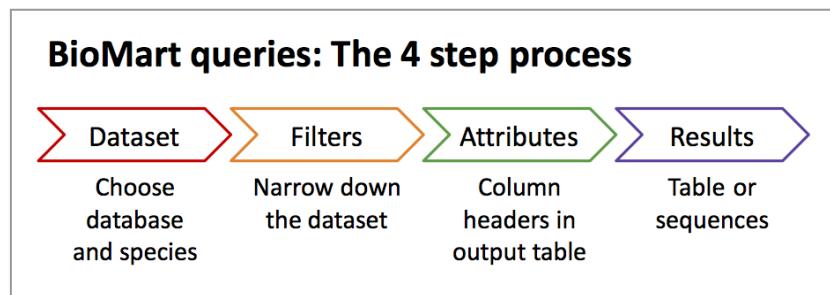


## Exercise: Ensembl Fungi BioMart

Links to be clicked shown in blue, text to be entered shown in red.

Follow these instructions to guide you through BioMart to answer the following query:

- (a) How many genes within the 14:1128520-1142558 region are found in *Fusarium solani* that do not have an orthologue in *Fusarium verticillioides*?
- (b) Export the gene name, locations and GO terms associated with these genes
- (c) Export their cDNA sequences



Click on [BioMart](#) in the top header of any [fungi.ensembl.org](http://fungi.ensembl.org) page or enter <https://fungi.ensembl.org/biomart/martview/> into your browser.

**NOTE:** These answers were determined using BioMart Ensembl Fungi 58

**Step 1a:** Choose [Ensembl Fungi Genes 58](#) as the database

The screenshot shows the Ensembl Fungi BioMart interface. At the top, there's a navigation bar with links for BLAST, BioMart, FTP, and Docs & FAQs. Below the navigation bar, there are buttons for New, Count, and Results. On the left, a 'Dataset' section displays '[None selected]'. A dropdown menu titled '✓ - CHOOSE DATABASE -' is open, showing two options: 'Ensembl Fungi Genes 58' (which is highlighted in blue) and 'Ensembl Fungi Variations 58'.

**Step 1b:** Choose [Fusarium solani](#) genes (v2.0) as the dataset

The screenshot shows the Ensembl Fungi BioMart interface. On the left, there's a sidebar with buttons for 'New', 'Count', and 'Results'. The main area has a 'Dataset' dropdown set to 'Ensembl Fungi Genes 58' and a 'Filters' dropdown set to 'Fusarium solani genes (v2.0)'. At the top right, there are links for BLAST, BioMart, FTP, Docs & FAQs, and options for URL, XML, Perl, and Help.

## Step 2: Choose appropriate filters

We want to narrow down the dataset of all *F. solani* genes to a subset of genes matching our filters. We are interested in *F. solani* genes that **do not** have an orthologue with *F. verticillioides*. We need to filter the dataset to find these genes.

The screenshot shows the same interface as above, but with the 'Filters' section expanded. A callout box labeled 'Step 2a: Click on Filters' points to the 'Filters' button in the sidebar. Another callout box labeled 'Step 2b: Expand the MULTI SPECIES COMPARISONS section' points to the 'MULTI SPECIES COMPARISONS' section in the main panel. The 'Filters' dropdown now shows '[None selected]'. The main panel has a header 'Please restrict your query using criteria below' and a note '(If filter values are truncated in any lists, hover over the list item to see the full text)'. The 'MULTI SPECIES COMPARISONS' section contains several checkboxes: 'GENE', 'PATHOGEN PHENOTYPES (PHI-BASE)', 'GENE ONTOLOGY', and 'MULTI SPECIES COMPARISONS'. Under 'MULTI SPECIES COMPARISONS', there is a checkbox for 'Homologue filters' and a dropdown menu for 'Paralogous Fusarium solani Genes' with options 'Only' (selected) and 'Excluded'.

