
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>

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.

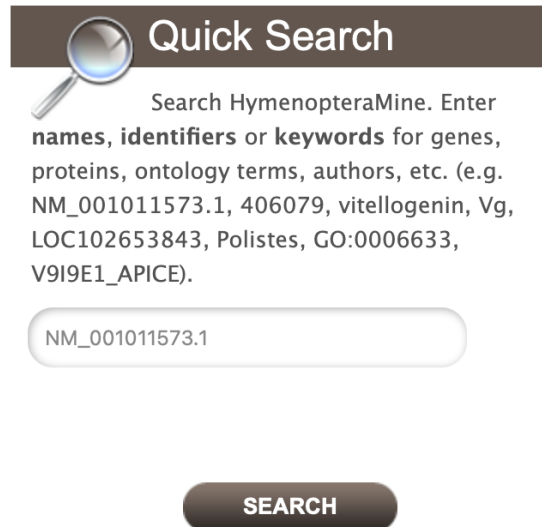
HGD Blast – Links to a Blast page where users may BLAST their sequence(s) of interest with the Hymenoptera species reference genome, CDS sequences, and protein sequences.

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.

The image shows a 'Quick Search' interface. It features a dark brown header bar with a magnifying glass icon and the text 'Quick Search'. Below this, there is a text input area with a light gray background and rounded corners. The input area contains the text 'NM_001011573.1'. To the right of the input area, there is a dark brown button with the word 'SEARCH' in white capital letters. The text 'Search HymenopteraMine. Enter names, identifiers or keywords for genes, proteins, ontology terms, authors, etc. (e.g. NM_001011573.1, 406079, vitellogenin, Vg, LOC102653843, Polistes, GO:0006633, V9I9E1_APICE).' is displayed above the input area.

Quick Search

Search HymenopteraMine. Enter names, identifiers or keywords for genes, proteins, ontology terms, authors, etc. (e.g. NM_001011573.1, 406079, vitellogenin, Vg, LOC102653843, Polistes, GO:0006633, V9I9E1_APICE).

SEARCH

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.

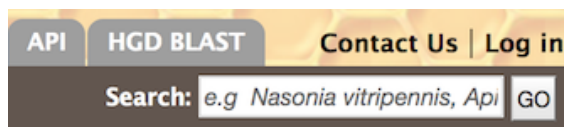


Fig. 2: Quick Search from any page

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.

Search our database by keyword

LCCH3 Search

Examples

- Search this entire website. Enter **identifiers, names or keywords** for genes, pathways, authors, ontology terms, etc. (e.g. *eve*, *embryo*, *zen*, *allele*)
- Use **OR** to search for either of two terms (e.g. *fly OR drosophila*) or quotation marks to search for phrases (e.g. *"dna binding"*).
- Boolean search syntax** is supported: e.g. *dros** for partial matches or *fly AND NOT embryo* to exclude a term

Search results 1 to 19 out of 19 for LCCH3

Categories

Hits by Category

- mRNA: 10
- Gene: 3
- Protein: 3
- Publication: 3

Hits by Organism

- A. mellifera: 9
- D. melanogaster: 3
- N. vitripennis: 3
- F. arisanus: 1

Type	Details	Score
Gene	100116310 LCCH3 Source: RefSeq Biotype: protein_coding Description: ligand-gated chloride channel homolog 3 Length: 14499 FASTA Chromosome: 3: 21105173-21119671 Location: Nasonia vitripennis Organism: Nvit_psr_1.1 Assembly:	*****
Gene	32554 Lcch3 Source: RefSeq Biotype: protein_coding Description: Ligand-gated chloride channel homolog 3 Length: 3425 FASTA Chromosome: X: 15925105-15928529 Location: Drosophila melanogaster Organism: Release_6_plus_ISO1_MT Assembly:	*****
Gene	412740 LCCH3 Source: RefSeq Biotype: protein_coding Description: ligand-gated chloride channel homolog 3 Length: 13343 FASTA Chromosome: LG9: 1324939-1338281 Location: Apis mellifera Organism: Amel_HAV3.1 Assembly:	*****
mRNA	NM_206746.2 Source: RefSeq Length: 1903 FASTA Assembly: Release_6_plus_ISO1_MT	*****
mRNA	XM_026442616.1 Source: RefSeq Length: 2027 FASTA Assembly: Amel_HAV3.1	*****

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

Search results 1 to 3 out of 3 for *LCCH3*
Category restricted to Gene 

0.12s

Categories

Category: *Gene*
« show all

Hits by Organism

- A. mellifera: 1
- D. melanogaster: 1
- N. vitripennis: 1

Type	Details	Score
<input type="checkbox"/> Gene	100116310 LCCH3 Source: RefSeq Biotype: protein_coding Description: ligand-gated chloride channel homolog 3 Length: 14499 FASTA Chromosome: 3: 21105173-21119671 Location: Nasonia vitripennis Organism: Nvit_psr_1.1 Assembly:
<input type="checkbox"/> Gene	32554 Lcch3 Source: RefSeq Biotype: protein_coding Description: Ligand-gated chloride channel homolog 3 Length: 3425 FASTA Chromosome: X: 15925105-15928529 Location: Drosophila melanogaster Organism: Release_6_plus_ISO1_MT Assembly:
<input type="checkbox"/> Gene	412740 LCCH3 Source: RefSeq Biotype: protein_coding Description: ligand-gated chloride channel homolog 3 Length: 13343 FASTA Chromosome: LG9: 1324939-1338281 Location: Apis mellifera Organism: Amel_HAv3.1 Assembly:

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

GENES	PROTEINS	HOMOLOGY	FUNCTION	ENTIRE GENE SET	ALIAS AND DBXREF
-------	----------	----------	----------	-----------------	------------------

Gene models in HymenopteraMine come from multiple sources including NCBI RefSeq, Ensembl Rapid Release and consortium official gene sets. Consortium data sources are labeled using the original consortium official gene set names, and may have "HGD" appended, if the data source has ids assigned by HGD. Use "ALIAS AND DBXREF" templates to convert identifiers.

Query for genes:


- Gene ID ➡ Coding Sequences
- Gene ID ➡ Transcript id(s)
- Gene ➡ Transcript + Exon
- Gene ➡ Chromosomal location
- Gene ID ➡ Gene Symbol and Description
- Gene Symbol ➡ Gene ID
- Organism and Chromosome ➡ Genes
- RefSeq Gene Set ➡ Symbols and Descriptions

» [More queries](#)

popular templates

Fig. 5: Popular templates

Please contact us if you would like any additional template queries or if you have a concern about a query not completing.




Templates

Templates are predefined queries, each has a simple form and a description. Click on a template to run it, you can search for templates by keyword and filter them by category.


Filter: Filter: -- all categories --

Actions: Options: ☒ Show descriptions ☐ Show Tags

You are not logged in. [Log in](#) to mark items as favourites .


- ☐ **Gene ID --> GO Terms**
 Given a gene id, retrieve GO terms.
- ☐ **Organism --> All Gene IDs --> Gene Alias IDs**
 Retrieve gene ids and their aliases for an organism. Multiple gene sets may be returned if the gene source is not selected. See the Data Source page for gene source names. This template can be used for *A. dorsata*, *A. florea*, *A. mellifera*, *B. impatiens*, *B. terrestris*, *L. humile*, *P. barbatus*, *W. auropunctata*.
- ☐ **Gene ID --> Homologues**
 Given a gene id, retrieve all homologues.
- ☐ **Organism --> Homologues**
 Retrieve homologues for an organism.
- ☐ **Alias ID --> Gene ID**
 Given an Alias ID, retrieve Gene ID, optionally constrained by gene source. This template can be used for *A. dorsata*, *A. florea*, *A. mellifera*, *B. impatiens*, *B. terrestris*, *L. humile*, *P. barbatus*, *W. auropunctata*.
- ☐ **Gene Symbol --> GO Terms**
 Given a Gene Symbol, retrieve GO terms.
- ☐ **Gene ID --> Pathways**
 Given a gene id, retrieve pathways.




