BovineMine Documentation

Release 1.6

Deepak Unni, Aditi Tayal, Colin Diesh, Christine Elsik, Darren Hag

Table of Contents

1	Overview of BovineMine	3
2	2.2.4 Optional filters 2.3 QueryBuilder	5 7 9 11 12 12 12 14 14 15
3	3.1 Summary 3.2 Transcripts 3.3 Proteins 3.4 Function 3.5 Homology 3.6 Interactions 3.7 Publications	23 24 24 25 25 25 27 27 27
4	Genomic Regions Search	29
5	5.1 Creating Lists	33 33 36
6	MyMine	39
7	API	41

8	Data Sources	43
9	How to cite	45

BovineMine is a data mining resource for Bos taurus reference genome assemblies (ARS-UCD1.2 and UMD3.1.1). Powered by InterMine, this platform provides access to a number of datasets from a variety of source. It also provides customized bioinformatics tools that researchers can use to create their own custom datasets. BovineMine is part of the Bovine Genome Database



Main site: http://bovinegenome.org/bovinemine

Bovine Genome Database YouTube Channel with BovineMine help videos: Bovine Genome Database on YouTube Link to the available datasets in BovineMine: http://bovinegenome.org/bovinemine/dataCategories.do

For more generic examples on how to use InterMine, click here. These are tutorials created by FlyMine that showcase the different features of InterMine.

The Bovine Genome Database is supported by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613689, and has been supported by grants 2007-35616-17882 and 2010-65205-20407 from the USDA National Institute of Food and Agriculture. BGD is hosted at the University of Missouri. If you have comments or if you wish to report a problem, please contact the Bovine Genome Database Administrator.

Table of Contents 1

2 Table of Contents

CHAPTER 1

Overview of BovineMine

This section provides a brief overview of the layout for BovineMine.



The navigation panel highlights different functionalities of BovineMine.

Home - The home page for BovineMine

MyMine - The MyMine serves as a portal for users for account management. When logged in to Bovine-Mine users can access their saved templates, most recent queries and saved lists.

Templates - List of templates that users can select from based on the nature of their query. Each template is a predefined query with a simple form containing a description of what input is expected and the type of output that will be generated.

Lists - Allows users to upload lists of genes on which they can perform enrichment analyses and export the results. Users that log in to BovineMine can save their lists for future use.

QueryBuilder - A flexible interface that allows users to create their own custom query template while browsing the BovineMine data models. Queries can be exported in a variety of formats to share with other users.

Regions - The Genomic Region Search tool where users can enter a series of genomic coordinates, specify flanking regions and fetch all features that fall within the given interval. The result can be exported or saved as a list for further analyses.

Data Sources - Provides a summary of all the data loaded into BovineMine including their release versions, associated publications and links to source sites.

Help - Links to the BovineMine help docs and tutorials

API - Describes the InterMine API that allows users to programmatically access BovineMine.

Navigation and Searching in BovineMine

There are several ways for users to query BovineMine.

2.1 Quick Search

Quick Search allows users to search keywords from any of the BovineMine datasets. There is a Quick Search box on the BovineMine home page or a smaller search box in the upper right corner of all pages.

We are currently working to fix our Quick Search tool and hope to have it running again soon.

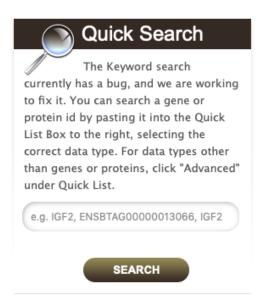


Fig. 1: Quick Search from home page

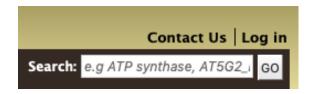
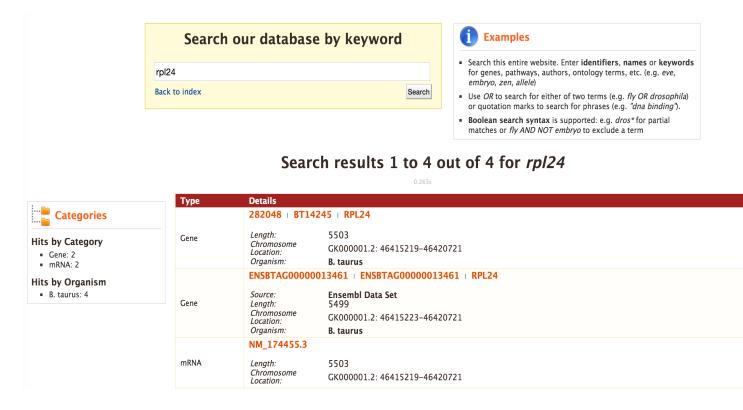


Fig. 2: Quick Search from any page

Quick Searches can be conducted with a number of identifiers including gene names, transcripts, pathways, gene identifiers or organisms. The wildcard character * can be used to retrieve all results that match a particular search query.

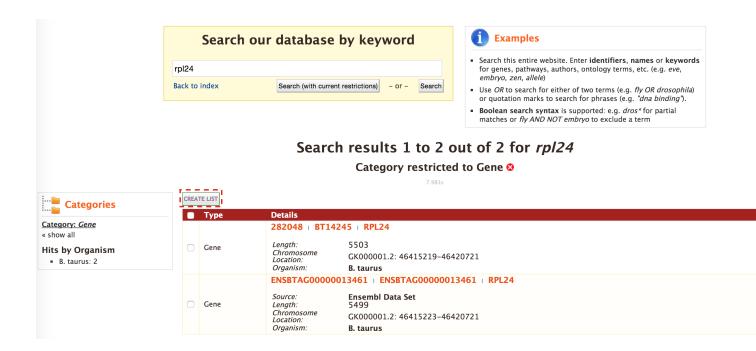
As an example, we will consider gene **rpl24**. Enter *rpl24* in the search box and click **Search**. The results page is tabulated and shows a summary about your query, as shown below.



Users can filter the results based on **Category** and **Organism**. The score column in the result table indicates the similarity of your query to each of the result fetched by BovineMine.

The results page can also be converted to a list. To enable this feature click on Gene in Hits by Category.

This will filter the results for the feature type **Gene** and checkboxes will be available for users to select genes they would like to add to their list. Once the genes are selected, click on **CREATE LIST**.



2.2 Templates

Templates or predefined queries are another search method within BovineMine. Popular templates are displayed on the home page, grouped by category (e.g., Genes, Proteins, Interactions) and the complete list can be seen by clicking the **Templates** menu tab.

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

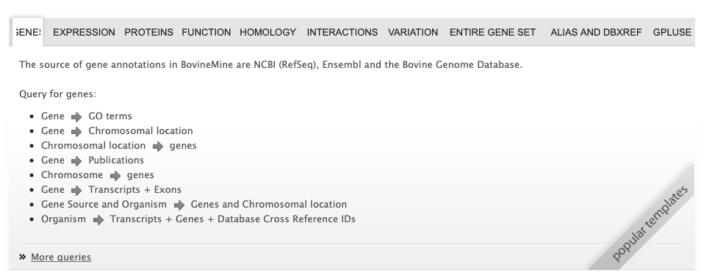


Fig. 3: Popular templates

As an example, the **Gene -> Homologues** template queries BovineMine to retrieves all homologues for a given gene. Here, we will do a search for the gene *GSTM1*.

