

# Welcome to the GrameneMart Tutorial

A tool for batch data sequence retrieval

1. Select a Gramene dataset to search against.
2. Add filters to the dataset to increase its specificity.
3. Choose the fields to include in the report.
4. Generate a batch report in a format that can be imported into local tools, such as Excel.

# Tutorial Tips



If you are viewing this tutorial with Adobe Acrobat Reader, click the “pages” on the left hand side of the Reader for easier navigation.

*Note! Although we continually work to make Gramene compatible with all browsers, there are problems with some browser versions. If you're having difficulty viewing Gramene, try using a different browser. Please report any problems with browsers through Gramene Feedback.*



GrameneMart is based upon BioMart.

The “Mart” tool is particularly suited for providing the 'data mining like” searches of complex descriptive (e.g. biological) data, and is optimized for large databases, such as genomic sequence or microarray experiments.

BioMart software is completely Open Source, licensed under the LGPL, and freely available to anyone without restrictions

## GrameneMart

- Built upon BioMart, GrameneMart allows you to search the genes from many of the available genomes at Gramene.
- These genes can be filtered by chromosome or region, as well as other gene and protein data available for these genes.
- Results can be downloaded

# Gramene Home Page

## GRAMENE

*A Resource for Comparative Grass Genomics*

v23 (November 2006)

[Search](#)
[Genomes](#)
[Species](#)
[Download](#)
[Resources](#)
[About](#)
[Help](#)
[Feedback](#)

### Quick Search

All Available

Search

Search a single module or all available modules plus online documentation.

[Diversity](#), [Pathways](#), [BLAST](#) and [Mart](#) not available in this search.

### Have Questions...?

- Gramene now has [tutorials](#) for every module, also recommended for experienced users.
- Ask questions through [Feedback](#) or [Email](#).
- See [FAQ](#) for questions and answers.

Gramene Tip:

The AGI Maize FPC maps cannot be viewed directly, because they are not reference maps. If you are looking for them, start with a maize genetic map or the rice sequence map, then add the [maize FPC maps](#) as comparative maps.

- [Browse All Tips](#)

### Quick Start

Sequenced genomes for [Rice](#), [Maize](#) & [Arabidopsis](#); Look for [rice/maize synteny](#); Narrow search by [GeneMart](#); Search for sequence alignment with [BLAST](#); search by [Gene Ontology](#).

Search for [Fam](#) or [ProSite](#) or Browse by Gene Ontology using [GO Slim](#).

Search for physical maps for [Rice](#), [Maize](#), [Wheat](#), [Barley](#), [Oats](#), [Sorghum](#), and other grasses, or use the [Map Viewer \(CMap\)](#) to compare maps of different types and species.

Search for Genetic markers (RFLPs, SSRs, etc.), DNA Probes (Primers, Overgos, etc.), Genomic Islands, etc.), and Sequences (GSSs, ESTs, etc.); Use the Simple Sequence Repeat (SSR) Search; or search by species, including [Rice \(Oryza sativa\)](#), [Maize](#), [Sorghum](#) and [Others](#).

**All-GrameneMart**

- TRAITS**: Search the [Genes](#) or [QTL](#) database for important phenotype-related loci such as [Rice Genes](#), [Rice QTL](#), [Maize QTL](#). Don't forget to check the [traits in Ontologies](#).
- GENETIC DIVERSITY** **NEW**: Search for SNP and SSR markers in the [germplasm](#) database.
- BIOCHEMICAL PATHWAYS** **NEW**: Search for [metabolic pathways](#) of the [metabolic network](#). Compare [rice](#) and [maize](#) pathways.
- LITERATURE**: Search the literature for your [interests](#).
- SUBMISSION**: Submit a [Rice Gene](#) or [Ontology Term](#) to Gramene.

### Featured News

- NEW** Gramene Release 23 [release notes](#).
- NEW** [MaizeSequence.org](#) now available!
- NEW** [Gramene November Newsletter](#)
- [Rice News Worldwide](#) from IRRI

### Visit with us at

- [Data Warehouse Technologies in Bioinformatics](#), December 4-6, 2006.
- [Plant and Animal Genome XV Conference](#), January 13-17, 2007. ([Gramene Workshop](#))

[View Previous Gramene Presentations](#)

[Calendar](#)

Gramene is a curated, open-source, web-accessible data resource for comparative genome analysis in the grasses. Our goal is to facilitate the study of cross-species homology relationships using information derived from public projects involved in genomic and EST sequencing, protein structure and function analysis, genetic and physical mapping, interpretation of biochemical pathways, gene and QTL localization and descriptions of phenotypic characters and mutations.

*Note! Although we continually work to make Gramene compatible with all browsers, if you're having difficulty viewing Gramene, try using a different browser. Please report any problems with browsers through [Gramene Feedback](#).*

SPECIES

Last modified: Wed Nov 8 13:09:58 2006

[Home](#)
[Site Map](#)
[About](#)
[Cite](#)

10/18/07

## Select a dataset

The image shows two screenshots of a web interface. The top screenshot shows the 'Dataset' dropdown menu open, with 'Gramene 26' selected. The bottom screenshot shows the 'Dataset' dropdown menu open, with 'Gramene 26' selected, and a list of available datasets within that database.

1) Select a **database** for this query, and the datasets will open up.

Select a **dataset** for this query, these are the available datasets within the database you selected..

*After selecting your dataset, you may need to wait a few moments for this page to load.*

## “Count”

Click the “Count” button at any point to see how many results you have. Clicking on Count as you apply each filter will update the summary .