Fig. 6: List of templates on the Templates page



GO Term → Gene

Given a GO term, retrieve genes annotated with that GO Term for a selected organism.

Ontology Term > Ontology Term
 LOOKUP: 

Organism > Short Name
   

[web service URL](#) | [Perl](#) | [Python](#) | [Ruby](#) | [Java](#) [\[help\]](#) | [export XML](#)

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

Showing 1 to 25 of 1,192 rows Rows per page: 25

GO Annotation Ontology Term Identifier	GO Annotation Ontology Term Name	GO Annotation Ontology Term Description	Code Code	Gene Gene ID	Gene Gene Symbol	Organism Organism Short Name	Data Sets Data Sets Name	Data Source 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.

page 1

Python
 Perl
 Java
 Ruby
 JavaScript
 XML

Gene
Symbol

Organism
Short Name

Data Source
Name

LOC105616992 A. cephalotes UniProt

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.

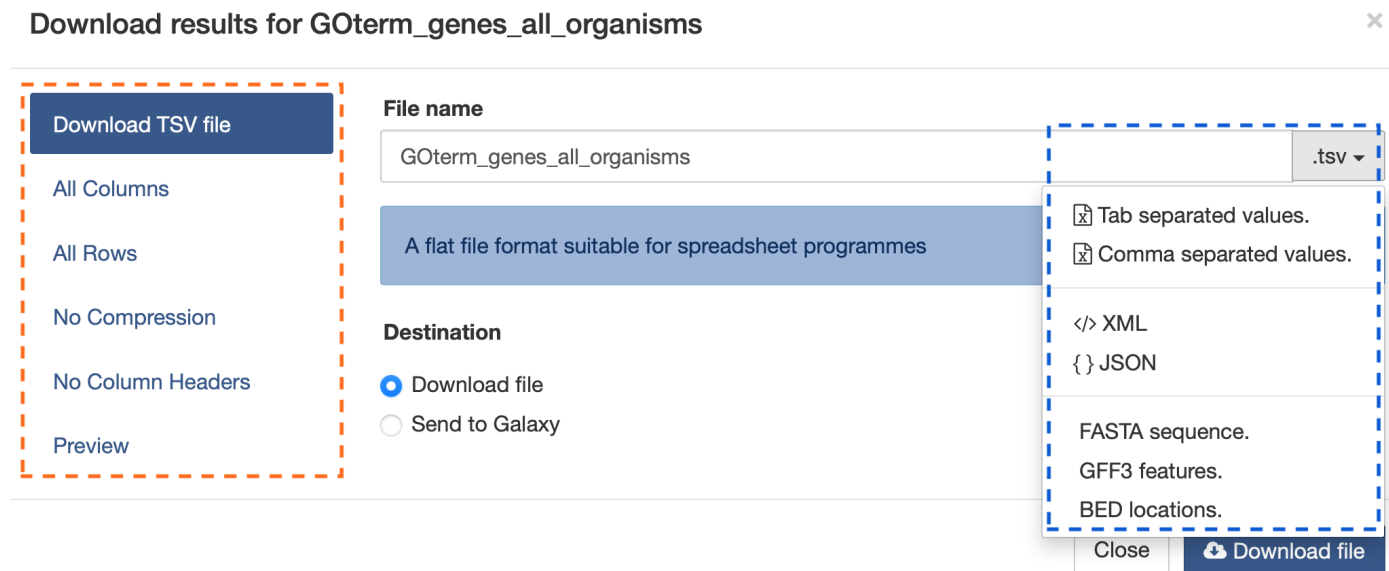


Fig. 10: Download results options

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.

Organism > Gene IDs with their Alias IDs
 Retrieve gene ids and their aliases for an organism. Multiple gene sets may be returned if the gene source is not selected. See the Data Source page for gene source names. This template can be used for *A. dorsata*, *A. florea*, *A. mellifera*, *B. impatiens*, *B. terrestris*, *L. humile*, *P. barbatus*, *W. auropunctata*.

Organism > Short Name
 =

Gene > Source
 optional ☐ =
 ON | OFF ☐ constrain to be IN

[Show Results](#) [Edit Query](#)

[web service URL](#) [Perl](#) | [Python](#) | [Ruby](#) | [Java](#) [\[help\]](#) [export XML](#)

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

QueryBuilder
 Advanced users can use a flexible query interface to construct their own data mining queries. The QueryBuilder lets you view the data model, apply constraints and select output. You can also export queries to share them with others.

[Browse data model](#) [Import query from XML](#) [Login to view saved queries](#)

Select a Data Type to Begin a Query

Click on a class name for a description or double click on a class name to create a new query starting at that class

Gene
 Protein

 Alias Name
 Annotatable
 Author
 Bio-Entity
 CDS
 Chromosome
 Coding Sequence

[Select](#)

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

Select a Data Type to Begin a Query

Click on a class name for a description or double click on a class name to create a new query starting at that class

Gene
 Protein

 Alias Name
 Annotatable
 Author
 Bio-Entity
 CDS
 Chromosome
 Coding Sequence

Gene: "Gene" is an abstract term used to describe a collection of transcripts and related regulatory features.

[Select](#)

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.

Model browser

Browse through the classes and attributes. Click on **SUMMARY** links to add summary of fields to the results table or on **SHOW** links to add individual fields to the results. Use **CONSTR** links to constrain a value in the query.

Gene **SUMMARY** **CONSTR**

- Biotype **SHOW** **CONSTR**
- Description **SHOW** **CONSTR**
- Exception **SHOW** **CONSTR**
- Length **Integer** **SHOW** **CONSTR**
- Name **SHOW** **CONSTR**
- Note **SHOW** **CONSTR**
- Gene ID **SHOW** **CONSTR**
- Score Double **SHOW** **CONSTR**
- Source **SHOW** **CONSTR**
- Symbol **SHOW** **CONSTR**
- Aliases Alias Name **SUMMARY** **CONSTR**
- CDSs CDS **SUMMARY** **CONSTR**
- Chromosome Chromosome **SUMMARY** **CONSTR**
- Chromosome Location Location **SUMMARY** **CONSTR**
- Coding Sequences Coding Sequence **SUMMARY** **CONSTR**
- Cross References Cross Reference **SUMMARY** **CONSTR**
- Data Sets Data Set **SUMMARY** **CONSTR**
- Exons Exon **SUMMARY** **CONSTR**

☐ Show empty fields

Fields selected for output

Columns to Display

Use the **SHOW** or **SUMMARY** links to add fields to the results table.

No fields selected for output

Query Overview

no fields constrained

Constraint logic:

no constraints

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.

Model browser

Browse through the classes and attributes. Click on [SUMMARY](#) links to add summary of fields to the results table or on [SHOW](#) links to add individual fields to the results. Use [CONSTRAIN](#) links to constrain a value in the query.