2.2. Templates 7

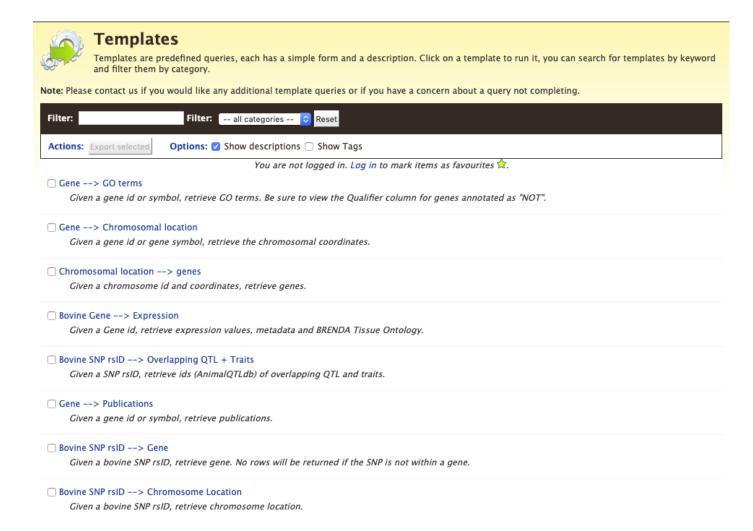


Fig. 4: Full list of templates on Templates page

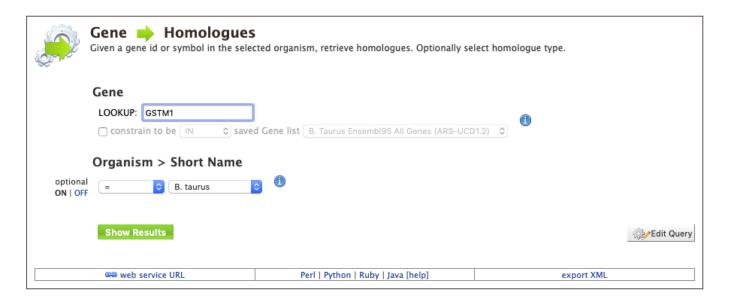


Fig. 5: Example: Gene -> Homologue

The results page displays all of the homologues for that query gene. When logged in to BovineMine, users can save their results as a list for further analyses by clicking on the **Save as List** button above the results table then choosing columns to save in their list. See the lists section for more detail on creating and saving lists. Note the "Trail: Query" text at the upper left of the results table. Clicking on the "Query" link will bring you back to the query that generated the table to allow for edits without having to start with a new template.

2.2.1 Generate query code

The code for each template query can be retrieved 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.

2.2.2 Download results

The search results from a template query can 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. 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. 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.

2.2. Templates 9

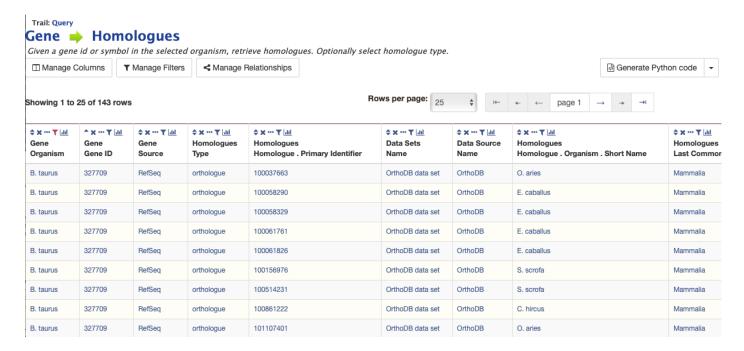


Fig. 6: Example: Gene -> Homologue template search results, identifier for Gene GSTM1

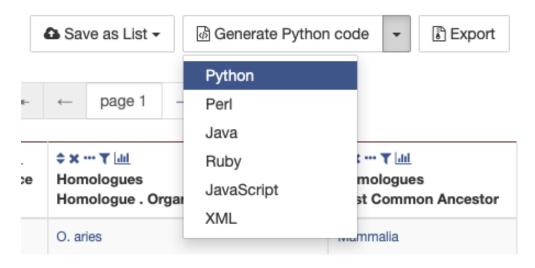


Fig. 7: Options for generating code from template query

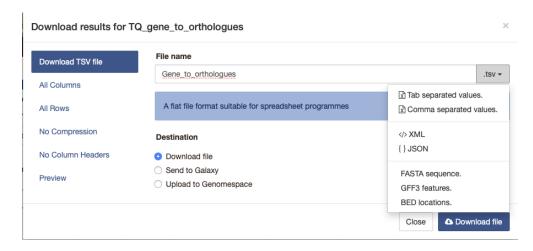


Fig. 8: Options for downloading results from template query

2.2.3 Customize output and manage columns

To customize the results table layout, click the **Manage Columns** button. This allows users to rearrange, remove or order columns. Filters can be edited or removed by clicking the **Manage Filters** button. To specify the entity relationships within the query and change the way the results are presented in the table, click **Manage Relationships**. Clicking on the **Manage Relationships** option bring also brings up a blue information panel ("What does this do?") that provides more detailed information.

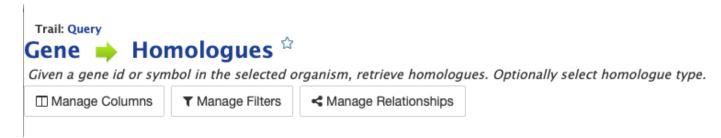


Fig. 9: Options for customizing the results of a template query.

To further manage column data, each column has a set of icons in its header. Mousing over any icon will reveal what it does. The sort icon resembles two triangles and allows you to sort a column in either direction (ascending/descending; a->z, z->a). To delete any column from your table, click on the "x" icon. Alternatively, the three dots "..." can be selected to "Toggle column visibility" and hide the column. Once hidden, the icon turns into a double arrow that can be selected to expand the column into its original form. The next icon resembling a funnel can be selected to edit or remove any currently active filters. A summary of the data within a column can be viewed by clicking on the icon that resembles a graph. The data within each summary can be selected for further filtering or downloading.

2.2. Templates 11



Fig. 10: Column header icons available to edit the results of a template query.

2.2.4 Optional filters

Some templates have optional filters that are disabled by default. For the Gene -> Homologue query, there is an optional filter to specify the organism used in the query. To enable the filter, click **ON** below the **Organism > Short Name** label.