**Dataset 6727 / 16163 Genes**  
Fusarium solani genes (v2.0)

**Filters**  
Orthologous Fusarium verticillioides Genes: Excluded

**Attributes**  
Gene stable ID  
Transcript stable ID

**Dataset**  
[None Selected]

Please restrict your query using criteria below  
(If filter values are truncated in any lists, hover over the list item to see the full text)

REGION:  
 GENE:  
 PATHOGEN PHENOTYPES (PHI-BASE):  
 GENE ONTOLOGY: **Step 2c: Choose Orthologous Fusarium verticillioides Genes**  
 MULTI SPECIES CO: **Step 2d: Select the Excluded option**

Homologue filters  
 Orthologous Fusarium verticillioides Genes  
 Only  
 Excluded

PROTEIN DOMAIN

**Dataset 4 / 16163 Genes**  
Fusarium solani genes (v2.0)

**Filters**  
Orthologous Fusarium verticillioides Genes: Excluded  
Chromosome/scaffold: 14  
Start: 1128520  
End: 1142558

**Attributes**  
Gene stable ID  
Transcript stable ID

**Dataset**  
[None Selected]

**Update Count**

Please restrict your query using criteria below  
(If filter values are truncated in any lists, hover over the list item to see the full text)

REGION: **Step 2e: Expand the REGION section**

Chromosome/scaffold

Coordinates  
Start  
End

**Step 2f: Enter Start/End coordinates**

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
sca\_16\_unmapped  
sca\_21\_unmapped  
sca\_35\_unmapped  
sca\_45\_unmapped  
sca\_51\_unmapped  
sca\_66\_unmapped

1128520  
1142558

Using the [Count](#) function we can see that there are 4 *F. solani* genes (out of a total of 16,163) in the 14:1128520-1142558 region that do not have an orthologue in *F. verticillioides*.

### Step 3: Select Attributes

Attributes (our desired output) are defined by what we would like to learn about the data. We want to find out more information about these genes, including:

1. Gene name
2. Locations
3. Associated GO terms
4. cDNA sequences

There are four main attribute types: Features, Structures, Homologues and Sequences. BioMart allows querying only one type at a time. We can answer points 1-3 in a single query as they can all be found under [Features](#), but we will need to build a second query to answer point 4 ([Sequence](#) type).

**Dataset 4 / 16163 Genes**  
Fusarium solani genes (v2.0)  
**Filters**  
Orthologous Fusarium verticillioides Genes: Excluded  
Chromosome/scaffold:  
Start: 1128520  
End: 1142558  
**Attributes**  
Gene stable ID  
Transcript stable ID  
Chromosome/scaffold name  
Gene start (bp)  
Gene end (bp)  
Gene name  
**Dataset**  
[None Selected]

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

Features    Homologues  
 Structures    Sequences

**Step 3b:** In the [Features](#) category, expand the [GENE](#) section

**GENE:**

- Gene stable ID
- Transcript stable ID
- Protein stable ID
- Exon stable ID
- Gene description
- Chromosome/scaffold name
- Gene start (bp)
- Gene end (bp)
- Strand
- Karyotype band
- Transcript start (bp)
- Transcript end (bp)

- Transcription start site (TSS)
- Transcript length (including UTRs and CDS)
- Ensembl Canonical
- Gene name
- Source of gene name
- Transcript count
- Gene % GC content
- Gene type
- Transcript type
- Source (gene)
- Source (transcript)
- Gene Synonym

**EXTERNAL:**

**PROTEIN DOMAINS AND FAMILIES:**

Make sure that [Features](#) is selected at the top of the page. Expand the [GENE](#) section, select [Chromosome/scaffold name](#), [Gene start](#) and [Gene end](#), and [Gene name](#).

**Dataset 4 / 16163 Genes**  
Fusarium solani genes (v2.0)  
**Filters**  
Orthologous Fusarium verticillioides Genes: Excluded  
Chromosome/scaffold: 14  
Start: 1128520  
End: 1142558  
**Attributes**  
Gene stable ID  
Transcript stable ID  
Chromosome/scaffold name  
Gene start (bp)  
Gene end (bp)  
Gene name  
GO term accession  
GO term name  
**Dataset**  
[None Selected]

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

Features    Homologues (Max select 6 orthologues)  
 Structures    Sequences

**Step 3c:** Expand the [EXTERNAL](#) section

**GENE:**

**EXTERNAL:**

**GO**

- GO term accession
- GO term name
- GO term definition

**GOSlim GOA**

- GOSlim GOA Accession(s)