- Gene
 - Biotype
 - SHOW
 - CONSTRAIN
 - Description
 - SHOW
 - CONSTRAIN
 - Exception
 - SHOW
 - CONSTRAIN
 - Length Integer
 - SHOW
 - CONSTRAIN
 - Name
 - SHOW
 - CONSTRAIN
 - Note
 - SHOW
 - CONSTRAIN
 - Gene ID
 - SHOW
 - CONSTRAIN
 - Score Double
 - SHOW
 - CONSTRAIN
 - Source
 - SHOW
 - CONSTRAIN
 - Symbol
 - SHOW
 - CONSTRAIN
 - Aliases Alias Name
 - SUMMARY
 - CONSTRAIN
 - CDSs CDS
 - SUMMARY
 - CONSTRAIN
 - Chromosome Chromosome
 - SUMMARY
 - CONSTRAIN
 - Chromosome Location Location
 - SUMMARY
 - CONSTRAIN
 - Coding Sequences Coding Sequence
 - SUMMARY
 - CONSTRAIN
 - Cross References Cross Reference
 - SUMMARY
 - CONSTRAIN
 - Data Sets Data Set
 - SUMMARY
 - CONSTRAIN

☐ Show empty fields

Query Overview

Gene

- Gene ID
- Symbol

Constraint logic:

no constraints

Fields selected for output

Columns to Display

Use the [SHOW](#) or [SUMMARY](#) links to add fields to the results table. Click and drag the blue output boxes to choose the output column order. Click [\[1\]](#) to choose a column to sort results by, click again to select ascending [\[1\]](#) or descending [\[2\]](#). Use the [REMOVE ALL](#) link to remove all fields from the results table.

[REMOVE ALL](#)

Gene > Gene ID (no description) [\[1\]](#)

Gene > Symbol (no description) [\[2\]](#)

[Show results](#)

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

Constrain for Gene > Biotype

Constrain

Choose a filter

Filter query results on this field having a specific value

Gene > Biotype

= [\[1\]](#) protein_coding [\[2\]](#)

OR ☐ Contained in list: IN [\[1\]](#) A. mellifera OGSv3.2 genes mapping to more than one chromosome [\[2\]](#) [Add to query](#)

Filter query results on this field having any value or not

☒ Has no value ☐ Has a value [Add to query](#)

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

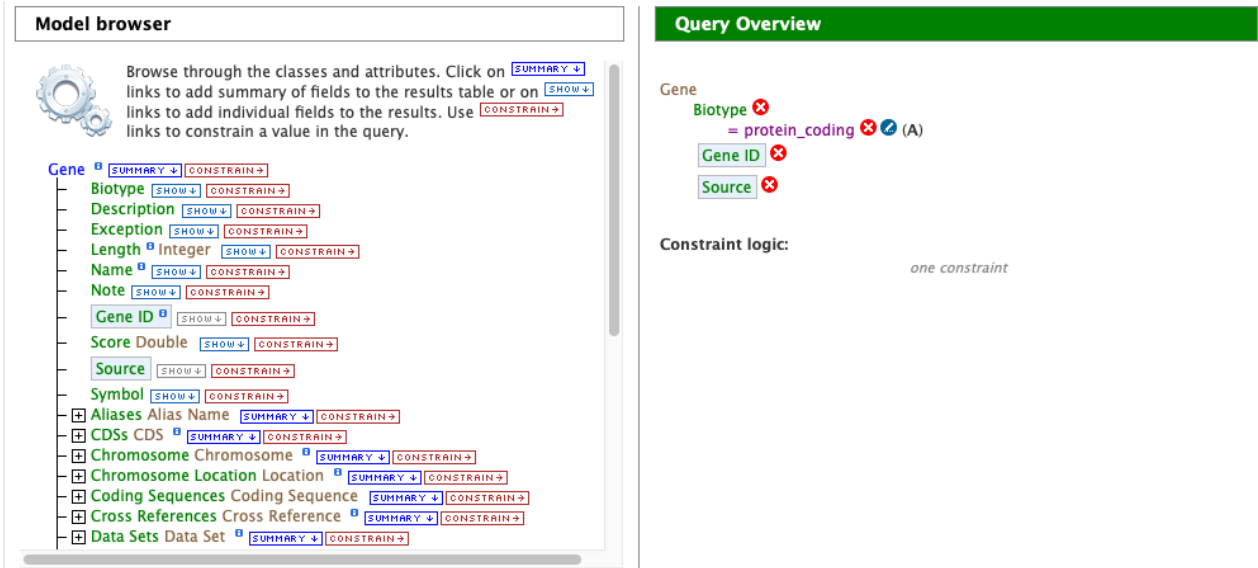


Fig. 16: Constraint added to query

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.

Trail: [Query](#) > Results

Manage Columns

Manage Filters

Manage Relationships

Save as List

Generate Python code

Export

Showing rows 1 to 25 of 1,462,000

Rows per page: 25

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

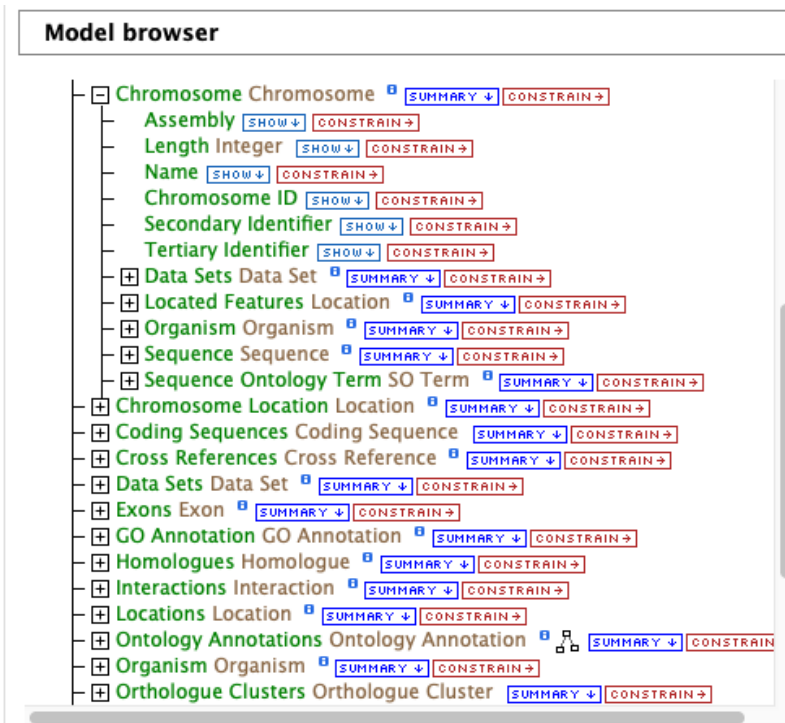


Fig. 19: Step 1: View attributes of the Chromosome feature class

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*.

Model browser

Browse through the classes and attributes. Click on **SUMMARY** links to add summary of fields to the results table or on **SHOW** links to add individual fields to the results. Use **CONSTRAIN** links to constrain a value in the query.