The screenshot shows a web interface with a top navigation bar containing buttons for 'New', 'Count', and 'Results'. Below this, a sidebar on the left contains sections for 'Dataset' (showing '58406 / 58406 Genes'), 'Filters' (with '[None selected]'), 'Attributes' (listing 'Ensembl Gene ID' and 'Ensembl Transcript ID'), and another 'Dataset' section (showing '[None Selected]'). The main content area on the right features a dropdown menu currently set to 'Gramene 26' and another dropdown menu below it showing 'Oryza sativa genes (TIGR5)'. Green callout boxes with arrows point to specific elements: one points to the 'Count' button, another points to the 'Dataset' section in the sidebar, a third points to the 'Filters' section, and a fourth points to the 'Attributes' section.

*You have the whole dataset of 58406 Oryza sativa japonica genes, because no filters have been applied at this point.*

Click “Filters” to see your options (next slide)

*Default results options may display here under “Attributes”*



After each filter selection click on “**Count**” to update your summary, but remain on the filter page.

## Formulate the Filters:

New Count Results XML Perl Help

Dataset 58406 / 58406 Genes

Filters  
[None selected]

Attributes  
[None selected]

Dataset  
[None Selected]

Please restrict your query using criteria below

- ⊕ REGION:
- ⊕ GENE:
- ⊕ GENE ONTOLOGY:
- ⊕ MULTISPECIES COMPARISONS:
- ⊕ PROTEIN:

*Remember the summary shows the number of records matching your query and filters. Continue to add filters until the number becomes a manageable size.*

Click on the expand icon in front of a filter to open that category to be able to apply aspects of that filter.



## Formulate the Filters: Optionally Filter by Region

**New Count Results** XML Perl Help

Please restrict your query using the filters below

**Dataset** 4198 58406 Genes

**Filters**

Chromosome 12

**Attribute**

[None selected]

**Dataset**

[None Selected]

**REGION:**

☒ Chromosome

☐ Base pair

Gene Start (bp)

Gene End (bp)

☐ Genomic features

Type of feature

ID(s)

**GENE:**

**PHENOLOGY:**

**SPECIES COMPARISONS:**

1

10

11

**✓ 12**

2

3

4

5

6

7

8

9

chloroplast

mitochondrion

AC099402

AE017048

AE017081

AE017082

AE017085

AE017091

AE017097

AE017099

AE017100

AE017101

AE017105

AE017111

AE017114

AE017115

AE017116

AE017117

AE017120

AE017121

AE017122

2 - Then click on "count" to update your summary. There are 4198 genes on chromosome 12 of the Oryza sativa japonica genome.

1 - For example - select the region of your choice

## Formulate the Filters: Optionally Filter by Region

**Dataset** 4198 / 58406 Genes

**Filters**

Chromosome: 12

**Attributes**

[None selected]

**Dataset**

[None Selected]

Please restrict your query using criteria below

☒ **REGION:**

☒ Chromosome 12

☐ Base pair

Gene Start (bp) 1

Gene End (bp) 10000000

☐ Genomic features

Type of feature FPContig name

ID(s)

Choose File no file selected

You can search for all genes between two individual BasePairs

You can search for all genes on one or more FPContigs

## Formulate the Filters: Optionally Filter by Gene

If you wish to filter by these criteria, **select the box** and **enter specific data**

The screenshot shows a web interface for filtering genes. The top navigation bar includes 'New', 'Count', 'Results', 'Perl', and 'Help'. The left sidebar displays 'Dataset 4198 / 58406 Genes' and a 'Filters' section with 'Chromosome: 12' and 'Attributes' (Ensembl Gene ID, Ensembl Transcript ID). The main area is titled 'GENE:' and contains several filter sections:

- ID LIST FILTERS:** Includes a checkbox and a help icon. A callout points to the 'ID list limit' checkbox, stating: 'If you wish to filter by these criteria, **select the box** and **enter specific data**'. Below this is a dropdown menu set to 'with Ensembl Gene ID(s)' and radio buttons for 'Only' (selected) and 'Excluded'.
- ID list limit:** Includes a checkbox and a help icon. A callout points to the 'Choose File' button, stating: 'Import values from a file'. Below this is a text input field and a 'Choose File' button labeled 'no file selected'.
- Transcript count >=**: Includes a checkbox and a text input field.
- Gene type**: Includes a checkbox and a dropdown menu with options: 'miRNA', 'protein\_coding', 'pseudogene', 'siRNA', and 'transposon\_pseudogene'.

## Formulate the Filters: Optionally Filter by Gene Ontology

If you wish to filter by gene ontology, **select the box** and **enter specific data**

The screenshot shows the EBI QuickGO interface. At the top, there are tabs for 'New', 'Count', and 'Results'. Below these, there are links for 'XML', 'Perl', and 'Help'. The main content area is divided into two columns. The left column contains a sidebar with 'Dataset 23 / 58406 Genes', 'Filters', 'Attributes', and 'Dataset'. The 'Filters' section shows 'Chromosome: 12', 'Evidence code (Molecular function) : IEA', and 'Molecular function : GO:0030234'. The 'Attributes' section shows 'Ensembl Gene ID' and 'Ensembl Transcript ID'. The 'Dataset' section shows '[None Selected]'. The right column contains the 'GENE ONTOLOGY' section, which has three rows. Each row has a checkbox, a label, and a field for the GO term. The first row is 'Evidence code (Molecular function)' with a checked checkbox, 'IEA' in a dropdown, and a 'browse' button. The second row is 'Molecular function' with a checked checkbox, 'GO:0030234' in a text box, and a 'browse' button. The third row is 'Evidence code (Biological process)' with an unchecked checkbox, 'IEA' in a dropdown, and a 'browse' button. The fourth row is 'Biological process' with an unchecked checkbox, an empty text box, and a 'browse' button. The fifth row is 'Evidence code (Cellular component)' with an unchecked checkbox, 'IEA' in a dropdown, and a 'browse' button. The sixth row is 'Cellular component' with an unchecked checkbox, an empty text box, and a 'browse' button.

Then click on “count” to update your summary. There are 23 genes on chromosome 12 of with our selected ontology.