**Step 3d:** Select [GO term accession](#) and [GO term name](#)

**GOSlim GOA Description**

**Pathogen Phenotypes (source: PHI-base)**

- PHI-base ID
- Host

**External References (max 3)**

- European Nucleotide Archive ID
- INSDC protein ID
- MEROPS - the Peptidase Database ID
- NCBI gene (formerly Entrezgene) description
- NCBI gene (formerly Entrezgene) accession
- NCBI gene (formerly Entrezgene) ID
- PDB ID

**Step 3d:** Select [GO term accession](#) and [GO term name](#)

- RefSeq peptide predicted ID
- RFAM ID
- STRING ID
- tRNAscan-SE ID
- UniParc ID
- UniProtKB/Swiss-Prot ID
- UniProtKB/TrEMBL ID

Expand the [EXTERNAL](#) section. This section contains lots of identifiers from databases outside of Ensembl. Select [GO term accession](#) and [GO term name](#).

**Step 4:** Get results!

You will retrieve your BioMart results in tabular format. Notice the order of the columns - these are in the same order in which you selected your [Attributes](#).

You can download the data if you like. The output table shows only 10 first rows by default.

Gene stable ID	Transcript stable ID	Chromosome/scaffold name	Gene start (bp)	Gene end (bp)	Gene name	GO term accession	GO term name
NechaG73960	NechaT73560	14	1129115	1131280	PEP5	GO:0016021	integral component of membrane
NechaC73960	NechaT73560	14	1129115	1131280	PEP5	GO:0022857	transmembrane transporter activity
NechaC73960	NechaT73560	14	1129115	1131280	PEP5	GO:0055085	transmembrane transport
NechaG73960	NechaT73560	14	1129115	1131280	PEP5	GO:0016020	membrane
NechaC64937	NechaT64937	14	1131753	1133840	PDA1	GO:0016021	integral component of membrane
NechaC64937	NechaT64937	14	1131753	1133840	PDA1	GO:0004497	monooxygenase activity
NechaC64937	NechaT64937	14	1131753	1133840	PDA1	GO:0020037	heme binding
NechaG64937	NechaT64937	14	1131753	1133840	PDA1	GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
NechaC64937	NechaT64937	14	1131753	1133840	PDA1	GO:0005508	iron ion binding
NechaC64937	NechaT64937	14	1131753	1133840	PDA1	GO:0016020	membrane

You can click on the location links and explore the synteny between the two species on the Ensembl Fungi browser.

**What about the last point? ‘Export their cDNA sequences?’**

In the [Attributes](#) section there are some ‘radio buttons’. If you’d like to export Sequence data, you need to build a separate query.

**Step 3.2:** Let’s go back to step 3: Selecting attributes

From the results page, click back to [Attributes](#) in the left-hand navigation panel – there’s no need to start from scratch.

**Dataset 4 / 16163 Genes**  
Fusarium solani genes (v2.0)  
**Filters**  
Orthologous Fusarium verticillioides Genes: Excluded Chromosome/scaffold: 14  
Start: 1128520  
End: 1142558  
**Attributes**  
Gene stable ID  
Transcript stable ID  
Gene name  
cDNA sequences

**Dataset**  
[None Selected]

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

**Step 3.2b: Click on Sequences**

**SEQUENCES:**

**Sequences (max 1)**

**Step 3.2a: Click on Attributes again**

**Upstream flank**

**Downstream flank**

**HEADER INFORMATION:**

**Step 3.2c: Select cDNA sequences**

○ Features ○ Homologues (Max selected)  
○ Structures ○ Sequences

○ 5' UTR  
○ 3' UTR  
○ Exon sequences  
● cDNA sequences  
○ Coding sequence  
○ Peptide

Also expand the **HEADER INFORMATION** section and select **Gene name**.

#### Step 4.2: View results for the sequences

**Dataset 4 / 16163 Genes**  
Fusarium solani genes (v2.0)  
**Filters**  
Orthologous Fusarium verticillioides Genes: Excluded Chromosome/scaffold: 14  
Start: 1128520  
End: 1142558  
**Attributes**  
Gene stable ID  
Transcript stable ID  
Gene name  
cDNA sequences