Gene **SUMMARY** **CONSTRAIN**

Biotype **SHOW** **CONSTRAIN**

Description **SHOW** **CONSTRAIN**

Exception **SHOW** **CONSTRAIN**

Length Integer **SHOW** **CONSTRAIN**

Name **SHOW** **CONSTRAIN**

Note **SHOW** **CONSTRAIN**

Gene ID **SHOW** **CONSTRAIN**

Score Double **SHOW** **CONSTRAIN**

Source **SHOW** **CONSTRAIN**

Symbol **SHOW** **CONSTRAIN**

Aliases Alias Name **SUMMARY** **CONSTRAIN**

CDSs CDS **SUMMARY** **CONSTRAIN**

Chromosome Chromosome **SUMMARY** **CONSTRAIN**

Assembly **SHOW** **CONSTRAIN**

Length Integer **SHOW** **CONSTRAIN**

Name **SHOW** **CONSTRAIN**

Chromosome ID **SHOW** **CONSTRAIN**

Query Overview

Gene

Biotype **×**

= protein_coding **×** **⚙** (A)

Gene ID **×**

Symbol **×**

Chromosome Chromosome **×** **⚙**

Chromosome ID **×**

= NW_003791143.1 **×** **⚙** (B)

Constraint logic: A and B

A and B **✎**

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.

Trail: Query > Results

Manage Columns

Manage Filters

Manage Relationships

Save as List

Generate Python code

Export

Showing rows 1 to 25 of 625

Rows per page: 25

⏪

⏩

⏴

⏵

page 1

⏴

⏵

⏴

⏵

Gene	Gene ID
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**.

The screenshot displays the HymenopteraMine QueryBuilder interface. On the left, the **Model browser** shows a hierarchical tree of biological features. The **Exons** feature class is expanded, showing attributes like **DB Identifier** and **Length**, which have 'SHOW' buttons next to them. Below the tree is a checkbox for 'Show empty fields'. On the right, the **Query Overview** section shows the current query logic: **Gene** (with constraints for **Biotype** and **Gene ID**) joined to **Exons** (with constraints for **Chromosome**, **Chromosome ID**, and **Exon collection**). The **Constraint logic** is shown as 'A and B'. Below this, the **Fields selected for output** section shows four columns: **Gene > Gene ID**, **Gene > Symbol**, **Gene > Exons > DB identifier**, and **Gene > Exons > Length**. Each column has a 'SHOW' button and a blue square icon for sorting.

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.

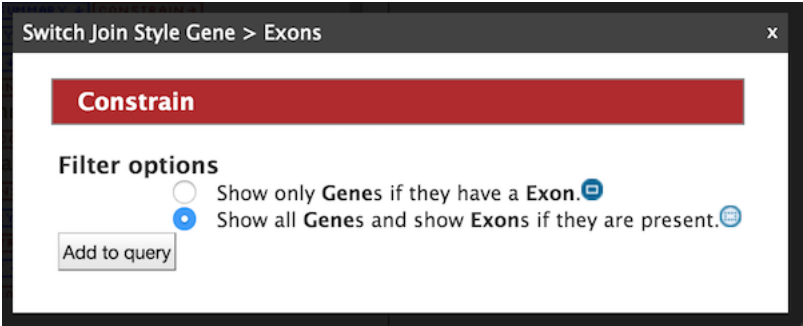


Fig. 23: Step 2: Change the join style

Trail: [Query](#) > Results

Manage Columns

Manage Filters

Manage Relationships

Save as List

Generate Python code

Export

Showing rows 1 to 25 of 625

Rows per page: 25

page 1

Gene Gene ID	Gene Symbol	Gene Exons
100862953	LOC100862953	12 Exons
100862997	LOC100862997	12 Exons
100862998	LOC100862998	29 Exons
100863028	LOC100863028	40 Exons
100863029	LOC100863029	12 Exons
100863069	LOC100863069	4 Exons

Fig. 24: Step 3: Display query results

Trail: [Query](#) > Results

Manage Columns

Manage Filters

Manage Relationships

Save as List

Generate Python code

Export

Showing 1 to 25 of 625 rows

Rows per page: 25

⏪

⏴

⏴

page 1

⏵

⏴

⏵

<div><div>⌕ ⌕ ⌕ ⌕ ⌕ ⌕</div><div>Gene</div><div>Gene ID</div></div>	<div><div>⌕ ⌕ ⌕ ⌕ ⌕ ⌕</div><div>Gene</div><div>Symbol</div></div>	<div><div>⌕ ⌕ ⌕ ⌕ ⌕ ⌕</div><div>Gene</div><div>Exons</div></div>																										
100862953	LOC100862953	<div>12 Exons</div>																										
100862997	LOC100862997	<div>12 Exons</div> <table><tr><th>DB identifier</th><th>Length</th></tr><tr><td>exon-XM_012490769.2-1</td><td>475</td></tr><tr><td>exon-XM_012490769.2-2</td><td>112</td></tr><tr><td>exon-XM_012490769.2-3</td><td>95</td></tr><tr><td>exon-XM_012490769.2-4</td><td>179</td></tr><tr><td>exon-XM_012490769.2-5</td><td>184</td></tr><tr><td>exon-XM_012490769.2-6</td><td>274</td></tr><tr><td>exon-XM_003696386.3-1</td><td>470</td></tr><tr><td>exon-XM_003696386.3-2</td><td>112</td></tr><tr><td>exon-XM_003696386.3-3</td><td>95</td></tr><tr><td>exon-XM_003696386.3-4</td><td>179</td></tr><tr><td>exon-XM_003696386.3-5</td><td>328</td></tr><tr><td>exon-XM_003696386.3-6</td><td>274</td></tr></table>	DB identifier	Length	exon-XM_012490769.2-1	475	exon-XM_012490769.2-2	112	exon-XM_012490769.2-3	95	exon-XM_012490769.2-4	179	exon-XM_012490769.2-5	184	exon-XM_012490769.2-6	274	exon-XM_003696386.3-1	470	exon-XM_003696386.3-2	112	exon-XM_003696386.3-3	95	exon-XM_003696386.3-4	179	exon-XM_003696386.3-5	328	exon-XM_003696386.3-6	274
DB identifier	Length																											
exon-XM_012490769.2-1	475																											
exon-XM_012490769.2-2	112																											
exon-XM_012490769.2-3	95																											
exon-XM_012490769.2-4	179																											
exon-XM_012490769.2-5	184																											
exon-XM_012490769.2-6	274																											
exon-XM_003696386.3-1	470																											
exon-XM_003696386.3-2	112																											
exon-XM_003696386.3-3	95																											
exon-XM_003696386.3-4	179																											
exon-XM_003696386.3-5	328																											
exon-XM_003696386.3-6	274																											
100862998	LOC100862998	<div>29 Exons</div>																										
100863028	LOC100863028	<div>40 Exons</div>																										
100863029	LOC100863029	<div>12 Exons</div>																										

Fig. 25: Step 4: Expand information on exons

Trail: [Query](#) > Results

☐ Manage Columns

Showing 1 to 25 of 9,689 rows

Rows per page: 25

Gene Gene ID	Gene Symbol	Exons DB identifier	Exons Length
100862953	LOC100862953	exon-XM_031918230.1-1	71
100862953	LOC100862953	exon-XM_031918230.1-10	98
100862953	LOC100862953	exon-XM_031918230.1-11	106
100862953	LOC100862953	exon-XM_031918230.1-12	342
100862953	LOC100862953	exon-XM_031918230.1-2	158
100862953	LOC100862953	exon-XM_031918230.1-3	103
100862953	LOC100862953	exon-XM_031918230.1-4	111
100862953	LOC100862953	exon-XM_031918230.1-5	141
100862953	LOC100862953	exon-XM_031918230.1-6	190
100862953	LOC100862953	exon-XM_031918230.1-7	115

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.)

Search for features within Genomic Regions

Search for features that overlap a list of genome coordinates you enter or upload, e.g. `chr1:29733..37349`

Warning: After running your search, using your browser back key to revise the search may cause unexpected changes to the parameters. It would be better to start a new search.