Use the ‘browse’ button to go to EBI’s **QuickGO** to find your GO term (*see next slide*).

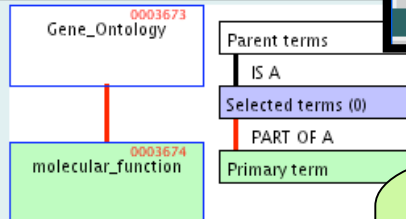
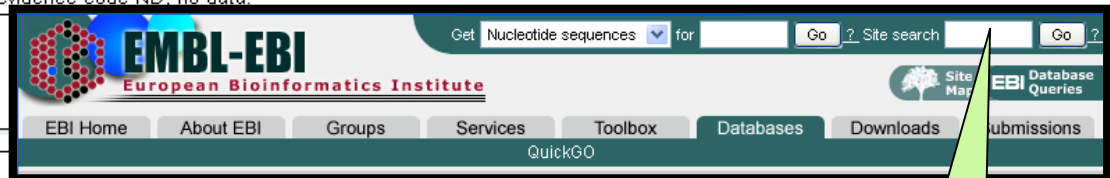
# QuickGO

QuickGO **GO Term GO:0003674**

[?] = h

<b>Term ID</b> [?]	GO:0003674
<b>Name</b> [?]	molecular_function
<b>Last updated</b> [?]	2001-03-30 04:29:44.0
<b>Definition</b> [?]	Elemental activities, such as catalysis or binding, describing the actions of a gene product at the molecular level. A given gene product may exhibit one or more molecular functions
<b>Synonyms</b> [?]	molecular function unknown
<b>Comment</b> [?]	Note that this term should be used for the annotation of gene products whose molecular function is unknown by annotating to this node with the evidence code ND, no data.
<b>Hierarchy</b> [?]	<ul style="list-style-type: none"> <li>View this term's parents in a denormalised tree.</li> <li>View with neither graph nor tree.</li> <li>Hide all selected terms except the primary one</li> <li>Add more terms to the selection with a search</li> </ul>

You should be on the ontology you entered from



<b>Child terms</b> [?]	<p>GO:0016209: antioxidant activity</p> <p>GO:0005488: binding</p> <p>GO:0003824: catalytic activity</p> <p>GO:0030188: chaperone regulator activity</p> <p>GO:0042056: chemoattractant activity</p> <p>GO:0045499: chemorepellant activity</p> <p>GO:0031992: energy transducer activity</p> <p>GO:0030234: enzyme regulator activity</p> <p>GO:0003774: motor activity</p> <p>GO:0045735: nutrient reservoir activity</p> <p>GO:0008369: obsolete molecular function</p> <p>GO:0031386: protein tag</p> <p>GO:0004871: signal transducer activity</p> <p>GO:0005198: structural molecule activity</p>
------------------------	---

Browse to find most appropriate term. (Click a term to see it's child terms) Note the term ID of choice (copy) and return to filter page (paste).

Or use the search function

## Formulate the Filters: Optionally Filter by Species Homologs

New Count Results XML Perl Help

Dataset 23 / 58406 Genes

Filters

Chromosome: 12  
Evidence code (Molecular function) : IEA  
Molecular function : GO:0030234

Attributes

Ensembl Gene ID  
Ensembl Transcript ID

Dataset  
[None Selected]

GENE:

GENE ONTOLOGY:

☒ Evidence code (Molecular function) IEA

☒ Molecular function GO:0030234 browse

☐ Evidence code (Biological process) IEA

☐ Biological process browse

☐ Evidence code (Cellular component) IEA

☐ Cellular component

MULTI SPECIES COMPARISONS:

☐ Homolog filters

Orthologous Maize Genes

Paralogous Rice (Japonica) Genes  
Orthologous Arabidopsis Genes  
Orthologous Rice (Indica) Genes  
Orthologous Poplar Genes  
Orthologous Sorghum Genes  
Orthologous Maize Genes  
Orthologous Drosophila Genes  
Orthologous Elegans Genes  
Orthologous Human Genes  
Orthologous Yeast Genes

Only Excluded

If you wish to compare with other available species, **select the box** and indicate whether to **compare only** or **exclude** that species.



## Formulate the Filters: Optionally Filter by Protein Information

The screenshot shows a web interface with a top navigation bar containing 'New', 'Count', 'Results', 'XML', 'Perl', and 'Help'. On the left, a sidebar displays 'Dataset 23 / 58406 Genes' and a 'Filters' section with 'Chromosome: 12', 'Evidence code (Molecular function) : IEA', and 'Molecular function : GO:0030234'. Below this is an 'Attributes' section with 'Ensembl Gene ID' and 'Ensembl Transcript ID'. The main content area is titled 'MULTI SPECIES COMPARISONS.' and includes a 'PROTEIN:' section with two checkboxes: 'Limit to genes with these family or domain IDs:' and 'Limit to genes with these family or domain IDs:'. A dropdown menu is open, showing options: 'with PROFILE ID(s)', 'with PFAM ID(s)', 'with coil(s)', 'with low complexity region(s)', 'with PRINTS ID(s)', 'with PROSITE ID(s)', and 'with InterPro ID(s)'. Below the dropdown is a text input field for 'PFAM ID(s)' and a 'Choose File' button. Another dropdown menu is open, showing options: 'PFAM ID(s)', 'PRINTS ID(s)', 'PROFILE ID(s)', 'PROSITE ID(s)', 'Ensembl Family ID(s)', and 'Interpro ID(s)'. A red arrow points from the 'Choose File' button to the second dropdown menu. Two green callout boxes provide instructions: one points to the 'PROTEIN:' section and the other points to the 'Choose File' button.

