# 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 Help

The hand icon indicates a link that allows you to go to the same page in your web browser.

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

Action Options are noted in this type of font.

Notes or comments use this style font.

bio mart

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

# Gramene Home Page

GrameneMart can also be accessed through genomes.

Click here to open GrameneMart Home Page

GRAMENE A Resource for Comparative Grass Genomics v20 (March 2006)

**Quick Search**

**Have Questions...?**

- Gramene now has **tutorials** for every module.
- Ask questions through Feedback or Email

**Gramene Tips:**

The SSR Marker Search is listed under the Marker database.

**All Tips**

**Quick Start**

...nced genomes for Rice, ... ...h with GrameneMart; Sea...

ProSite or Browse by Gene Ontology using GO Slim.

...ical maps for Rice, Maize, Wheat, Barley, Oats, Sorghum, and ...parative Map Viewer (CMap) to compare maps of different

**MOLECULAR MARKERS**: Use the Simp... Sequence Repeat Identification Tool (SSRIT); or search by marker type or species, includi...

**TRAITS**: Search the Genes or QTL da... Rice Genes, Rice QTL, Maize QTL. Do...

**LITERATURE**: Search the literature for your ...

**SUBMISSION**: Submit a Rice Gene or Ontology Term to Gramene.

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.

**...ed News**

...elease 20 for ... See the ...s.

...ews Archive

- RTWG Gramene workshop materials are available.
- Rice News Worldwide from IRRI

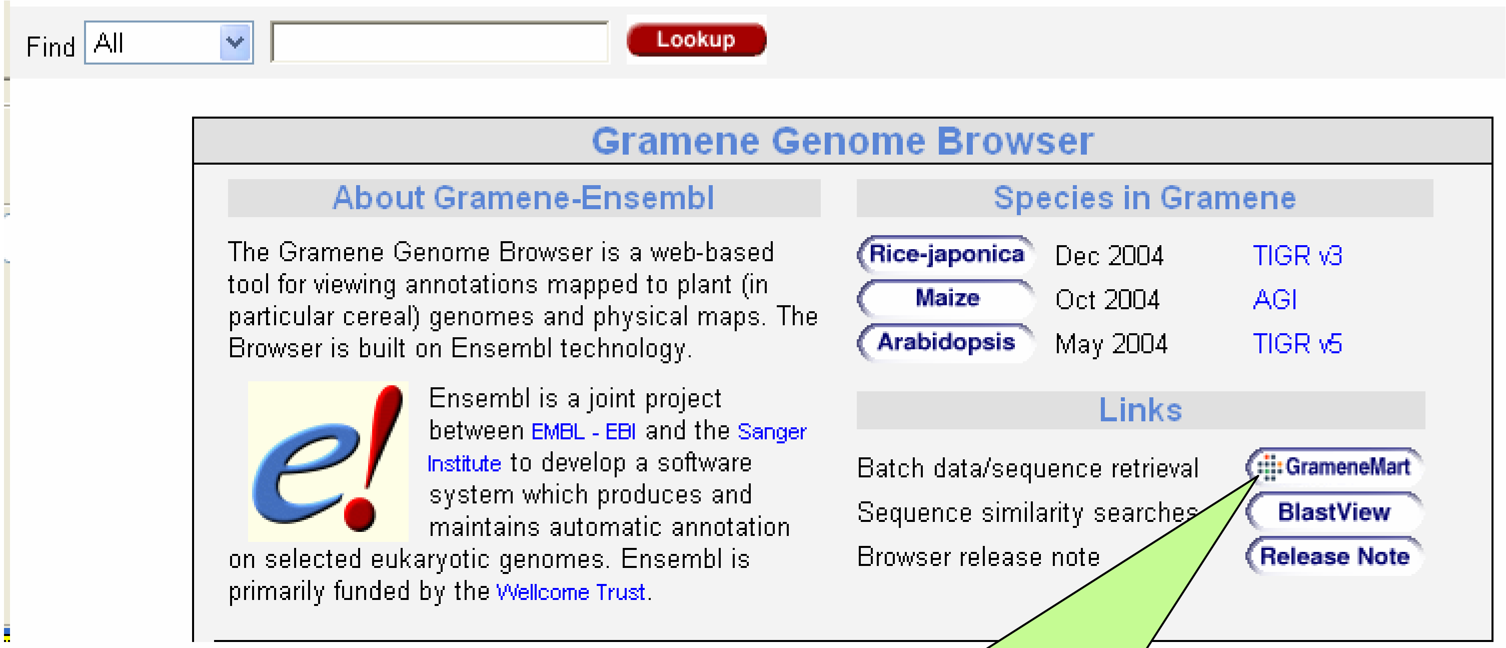
**Visit with us at**

- 48th Maize Genetics Conference, March 9-12, 2006, Asilomar, Pacific Grove, CA

Gramene Calendar

Last modified: Thu Mar 2 15:54:55 2006

# Genomes Home Page



## Gramene Genome Browser

### About Gramene-Ensembl

The Gramene Genome Browser is a web-based tool for viewing annotations mapped to plant (in particular cereal) genomes and physical maps. The Browser is built on Ensembl technology.

Ensembl is a joint project between EMBL - EBI and the Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes. Ensembl is primarily funded by the Wellcome Trust.

### Species in Gramene

| | | |
|---|---|---|
| Rice-japonica | Dec 2004 | TIGR v3 |
| Maize | Oct 2004 | AGI |
| Arabidopsis | May 2004 | TIGR v5 |

### Links

| | |
|---|---|
| Batch data/sequence retrieval | GrameneMart |
| Sequence similarity searches | BlastView |
| Browser release note | Release Note |

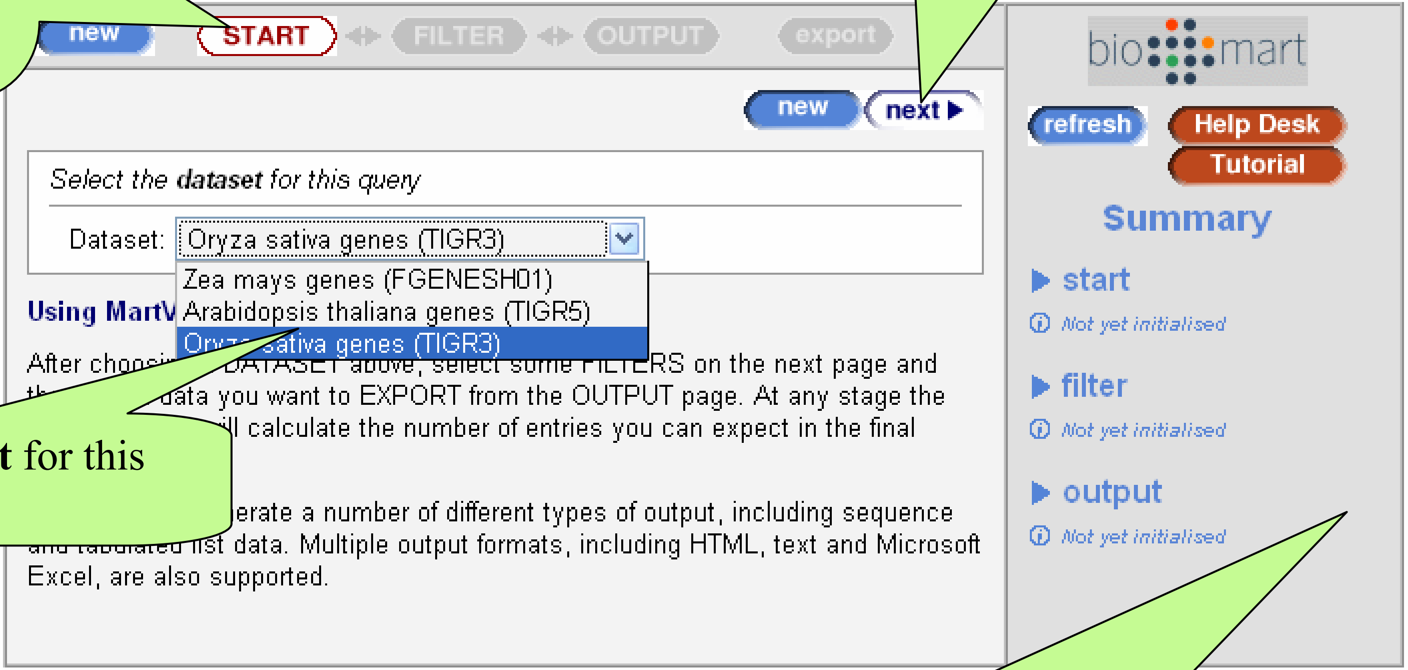
You can also enter Gramene Mart through the Genomes page

Gramene is a growing database, and results from your search may be slightly different than pictured here. However, the organization of the database remains the same.