[More genome coordinates help](#)

- Select Organism:
- Select Assembly:
- Select Feature Types:

☐ CDS ☐ Exon ☐ Gene ☐ mRNA ☐ Pseudogene ☐ snRNA ☐ Transcript

☐ Guide RNA ☐ lncRNA ☐ Primary Transcript ☐ rRNA ☐ tRNA

☐ miRNA ☐ Pseudogenic Exon ☐ snoRNA
- Type/Paste in genomic regions in ☒ base coordinate ☐ interbase coordinate

(example for input format `chr:1..1000`)
(example for input format `chr:1-1000`)
(example for tab delimited input format)

or Upload genomic regions from a .txt file...
 no file selected
- Extend your regions at both sides:

1k 10k 100k 1M 10M
- ☐ Check this box to perform a strand-specific region search (search + strand if region start<end; search - strand if region end<start)

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

Search for features within Genomic Regions

Search for features that overlap a list of genome coordinates you enter or upload, e.g. chr1:29733..37349

Warning: After running your search, using your browser back key to revise the search may cause unexpected changes to the parameters. It would be better to start a new search.

[More genome coordinates help](#)

- Select Organism:
- Select Assembly:
- ☒ Select Feature Types:

<input type="checkbox"/> CDS	<input type="checkbox"/> Exon	<input checked="" type="checkbox"/> Gene
<input type="checkbox"/> Guide RNA	<input type="checkbox"/> lncRNA	<input type="checkbox"/> mRNA
<input type="checkbox"/> miRNA	<input type="checkbox"/> Primary Transcript	<input type="checkbox"/> Pseudogene
<input type="checkbox"/> Pseudogenic Exon	<input type="checkbox"/> rRNA	<input type="checkbox"/> snRNA
<input type="checkbox"/> snoRNA	<input type="checkbox"/> tRNA	<input type="checkbox"/> Transcript
- Type/Paste in genomic regions in ☒ base coordinate ☐ interbase coordinate

(example for input format chr:1..1000)▼
 (example for input format chr:1-1000)▼
 (example for tab delimited input format)▼

LG5:900000..930000

or Upload genomic regions from a .txt file...
 no file selected
- Extend your regions at both sides:

1k

10k

100k

1M

10M
- ☐ Check this box to perform a strand-specific region search (search + strand if region start<end; search - strand if region end<start)

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 save the results in GFF3 or BED format. The FASTA sequences of the features may also be downloaded.

Selected organism: *A. mellifera*

Selected assembly: *Amel_HAv3.1*

Selected feature types: Gene

Hide

Export data for all features within all regions:

TABCSVGFF3BEDFASTA

Export entire sequences for all regions:

FASTA...

Create list by feature type:

Gene

Go

LG5:900000..930000

Export sequence for entire region:

FASTA...

TABCSVGFF3BEDFASTA

Create List by

Gene

Go

GENOME REGION	FEATURE	FEATURE TYPE	LOCATION
	LOC411919 411919	Gene a	LG5:814392..995408
	GB44402	Gene a	LG5:842039..995430
	GB47005	Gene a	LG5:926438..926695

<< First < Prev 1 – 1 of 1 Next > Last >>

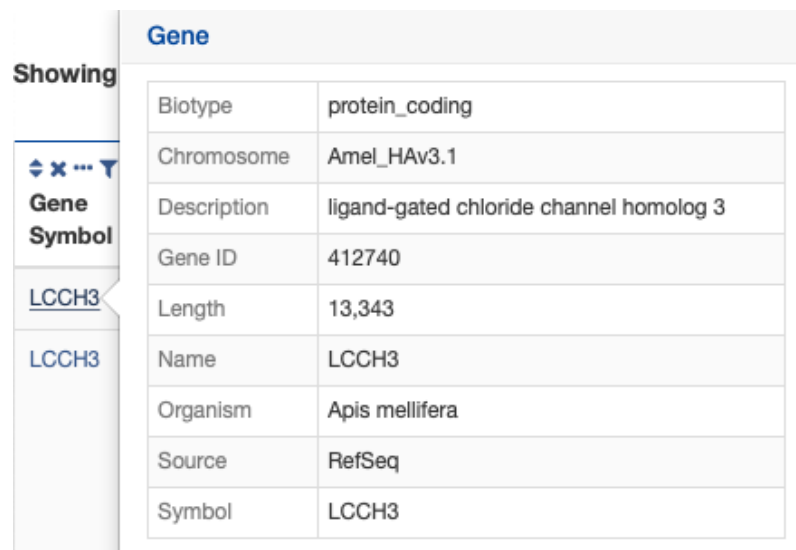
Fig. 29: Example: Genomic regions search results

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.



Gene	
Biotype	protein_coding
Chromosome	Amel_HAv3.1
Description	ligand-gated chloride channel homolog 3
Gene ID	412740
Length	13,343
Name	LCCH3
Organism	Apis mellifera
Source	RefSeq
Symbol	LCCH3

Fig. 1: Example: Summary window for LCCH3

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	<i>Apis mellifera</i>
GB45542	GB45542	amel_OGSv3.2	412740	RefSeq	<i>Apis mellifera</i>

Alias IDs (4 rows)