**Dataset**  
[None Selected]

**Step 4.2a: Click on Results again**

Email notification to

View 10 rows as FASTA Unique results only **Step 4.2b: Click on Go**

```
>NechatG73960|NechatT73950|PEP5
GCCTATCCAAGTTCAAGCAATGTCGGAGCATCGCGACAAGGTCGACGACGAGCCAAACGC
AGTTGGGTGAACTCCTCGCTGGAAACGAGGATGAAAACATACAAAGGGCTTACATGGAA
ATCTGUTTCGCGCTATCTGCTATTCGTTGCAACTGCACAGGGTACCTCGCTATTTGT
CTGGATCCTCAAGGGCTGATCATCTGGCTTACCTTCACGGCACCCCCAAATTGGCG
GTGGATTTCTGGGGTCGAGAGTACCCATCCTCATTTCCACAGGCCCTCTGCTCATCGG
TTCGCTTCTGATAAGCAGGGGGTTTCCATGACCATGGCGATTGCTGAGTGCTGCTGGTC
GGTAGGATCAGGATCGTCCTCATTACTGTATGAGTTGATCAGAGATCATGCCAGGG
CGATATCGAACATTCGCTCAAGCCGGACTCAACATGGCCAAATTCACTGGAGGAATATTC
ACACTGCTCGGGCTTCGCTCTGCAAGAAGAGTGACGAAGGCTTCGGGTTGCTAG
TACATTACACACATTTTCCCATCACTTTTACCTTACCTCTCTCTTCAACCCG
CCCCAAACGGCCCTCGAACACTTCTGCGGAATGATTCTGGCTGACATGGGG
ATAGGATACATGCTATTCTCTATCCGGAATGATTCTGTTGCTCAGCATGGCCTGACATGGGG
AAACACCCCTATCCGGAAAGGATGCTCATGTCGCTCAGCACCTTATGTCGGGCTAGGT
TTTATGCGCCCTACTGCTATGTTGGAGGTCAAGAGAAGGACGGATTCTGTCACCATGCT
CTTTTGAAACGGCCGCAACTTCCGCTTCCGCTTATCCGTAATTTCGCTCTTGGAGGATT
GCATTCTACCGCTGCAACAACTTCTCCCTTCGTAATTTCGCTCTTGGAGGAT
CAATTCAAGGGCCGGTTTGTCTCCATCGTCTCATGGCGGGGAGGTTCATCCGTG
GTTGGGGCTTACTCTTCAAGAAAAGGGTTCGACGCCCTTGATGCTGGGAATG
GGGCTTTCTCAGCTATTGCTATGGCATGATGGTACCATCAAGCTAACCGTCAAGCTGCC
```

What did you learn about these genes in this exercise?

Could you learn these things from the Ensembl browser? Would it take longer?

For more details on BioMart, have a look at this publication:

Kinsella RJ, Kähäri A, Haider S, et al. [Ensembl BioMarts: a hub for data retrieval across taxonomic space](#). Database : the Journal of Biological Databases and Curation. 2011;2011:bar030. DOI: 10.1093/database/bar030. PMID: 21785142; PMCID: PMC3170168.

## Additional BioMart Exercise 1 – Export orthologues

Use Ensembl Fungi BioMart to retrieve all *Zymoseptoria tritici* genes associated with the GO term ‘detoxification’ located on chromosome 1. Export the gene IDs, names, homology type and confidence of their orthologues in *Blumeria graminis*, *Botrytis cinerea*, *Cryptococcus neoformans* and *Saccharomyces cerevisiae*.

- (a) Do all of these *Z. tritici* genes have an orthologue in the other species? Which of these species are pathogenic? Do you see a correlation?
- (b) Can you find an orthologue in *Cryptococcus neoformans* with high orthology confidence? What is the Gene ID? We will explore more about this orthologue in the exercise section for the Evolutionary Analysis module.

### Exercise 1 answers

You can open BioMart by clicking [BioMart](#) in the navigation bar at the top of any Ensembl Fungi page, or by entering the URL <https://fungi.ensembl.org/biomart/martview/> in your browser. Using the four-step process described above, we need to select the following in BioMart options in the left-hand panel:

1. Dataset: Click on **CHOOSE DATABASE** and select [Ensembl Fungi Genes 58](#) from the drop-down menu. Click on **CHOOSE DATASET** and select [Zymoseptoria tritici genes \(MG2\)](#) from the drop-down menu.

2. Filters: Open the **REGION** tab and select **1** under **Chromosome/scaffold**. Open the **GENE ONTOLOGY** tab and enter **detoxification** under **GO Term Name**. Click on the **Count** button in the top left-hand corner. Your filter should apply to 19/11,091 genes.