# Select a Dataset

Mart tools are structured as a series of pages to be filled in successively. The steps are outlined at as **headers** at the top of MartView

2. Press **Next**. This will move you to the next step, and highlight the headers one by one.



*Select the **dataset** for this query*

Dataset: Oryza sativa genes (TIGR3)

- Zea mays genes (FGENESH01)
- Arabidopsis thaliana genes (TIGR5)
- Oryza sativa genes (TIGR3)

**Using MartV**

After choosi DATASET above, select some FILTERS on the next page and th data you want to EXPORT from the OUTPUT page. At any stage the ill calculate the number of entries you can expect in the final

erate a number of different types of output, including sequence and tabulated list data. Multiple output formats, including HTML, text and Microsoft Excel, are also supported.

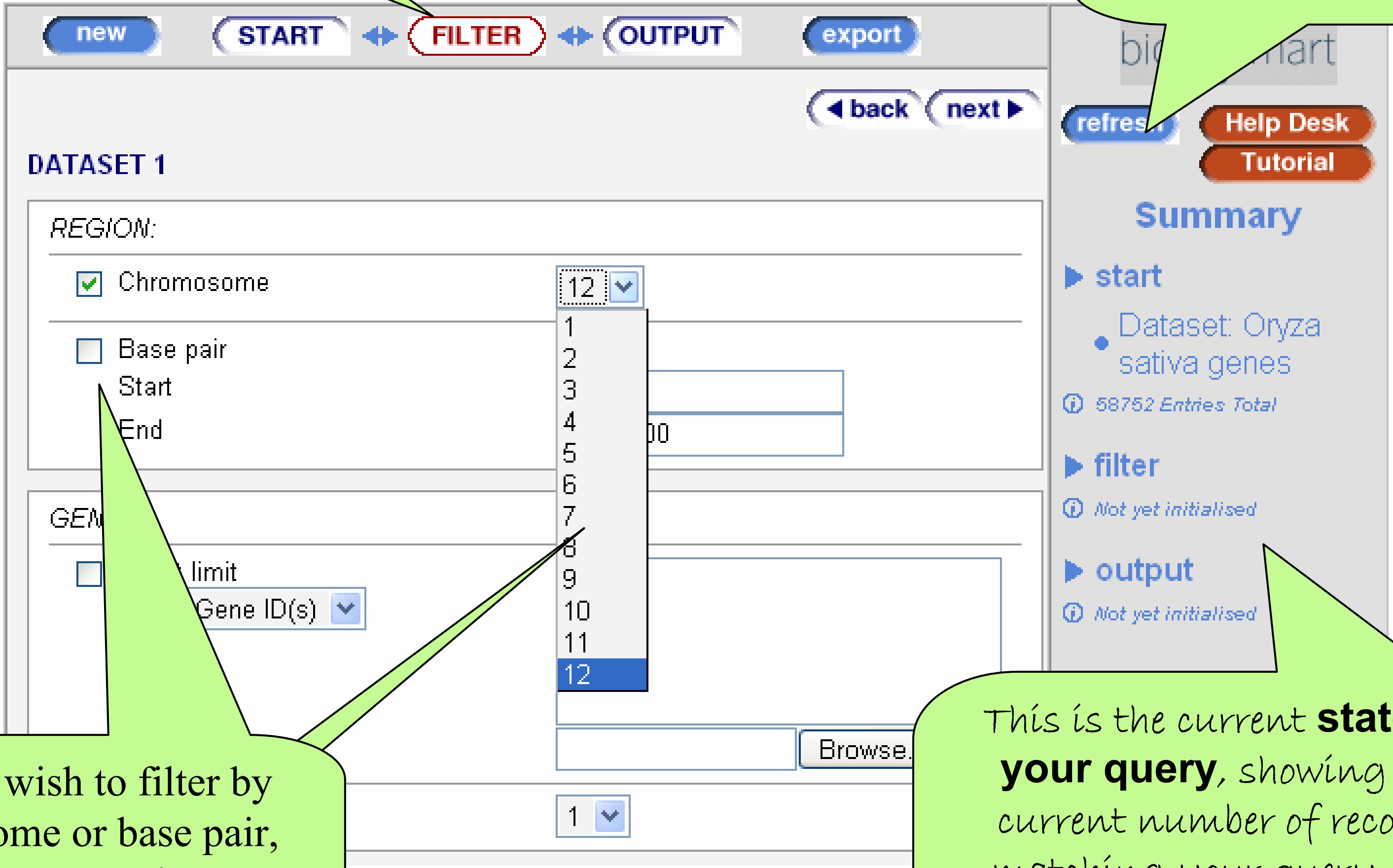
1. Select a **dataset** for this query.

The **Summary** section at the right hand side is updated accordingly. It is always possible to go back to a previous step and amend the options. Clicking on 'refresh' will update the summary.

# Formulate the Filters:
## Filter by Region

Apply a filter to narrow your search

After each filter selection click on "**refresh**" to update your summary, but remain on the filter page.



DATASET 1

REGION:

☑ Chromosome 12

☐ Base pair

Start

End

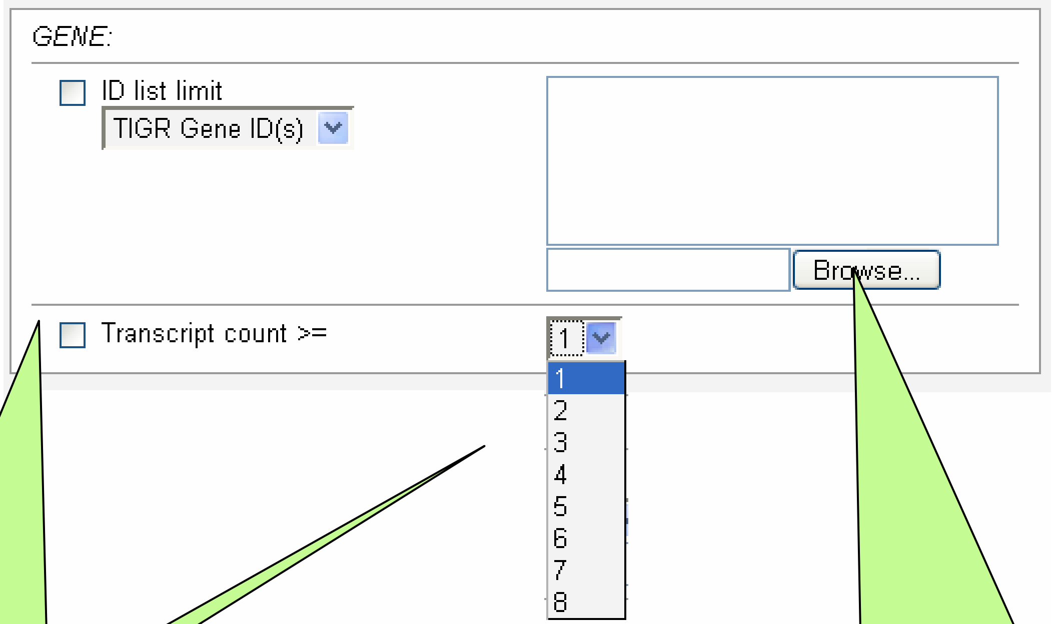
Summary

- start
  - Dataset: Oryza sativa genes
  - 58752 Entries Total
- filter
  - Not yet initialised
- output
  - Not yet initialised

1. If you wish to filter by chromosome or base pair, **select the box** and **enter specific data**

This is the current **state of your query**, showing the current number of records matching your query and filters. Continue to add filters until the number becomes a manageable size.