**If you wish to filter by protein, **select the box**, make a selection from the **drop down menu** and **enter specific data****

**Use the Choose File button to import values from a file**



## Formulate the Filters: Optionally Filter by SNP information

The screenshot shows a web interface for SNP filtering. On the left, a sidebar contains 'Dataset 23 / 58406 Genes', 'Filters' (with sub-items: Chromosome: 12, Evidence code (Molecular function): IEA, Molecular function: GO:0030234), 'Attributes' (with sub-items: Ensembl Gene ID, Ensembl Transcript ID), and 'Dataset' (with sub-item: [None Selected]). The main panel is titled 'MULTI SPECIES COMPARISONS.' and has tabs for 'New', 'Count', 'Results', 'XML', 'Perl', and 'Help'. It contains several filter sections: 'Homolog filters' (with a dropdown for 'Paralogous Rice (Japonica) Genes' and radio buttons for 'Only' and 'Excluded'), 'PROTEIN:' (with two 'Limit to genes with these family or domain IDs:' sections, one for 'with PROFILE ID(s)' and one for 'PFAM ID(s)', each with a dropdown and radio buttons), 'SNP:' (with 'Snp type' (dropdown), 'Synonymous status' (dropdown), and 'Associated with validated SNPs' (radio buttons)). A dropdown menu is open for the 'SNP:' section, showing options: 'Coding' (selected), 'Intronic', '5utr', '5upstream', '3utr', and '3downstream'. Another dropdown menu is open for the 'Synonymous status' filter, showing options: 'Frameshifting SNPs' (checked), 'Synonymous SNPs', 'Non-synonymous SNPs', 'Stop gained SNPs', 'Stop lost SNPs', 'Splice site SNPs', and 'Essential splice site SNPs'. A red arrow points from the 'Essential splice site SNPs' option to the 'Frameshifting SNPs' option in the 'Synonymous status' dropdown.

Dataset 23 / 58406 Genes

**Filters**

Chromosome: 12  
Evidence code (Molecular function): IEA  
Molecular function: GO:0030234

**Attributes**

Ensembl Gene ID  
Ensembl Transcript ID

**Dataset**

[None Selected]

**MULTI SPECIES COMPARISONS.**

**Homolog filters**

Paralogous Rice (Japonica) Genes ☒ Only ☐ Excluded

**PROTEIN:**

Limit to genes with these family or domain IDs: with PROFILE ID(s) ☐ Only ☐ Excluded

Limit to genes with these family or domain IDs: PFAM ID(s) ☐ Only ☐ Excluded

**SNP:**

Snp type ☐ Only ☐ Excluded

Synonymous status ☒ Only ☐ Excluded

Associated with validated SNPs ☒ Only ☐ Excluded

**Frameshifting SNPs**

- Synonymous SNPs
- Non-synonymous SNPs
- Stop gained SNPs
- Stop lost SNPs
- Splice site SNPs
- Essential splice site SNPs

## After Filtering, go to Attributes

**New** **Count** **Results** **XML** **Perl** **Help**

**Dataset** 23 / 58406 Genes

**Filters**

Chromosome: 12  
Evidence code (Molecular function)  
: IEA  
Molecular function : GO:0030234

**Attributes**

Ensembl Gene  
Ensembl Transcript

**Dataset**  
[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

☒ **Features** ☐ **Homologs**  
☐ **Structures** ☐ **Sequences**  
☐ **SNPs**

⊕ **GENE:**  
⊕ **EXTERNAL:**  
⊕ **PROTEIN:**  
⊕ **GENOMIC REGION:**

Output fields can be selected from Features, Structures, SNPs, Homologs or Sequences. Select to navigate between them. Each type must be run individually.

After your output is a manageable size, click on “**Attributes**” to select your output options

# Attributes - Features

The screenshot shows a web interface for a bioinformatics tool. At the top, there are tabs for 'New', 'Count', and 'Results'. Below these, there's a section for 'Dataset 23 / 58406 Genes' with filters for 'Chromosome: 12', 'Evidence code (Molecular function): IEA', and 'Molecular function: GO:0030234'. The 'Attributes' section is highlighted with a green box. It contains a list of attributes with checkboxes: 'Ensembl Gene ID', 'Ensembl Transcript ID', 'Gene Start (bp)', 'Gene End (bp)', 'Prosite ID', 'PFAM ID', and 'Description'. A green callout points to this section, stating: 'Select **output options** you want by clicking on the box to the left of an option to check it. Selected items will show in summary table.'

Below the 'Attributes' section, there's a 'Dataset' section with a dropdown menu showing '[None Selected]'. To the right, there's a 'Features' section with radio buttons for 'Features', 'Structures', and 'SNPs'. The 'Features' radio button is selected and circled in red. A green callout points to this section, stating: 'When you are happy with the query you can preview the results by clicking the Results button (*see next slide*)'. Below the 'Features' section, there's a 'Homologs' section with radio buttons for 'Homologs' and 'Sequences'. The 'Homologs' radio button is selected. Below the 'Homologs' section, there's a 'Protein' section with a dropdown menu for 'Protein' and a list of 'Family Attributes' with checkboxes: 'Family Description', 'Ensembl Family ID', 'PFAM ID', 'Interpro ID', and 'Signal domain'. Below the 'Protein' section, there's a 'Transmembrane and Signal Domain Attributes' section with checkboxes for 'Transmembrane' and 'Signal domain'. Below the 'Transmembrane and Signal Domain Attributes' section, there's a 'Gene' section with a dropdown menu for 'Gene' and a list of 'Feature class', 'Subtype category', and 'Subtype description' with checkboxes.

By default this shows the first 10 results in hyper-linked HTML format but the number and format can be altered using the drop downs above this preview.