3. Attributes: Select [Homologues](#) from the options on the top. Open the [GENE](#) tab and select [Gene name](#). Open the [ORTHOLOGUES \[A-E\]](#) tab and select the following options:

- [Blumeria graminis gene stable ID](#)
- [Blumeria graminis gene name](#)
- [Blumeria graminis homology type](#)
- [Blumeria graminis orthology confidence \[0 low, 1 high\]](#)
- [Botrytis cinerea B05.10 gene stable ID](#)
- [Botrytis cinerea B05.10 gene name](#)
- [Botrytis cinerea B05.10 homology type](#)
- [Botrytis cinerea B05.10 orthology confidence \[0 low, 1 high\]](#)
- [Cryptococcus neoformans var. neoformans JEC21 gene stable ID](#)
- [Cryptococcus neoformans var. neoformans JEC21 gene name](#)
- [Cryptococcus neoformans var. neoformans JEC21 homology type](#)
- [Cryptococcus neoformans var. neoformans JEC21 orthology confidence \[0 low, 1 high\]](#)

Open the [ORTHOLOGUES \[P-T\]](#) tab and select the following options:

- [Saccharomyces cerevisiae gene stable ID](#)
- [Saccharomyces cerevisiae gene name](#)
- [Saccharomyces cerevisiae homology type](#)
- [Saccharomyces cerevisiae orthology confidence \[0 low, 1 high\]](#)

The screenshot shows the Ensembl Fungi BioMart search interface. On the left, there are three tabs: 'New', 'Count', and 'Results'. The 'Results' tab is selected. The main area has a sidebar with 'Dataset' (Zymoseptoria tritici genes (MG2)), 'Filters' (Chromosome/scaffold: 1, GO Term Name [e.g. regulation of biological process]: detoxification), and 'Attributes' (Gene stable ID, Transcript stable ID, Gene name, Blumeria graminis gene stable ID, Blumeria graminis gene name, Blumeria graminis homology type, Blumeria graminis orthology confidence [0 low, 1 high]). The main panel displays two sets of orthologous genes:

- Saccharomyces cerevisiae Orthologues:**
  - Pyrenopeltis tritici-repentis Pt-1C-BFP chromosome/scaffold end (bp) checkbox.
  - Pyrenopeltis tritici-repentis Pt-1C-BFP orthology confidence [0 low, 1 high] checkbox.
  - Query protein or transcript ID checkbox.
  - Last common ancestor with Saccharomyces cerevisiae checkbox.
  - Saccharomyces cerevisiae homology type checkbox.
  - %id. target Saccharomyces cerevisiae gene identical to query gene checkbox.
  - %id. query gene identical to target Saccharomyces cerevisiae gene checkbox.
  - Saccharomyces cerevisiae orthology confidence [0 low, 1 high] checkbox.
- Schizosaccharomyces cryophilus Orthologues:**
  - Schizosaccharomyces cryophilus gene stable ID checkbox.
  - Schizosaccharomyces cryophilus gene name checkbox.
  - Schizosaccharomyces cryophilus protein or transcript stable ID checkbox.
  - Schizosaccharomyces cryophilus chromosome/scaffold name checkbox.
  - Query protein or transcript ID checkbox.
  - Last common ancestor with Schizosaccharomyces cryophilus checkbox.
  - Schizosaccharomyces cryophilus homology type checkbox.
  - %id. target Schizosaccharomyces cryophilus checkbox.

At the bottom, it says 'Ensembl Genomes release 58 - January 2024 (c) EBI'.

4. Results: Click on the Results button in the top left-hand corner to view your output table. Select All from the drop-down menu to open the full table in a new tab.