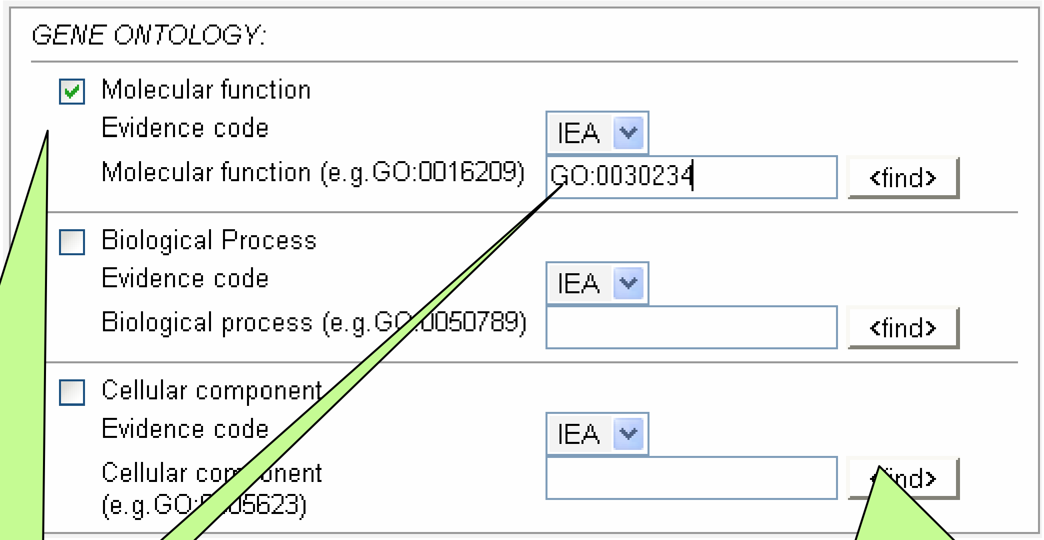
# Filter by Gene



2. If you wish to filter by gene or transcript count **select the box** and **enter specific data**

Use the 'Browse' button to import values from a file

# Filter by Gene Ontology



GENE ONTOLOGY:

☑ Molecular function

Evidence code: IEA

Molecular function (e.g.GO:0016209) GO:0030234 <find>

☐ Biological Process

Evidence code: IEA

Biological process (e.g.GO:0050789) <find>

☐ Cellular component

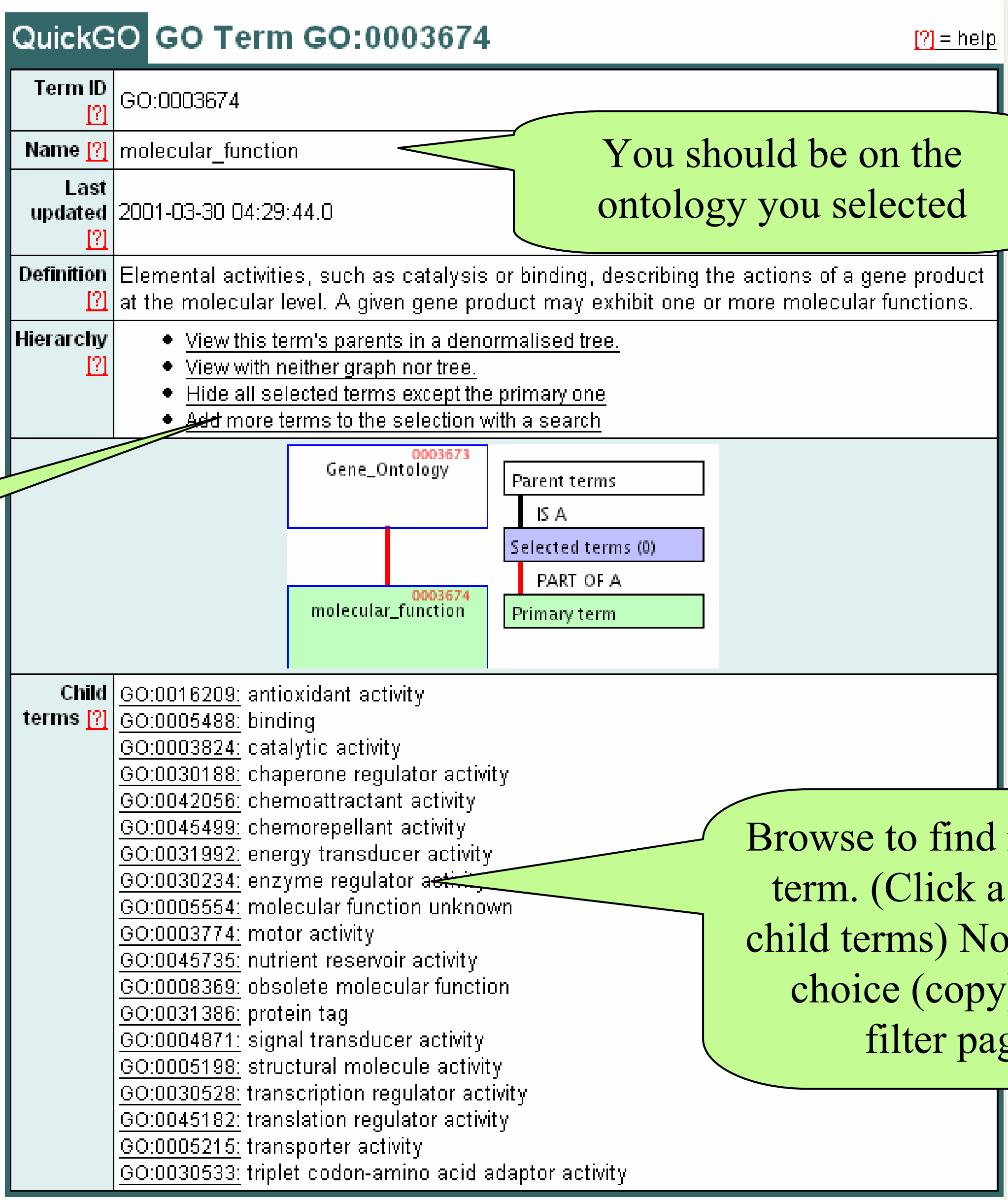
Evidence code: IEA

Cellular component (e.g.GO:0005623) <find>

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

Use the 'find' button to go to EBI's **QuickGO** to define your GO term (see next slide).

# QuickGO



**QuickGO** **GO Term GO:0003674**

[?] = help

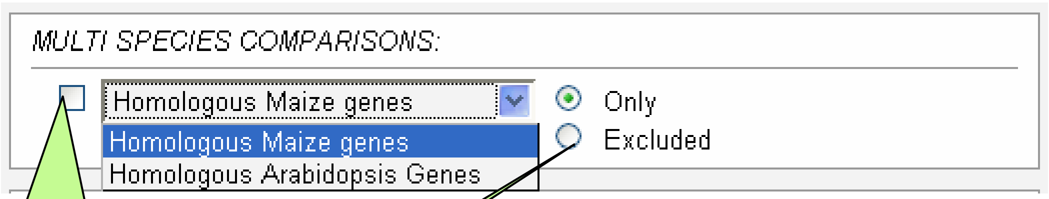
| | |
|---|---|
| **Term ID** [?] | GO:0003674 |
| **Name** [?] | molecular_function |
| **Last updated** [?] | 2001-03-30 04:29:44.0 |
| **Definition** [?] | Elemental activities, such as catalysis or binding, describing the actions of a gene product at the molecular level. A given gene product may exhibit one or more molecular functions. |
| **Hierarchy** [?] | - View this term's parents in a denormalised tree.<br>- View with neither graph nor tree.<br>- Hide all selected terms except the primary one<br>- Add more terms to the selection with a search |
| **Child terms** [?] | GO:0016209: antioxidant activity<br>GO:0005488: binding<br>GO:0003824: catalytic activity<br>GO:0030188: chaperone regulator activity<br>GO:0042056: chemoattractant activity<br>GO:0045499: chemorepellant activity<br>GO:0031992: energy transducer activity<br>GO:0030234: enzyme regulator activity<br>GO:0005554: molecular function unknown<br>GO:0003774: motor activity<br>GO:0045735: nutrient reservoir activity<br>GO:0008369: obsolete molecular function<br>GO:0031386: protein tag<br>GO:0004871: signal transducer activity<br>GO:0005198: structural molecule activity<br>GO:0030528: transcription regulator activity<br>GO:0045182: translation regulator activity<br>GO:0005215: transporter activity<br>GO:0030533: triplet codon-amino acid adaptor activity |

You should be on the ontology you selected

Or use the search function

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).