## Results -Features

The screenshot shows a web interface for viewing gene search results. The top navigation bar includes links for 'New', 'Count', 'Results', 'XML', 'Perl', and 'Help'. On the left, a sidebar contains 'Dataset 23 / 58406 Genes', 'Filters' (Chromosome: 12, Evidence code: IEA, Molecular function: GO:0030234), 'Attributes' (Ensembl Gene ID, Ensembl Transcript ID, Gene Start (bp), Gene End (bp), Description), and 'Dataset' ([None Selected]). The main area displays a table of results with columns for Ensembl Gene ID, Ensembl Transcript ID, Gene Start (bp), and Description. Above the table, there are controls for 'Export all results' (set to TSV), 'Email notification', 'View' (set to 10), and 'Format' (set to HTML). A 'Unique results only' checkbox is also present. Callouts highlight these features: one points to the filters and attributes sidebar, another points to the 'View' and 'Format' dropdowns, and a third points to the 'Unique results only' checkbox.

Export all results: TSV ☐ Unique results only

Email notification:

View: 10  Unique results only ☐

Ensembl Gene ID	Ensembl Transcript ID	Gene Start (bp)	Description
LOC_Os12g36210	LOC_Os12g36210.1	22159864	subtilisin putative,
LOC_Os12g12010	LOC_Os12g12010.1	6582236	verticillium precursor
LOC_Os12g16890	LOC_Os12g16890.1	9686394	prolamin expressed
LOC_Os12g37480	LOC_Os12g37480.1	22966329	18 enzyme inhibitor/ pectinesterase/ pectinesterase inhibitor, putative, expressed
LOC_Os12g10930	LOC_Os12g10930.1		NLOE, putative
LOC_Os12g05900	LOC_Os12g05900.1		ac GTPase activating protein 1, putative, expressed
LOC_Os12g05900	LOC_Os12g05900.1		ac GTPase activating protein 1, putative, expressed
LOC_Os12g39210	LOC_Os12g39210.1		yclin-A2, putative, expressed
LOC_Os12g30020	LOC_Os12g30020.1		yclin-T1, putative, expressed
LOC_Os12g30020	LOC_Os12g30020.1		yclin-T1, putative, expressed

If you want to save your results, options in the top section of the page allow you to select an export format.

In addition you can choose to show unique results in cases where there is redundancy due to the data model but note this can affect how long it takes to retrieve your results.

*Hyperlinks to Genome data*

# Export Results

Note for particularly intensive queries server time-outs can be a problem, in which case use one of the Web file options.

New Count Results XML Perl Help

Dataset 23 / 58406 Genes

Filters

Chromosome: 12

Export all results to File

Email notification to

HTML  
CSV  
TSV  
XLS  
ADF

☐ Unique results only Go

HTML is web-page format, with active links to Genomes data  
CSV - comma separated values  
TSV - Tab separated values  
XLS - excel file  
ADF -



# Query Script

New

Count

Results

XML

Perl

Help

Dataset 23 / 58406 Genes

Filters

Chromosome: 12  
Evidence code (Molecular function): IEA  
Molecular function : GO:0030215

Attributes

Ensembl Gene ID  
Ensembl Transcript ID  
Gene Start (bp)  
Gene End (bp)  
Prosit ID  
PFAM ID

☒ Features
☐ Homologs
☐ Structures
☐ Sequences

```

# An example script demonstrating the use of BioMart API.
# This perl API representation is only available for configuration versions >= 0.5
use strict;
use BioMart::Initializer;
use BioMart::Query;
use BioMart::QueryRunner;

my $confFile = "PATH TO YOUR REGISTRY FILE UNDER biomaart-perl/conf/. For Biomaart Central Registry navigate to
http://www.biomaart.org/biomaart/martservice?type=registry";

#
# NB: change action to 'clean' if you wish to start a fresh configuration
# and to 'cached' if you want to skip configuration step on subsequent runs from the same registry
#

registryFile'=>$confFile, 'action'=>$action);

registry,'virtualSchemaName'=>'default');

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE Query>
<Query virtualSchemaName = "default" formatter = "TSV" header = "0"
uniqueRows = "0" count = "" datasetConfigVersion = "0.6" >

  <Dataset name = "ojaponica_gene_ensembl" interface="default">
    <Filter name = "evidence_code" value = "IEA"/>
    <Filter name = "mol_function" value = "GO:0030215"/>
    <Filter name = "chromosome_name" value = "12"/>
    <Attribute name = "ensembl_gene_id" />
    <Attribute name = "ensembl_transcript_id" />
    <Attribute name = "start_position" />
    <Attribute name = "end_position" />
    <Attribute name = "prosit" />
    <Attribute name = "pfam" />
    <Attribute name = "description" />
  </Dataset>
</Query>

my $query_runner = BioMart::QueryRunner->new(
##### GET COUNT #####
# $query->count(1);

```

A summary of your current query in BioMart Query XML format can be achieved by clicking the XML button on the toolbar. This is useful for people accessing BioMart via the martservice webservice interface. Similarly a BioMart perl API script to run the same query can be generated by clicking the Perl button. For more detailed help please see the documentation at [www.biomaart.org](http://www.biomaart.org).

10/18/07

my \$query\_runner = BioMart::QueryRunner->new(
##### GET COUNT #####
# \$query->count(1);

# Attributes - Structures

**Dataset 23 / 58406 Genes**

**Filters**

Chromosome: 12  
Evidence code (Molecular function) : IEA  
Molecular function : GO:0030234

**Attributes**

Ensembl Gene ID  
Chromosome  
Biotype  
Gene Start (bp)  
Gene End (bp)  
Ensembl Exon ID

**Dataset**

[None Selected]

**Features**

☒ Structures  
☐ Homologs  
☐ Sequences  
☐ SNPs

**GENE:**

**Ensembl Attributes**

☒ Ensembl Gene ID  
☐ Ensembl Transcript ID  
☐ Ensembl Peptide ID  
☒ Chromosome  
☒ Gene Start (bp)  
☒ Gene End (bp)  
☐ Transcript Start (bp)  
☐ Transcript End (bp)  
☐ Strand