Manage Columns

Manage Filters

Manage Relationships

Save as List ▾

Generate Python code ▾

Export

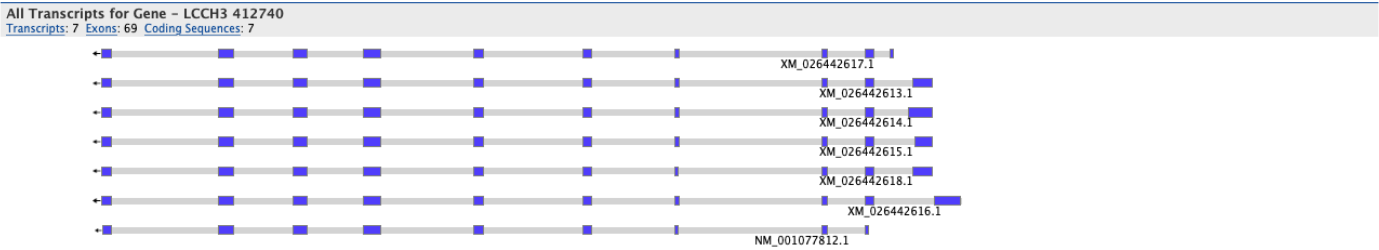
Showing 1 to 4 of 4 rows

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	ATGCATCACAGGATGTGGTTGCAGCAGATCTTATCCTGCTGCAGATGATTCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCGATAGATTGGAA AATGTGACGCAAAACATATCAGCAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTCGGCGGAGAACCCCTGTTAGTCGGGATGGATCT...
XM_026442613.1-CDS	XP_026298398.1	1485	ATGCATCACAGGATGTGGTTGCAGCAGATCTTATCCTGCTGCAGATGATTCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCGATAGATTGGAA AATGTGACGCAAAACATATCAGCAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTCGGCGTGTGTTCAITCCAGGAGAACCCCTGTT...
XM_026442614.1-CDS	XP_026298399.1	1485	ATGCATCACAGGATGTGGTTGCAGCAGATCTTATCCTGCTGCAGATGATTCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCGATAGATTGGAA AATGTGACGCAAAACATATCAGCAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTCGGCGTGTGTTCAITCCAGGAGAACCCCTGTT...
XM_026442615.1-CDS	XP_026298400.1	1485	ATGCATCACAGGATGTGGTTGCAGCAGATCTTATCCTGCTGCAGATGATTCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCGATAGATTGGAA AATGTGACGCAAAACATATCAGCAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTCGGCGTGTGTTCAITCCAGGAGAACCCCTGTT...
XM_026442616.1-CDS	XP_026298401.1	1485	ATGCATCACAGGATGTGGTTGCAGCAGATCTTATCCTGCTGCAGATGATTCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCGATAGATTGGAA AATGTGACGCAAAACATATCAGCAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTCGGCGTGTGTTCAITCCAGGAGAACCCCTGTT...
XM_026442617.1-CDS	XP_026298402.1	1485	ATGCATCACAGGATGTGGTTGCAGCAGATCTTATCCTGCTGCAGATGATTCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCGATAGATTGGAA AATGTGACGCAAAACATATCAGCAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTCGGCGTGTGTTCAITCCAGGAGAACCCCTGTT...
XM_026442618.1-CDS	XP_026298403.1	1470	ATGCATCACAGGATGTGGTTGCAGCAGATCTTATCCTGCTGCAGATGATTCTAATTGCTTGGGCCAGTCTTGAAAACACAGGAATATCGATAGATTGGAA AATGTGACGCAAAACATATCAGCAATCCTTGATGGTTACGATATTCGATTAAGGCCAAATTCGGCGTGTGTTCAITCCAGGAGAACCCCTGTT...

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	MHHRMWLQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILGDYDRLRPNFGVVSFPGEPLLVGMDLTASFDAISEVNMODYTITMYLNQYWKDERLAFSQEEVLTLSGDFAEKIWWPDTF FANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPPLDSQNCVIEISYGYTVLDVWMYWKETPVRG...
XP_026298398.1	XM_026442613.1	494	MHHRMWLQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILGDYDRLRPNFGVVSFPGEPLLVGMDLTASFDAISEVNMODYTITMYLNQYWKDERLAFSQEEVLTLSGDFAEKIWW VPDTFFANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPPLDSQNCVIEISYGYTVLDVWMYWKETPVRG...
XP_026298399.1	XM_026442614.1	494	MHHRMWLQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILGDYDRLRPNFGVVSFPGEPLLVGMDLTASFDAISEVNMODYTITMYLNQYWKDERLAFSQEEVLTLSGDFAEKIWW VPDTFFANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPPLDSQNCVIEISYGYTVLDVWMYWKETPVRG...
XP_026298400.1	XM_026442615.1	494	MHHRMWLQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILGDYDRLRPNFGVVSFPGEPLLVGMDLTASFDAISEVNMODYTITMYLNQYWKDERLAFSQEEVLTLSGDFAEKIWW VPDTFFANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPPLDSQNCVIEISYGYTVLDVWMYWKETPVRG...
XP_026298401.1	XM_026442616.1	494	MHHRMWLQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILGDYDRLRPNFGVVSFPGEPLLVGMDLTASFDAISEVNMODYTITMYLNQYWKDERLAFSQEEVLTLSGDFAEKIWW VPDTFFANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPPLDSQNCVIEISYGYTVLDVWMYWKETPVRG...
XP_026298402.1	XM_026442617.1	494	MHHRMWLQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILGDYDRLRPNFGVVSFPGEPLLVGMDLTASFDAISEVNMODYTITMYLNQYWKDERLAFSQEEVLTLSGDFAEKIWW VPDTFFANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPPLDSQNCVIEISYGYTVLDVWMYWKETPVRG...
XP_026298403.1	XM_026442618.1	489	MHHRMWLQIFILLQMIHLIAWASLENTGISDRLENVTQTISRILGDYDRLRPNFGVVSFPGEPLLVGMDLTASFDAISEVNMODYTITMYLNQYWKDERLAFSQEEVLTLSGDFAEKIWWPDTF FANDKNSFLHDVTERNKLVRLSGDGSVTYGMRFTTTLACMMDLHYYPPLDSQNCVIEISYGYTVLDVWMYWKETPVRG...

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.

Proteins

Curated comments from UniProt			Hide proteins
Type	Comment	Proteins	
Q0GQR5_APIME	similarity	Belongs to the ligand-gated ion channel (TC 1.A.9) family. Gamma-aminobutyric acid receptor (TC 1.A.9.5) subfamily.	

1 Proteins			
Trail: Gene			
<input type="checkbox"/> Manage Columns	<input type="checkbox"/> Manage Filters	<input type="checkbox"/> Manage Relationships	<input type="button" value="Save as List"/> <input type="button" value="Generate Python code"/> <input type="button" value="Export"/>

Showing 1 to 1 of 1 row

Proteins DB identifier	Proteins Primary Accession	Proteins Organism	Proteins Length
Q0GQR5_APIME	Q0GQR5	Apis mellifera	489






















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		
presynaptic active zone		IEA
chemical synaptic transmission, postsynaptic		IEA
presynaptic membrane		IEA
nucleus		IEA
cell periphery		IEA
endomembrane system		IEA
postsynapse		IEA
postsynaptic specialization		IEA
plasma membrane		ECO:0007669
intracellular anatomical structure		IEA
dendritic tree		IEA
molecular function		
GABA-A receptor activity		ECO:0007669
extracellular ligand-gated monoatomic ion channel activity		ECO:0007669
chloride channel activity		ECO:0007669
biological process		
cellular response to stimulus		IEA
chemical synaptic transmission, postsynaptic		IEA
signaling		IEA
extracellular ligand-gated monoatomic ion channel activity		ECO:0007669
cell communication		IEA
chloride channel activity		ECO:0007669
regulation of presynaptic membrane potential		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 Name	 Pathways URL
ame04080	Neuroactive ligand-receptor interaction - Apis mellifera (honey bee)	https://www.genome.jp/pathway/ame04080

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. echinator	LOC105145387 LOC105149353
A. florea	LOC100869944 LOC100871112
A. gifuensis	LOC122856609 LOC122852242
A. haemorrhoea	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	
<input type="checkbox"/> Manage Columns <input type="checkbox"/> Manage Filters <input type="checkbox"/> Manage Relationships	<input type="button" value="Save as List"/> <input type="button" value="Generate Python code"/> <input type="button" value="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

<div><div></div><div></div></div> Gene DB identifier	<div><div></div><div></div></div> Publications PubMed ID	<div><div></div><div></div></div> Publications First Author	<div><div></div><div></div></div> Publications Title	<div><div></div><div></div></div> Publications Journal	<div><div></div><div></div></div> Publications Volume	<div><div></div><div></div></div> Publications Year	<div><div></div><div></div></div> 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.

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, Mocsl, 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.

Upload

View

Search:

1 Upload list of identifiers

2 Verify identifier matches

List analysis

List

• Gene

• Organism

• Identifier

Create a new list

Select the type of list to create and either enter in a list of identifiers or upload identifiers from a file. A search will be performed for all the identifiers in your list.

• Separate identifiers by a **comma**, **space**, **tab** or **new line**.

• Qualify any identifiers that contain whitespace with double quotes like so: "even skipped".

Select Type:

for Organism:

Type/Paste in identifiers

(click to see an example)

or Upload identifiers from a .txt file...