The screenshot shows the Ensembl Fungi BioMart search interface with the 'Results' tab selected. The sidebar includes 'Dataset' (Zymoseptoria tritici genes (MG2)), 'Filters' (Chromosome/scaffold: 1, GO Term Name [e.g. regulation of biological process]: detoxification), and 'Attributes' (Gene stable ID, Transcript stable ID, Gene name, Blumeria graminis gene stable ID, Blumeria graminis gene name, Blumeria graminis homology type, Blumeria graminis orthology confidence [0 low, 1 high]). The main panel has export options ('File', 'TSV', 'Unique results only', 'Go') and an email notification field. It also allows selecting rows (10, 20, 50, 100, 150, 200, All) and viewing them as HTML or TSV. A table is displayed with columns: Gene stable ID, Transcript stable ID, Blumeria graminis gene stable ID, Blumeria graminis gene name, Blumeria graminis homology type, Blumeria graminis orthology confidence [0 low, 1 high], Botrytis cinerea B05.1 gene stable ID, and Botrytis cinerea B05.1 gene name. The table contains several rows of data, such as Mycg3G90087, Mycg3T900B7, BLGH\_05232, ortholog\_oneZone, 1, Bcin03g01480, Bcpox, etc.

Ensembl Genomes release 58 - January 2024 (c) EBI

No, not all *Z. tritici* genes located on chromosome 1 with the associated GO term ‘detoxification’ have an orthologue in the other species. *B. graminis* causes powdery mildew on grasses (e.g. cereals), *B. cinerea* is known to cause botrytis bunch rot in grape and *C. neoformans* is the causative agent of cryptococcosis and cryptococcal meningitis. Do you see a correlation?

- (a) CNM01690 in *C. neoformans* has high orthology confidence.

## Additional BioMart Exercise 2 – Finding genes by protein domain

Generate a list of all *Magnaporthe oryzae* (MG8) genes on chromosome 4 that are annotated to contain Transmembrane domains/helices. Include the Ensembl gene stable ID and description.

### Exercise 2 answers

Click on the [New](#) button in the top left-hand corner to start a new BioMart query. Using the four-step process described above, we need to select the following in BioMart options in the left-hand panel:

1. Dataset: Click on [CHOOSE DATABASE](#) and select [Ensembl Fungi Genes 58](#) from the drop-down menu. Click on [CHOOSE DATASET](#) and select [Magnaporthe oryzae genes \(MG8\)](#) from the drop-down menu.

The screenshot shows the Ensembl Fungi BioMart interface. The top navigation bar includes links for BLAST, BioMart, FTP, and Docs & FAQs, along with URL, XML, Perl, and Help buttons. The main interface has three tabs at the top: New (selected), Count, and Results. On the left, there's a sidebar with sections for Dataset, Filters, and Attributes. Under Dataset, 'Ensembl Fungi Genes 58' is selected. Under Filters, '[None selected]' is shown. Under Attributes, 'Gene stable ID' and 'Transcript stable ID' are listed. The main panel is currently empty, indicating no results have been returned yet. At the bottom, a footer bar displays 'Ensembl Genomes release 58 - January 2024 (c) EBI'.

2. Filters: Open the [REGION](#) tab and select [4](#) under [Chromosome/scaffold](#). Open the [PROTEIN DOMAINS AND FAMILIES](#) tab and select [With transmembrane helices - Only](#) under [Limit to genes....](#). Click on the [Count](#) button in the top left-hand corner. Your filter should apply to 297/13,470 genes.

## Database queries: Ensembl Fungi (BioMart)

The screenshot shows the Ensembl Fungi BioMart search interface. On the left, a sidebar displays the dataset information: "Dataset 297 / 13470 Genes" and "Magnaporthe oryzae genes (MG8)". Under "Filters", it shows "Chromosome/scaffold: 4" and "With Transmembrane helices: Only". Under "Attributes", it lists "Gene stable ID" and "Transcript stable ID". The main panel has a header "Please restrict your query using criteria below" with a note "(If filter values are truncated in any lists, hover over the list item to see the full text)". It contains several filter sections: "REGION", "GENE", "PATHOGEN PHENOTYPES (PHI-BASE)", "GENE ONTOLOGY", "MULTI SPECIES COMPARISONS", and "PROTEIN DOMAINS AND FAMILIES". In the "PROTEIN DOMAINS AND FAMILIES" section, there are two checkboxes: "Limit to genes ..." (checked) and "Limit to genes with these family or domain IDs [Max 500 advised]" (unchecked). Below these is a dropdown menu for "Interpro ID(s)" containing "IPR000008" and a file upload input field with "Choose file" and "No file chosen". At the bottom of the main panel, a footer bar reads "Ensembl Genomes release 58 - January 2024 (c) EBI".

3. Attributes: Select **Features** from the options on the top. Open the **GENE** tab, unselect **Transcript stable ID** and select **Gene description**.