**EXON:**

**Exon Attributes**

☐ Exon Rank  
☒ Ensembl Exon ID  
☐ Exon Start (bp)  
☐ Exon End (bp)  
☐ Coding Start

**Export all results to**

File  TSV ☐ Unique results only

**Email notification to**

**View**

20 rows as CSV ☐ Unique results only

**Dataset 23 / 58406 Genes**

**Filters**

Chromosome: 12  
Evidence code (Molecular function) : IEA  
Molecular function : GO:0030234

**Attributes**

Ensembl Gene ID  
Chromosome  
Biotype  
Gene Start (bp)  
Gene End (bp)  
Ensembl Exon ID

**Dataset**

[None Selected]

Ensembl Gene ID,Chromosome,Biotype,Gene Start (bp),Gene End (bp),Ensembl Exon ID  
LOC\_Os12g36210,12,protein\_coding,22159864,22160429,LOC\_Os12g36210.1.exon1  
LOC\_Os12g12010,12,protein\_coding,6582236,6585235,LOC\_Os12g12010.1.exon1  
LOC\_Os12g16890,12,protein\_coding,9686394,9687123,LOC\_Os12g16890.1.exon1  
LOC\_Os12g37480,12,protein\_coding,22966329,22967418,LOC\_Os12g37480.1.exon1  
LOC\_Os12g10930,12,protein\_coding,5903561,5906601,LOC\_Os12g10930.1.exon3  
LOC\_Os12g10930,12,protein\_coding,5903561,5906601,LOC\_Os12g10930.1.exon2  
LOC\_Os12g10930,12,protein\_coding,5903561,5906601,LOC\_Os12g10930.1.exon1  
LOC\_Os12g05900,12,protein\_coding,2707478,2712514,LOC\_Os12g05900.1.exon5  
LOC\_Os12g05900,12,protein\_coding,2707478,2712514,LOC\_Os12g05900.1.exon4  
LOC\_Os12g05900,12,protein\_coding,2707478,2712514,LOC\_Os12g05900.1.exon3  
LOC\_Os12g05900,12,protein\_coding,2707478,2712514,LOC\_Os12g05900.1.exon2  
LOC\_Os12g05900,12,protein\_coding,2707478,2712514,LOC\_Os12g05900.1.exon1  
LOC\_Os12g39210,12,protein\_coding,24109067,24111320,LOC\_Os12g39210.1.exon1  
LOC\_Os12g39210,12,protein\_coding,24109067,24111320,LOC\_Os12g39210.1.exon2  
LOC\_Os12g39210,12,protein\_coding,24109067,24111320,LOC\_Os12g39210.1.exon3  
LOC\_Os12g39210,12,protein\_coding,24109067,24111320,LOC\_Os12g39210.1.exon4

10/18/07



# Attributes - SNPs

**New** **Count** **Results** **XML** **Perl** **Help**

**Dataset 23 / 58406 Genes**

**Filters**

Chromosome: 12  
Evidence code (Molecular function): IEA  
Molecular function : GO:0030234

**Attributes**

Synonymous SNP count  
Non-synonymous SNP count

**Dataset**  
[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

☐ Features ☐ Homologs  
☐ Structures ☐ Sequences  
☒ **SNPs**

**Gene SNP Attributes**

☒ Synonymous SNP count ☒ Non-synonymous SNP count

**New** **Count** **Results** **XML** **Perl** **Help**

**Dataset 23 / 58406 Genes**

**Filters**

Chromosome: 12  
Evidence code (Molecular function): IEA  
Molecular function : GO:0030234

**Attributes**

Synonymous SNP count  
Non-synonymous SNP count

**Dataset**  
[None Selected]

Export all results to   ☐ Unique results only

Email notification to

View  rows as  ☐ Unique results only

	1
	2
3	4
5	2
5	2
	4
3	8
15	4
14	5
21	10
11	4
11	12
3	1
1	2
5	3
1	6
1	

10/18/07

# Attributes - Homologs

**New** **Count** **Results** **XML** **Perl** **Help**

☐ Features ☒ **Homologs** ☐ Structures ☐ Sequences ☐ SNPs

**Dataset** 23 / 58406 Genes

**Filters**

Chromosome: 12  
Evidence code (Molecular function) : IEA  
Molecular function : GO:0030234

**Attributes**

Ensembl Gene ID  
Ensembl Transcript ID

**Dataset**

[None Selected]

**GENE:**

**Ensembl Attributes**

☒ Ensembl Gene ID  
☒ Ensembl Transcript ID  
☐ Ensembl Peptide ID  
☐ Chromosome Name  
☐ Gene Start (bp)  
☐ Gene End (bp)  
☐ Strand

☐ External Gene ID  
☐ External Gene DB  
☐ Ensembl Peptide length  
☐ Transcript count  
☐ % GC content  
☐ Description

**PARALOGS:**

**Rice (Japonica) Paralogs Attributes**

☐ Rice (Japonica) Paralogs Chr Start (bp)  
☐ Rice (Japonica) Paralogs Ensembl Gene ID  
☐ Rice (Japonica) Paralogs External ID  
☐ Rice (Japonica) Paralogs External DB  
☐ Rice (Japonica) Paralogs Chromosome  
☐ Rice (Japonica) Paralogs Gene Start (bp)  
☐ Rice (Japonica) Paralogs Gene End (bp)  
☐ Ensembl Peptide ID

☐ % Coverage  
☐ % Identity  
☐ Rice (Japonica) Paralog Ensembl Peptide ID  
☐ Rice (Japonica) Paralog % Coverage  
☐ Rice (Japonica) Paralog % Identity  
☐ Ancestor

