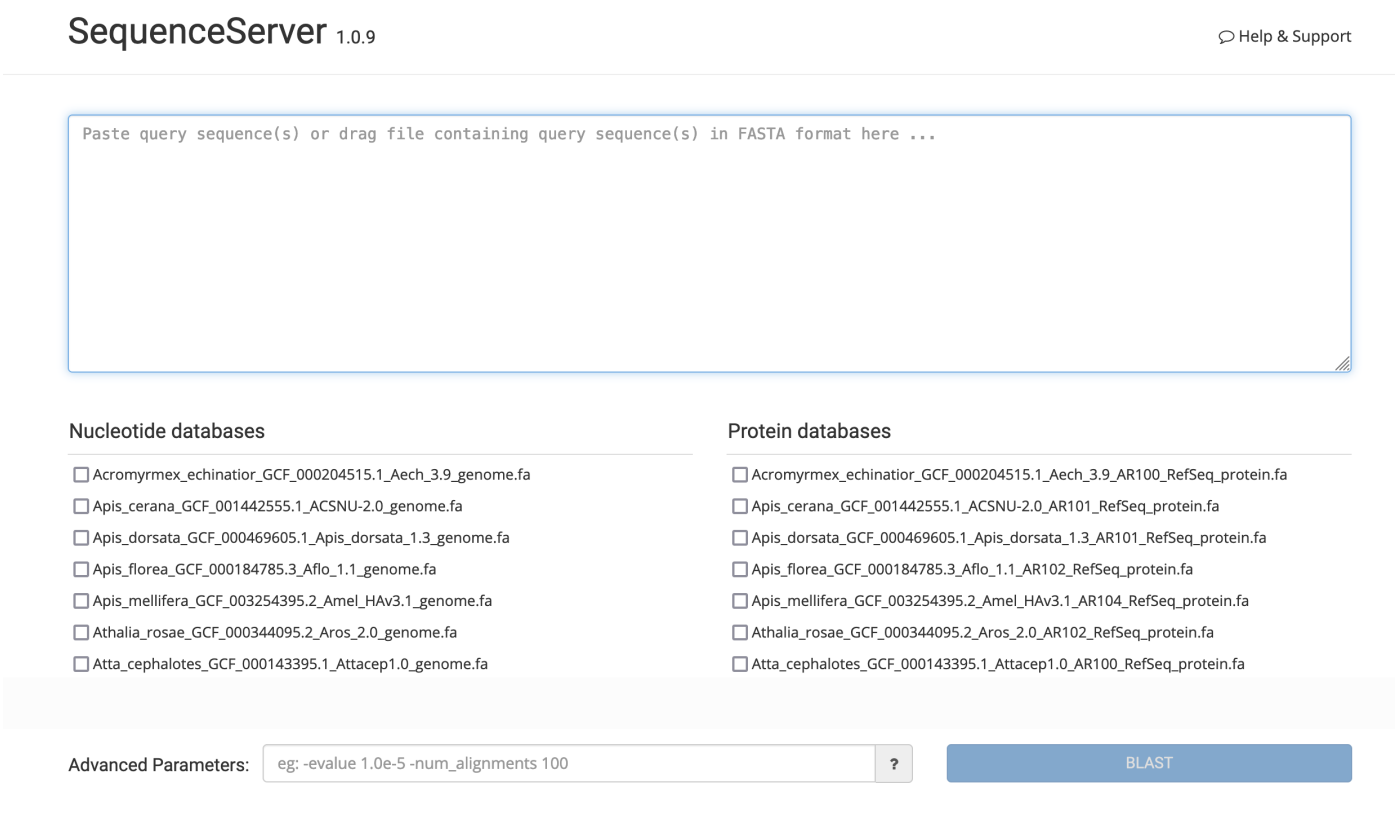


Fig. 1: BLAST search

CHAPTER 9

How to Cite

Please cite the use of HymenopteraMine as follows:

Elsik CG, Tayal A, Diesh CM, Unni DR, Emery ML, Nguyen HN, Hagen DE. Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine. Nucleic Acids Research 2016 Jan 4;44(D1):D793-800. doi: 10.1093/nar/gkv1208. Epub 2015 Nov 17. PubMed PMID: 26578564.

Elsik, C., Tayal, A., Diesh, C., Unni, D., Emery, M., Nguyen, H., & Hagen, D. (2016, Jan 4). HymenopteraMine. Retrieved [Date], from http://hymenopteragenome.org/hymenopteramine/begin.do.

Please cite the use of a specific page in HymenopteraMine as follows:

Elsik, C., Tayal, A., Diesh, C., Unni, D., Emery, M., & Hagen, D. (2016, Jan 4). [Title of page]. Retrieved [Date], from [link of page].

Please cite the use of any other tools and data from the HGD website as follows:

Elsik CG, Tayal A, Diesh CM, Unni DR, Emery ML, Nguyen HN, Hagen DE. Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine. Nucleic Acids Research 2016 Jan 4;44(D1):D793-800. doi: 10.1093/nar/gkv1208. Epub 2015 Nov 17. PubMed PMID: 26578564.

In addition, please cite the use of genome consortium data with the appropriate genome consortium publication.