# Compare Species



*MULTI SPECIES COMPARISONS:*

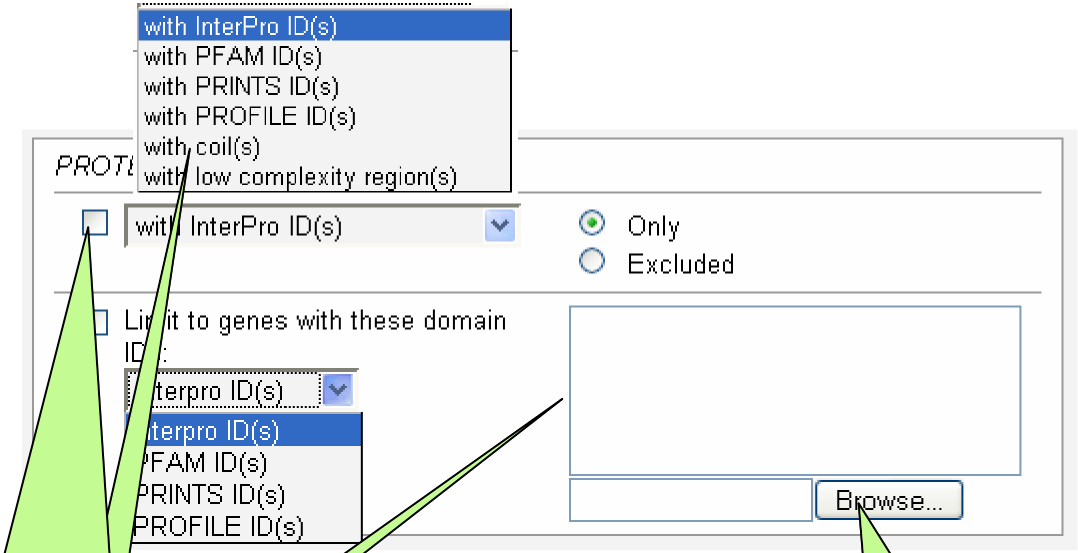
☐ Homologous Maize genes

● Only

○ Excluded

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

# Filter by Protein

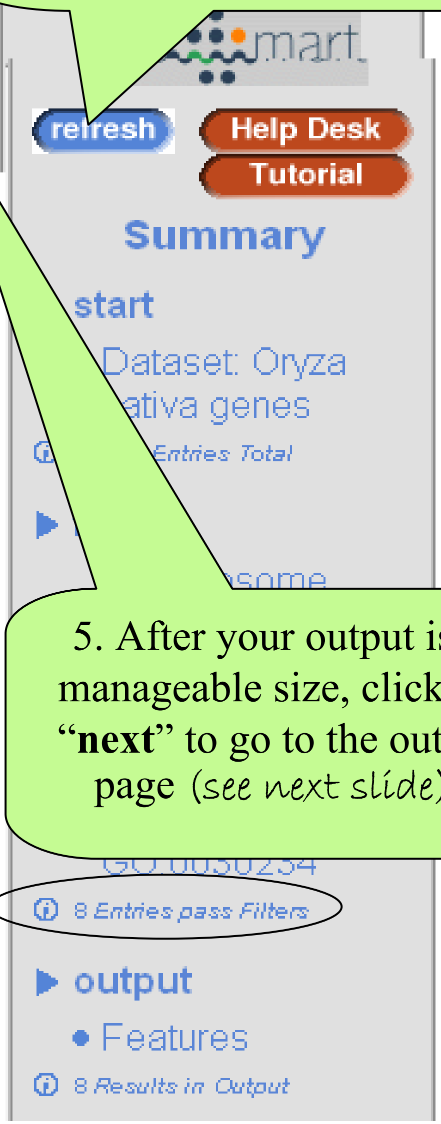


4. If you wish to filter by protein, **select the box,** make a selection from the **drop down menu** and **enter specific data**

Use the 'Browse' button to import values from a file

# Apply the filter

After selecting a filter click on "**refresh**"



Summary

start

Dataset: Oryza ativa genes

Entries Total

GO 0030254

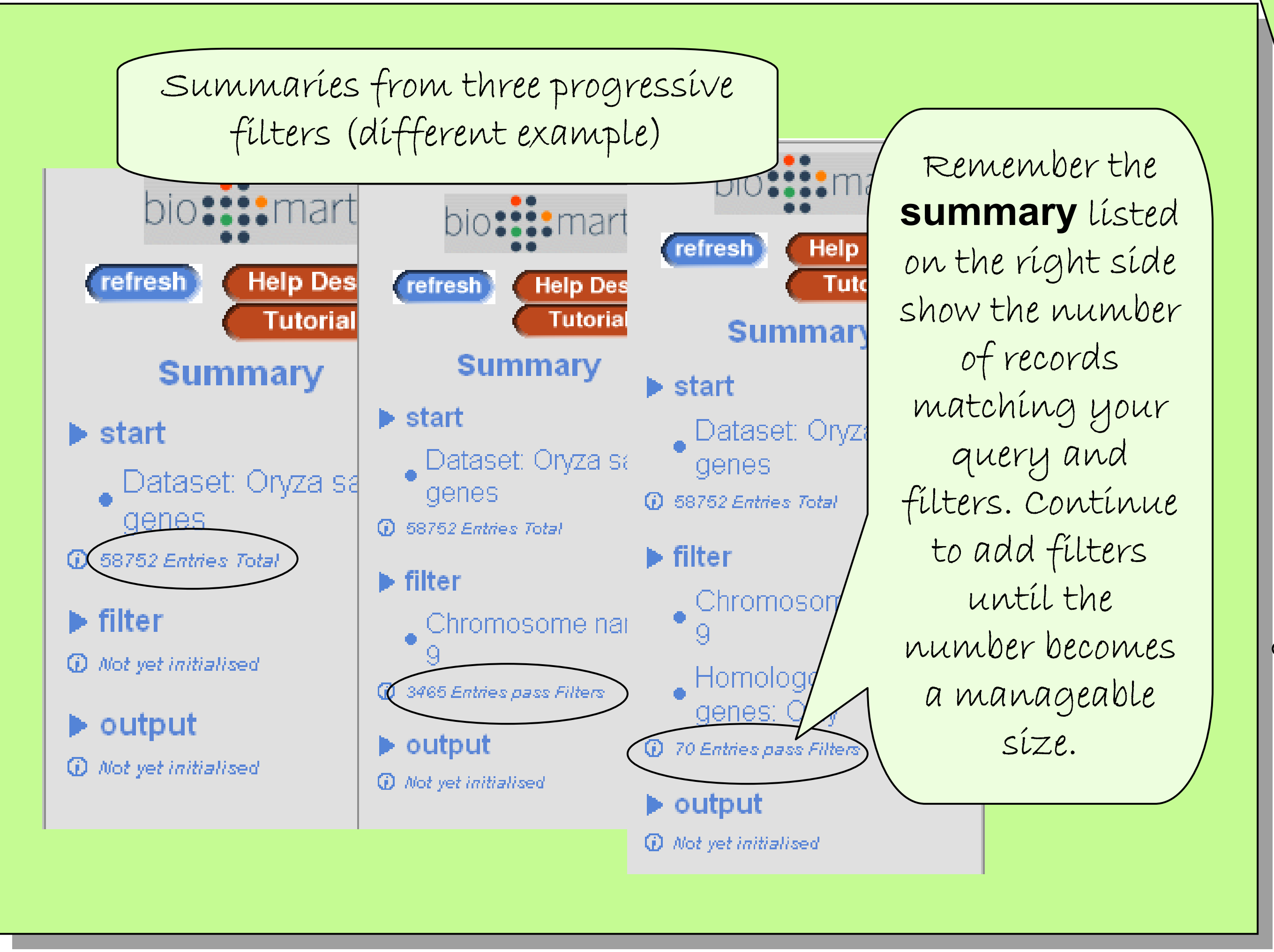
8 Entries pass Filters

output

- Features

8 Results in Output

5. After your output is a manageable size, click on "**next**" to go to the output page (see next slide).



Summaries from three progressive filters (different example)