**ARABIDOPSIS ORTHOLOGS:**

**RICE (INDICA) ORTHOLOGS:**

**Maize ORTHOLOGS:**

**POPLAR ORTHOLOGS:**

**SORGHUM ORTHOLOGS:**

**C.ELEGANS ORTHOLOGS:**

**DROSOPHILA ORTHOLOGS:**

**HUMAN ORTHOLOGS:**

**YEAST ORTHOLOGS:**

An example of options for homologs

Options vary with attribute type

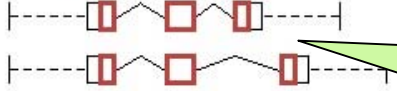
**New Count Results** **XML Perl Help**

Please select columns to be included in the output and hit 'Results' when ready

☐ Features
 ☐ Homologs
 ☐ Structures
 ☒ Sequences
 ☐ SNPs

☐ SEQUENCES:

**Sequences (max 1)**



☐ Unspliced (Transcript)
 ☐ 5' UTR
 ☐ Unspliced (Gene)
 ☐ 3' UTR
 ☐ Flank (Transcript)
 ☐ Exon sequences
 ☐ Flank (Gene)
 ☐ cDNA sequences
 ☐ Flank-coding region (Transcript)
 ☐ Coding sequence
 ☐ Flank-coding region (Gene)
 ☒ Peptide

**Upstream flank**

☐ Upstream flank

**Downstream flank**

☐ Downstream flank

**Dataset 23 / 58406 Genes**

**Filters**

Chromosome: 12  
Evidence code (Molecular function) : IEA  
Molecular function : GO:0030234

**Attributes**

Peptide  
Ensembl Gene ID  
Chromosome  
Biotype

**Dataset**

[None Selected]

## Attributes - Sequences

An example of options for Sequences. The diagram will have you're the area of your option highlighted in red.

**New Count Results** **XML Perl Help**

**Dataset 23 / 58406 Genes**

**Filters**

Chromosome: 12  
Evidence code (Molecular function) : IEA  
Molecular function : GO:0030234

**Attributes**

Ensembl Gene ID  
Chromosome  
Biotype  
Peptide

**Dataset**

[None Selected]

Export all results to   ☐ Unique results only

Email notification to

View  rows as  ☐ Unique results only

```

>LOC_Os12g17550|12|protein_coding
MADSSSSMTNQLRETQKQETTIRAHILKPNTSNHHHHQQQQQQQQQQQQQGGEDLSKKK
QEEETSPPRPVGGHGGGTWRSDLRIPGLVEMELCIMLECCSQERTYLPYYGLLAQRL
CLINKVYQKNFEKFAKQYSMIDRLDTNKLGNVANFFAHLATDALPWHVLAIRLTEED
TTSSSRIFIKILFHELSDHLGIRQLNKRSLDPKMDYFDSIFLMDHPKNTFRWINFFTSI
GLGGITETLREYQCLQCNNRSQNVLMRVAETQKKKNRMAEPRFPPELTKIALLCLLL
PLFSPATASIPSAASTSSDRSCIADERAALLAIKATFFDPNSRLASWQEDCCSWGVRC
SNRTGHVIKRLRGNTDDCLSFYGDKLRGEMSYSLVSLQKRLYLDLSCNNFNWSQIPVFL
GSLPSLRYLNLGYGFFYGSVPPQLGNLSKLAYLDLTSYSYNQLYSVALSWLSLSSLKHL
VMNHVNLTTAVDWVDEINMLPALKVLYLKQCGLRKTVPFLLRSNITGLEVLDISGNRFHT
KIAPNWFWNITSLSALDIRSCGFFGSIPDEIGRMASLEEVYFQGNLMSTMIPSSFKNLC
NLKVLDLRSTNTTGDIRELIEKLPNCHWNKLQQLGSLSYNNIGTLPNWEPLANLTVLLL
SNTNISGAMPSSIWALTKLNLIDLCSNKLNGTVREDQLGNLNLVYLGNGTHLQIKASS
DWIPPFKLQFSGNLPLWMKKFLPLSLRLRSNMFSGHIPTELTRIDQLQFLDLAENYF
SGSIPDSLVNLSAMARTSGYSVLLDEVIATGQGAILNFSWNLINGEIPETIGLQKLESL
DLSHNELSGEIPSSMQDLNALGTMNLSYNNLSGRIPRGNTMGSSYDASSYIGNIGLCPPL
TRNCSGNATSKDLPRNHVDLEHISLYLGMATGFVLSLVVCLLLPKTSWRKSYFPMFVDR
QQKKIYVSVKIRSAVLKRKLGNH*
>LOC_Os12g10930|12|protein_coding
MSSFSCRVAHHLTSLTVMHILLVQVATPTLADRTTTSIVTTPVLCLEQASALLQLKGS
FNVTAGDYSTVFRSWVAGACDCHWEGVHCDGADGRVTSIDLGCCHHLQADSVHPALFRITS
TKHLDLSCNNESMSTLRECEQVRLRLMHLDLGNTNLAGCEVACVGTGNTNHLVLTSLKE
  
```

10/18/07



# Help file

The screenshot shows the Ensembl BioMart interface. At the top, there is a toolbar with buttons for 'New', 'Count', 'Results', and 'Help'. A green callout bubble with the text 'Click on "help" for help document' points to the 'Help' button. Below the toolbar, the left panel shows 'Dataset 23 / 58406 Genes' with sections for 'Filters' (Chromosome: 12, Evidence code (Molecular IEA), Molecular function : GO:00), 'Attributes' (Ensembl Gene ID, Chromosome, Biotype, Peptide), and 'Dataset' ([None Selected]). The main content area is titled 'Ensembl MartView' and contains a 'Help with help!' section. This section includes 'General' (Full text search), 'Entry Points' (Ensembl BLASTView, Ensembl ExaSearch, Ensembl HelpView, Ensembl Homepage, Ensembl TextView, Ensembl User Accounts, Ensembl User Groups), and 'Data Displays' (Ensembl AlignSliceView, Ensembl AlignView, Ensembl ChromoView, Ensembl ContigView, Ensembl CytoView, Ensembl DomainView, Ensembl DotterView, Ensembl ExonView, Ensembl FamilyView, Ensembl FeatureView). The right panel contains a search bar for 'Search Help: martview', a 'Highlight search term(s)' checkbox, and a 'Go' button. Below this is a navigation bar with links for 'HOME', 'BLAST', 'SITEMAP', and 'Contact helpdesk'. The main content area also includes a 'FASTA' dropdown, a 'Unique results only' checkbox, and a 'Go' button. The footer contains copyright information: '© 2007 WTSI / EBI. Ensembl is available to download for public use - please see the code licence for details.'