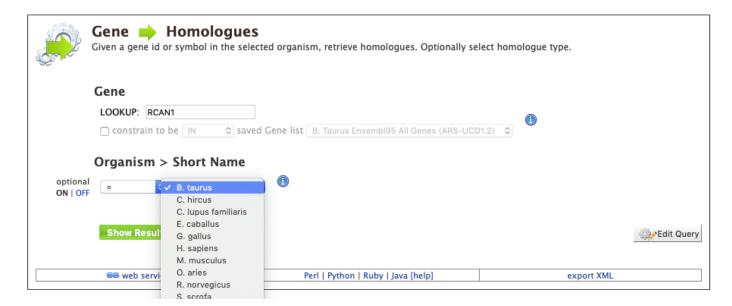


Fig. 11: Using the optional organism filter in the Gene -> Homologue query template.

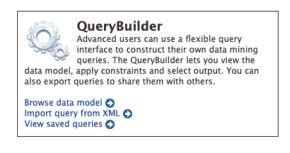
2.3 QueryBuilder

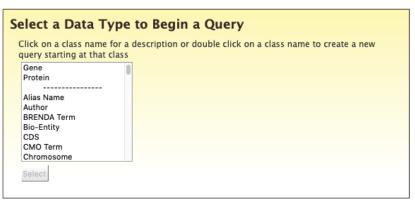
The provided templates are suitable for many different types of searches, new queries may be built from scratch using the **QueryBuilder**. The possibilities of queries using the QueryBuilder are endless. The output may be formatted exactly as desired, and the query constraints may be chosen to perform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

To begin, select a **Data Type**. For example, select **Gene** as a Data Type then click the **Select** button to be taken to the Model browser.

2.3.1 Model browser

After selecting a data type, the **Model builder** appears displaying the attributes for the chosen feature class **Gene**.





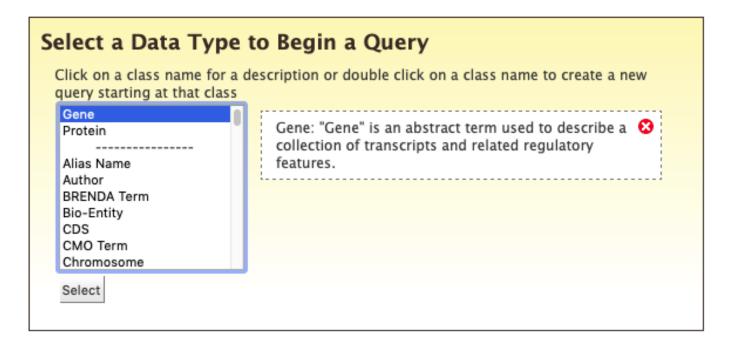


Fig. 12: Selecting data type as **Gene** in QueryBuilder.

2.3. QueryBuilder 13

First lets select **Gene** as a Data Type in the QueryBuilder. Then click on **Select**. This will take you to a Model browser where you can select the attributes for the feature class 'Gene', which you would want to be shown in your results.

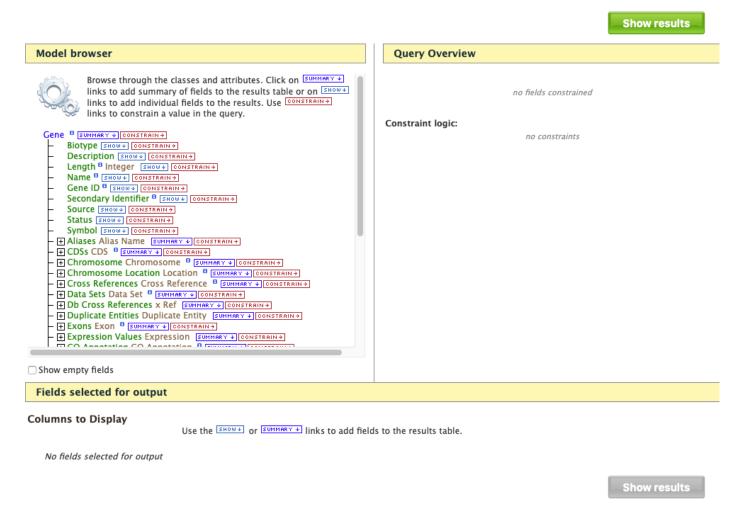


Fig. 13: Model browser with "Gene" selected as the data type.

2.3.2 QueryBuilder Examples

The following three examples provide details as to how to use the QueryBuilder using "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 **Biotype**, **Gene ID** and **Symbol**, which will add these fields to the query. Notice that these two fields appear below the data type **Gene** in the Query Overview section.

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

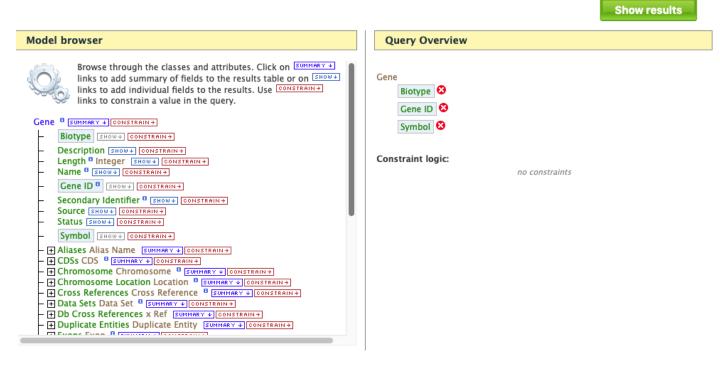


Fig. 14: Model browser with "Gene" selected as the data type. Example 1, Step 1: Select fields to be added to the query

constraint from the query. Clicking on the blue pencil icon next to a constraint brings up the constraint editing window where changes may be made to the query filters.

Lastly, click on **Show Results** above the Model Browser. The resulting table contains all protein coding genes in the database, with Gene ID, Gene Symbol and Gene Biotype as the table columns. Because the Biotypes should all be the same (protein coding), that column can be deleted by clicking the "x" above it. Alternatively, the three dots "..." can be selected to "Toggle column visibility" and hide the column. See the *Customize output and manage columns* section to review ways to manage column data.

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

Users can customize the previously run query by adding another constraint for **Chromosome**. Note the "Trail: Query" text at the upper left of the results table. Clicking on the "Query" link will bring you back to the query that generated the table to allow for edits without having to build a new query. In the Model browser, click on the "+" (plus sign) next to the Chromosome feature class to display its attributes.

Next click on the **Chromosome ID** attribute and in the text box of the pop-up window select "=" (equals). Enter **NC_019465.1** then click on **Add to Query**, which adds the additional constraint to the query.

Click on **Show results** and the query will result in all protein-coding genes on the Chromosome with constrained ID. Note that the number of results has been reduced with the addition of the constraint.

2.3. QueryBuilder 15

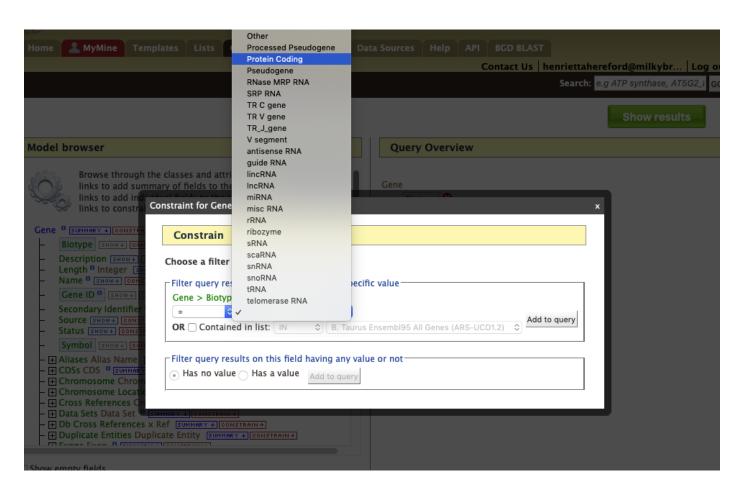


Fig. 15: Example 1, Step 2 Adding a constraint to the query on Biotype.