Summary

start

- Dataset: Oryza sa genes

58752 Entries Total

filter

Not yet initialised

output

Not yet initialised

Summary

start

- Dataset: Oryza sa genes

58752 Entries Total

filter

- Chromosome nai 9

3465 Entries pass Filters

output

Not yet initialised

Summary

start

- Dataset: Oryz genes

58752 Entries Total

filter

- Chromosom 9
- Homolog genes: C

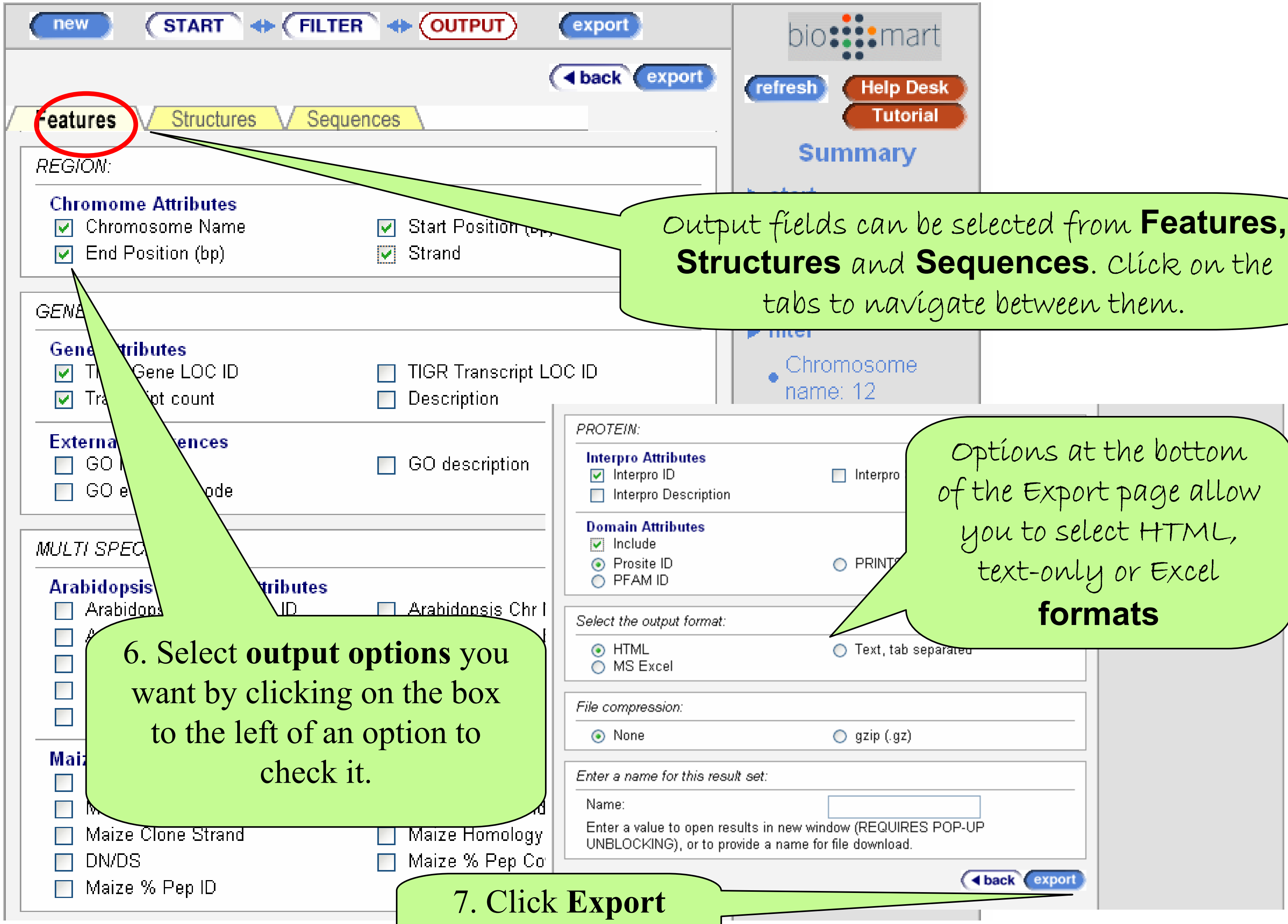
70 Entries pass Filters

output

Not yet initialised

Remember the **summary** listed on the right side show the number of records matching your query and filters. Continue to add filters until the number becomes a manageable size.

# Features Output



REGION:

**Chromome Attributes**

☑ Chromosome Name ☑ Start Position (bp)

☑ End Position (bp) ☑ Strand

GENE

**Gene tributes**

☑ TI Gene LOC ID ☐ TIGR Transcript LOC ID

☑ Tra pt count ☐ Description

**Externa ences**

☐ GO I ☐ GO description

☐ GO e ode

MULTI SPEC

**Arabidopsis tributes**

☐ Arabidops ID ☐ Arabidopsis Chr I

**Mai**

☐ Maize Clone Strand ☐ Maize Homology

☐ DN/DS ☐ Maize % Pep Co

☐ Maize % Pep ID

Chromosome name: 12

PROTEIN:

**Interpro Attributes**

☑ Interpro ID ☐ Interpro

☐ Interpro Description

**Domain Attributes**

☑ Include

● Prosite ID ○ PRINT

○ PFAM ID

Select the output format:

● HTML ○ Text, tab separated

○ MS Excel

File compression:

● None ○ gzip (.gz)

Enter a name for this result set:

Name:

Enter a value to open results in new window (REQUIRES POP-UP UNBLOCKING), or to provide a name for file download.

Output fields can be selected from **Features, Structures** and **Sequences**. Click on the tabs to navigate between them.

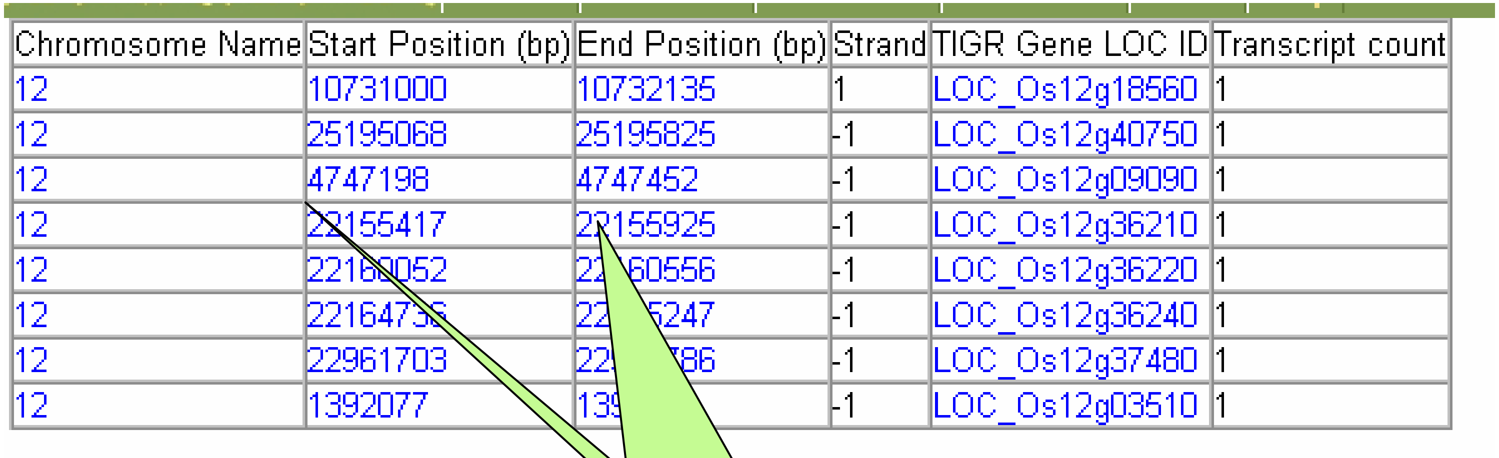
Options at the bottom of the Export page allow you to select HTML, text-only or Excel **formats**