Click on "help" for help document

Ensembl MartView

Help with help!

- General
  - Full text search
- Entry Points
  - Ensembl BLASTView
  - Ensembl ExaSearch
  - Ensembl HelpView
  - Ensembl Homepage
  - Ensembl TextView
  - Ensembl User Accounts
  - Ensembl User Groups
- Data Displays
  - Ensembl AlignSliceView
  - Ensembl AlignView
  - Ensembl ChromoView
  - Ensembl ContigView
  - Ensembl CytoView
  - Ensembl DomainView
  - Ensembl DotterView
  - Ensembl ExonView
  - Ensembl FamilyView
  - Ensembl FeatureView

Search Help: martview  
Highlight search term(s) ☐ Go

HOME · BLAST · SITEMAP [Contact helpdesk](#)

FASTA ☐ Unique results only Go

© 2007 WTSI / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.

# FAQ

The screenshot shows the GRAMENE website interface. At the top, the 'GRAMENE Help' logo is on the left, and a 'Documents' search bar is on the right. A horizontal navigation menu includes 'Search', 'Genomes', 'Species', 'Download', 'Resources', 'About', 'Help', and 'Feedback'. A dropdown menu is open under 'Help', listing 'Site Map', 'Help Documents', 'Release Notes', 'Tutorials and Exercises', 'Glossary', 'FAQs' (highlighted with a mouse cursor), 'Gramene Documentation', 'Workshop Materials', and 'Mailing Lists'. A callout bubble points to the 'FAQs' link, stating 'Use Feedback to ask a question'. Another callout bubble points to the 'Search' button in the top right, stating 'Search FAQ'. On the left side, a vertical sidebar contains expandable sections: 'Data Curation', 'Maps and CMap', 'BLAST', 'Proteins', 'Ontologies', 'Genomes', 'Markers', 'QTL', and 'Genes'. A callout bubble points to the 'FAQs' section in this sidebar, stating 'FAQs may be able to answer some of your questions.' The main content area is titled 'GrameneMart' and contains a list of questions. A red rectangle highlights this section, and a callout bubble points to it, stating 'Click on a question or FAQ section to open and close the information.' The questions listed are: 'Do you have a list of all rice/arabidopsis orthologs with their respective ids.', 'How are the orthologues computed?', 'Could you please explain the one2many. Does it mean many in rice or ara', 'The peptide length for the arabidopsis genes is one more than the', 'I want to know how to download the corresponding locus ID and GenBank accession number (refseq pepase predicted on your website) of a gene. For example, LOC\_Os01g01340 V.S. RefSeq peptide predicted XP\_549822.1', 'How can I download promoter sequences for arabidopsis and rice (oryza sativa)', and 'How do I get the full set of rice genome annotations - for the entire genome? The BioMart interface seems to limit me to doing a single chromosome at a time.'

GRAMENE Help

Search Genomes Species Download Resources About Help Feedback

Site Map

Help Documents

Release Notes

Tutorials and Exercises

Glossary

FAQs

Gramene Documentation

Workshop Materials

Mailing Lists

Documents

FAQs may be able to answer some of your questions.

Use Feedback to ask a question

Search FAQ

Click on a question or FAQ section to open and close the information.

**GrameneMart**

**Do you have a list of all rice/arabidopsis orthologs with their respective ids**

We suggest using the GrameneMart searches to generate this list. If you are not familiar with this please follow the tutorials provided at <http://www.gramene.org/multi/martview> to get started. For your specific query select the search filters appropriately. Here is a simplified way using your query. -Go to <http://www.gramene.org/multi/martview> select the rice data set and click the NEXT button on top. By doing this you will be taken to the features section. -Click on the TIGR gene select the type of output you want -Click on EXPORT button. -This should give you the list of orthologs.

**How are the orthologues computed?**

**Could you please explain the one2many. Does it mean many in rice or ara**

**The peptide length for the arabidopsis genes is one more than the**

**I want to know how to download the corresponding locus ID and GenBank accession number (refseq pepase predicted on your website) of a gene. For example, LOC\_Os01g01340 V.S. RefSeq peptide predicted XP\_549822.1**

**How can I download promoter sequences for arabidopsis and rice (oryza sativa)**

**How do I get the full set of rice genome annotations - for the entire genome? The BioMart interface seems to limit me to doing a single chromosome at a time.**

**Pathways**

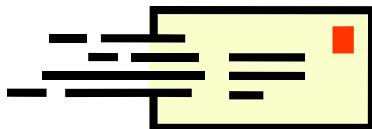
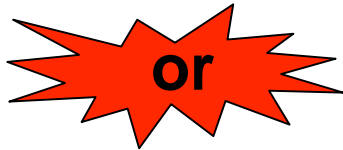
## For Further Information

- Data mining using EnSmart (oriented towards human data mining at Ensembl)
  - [www.ensembl.org/Homo\\_sapiens/helpview?se=1;kw=martview](http://www.ensembl.org/Homo_sapiens/helpview?se=1;kw=martview)
- BioMart documents
  - [www.biomart.org/install.html](http://www.biomart.org/install.html)

# Contact Gramene



Use the feedback button, located at the top of every page, to provide feedback or to ask questions about Gramene.



Email Gramene users at [gramene@gramene.org](mailto:gramene@gramene.org)