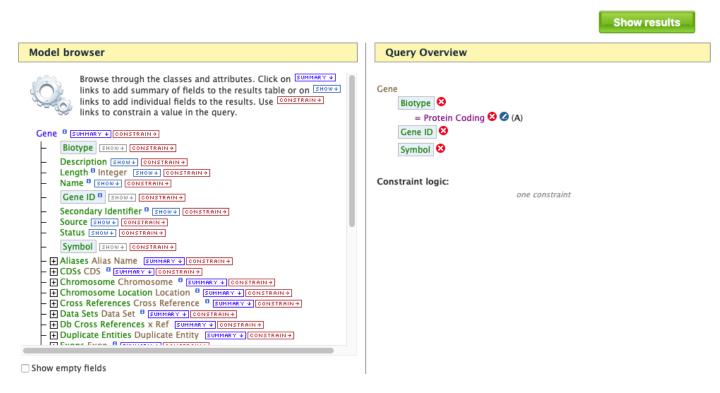


Fig. 16: Example 1, Step 2 Constraint has been added to the query on Biotype.

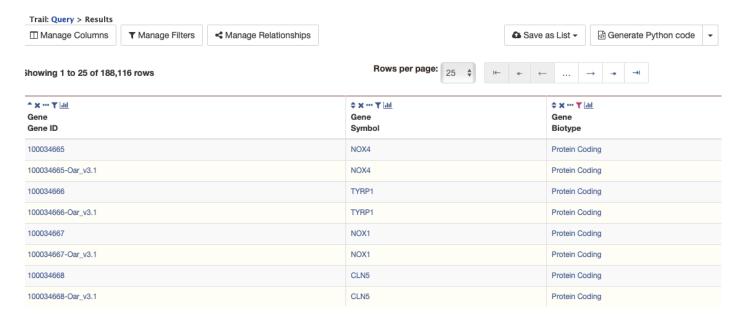


Fig. 17: Example 1, Step 3 Display query results.

2.3. QueryBuilder 17

Model browser Chromosome Chromosome | SUMMARY + CONSTRAIN + Assembly SHOW+ CONSTRAIN+ Length Integer SHOW → CONSTRBIN → Name SHOW# CONSTRAIN# Chromosome ID SHOW + CONSTRAIN+ Secondary Identifier SHOW+ CONSTRAIN+ Tertiary Identifier SHOW + CONSTRAIN + - + Data Sets Data Set ® SUMMARY ◆ CONSTRAIN → - + Organism Organism @ SUMMARY + CONSTRAIN+ - + Sequence Sequence B SUMMARY + CONSTRAIN+ - H Sequence Ontology Term SO Term | SUMMARY + CONSTRAIN+ + Chromosome Location Location B SUMMARY + CONSTRAIN+ - — — Cross References Cross Reference — SUMMARY → CONSTRAIN → + Data Sets Data Set | SUMMARY + CONSTRAIN+ → Db Cross References x Ref SUHMARY + CONSTRAIN + ■ Duplicate Entities Duplicate Entity [SUMMARY +] CONSTRAIN+ + Exons Exon | SUHMARY + CONSTRAIN+ Expression Values Expression | SUHMARY → CONSTRAIN → | - GO Annotation GO Annotation SUMMARY ◆ CONSTRAIN → - Homologues Homologue | SUHMARY + CONSTRAIN+ + Interactions Interaction | SUMMARY + CONSTRAIN+ + Located Features Location B SUMMARY + CONSTRAIN+ + Locations Location B SUMMARY + CONSTRAIN+ - Ontology Annotations Ontology Annotation Submary Constrain

Fig. 18: Example 2, Step 1 View attributes of Chromosome feature class

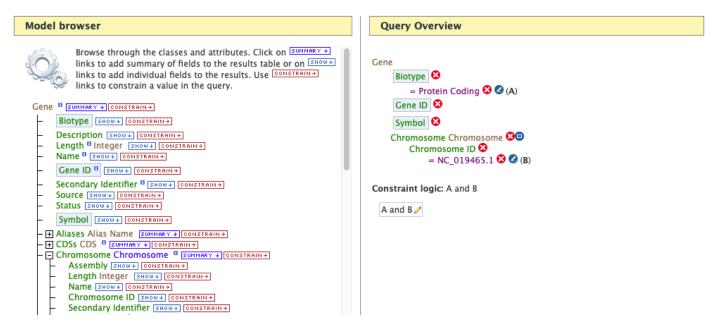


Fig. 19: Example 2, Step 2 Adding a constraint to the Chromosome ID

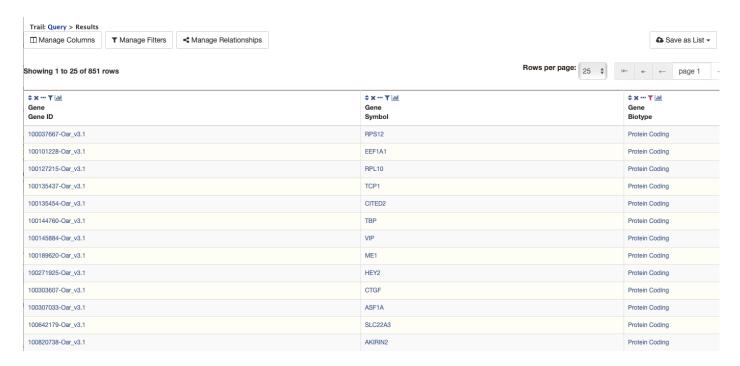


Fig. 20: Example 2, Step 3 Result table after constraining by Chromosome ID

2.3.2.3 Example 3: Querying for Protein Coding genes on a particular chromosome and their exons

This last example builds upon the previous queries to display all exons for each gene on a particular chromosome. Again, note the "Trail: Query" text at the upper left of the results table. Clicking on the "Query" link will bring you back to the query that generated the table to allow for edits without having to build a new query. Scroll down to **Exon** feature class and select the "+" (plus sign) next to Exons to display its attributes. Click **Show** next to **Length** and **Exon Identifier**.

The Query Overview shows the query in progress with the selected fields. Also notice that a third type of icon, a blue square, appears next to some attributes. Clicking on a blue square icon brings up a window where the query Join Style may be modified. When adding a constraint, you can decide whether you want to show only those results with the information (genes with exons) or all results and the constrained feature if it exists (e.g., all genes and indicate exons if they exist). Click on the blue square icon next to **Exon collection** to bring up the Switch Join Style window. The default option is to show only Genes if they have a exon (inner join). Change this to **Show all Genes and show Exons if they are present** (outer join) then click **Add to query**.