6. Select **output options** you want by clicking on the box to the left of an option to check it.

7. Click **Export** (see Slide 14)

# Features Results

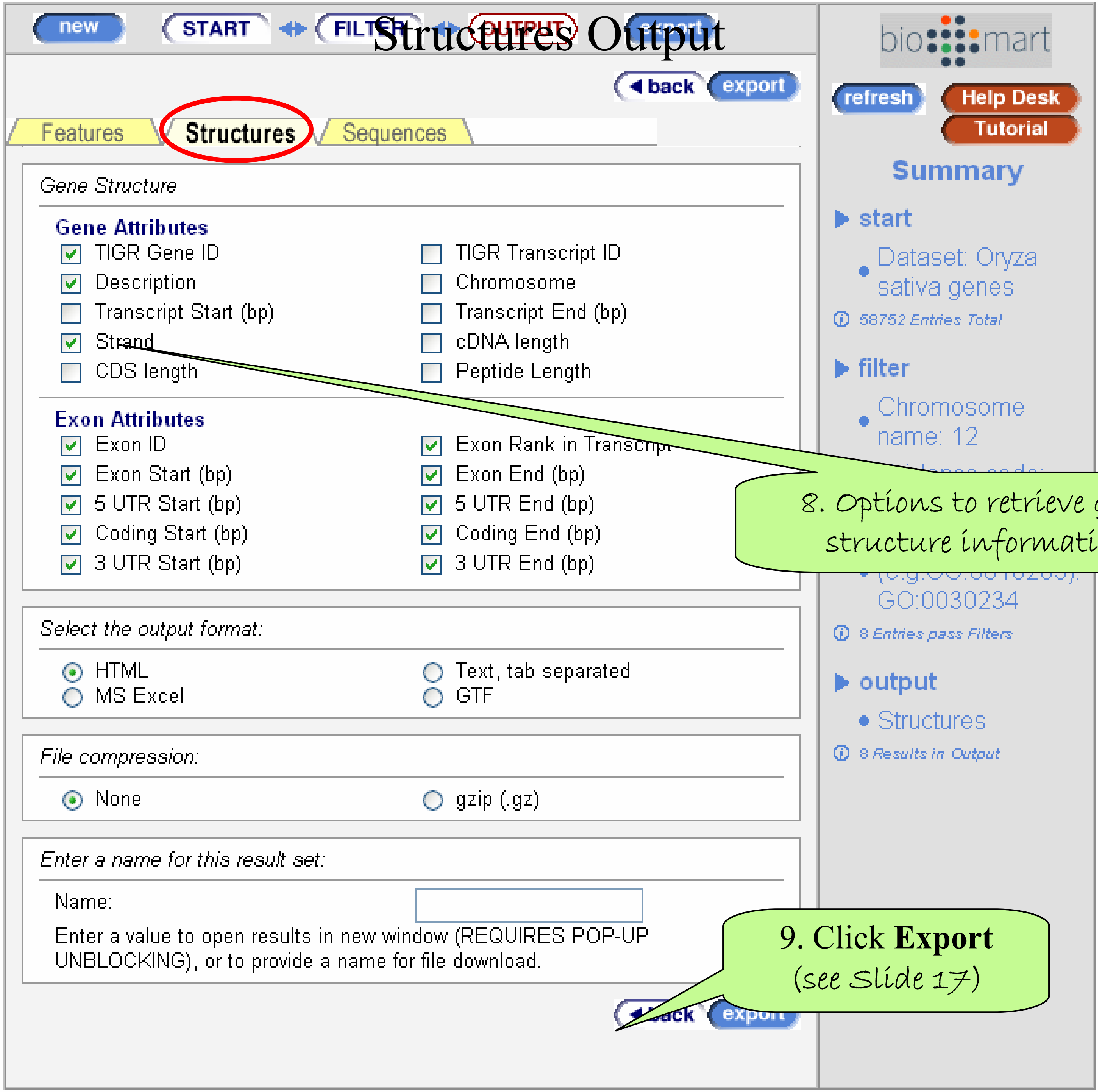
Each selected output field is a column in the table.



| Chromosome Name | Start Position (bp) | End Position (bp) | Strand | TIGR Gene LOC ID | Transcript count |
|---|---|---|---|---|---|
| 12 | 10731000 | 10732135 | 1 | LOC_Os12g18560 | 1 |
| 12 | 25195068 | 25195825 | -1 | LOC_Os12g40750 | 1 |
| 12 | 4747198 | 4747452 | -1 | LOC_Os12g09090 | 1 |
| 12 | 22155417 | 22155925 | -1 | LOC_Os12g36210 | 1 |
| 12 | 22160052 | [illegible] | -1 | LOC_Os12g36220 | 1 |
| 12 | 22164736 | [illegible] | -1 | LOC_Os12g36240 | 1 |
| 12 | 22961703 | [illegible] | -1 | LOC_Os12g37480 | 1 |
| 12 | 1392077 | [illegible] | -1 | LOC_Os12g03510 | 1 |

Links to Genome Browser

# Structures Output



*Gene Structure*

**Gene Attributes**

| | |
|---|---|
| ☑ TIGR Gene ID | ☐ TIGR Transcript ID |
| ☑ Description | ☐ Chromosome |
| ☐ Transcript Start (bp) | ☐ Transcript End (bp) |
| ☑ Strand | ☐ cDNA length |
| ☐ CDS length | ☐ Peptide Length |

**Exon Attributes**

| | |
|---|---|
| ☑ Exon ID | ☑ Exon Rank in Transcript |
| ☑ Exon Start (bp) | ☑ Exon End (bp) |
| ☑ 5 UTR Start (bp) | ☑ 5 UTR End (bp) |
| ☑ Coding Start (bp) | ☑ Coding End (bp) |
| ☑ 3 UTR Start (bp) | ☑ 3 UTR End (bp) |

*Select the output format:*

● HTML
○ Text, tab separated
○ MS Excel
○ GTF

*File compression:*

● None
○ gzip (.gz)

*Enter a name for this result set:*

Name:

Enter a value to open results in new window (REQUIRES POP-UP UNBLOCKING), or to provide a name for file download.