Choose File

No file chosen

☐ Match on case

Reset

Create List

Fig. 1: List upload form

You entered: 5 identifiers

We found: 5 Genes

Save a list of 5 Genes

Summary

Download summary

Direct Hits

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

List Analysis for Example gene list (5 Genes)

Manage Columns

Manage Filters

Manage Relationships

Save as List

Generate Python code

Export

Showing 1 to 5 of 5 rows

Gene ID	Gene Source	Gene Biotype	Gene Symbol	Gene Description	Gene Length	Gene Chromosome ID	Gene Chromosome Location Start	Gene Chromosome Location End	Gene Chromosome Strand	Gene Organism	Gene Assembly
31042	RefSeq	protein_coding	TRAM	TRAM,	2591	X	935232	937822	1	Drosophila melanogaster	Release_6_plus_ISO1
33045	RefSeq	protein_coding	mal	maroon-like	3536	X	20429675	20433210	1	Drosophila melanogaster	Release_6_plus_ISO1
39238	RefSeq	protein_coding	Mocs1	Molybdenum cofactor synthesis 1	2823	3L	11070526	11073348	1	Drosophila melanogaster	Release_6_plus_ISO1
46080	RefSeq	protein_coding	Sec61beta	Sec61 beta subunit	911	2R	14618771	14619681	1	Drosophila melanogaster	Release_6_plus_ISO1
GB41586	amel_OGSv3.2	NO VALUE	NO VALUE	NO VALUE	1067	LG11	14299553	14300619	1	Apis mellifera	Amel_HAv3.1

Orthologues

A. cephalotes (5) A. cerana (5) A. colombica (5) A. dorsata (7) A. echinator (5) A. florea (4) A. gifuensis (4) A. haemorrhoea (4) A. laboriosa (4) A. mellifera (4) A. nigerrimus (4) A. nigricornis (4) A. rosae (4) B. bifarius (4) B. brevipes (4) B. campestris (4) B. confusus (4) B. consobrinus (4) B. cullumanus (4) B. difficillimus (4) B. haemorrhoidalis (4) B. hortorum (4) B. hypnorum (3) B. impatiens (3) B. kinseyi (4) B. opulentus (4) B. pascuorum (4) B. picipes (4) B. polaris (4) B. pratorum (4) B. pyrosoma (4) B. sibiricus (4) B. skorikovi (4) B. soroensis (4) B. superbus (4) B. sylvestris (4) B. terrestris (4) B. turneri (3) B. vancouverensis nearcticus (4) B. vosnesenskii (4) B. waltoni (4) C. calcarata (4) C. cinctus (4) C. costatus (6) C. floridanum (4) C. floridanus (4) C. gigas (4) C. glomerata (4) C. insularis (4) C. obscurior (4) C. rybensis (4) C. solmsi marchali (4) D. aliolem (4) D. media (3) D. novaeangliae (4) D. quadricipes (4) D. saxonica (3) D. similis (4) E. continuus (3) E. lituratus (4) E. mexicana (4) F. arisanus (4) F. exsecta (4) F. varia (4) H. laboriosa (4) H. saltator (4) I. xanthorius (4) L. albipes (4) L. heterotoma (4) L. humile (4) L. lativentris (4) L. morio (4) M. demolitor (4) M. europaea (3) M. genalis (4) M. pharaonis (4) M. quadrifasciata (8) M. rotundata (4) N. fabriciana (3) N. fabricii (4) N. fulva (4) N. lecontei (4) N. melanderi (4) N. pinetum (4) N. spinosus (4) N. virginianus (4) N. vitripennis (4) O. abietinus (4) O. bicornis bicornis (4) O. biroii (4) O. brunneus (4) O. lignaria (4) P. barbatus (4) P. canadensis (4) P. dominula (4) P. fuscatus (4) P. gracilis (4) S. invicta (4) S. monilicornis (4) S. tumulorum (4) T. cornetzi (5) T. curvispinosus (3) T. pretiosum (4) T. septentrionalis (5) T. zetekii (5) V. canescens (4) V. czabro (4) V. emeryi (4) V. germanica (4) V. mandarinia (4) V. pensylvanica (4) V. velutina (4) V. vulgaris (4) W. auropunctata (4)

External Links

No external links.

Fig. 3: Example: List analysis for gene list

Widgets displaying properties of 'Example gene list'

Click to select widgets you would like to display:
☒ Gene Ontology Enrichment ☒ Publication Enrichment ☒ Pathway Enrichment

Important Notes for Enrichment Widgets to avoid false positives:

- If your gene list contains genes from more than one gene set, it is critical that you create new lists for each specific gene set and then redo the analyses separately by clicking the newly saved lists on your List View page. You can filter the table above for a particular gene set using the histogram icon above the Gene Source column in the table above. Then "Save as List".
- After you save the new list, you need to click on the list name (either shown at the top of this page or in your List View page), in order for the enrichments to be recalculated.
- You should change the background population from the default to one of the available gene sets or your own background gene list.

Gene Ontology Enrichment
GO terms enriched for items in this list.
Number of Genes in this list not analysed in this widget: 1
Test Correction: Holm-Bonferroni Max p-value: 0.05 Ontology: biological_process
Background population: Default Normalise by length: ☒ Normalise
View Download

GO Term	p-Value	Matches
<input type="checkbox"/> SRP-dependent cotranslational protein targeting to membrane, translocation [GO:0006616]	0.013245	2
<input type="checkbox"/> Mo-molybdopter in cofactor biosynthetic process [GO:0006777]	0.023170	2
<input type="checkbox"/> molybdenum incorporation into molybdenum-molybdopter in complex [GO:0018315]	0.023170	2
<input type="checkbox"/> Mo-molybdopter in cofactor metabolic process [GO:0019720]	0.023170	2
<input type="checkbox"/> metal incorporation into metallo-molybdopter in complex [GO:0042940]	0.023170	2
<input type="checkbox"/> molybdopter in cofactor biosynthetic process	0.025473	2

Publication Enrichment
Publications enriched for genes in this list.
Number of Genes in this list not analysed in this widget: 1
Test Correction: None Max p-value: 0.05 Background population: Default Change
View Download


Publication	p-Value	Matches
<input type="checkbox"/> The molybdoenzyme system of <i>Drosophila melanogaster</i> . I. Sulfite oxidase: identification and properties. Expression of the enzyme in maroon-like (mal), low-xanthine dehydrogenase (xrd), and cinnamon (cin) flies. [6800353]	7.587572e-9	2
<input type="checkbox"/> Sec61beta, a subunit of the protein translocation channel, is required during <i>Drosophila</i> development. [10564656]	2.011541e-4	1
<input type="checkbox"/> Diverse splicing mechanisms fuse the evolutionarily conserved bicistronic MOCS1A and MOCS1B open reading frames. [10917590]	2.011541e-4	1
<input type="checkbox"/> Comparison of the sequences of the <i>Aspergillus nidulans</i> hxB and <i>Drosophila melanogaster</i> mal-I genes with nIfS from	2.011541e-4	1

Pathway Enrichment
Pathways enriched for genes in this list
Number of Genes in this list not analysed in this widget: 1
Test Correction: None Max p-value: 0.05 DataSet: KEGG pathways d
Background population: Default Change
View Download

Pathways	p-Value	Matches
<input type="checkbox"/> Folate biosynthesis - <i>Drosophila melanogaster</i> (fruit fly)	2.790145e-4	2
<input type="checkbox"/> Protein processing in endoplasmic reticulum - <i>Drosophila melanogaster</i> (fruit fly)	0.003165	2
<input type="checkbox"/> Protein export - <i>Drosophila melanogaster</i> (fruit fly)	0.015862	1

Fig. 4: Example: Available widgets

Upload | View Search: e.g. *Nasonia vitripennis*, *A. g.* GC



Lists

View your own and public lists, search by keyword and compare or combine the contents of lists. Click on a list to view graphs and summaries in an analysis page, select lists using checkboxes to perform set operations. Click 'Upload' above to import a new list.