Then click "Show results" to run the new query.

The results table now lists a new column **Gene Exons**, which we added to the query. If we look at the second row in the table, it lists 7 exons. Click on the **7 exons** link to expand that entry. That column now has additional rows containing the **Exon identifier** and **Length** for each of the 7 exons.

2.3. QueryBuilder 19

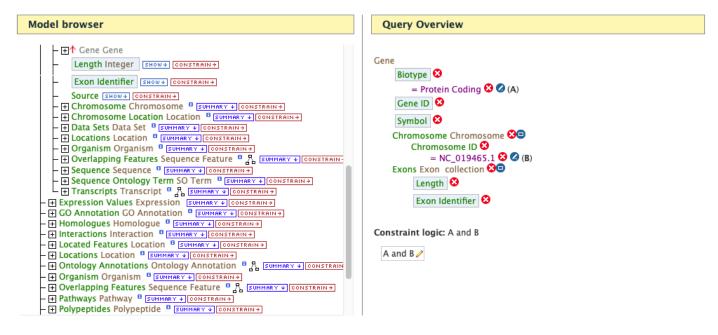


Fig. 21: Example 3, Step 1 Expand exon attributes and add fields to the query

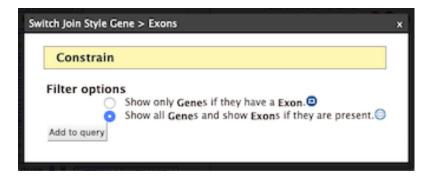


Fig. 22: Example 3, Step 2 Select join style for exons

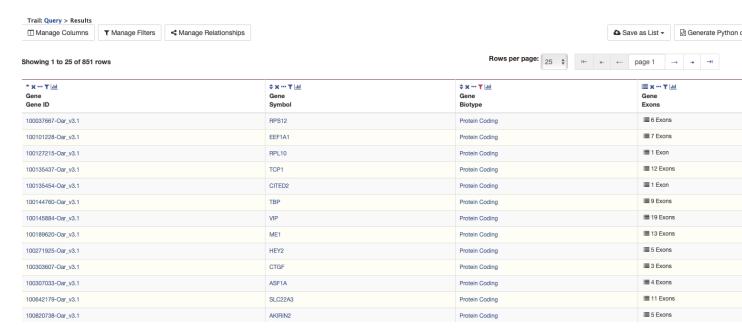


Fig. 23: Example 3, Step 3 Query results with exon constraint

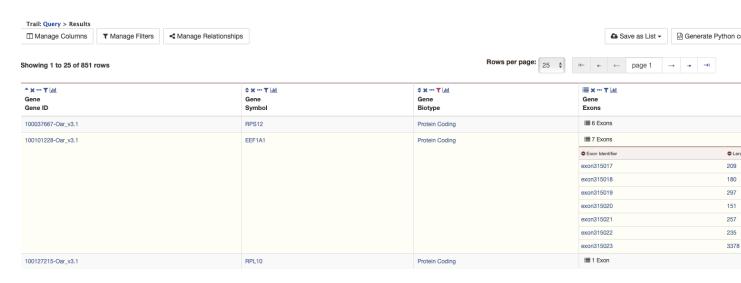


Fig. 24: Example 3, Step 4 Query results with exon column expanded

2.3. QueryBuilder 21

In changing the join style to an outer join, the exons have been grouped together by gene making it easier to determine how many exons are contained in each gene. If the same query is run with the default join (outer join) of **Show only Genes if they have an Exon**, the results table adds a new row for each new exon rather than grouping exons by gene.

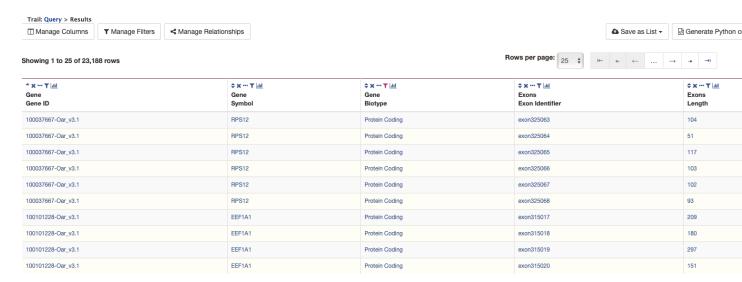


Fig. 25: Example 3, Step 5 Query results with default join style run for genes containing exons

Report Pages

All objects in BovineMine (e.g., gene, protein, transcript, publication) have report pages that can viewed after running a query. It allows users to view all available information for that object while providing links to related objects. As an example, we can revisit the templates example. In the list of templates under the **Templates** tab on the BovineMine home page, select Gene -> Homologues to query BovineMine to retrieve all homologues for a given gene. Enter "GSTM1" into the the LOOKUP search box then click **Show Results**. In the results table, note that every entry is contains a link. You can mouse over any link to bring up a summary of that object. If we hover over the first Gene ID, we can see a summary box that includes

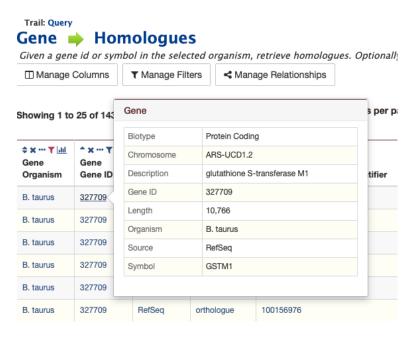


Fig. 1: Summary for gene entry in query results table

Clicking on that same item will bring up its report page that includes a comprehensive for gene GSTM1. The report page header shows the Gene ID and its Biotype, for this example, protein coding. The tabs at the top of the page in the

Quick Links menu bar quickly bring you to the data listed. The column on the right side of the report page displays external links to other Mines and databases.

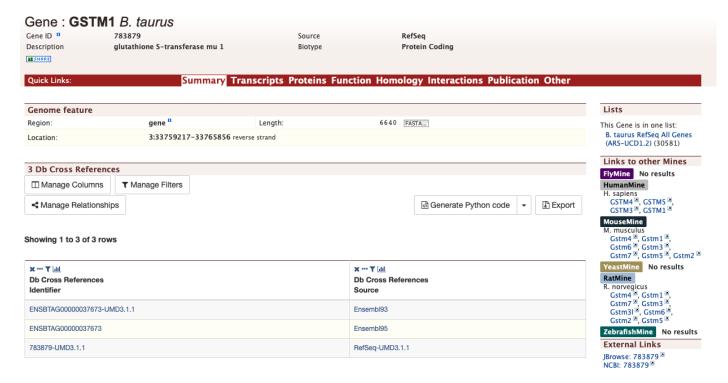


Fig. 2: Report page for protein-coding gene

The content of the report page is divided into categories based on the type of information provided for that particular object. Clicking on links within each category bring up more details about the objects of interest.