**Summary**

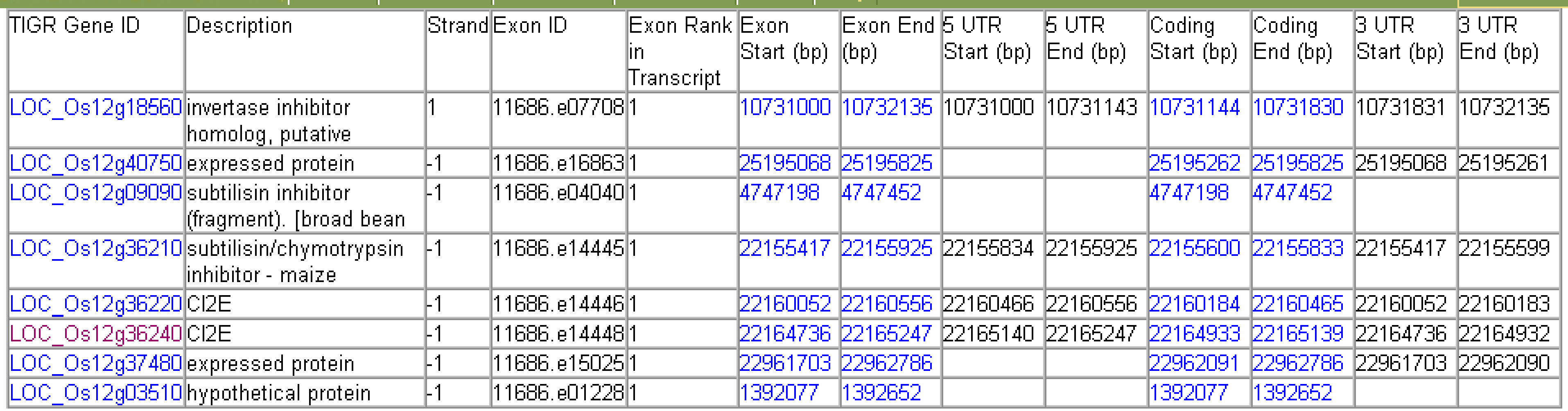
- start
  - Dataset: Oryza sativa genes
  - *58752 Entries Total*
- filter
  - Chromosome name: 12
  - GO:0030234
  - *8 Entries pass Filters*
- output
  - Structures
  - *8 Results in Output*

8. Options to retrieve gene structure information

9. Click **Export** (see Slide 17)

# Structures Results

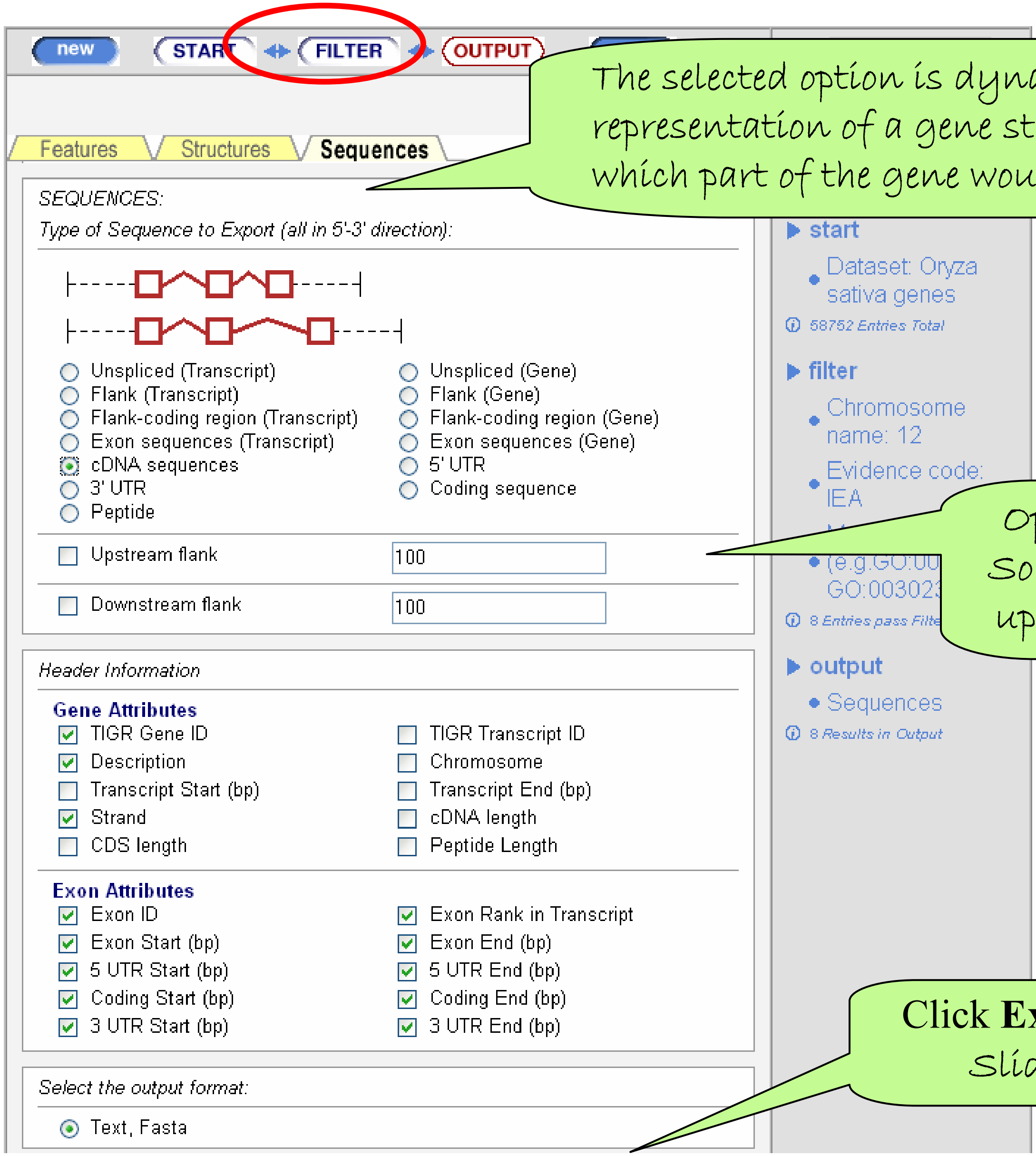
Each selected output field is a column in the table.



| TIGR Gene ID | Description | Strand | Exon ID | Exon Rank in Transcript | Exon Start (bp) | Exon End (bp) | 5 UTR Start (bp) | 5 UTR End (bp) | Coding Start (bp) | Coding End (bp) | 3 UTR Start (bp) | 3 UTR End (bp) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LOC_Os12g18560 | invertase inhibitor homolog, putative | 1 | 11686.e07708 | 1 | 10731000 | 10732135 | 10731000 | 10731143 | 10731144 | 10731830 | 10731831 | 10732135 |
| LOC_Os12g40750 | expressed protein | -1 | 11686.e16863 | 1 | 25195068 | 25195825 |  |  | 25195262 | 25195825 | 25195068 | 25195261 |
| LOC_Os12g09090 | subtilisin inhibitor (fragment). [broad bean | -1 | 11686.e04040 | 1 | 4747198 | 4747452 |  |  | 4747198 | 4747452 |  |  |
| LOC_Os12g36210 | subtilisin/chymotrypsin inhibitor - maize | -1 | 11686.e14445 | 1 | 22155417 | 22155925 | 22155834 | 22155925 | 22155600 | 22155833 | 22155417 | 22155599 |
| LOC_Os12g36220 | CI2E | -1 | 11686.e14446 | 1 | 22160052 | 22160556 | 22160466 | 22160556 | 22160184 | 22160465 | 22160052 | 22160183 |
| LOC_Os12g36240 | CI2E | -1 | 11686.e14448 | 1 | 22164736 | 22165247 | 22165140 | 22165247 | 22164933 | 22165139 | 22164736 | 22164932 |
| LOC_Os12g37480 | expressed protein | -1 | 11686.e15025 | 1 | 22961703 | 22962786 |  |  | 22962091 | 22962786 | 22961703 | 22962090 |
| LOC_Os12g03510 | hypothetical protein | -1 | 11686.e01228 | 1 | 1392077 | 1392652 |  |  | 1392077 | 1392652 |  |  |

Links to Genome Browser

# Sequences Output



SEQUENCES:

*Type of Sequence to Export (all in 5'-3' direction):*

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

☐ Upstream flank 100

☐ Downstream flank 100

*Header Information*

**Gene Attributes**

- ☑ TIGR Gene ID
- ☑ Description
- ☐ Transcript Start (bp)
- ☑ Strand
- ☐ CDS length
- ☐ TIGR Transcript ID
- ☐ Chromosome
- ☐ Transcript End (bp)
- ☐ cDNA length
- ☐ Peptide Length

**Exon Attributes**

- ☑ Exon ID
- ☑ Exon Start (bp)
- ☑ 5 UTR Start (bp)
- ☑ Coding Start (bp)
- ☑ 3 UTR Start (bp)
- ☑ Exon Rank in Transcript
- ☑ Exon End (bp)
- ☑ 5 UTR End (bp)
- ☑ Coding End (bp)
- ☑ 3 UTR End (bp)

*Select the output format:*

● Text, Fasta

**start**
- Dataset: Oryza sativa genes

*58752 Entries Total*

**filter**
- Chromosome name: 12
- Evidence code: IEA
- (e.g.GO:00
GO:00302

*8 Entries pass Filte*

**output**
- Sequences

*8 Results in Output*

The selected option is dynamically visualised in a schematic representation of a gene structure. The bold red lines indicate which part of the gene would be included in an export.