Filter:

Actions: Union | Intersect | Subtract | Asymmetric Difference | Copy Delete **Options:** ☒ Show descriptions ☐ Show Tags

You are not logged in. [Log in](#) to save lists permanently and to mark items as favourites .

☐ Example gene list 5 Genes

☐ A. mellifera OGSv3.2 genes mapping to more than one chromosome 22 Genes
These OGSv3.2 genes map to more than one chromosome, but due to a bug only one location is provided in this release of HymenopteraMine. These genes should have been eliminated from the database. You can use the Asymmetric Difference function at the top of the List Tool view page to remove these genes from your own lists. We apologize for the inconvenience.

☐ A. mellifera RefSeq genes mapping to more than one OGSv3.2 1401 Genes
These RefSeq genes overlap more than one OGSv3.2 gene, indicating a split/merge disagreement. This list can be used to remove these genes from your own lists using the Asymmetric Difference function at the top of the List Tool view page. As the RefSeq genes were computed on the Amel_HAv3.1 assembly, they should be considered more reliable than the OGSv3.2 genes. which were mapped from an older assembly.

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.



Your Lists

Your current lists

<input type="checkbox"/>	LIST NAME	DESCRIPTION	TYPE	NUMBER OF OBJECTS	DATE CREATED
<input type="checkbox"/>	Example gene list   Add tags Share with users		Gene	5 values	2023-01-09 16:55

Fig. 1: MyMine

CHAPTER 6

API

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

[Perl](#) | [Python](#) | [Ruby](#) | [Java](#)

Search:

Perl Web Service Client

The Perl web service client library makes it easy to run queries in HymenopteraMine directly from Perl programs. You can use these modules to construct any query you could run from web interface and fetch the results in a number of formats, including native Perl data structures and objects, and TSV/CSV strings.

Like all our code, it is open-source, coming licensed under the LGPL. For information on our API, visit our [wiki pages](#).

- Prerequisites**

You should install the [Perl webservice client library module](#) to get started. You can install it directly from CPAN (the comprehensive Perl archive network). This is a public repository of thousands of modules. Using CPAN will ensure you get the latest version, and that your dependencies are automatically managed for you.

To install the InterMine Perl client library type the following command into a shell:

```
> sudo cpan Webservice::InterMine
```
- Examples of using the downloaded script**

On each Template Query page and the QueryBuilder there is a link to get Perl code to run that particular query using the web service API. Just click the link, save the generated Perl script in a file and execute it. You can use the generated code as a starting point for your own programs. You can run the downloaded script by running the following command in a shell:

```
> perl path/to/downloaded/script.pl
```

If you get an error saying `Cannot find Webservice::InterMine in @INC`, or similar then see 'Prerequisites' above.

Feel free to edit the script – these are designed to be spring-boards to help you get where you want to. See the [Cookbook](#), which contains a set of short tutorial 'recipes' that demonstrate particular features of the Perl API, for ways to edit the scripts, and even extend their functionality. For a good reference to writing programs in Perl, see [here](#).

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 et al. - PubMed 34747465	HGD GO Annotation Download
	Gene Ontology Annotations from FlyBase	<i>Drosophila melanogaster</i>	FlyBase release F82022_05; 13 Sep 2022	Larkin et al. - PubMed 33219682	FlyBase Download
Pathways	Pathway Information from KEGG	<i>Acromyrmex echinator</i> <i>Aphidius gifuensis</i> <i>Apis laboriosa</i> <i>Apis mellifera</i> <i>Atta cephalotes</i> <i>Bombus bifarius</i> <i>Bombus impatiens</i> <i>Bombus pyrosoma</i> <i>Bombus terrestris</i> <i>Bombus vancouverensis nearcticus</i> <i>Bombus vosnesenskii</i> <i>Camponotus floridanus</i> <i>Cephus cinctus</i> <i>Ceratina calcarata</i> <i>Ceratosolen solmsi marchali</i> <i>Colletes gigas</i> <i>Cotesia glomerata</i> <i>Diachasma alloeum</i> <i>Dinoponera quadricaps</i> <i>Drosophila melanogaster</i> <i>Fopius arisanus</i> <i>Formica exsecta</i> <i>Harpegnathos saltator</i> <i>Leptopilina heterotoma</i> <i>Linepithema humile</i> <i>Megalopta genalis</i> <i>Microplitis demolitor</i> <i>Monomorium pharaonis</i> <i>Nasonia vitripennis</i> <i>Nomia melanderi</i> <i>Ooceraea biroi</i> <i>Osmia bicornis bicornis</i> <i>Pogonomyrmex barbatus</i> <i>Polistes canadensis</i> <i>Polistes fuscatus</i> <i>Pseudomyrmex gracilis</i> <i>Solenopsis invicta</i> <i>Trichogramma pretiosum</i> <i>Vespula pennsylvanica</i> <i>Vollenhovia emeryi</i>	KEGG release 104.0; 1 Oct 2022	Kanehisa et al. - PubMed 33125081	KEGG Download
	Pathway information inferred through orthologues from curated human pathways	<i>Drosophila melanogaster</i>	Reactome version 80; 5 Apr 2022	Gillespie et al. PubMed 34788843	Reactome Download
Interactions	Interactions from BioGRID	<i>Apis mellifera</i>	BioGRID version 4.4.213; 25 Aug 2022	Oughtred et al. - PubMed: 33070389	BioGRID Download
	Interactions from IntAct	<i>Drosophila melanogaster</i>			
	Interactions from IntAct	<i>Drosophila melanogaster</i>	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.

SequenceServer 1.0.9 [Help & Support](#)

Paste query sequence(s) or drag file containing query sequence(s) in FASTA format here ...

Nucleotide databases

- ☐ Acromyrmex_echinator_GCF_000204515.1_Aech_3.9_genome.fa
- ☐ Apis_cerana_GCF_001442555.1_ACSNU-2.0_genome.fa
- ☐ Apis_dorsata_GCF_000469605.1_Apis_dorsata_1.3_genome.fa
- ☐ Apis_florea_GCF_000184785.3_Aflo_1.1_genome.fa
- ☐ Apis_mellifera_GCF_003254395.2_Amel_HAv3.1_genome.fa
- ☐ Athalia_rosae_GCF_000344095.2_Aros_2.0_genome.fa
- ☐ Atta_cephalotes_GCF_000143395.1_Attacep1.0_genome.fa

Protein databases

- ☐ Acromyrmex_echinator_GCF_000204515.1_Aech_3.9_AR100_RefSeq_protein.fa
- ☐ Apis_cerana_GCF_001442555.1_ACSNU-2.0_AR101_RefSeq_protein.fa
- ☐ Apis_dorsata_GCF_000469605.1_Apis_dorsata_1.3_AR101_RefSeq_protein.fa
- ☐ Apis_florea_GCF_000184785.3_Aflo_1.1_AR102_RefSeq_protein.fa
- ☐ Apis_mellifera_GCF_003254395.2_Amel_HAv3.1_AR104_RefSeq_protein.fa
- ☐ Athalia_rosae_GCF_000344095.2_Aros_2.0_AR102_RefSeq_protein.fa
- ☐ Atta_cephalotes_GCF_000143395.1_Attacep1.0_AR100_RefSeq_protein.fa

Advanced Parameters: ?

BLAST

Fig. 1: BLAST search

CHAPTER 9

How to Cite

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