3.1 Summary

The **Summary** section near the top of the report provides information on the gene such as its length, chromosome location, and strand information. Users can also get the complete FASTA sequence of the gene by clicking on the FASTA tab.



3.2 Transcripts

The **Transcripts** section contains information about the gene model, such as transcripts and exons. Links to FASTA files are included where applicable.

Transcripts

All Transcripts for Gene – GSTM1 783879 Transcripts: 2 Exons: 16 Coding Sequence: 2			
Transcript	Exons		Coding Sequence
	exon632135	181 FASTA	
	exon632136	76 FASTA	
	exon632137	65 FASTA	
GSTM1 XM_010803234.3	exon632138	82 FASTA	XM_010803234.3-CDS
1263 FASTA	exon632139	101 FASTA	732 FASTA
	exon632140	96 FASTA	
	exon632141	111 FASTA	
	exon632142	551 FASTA	
	exon632143	54 FASTA	
	exon632144	76 FASTA	
	exon632145	65 FASTA	
GSTM1 NM_001080354.2	exon632146	82 FASTA	NM_001080354.2-CDS
1131 FASTA	exon632147	101 FASTA	657 FASTA
	exon632148	96 FASTA	
	exon632149	111 FASTA	
	exon632150	546 FASTA	

3.3 Proteins

The **Proteins** section provides information about the protein product of the gene. The comments section gives a brief description about the protein along with the UniProt accession and links to any outside data sets.

3.4 Function

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

- Cellular Component
- Molecular Function
- · Biological Process

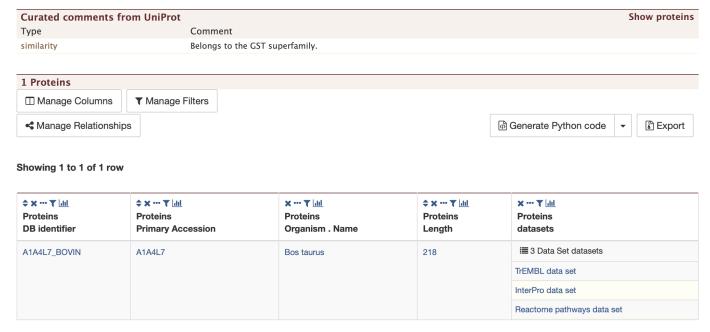
The GO terms are displayed along with the evidence code indicating how the annotations were derived. A results with Pathway information is also displayed if applicable.

3.5 Homology

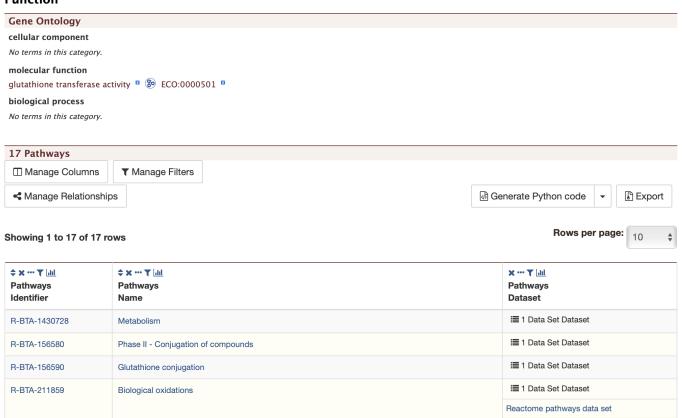
The **Homology** section provides information for all homologues. The first portion displays a summarized view of the homologues reported in different organisms. The next portion provides more detailed information about the homologue, the type of homologue and from which dataset the information was obtained all displayed in a results table.

3.3. Proteins 25

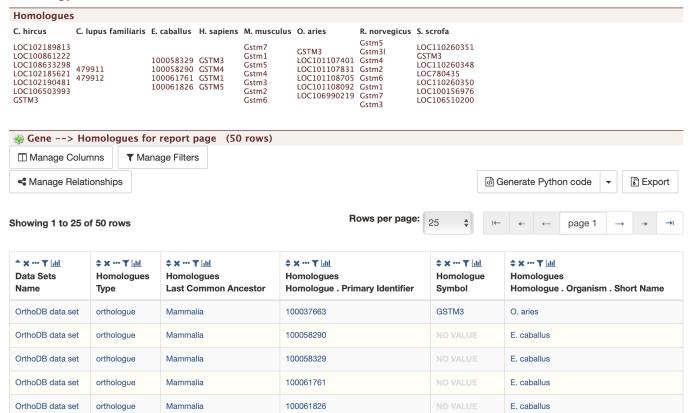
Proteins



Function



Homology



3.6 Interactions

The **Interactions** section provides interaction information. For GSTM1 there are no interaction information available but for genes that do have interaction information, a network is displayed showing all interactors for the current gene.

Interactions

```
Gene --> Interacting genes (0 rows)
```

3.7 Publications

The **Publications** section displays a table of publications related to the gene with links to full citations.

3.8 Other

This last section provides miscellaneous information that do not fit into any of the above categories. This example lists protein coding annotations and their sources.

3.6. Interactions 27

Publication



Showing 1 to 1 of 1 row

→ X ··· ▼ <u>lill</u> Publications Year	★ * * * ▼ .itll Publications First Author	\$\displays \cdot \times \tau_{\text{lil}} \\ Publications \\ Title		♦ x ··· ▼ ddl Publications Volume	♦ x ··· Y ddd Publications Pages	♦ x ··· ▼ [dil] Publications PubMed ID
2009	Zimin Aleksey V	A whole-genome assembly of the domestic cow, Bos taurus.	Genome Biol.	10	R42	19393038

Other

5 Data Sets

Protein coding gene annotations for Bos taurus ARS-UCD1.2 from NCBI RefSeq, KEGG pathways data set, TrEMBL data set, Bovine RNASeq Expression data, OrthoDB data set

CHAPTER 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 bar. 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_number:start..end
- 2. chromosome_number:start-end
- 3. chromosome_number start end (tab delimited)

Click on the input examples above the text input box (number 4) to view a representative set of coordinates in each format. Click the **Genome coordinates help** link near the top of the form for more detailed information on the input format requirements.

During a search, regions may be extended on either side of the genomic coordinates using the slider or by entering text in the field to the left of the slide bar. There is also the option to perform a strand-specific region search using the checkbox at the bottom of the form (number 6)

As an example, select *B. taurus* from the Select Organism drop-down, and ARS-UCD1.2 as the Assembly. Slick the box next to Select Feature Types to uncheck all of the boxes, then check the box next to Gene, and enter the following coordinates into the genomic regions search text field:

9:2000000..2800000

Click the search box to conduct the genomic regions search. If there are no overlaps within your search coordinates, the search can be done again with the search region extended using the slide bar or entering text into the search box (e.g., 10k).