Options for the sequences. Some options are dependent upon other options selected.

Click **Export** (see Slide 19)

# Sequences Results

```
>LOC_Os12g18560|invertase inhibitor homolog, putative|1|11686.e07708|1|10731000|10732135|10731000|10731143|10731144|1
AAGAAGCTAATTAACCCAAGCAACTAGCTATCTAGCTTAGCCTCGCTAAACCAACACCATCGTAAAAATCTCTTTGATAG
TTGACATCGAGGCAGTGATTAATTAAGTAGCTAGCTAGTTACAGGCACAAGGAGAGAAACACCAATGGCATCAATGGCGC
CATCGGCAATGGTGCTCATCGTCCTCCTCGTCCTGGTGGTTCTCCCGTCGAGCACTCTGTGCTCACGGGCGGGGCCTTCT
TCCAAGCACGGCCATGGCGGTGGCCACGCCAAGCGCGTGCCGCCACCGGCGT[illegible]GTACCGCCGCCGCCGCCGCCGCCACC
AGCGCCGGCGGCGCTGGTGCGTGCCACCTGCAACTCCACCTCCTACTACGACCTCT[illegible]GTCGCCGAGCTGTCCGCCGACC
CGTCGAGCGCCACGGCCGACGTGCGCGGACTGTCGTCCATCGCCGTCTCCGCCGCCGCC[illegible]AACGCATCCGGGGCGGCG
CAGGCGGCCTCGGCGCTGGCGAACGCGACCGACGCGGGGACGACGGCGGGCGTCGCCGGCGACC[illegible]GCGGCGCAGTCGT
ACAGAGGCTGCTCGCCACCTGCGCGGCCAAGTACGGCGACGCCCGCGACGCGCTCGCCGCGGCCAAG[illegible]CGATCGCGC
AGCAGGACTACGACATGGCGTCCGTGCACGTCAGCGCCGCCGCGGAGTACCCGCAGGTGTGCAGGACGCTC[illegible]GCGG
CAGAGCCCCGGAGACTACCCGCCGGAGCTCGCCGCGACAGAGGTGGCGCTCAGGCAGCTCTGCTCCGTCGCGCT[illegible]
CATCGCGCTCCTCAGCTCATCCAGCAACTAGCAGCTCTGCTTGTTACCGAGCTCAAGTTCACCCAACCAGCTAACTAC[illegible]
GCAATTCGTATAGGTACAAATGGTGCAAATATAGTACTGTATAATACTACTGCATAGAATACATATACGTGTAATGACAC
GTATTTATCTTTTTTTTTTTGCAAGGGGCACGTATATCAATTAATTGTGTGTCCCAATTAATTAGAGTCGAATCCACTTGA
TATGTTCTTTTGTTAATTTGTATTATCACTCCATAGAGGAGTTGCTGTAGTAGTGCAAAAGGTACATGCGGCCGCCGGCA
GTATGCATGTATTTCA

>LOC_Os12g40750|expressed protein|-1|11686.e16863|1|25195068|25195825|||25195262|25
ATGGGGTCCTCCTCATTACTACGCGGCACCGTCCTCCTCGTCGCCGCCGCCGTCGTCCTCGCATCGGCGCCCGTCTCGCC
CGGCGGCGGCGGGGTGGACACGGTGGCGGAGTCGTGCAACGCCATCAGGGACTTCGTGGAGGTGGCGTTCTGCGAGGCGG
CGCTGCGGTCGAGCGGCCCTGGCGCCGCCACGGCCGACCGCCACGGCCACCTCCTGATCGCGGCCGACCTGGCCGCGGCG
CGCGGGGACTCGGCGAGGGACGACGCCGTGGCGGCGGCGGCGGCGGCGCGGGCGAGCGGCGGCGGAGAGCGGGACGGGCT
GGAGGCGTGCGGCATCCTGTACGGGTCGTCGGTGGCGGCGGTGCGGCTGATGCGCGGGTACGCGGCGGCGCGGAGCTGGG
AGGCGGCGAGGTCGCTGCTGTGGCTGTCGGGGCAGGCCGGGATGGGGTGCGCGGCGGCGCTCGACTCGCCGGCGCCGGGG
ATGGACAGGGCCAACGAGGATTTCAGCCGGCTCACCACCATGGCCACCGCGCTGCTCAACAGCGTCGCCGCCACCAGCGG
CTGAGCTCTCCGCTAGCTCCGGCCACGGCTGACGTGGCATGACATCATCATCCCATCGATGCATGACGTGGCACTCTTTT
TTCTTTCTCTTTTTTTTTCTTTTGCCTTTTTTTTTATGGTAGTAGAGATGTTGTGGGAGTTGTTTGATGTAATGTAAGCT
GGTGACATAAATAAATAATGGCAATTGCATTTCAAATC

>LOC_Os12g09090|subtilisin inhibitor (fragment). [broad bean|-1|11686.e04040|1|4747198|4747452|||4747198|4747452||
ATGGAGAGAGGCGGAGCGGCGGCGGCTGCCGGCGGTGGCGTGGTGGTGTCGCCGTCGGCGGCGAAGGATTCGTGGCCGGA
GTTGGTCGGCGTGTCGTCGGAGGCGGCGAAGACGAAGATCGGAGAGGAGAGGCCGGAGTTGGACGTGCAGGTGGTCCCCG
CCGACGCCTTCGTCACCACGGACTACAACGCCGGCCGTGTCAGGGTCTTCGTCGACTCGGACGACAAGGTCGCCAGACCT
CCCAAGATCGGCTAA

>LOC_Os12g36210|subtilisin/chymotrypsin inhibitor - maize|-1|11686.e14445|1|22155417|22155925|22155834|22155925|22155
AAGCTAGCTACATATATATATATATATATATATATATATATACTCTCTCCAAAGTTTACATTGCTGATATATAGCTCATCAT
CATCAATCGATCATGGCCGGAGTTGTGAGGTCGGCGGCGTGCAGCGGCGGCGAGGGGGAGCGGAAGACGTCGTGGCCGGA
GGTGGTGGGGCTGTCGGCGGAGGAGGCGAAGAAGGTGATCCTCAAGGACATGCCCGACGCCGACATCGTCGTGCTTCCGG
CGGGGTCGCCGGTGACCCTCGACTTCCGCTCCAACCGCGTCCGCATCTTCGTTGACACCGTCGCCGGCACCCCCACCATC
GGCTAGCCTCCCTCATCTGCCTTCCTTCAGCGCCCGGGCCCCCGGCCATCTTTATCTTAATTAAATAATGATCACATGCA
CGCAAGTGATATGTACTGTGTGCATGTTTCGATCGATCCATCCCCATCTGAATAATGTTTTACCGTCAATGTTTTTTTCC
TATGAATAAATATGGAAGAATCTTCTCCC

>LOC_Os12g36220|CI2F|-1|11686.e14446|1|22160052|22160556|22160466|22160556|22160184|22160465|22160052|22160183
```

Lists selected sequences, with Identifications at the top

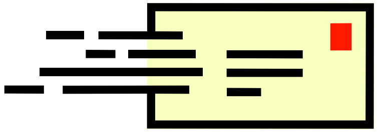
# For Further Information

- Data mining using EnSmart (oriented towards human data mining at Ensembl)
  - www.ensembl.org/Homo_sapiens/helpview?se=1;kw=martview
- BioMart documents
  - 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.

**or**



Email Gramene at gramene@gramene.org