The search results page presents a list of features present within the genomic interval that was searched. In this case, the feature type was limited to Gene. The results may be exported as tab-separated or comma-separated values. If they

Overla	p features	search f	rom a	new list	of Gen	omic Region	ıs
chromoson ids (e.g. Gk	ne identifier system. Th	e new assemblies v3.1 uses RefSeq	(ARS-UCD1.2 ids. To retriev	, ARS1, Oar_v4.0) e a map of chron	use digits, X, Y nosome identifi	00000. Be sure to use th / and MT. Bovine UMD3. ers, you can use the tem IE SET"	1.1 uses GenBank
Genome co	ordinates help						
1.	Select Organism: B. t	aurus 💠					
2.	Assembly: ARS-UCD1	2 🗘					
3.	✓ Select Feature Type	5:					
4.	✓ CDS B ✓ Exon B ✓ Gene B ✓ Primary Transcript B ✓ Probe B ✓ Indel B ✓ Lnc RNA B ✓ Ribozyme B ✓ SNP B Type/Paste in genomi (example for input for (example for tab deling)	c regions in 💿 bas mat chr:11000) mat chr:1-1000)	,	tide ⁰ Region ⁰ pt ⁰		✓ Sca RNA [□] ✓ Sequence Alteration [□] ✓ Substitution [□] ✓ tRNA [□] ✓ Tandem Repeat [□]	
	or Upload genomic rec		ile				
5.	Extend your regions a	t both sides:					
	0	1k 	10k 	100k 	1M I	10M 	
6.	Check this box to pe	erform a strand-si	oecific region	search (search +	strand if region	n start <end; search="" str<="" td="" –=""><td>and if region</td></end;>	and if region
							Reset Search

Fig. 1: Genomic Regions search form

verla	p features s	earch fro	m a new	list of	Gend	omic Regions	5
chromoson ids (e.g. GK	features that overlap a list ne identifier system. The n (000001.2). Sheep Oar_v3. .ssembly>Chromosome	ew assemblies (ARS 1 uses RefSeq ids.	S-UCD1.2, ARS1, To retrieve a map	Oar_v4.0) use of chromosor	digits, X, Y ne identifie	and MT. Bovine UMD3.1.1 ers, you can use the templ	uses GenBank
• Genor	ordinates help me regions in the following chromosome:startend, e. chromosome:start-end, e. chromosome:start-end, e. cab delimited base coordinate (e.g. BLAS citely select one. By default genome region needs to ta	g. <i>GK000001.2:15</i> .g. <i>GK000001.2:15</i> ST, GFF/GFF3) and i , the base coordina	00025000 000–25000 interbase coordi	nate (e.g. UCS)	C BED, Chad	do) systems are supported	l, users need to
1.	Select Organism: B. taur	us ≎					
2.	Assembly: ARS-UCD1.2	•					
	✓ Select Feature Types: CDS B Exon B ✓ Gene B Primary Transcript B Probe B Indel B Lnc RNA B Ribozyme B SNP B Type/Paste in genomic re (example for input forma (example for tab delimite 9:2000000.2800000	egions in o base co t chr:11000) v t chr:1-1000) v	SRNA B miRNA B Polypeptide B snRNA B snoRNA B QTL B rRNA B Repeat Region B Transcript B			Sca RNA B Sequence Alteration B Substitution B TRNA B Tandem Repeat B	
5.	or Upload genomic region Choose File no file select Extend your regions at bo	ted oth sides:		<i>//</i>	1M	10M	
					<u> </u>		
6.	Check this box to perfo end <start)< td=""><td>orm a strand-specif</td><td>ic region search (</td><td>searcn + stran</td><td>a if region</td><td>start<end; search="" stran<="" td="" –=""><td>Reset Search</td></end;></td></start)<>	orm a strand-specif	ic region search (searcn + stran	a if region	start <end; search="" stran<="" td="" –=""><td>Reset Search</td></end;>	Reset Search

Fig. 2: Genomic Regions search example with Bos taurus

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. Links within the features provide detailed reports. If users are interested in creating a list of particular features from the result page then they can filter based on feature type (if applicable), shown in red box, and click on G_0 .



Fig. 3: Genomic Regions search results

Lists

5.1 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 comma-separated identifiers into the Lists upload form under the **Lists** tab. Notice that they do not have to be in the same format. A Summary table is displayed with the results of searching for each of the five identifiers in the list.

```
105503010, KCNK1, STAT5A, BTG1, ENSBTAG00000006858
```

Leave the **Select Type** drop-down menu to **Gene** and the **Organism** drop-down to **Any**. Click on **Create List**. Note that you can also upload a list from a .txt file.

The summary table provides information regarding those identifiers that had a direct hit without any duplicates. If there are any duplicates, users can decide to add the relevant entries individually by clicking on the **Add** button under the **Action** column or choosing the **Add all** tab. Here we will click **Add all**. Once the selections have been added, the list can be saved by clicking the **Save a list of 46 Genes** button on the top of the summary table. Name the list by entering text into the **Choose a name for the list** box at the top of the results page.

After the list is saved, users are presented with a **List Analysis** page. This page provides users with widgets to perform analyses on gene lists that they have created.

The selection of widgets provided on the List Analysis page depend on the contents of the list. The available widgets for this list example include:

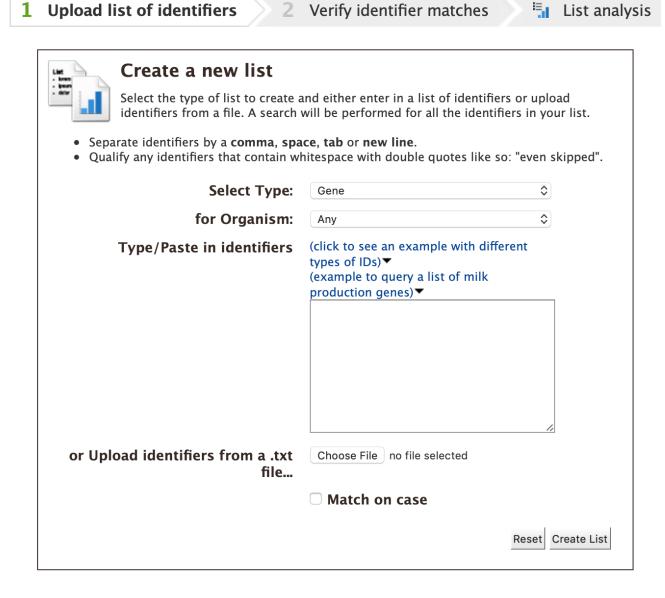


Fig. 1: List upload form

34 Chapter 5. Lists

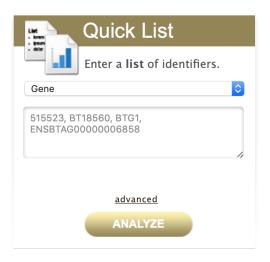


Fig. 2: Quick list from BovineMine home page

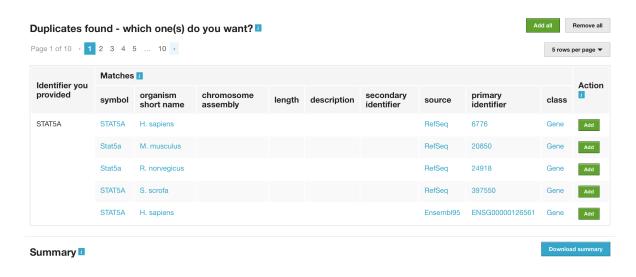


Fig. 3: List Example: Search results for list of identifiers

5.1. Creating Lists 35

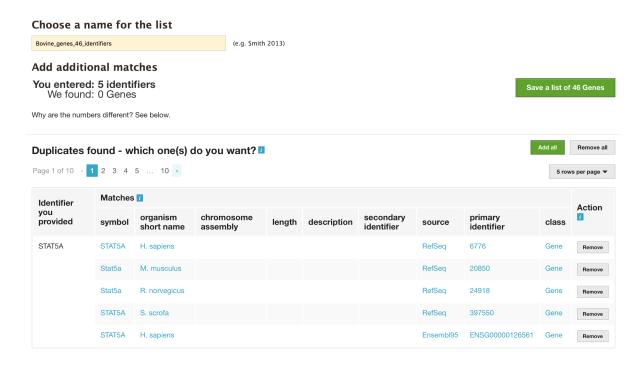


Fig. 4: List Example: Saving list of identifiers

- 1. Gene Ontology Enrichment
- 2. Publication Enrichment
- 3. Pathway Enrichment
- 4. Reaction Enrichment
- 5. Reaction Subsystem Enrichment
- 6. Orthologues

5.2 Saving Lists

To see your saved lists, click the **View** tab on the **Lists** page. If not logged in, lists will be saved temporarily during your current session. However, you must be logged in to save your lists permanently. Further analyses of lists can be done with the **Actions** links at the top of the list. The links become active once lists are selected for analyses. Saved lists may also be accessed from the **MyMine** menu tab.

36 Chapter 5. Lists



Showing rows 1 to 25 of 26

			♦ x ··· ▼ .iil Gene Source	÷ x ··· ▼ lill Gene Description		♦ x ··· ▼ .lil Gene Organism	♦ x ··· ▼ .itl Gene Assembly
100217400	NO VALUE	BTG1	RefSeq	B-cell translocation gene 1, anti-proliferative	2268	O. aries	Oar_v4.0
100217400-Oar_v3.1	NO VALUE	BTG1	RefSeq- Oar_v3.1	B-cell translocation gene 1, anti-proliferative	1235	O. aries	Oar_v3.1
100861579	NO VALUE	BTG1	RefSeq	BTG anti-proliferation factor	1536	C. hircus	ARS1
101115679	NO VALUE	KCNK1	RefSeq	potassium channel, two pore domain subfamily K, member 1	72744	O. aries	Oar_v4.0
101115679-Oar_v3.1	NO VALUE	KCNK1	RefSeq- Oar_v3.1	potassium channel, two pore domain subfamily K, member 1	140519	O. aries	Oar_v3.1

Fig. 5: List Example: Analysis for gene list

5.2. Saving Lists 37

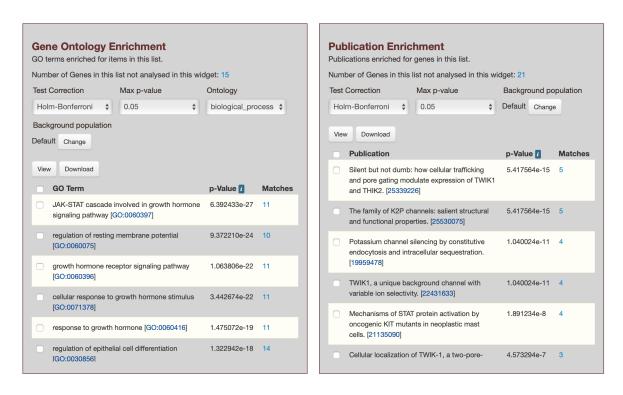


Fig. 6: List Example: Displayed widgets for list analysis

38 Chapter 5. Lists

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 - Lists saved by the user when logged in.

History - List of most recently run queries.

Queries - List of saved queries.

Templates - Templates created or marked as "favorite" by the user.

Password - Password reset form.

Account Details - User preferences form.



Your Lists

LIST NAME	DESCRIPTION	♦ TYPE	NUMBER OF OBJECTS	DATE CREATED
Gene_List_Bos_Example		Gene	5 values	2019-11-26 10:40

API

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

Perl | Python | Ruby | Java Search: e.g ATP synthase, AT5G2_1 GO

Perl Web Service Client

The Perl web service client library makes it easy to run queries in BovineMine 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.

42 Chapter 7. API

Data Sources

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

Data

This page lists all data sources loaded along with the date the data was downloaded and related publication. NOTE: All external links open in a new window.

Data Category	Organism	Data	Source	PubMed	Note
	B. taurus	NCBI Genes (RefSeq)	NCBI Bos taurus Annotation Release 106 (ARS-UCD1.2)	O'Leary et al – PubMed 26553804	NCBI FTP
	B. taurus	NCBI Genes (RefSeq)	NCBI Bos taurus Annotation Release 105 (UMD3.1.1)	O'Leary et al – PubMed 26553804	NCBI FTP
	C. hircus	NCBI Genes (RefSeq)	NCBI Capra hircus Annotation Release 102 (ARS1)	O'Leary et al - PubMed 26553804	NCBI FTP
	O. aries	NCBI Genes (RefSeq)	NCBI Ovis aries Annotation Release 102 (Oar_v4.0)	O'Leary et al – PubMed 26553804	NCBI FTP
	O. aries	NCBI Genes (RefSeq)	NCBI Ovis aries Annotation Release 101 (Oar_v3.1)	O'Leary et al – PubMed 26553804	NCBI FTP
	Homo sapiens	NCBI Genes (RefSeq)	NCBI Homo sapiens Annotation Release 109 (GRCh38.p12)	O'Leary et al – PubMed 26553804	NCBI FTP
Genes	Mus musculus	NCBI Genes (RefSeq)	NCBI Mus musculus Annotation Release 106 (GRCm38.p6)	O'Leary et al – PubMed 26553804	NCBI FTP
	Rattus norvegicus	NCBI Genes (RefSeq)	NCBI Rattus norvegicus Annotation Release 106 (Rnor_6.0)	O'Leary et al - PubMed 26553804	NCBI FTP
	Sus scrofa	NCBI Genes (RefSeq)	NCBI Sus scrofa Annotation Release 106 (Sscrofa11.1)	O'Leary et al - PubMed 26553804	NCBI FTP
	B. taurus	Official Gene Set v2.0 (UMD3.1.1)	Bovine Genome Database	Bovine Genome Sequencing and Analysis Consortium – PubMed 19390049	BGD Download

Fig. 1: BovineMine Data Sources table

ш	۸г	דכ	\Box	Y
\square /	٦г	РΤ	П	J

How to cite

Please cite the use of BovineMine as follows:

Shamimuzzaman M, Le Tourneau JJ, Unni DR, Diesh CM, Triant DA, Walsh AT, Tayal A, Conant GC, Hagen DE, Elsik CG, Bovine Genome Database: new annotation tools for a new reference genome, *Nucleic Acids Research*, gkz944, https://doi.org/10.1093/nar/gkz944 [Epub ahead of print]