The screenshot shows the Ensembl Fungi BioMart search interface. The sidebar remains the same as the previous screenshot. The main panel has a header "Please select columns to be included in the output and hit 'Results' when ready". It contains two groups of checkboxes: "Features" (selected) and "Structures" (unchecked), and "Homologues (Max select 6 orthologues)" (unchecked). Below this is a "GENE" tab with a sub-section "Ensembl". Under "Ensembl", there are two columns of checkboxes. The left column includes: "Gene stable ID" (checked), "Transcript stable ID" (unchecked), "Protein stable ID" (unchecked), "Exon stable ID" (unchecked), "Gene description" (checked), "Chromosome/scaffold name" (unchecked), "Gene start (bp)" (unchecked), "Gene end (bp)" (unchecked), "Strand" (unchecked), "Karyotype band" (unchecked), "Transcript start (bp)" (unchecked), "Transcript end (bp)" (unchecked), and "Transcription start site (TSS)" (unchecked). The right column includes: "Transcript length (including UTRs and CDS)" (unchecked), "Ensembl Canonical" (unchecked), "Gene name" (unchecked), "Source of gene name" (unchecked), "Transcript name" (unchecked), "Source of transcript name" (unchecked), "Transcript count" (unchecked), "Gene % GC content" (unchecked), "Gene type" (unchecked), "Transcript type" (unchecked), "Source (gene)" (unchecked), "Source (transcript)" (unchecked), and "Gene Synonym" (unchecked). At the bottom of the main panel, a footer bar reads "Ensembl Genomes release 58 - January 2024 (c) EBI".

4. Results: Click on the **Results** button in the top left-hand corner to view your output table. Select **All** from the drop-down menu to open the full table in a new tab.

**e!EnsemblFungi**

New Count Results

BLAST | BioMart | FTP | Docs & FAQs

★ URL | XML | Perl | Help

**Dataset 297 / 13470 Genes**  
Magnaporthe oryzae genes (MG8)

**Filters**  
Chromosome/scaffold: 4  
With Transmembrane helices: Only

**Attributes**  
Gene stable ID  
Gene description

**Dataset**  
[None Selected]

Export all results to  File  TSV  Unique results only  Go

Email notification to

View  10 rows as  HTML  Unique results only

Gene stable ID	Gene description
<a href="#">MGG_17084</a>	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4N801]
<a href="#">MGG_03684</a>	Mitochondrial distribution and morphology protein 38 [Source:UniProtKB/TrEMBL;Acc:G4N6R1]
<a href="#">MGG_09963</a>	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4N9P1]
<a href="#">MGG_03644</a>	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4N713]
<a href="#">MGG_06510</a>	Cytochrome b5 [Source:UniProtKB/TrEMBL;Acc:G4N6W6]
<a href="#">MGG_09720</a>	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4NAH4]
<a href="#">MGG_03721</a>	Urea transporter [Source:UniProtKB/TrEMBL;Acc:G4N6H1]
<a href="#">MGG_13659</a>	Dicarboxylic amino acid permease [Source:UniProtKB/TrEMBL;Acc:G4NAK4]
<a href="#">MGG_08498</a>	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4NAP3]
<a href="#">MGG_13624</a>	ABC transporter CDR4 [Source:UniProtKB/TrEMBL;Acc:G4N9L5]

Ensembl Genomes release 58 - January 2024 (c) EBI

## Additional BioMart Exercise 3 – Convert IDs

For a list of *Schizosaccharomyces pombe* UniProt (UniProtKB/Swiss-Prot) IDs, export the Gene name and description, as well as the PomBase IDs.

- (a) Do these 36 protein IDs correspond to 36 genes?

Input list of IDs:

Q92338	Q9US55	P78847	O74964
O13728	O14075	O94418	O14026
P49776	O94574	O94526	O74630
O74769	O94380	Q9UTG2	O14356
Q09170	P87172	O14326	O13339
Q9USK4	Q9USP5	Q9URZ3	P31411
O14040	Q9P7Y8	P42657	O13742
Q9Y804	Q9Y7Z8	P08647	O60159
094552	Q10331	O74335	O94280

### Exercise 3 answers

Click on the [New](#) button in the top left-hand corner to start a new BioMart query. Using the four-step process described above, we need to select the following in BioMart options in the left-hand panel:

- Dataset: Click on [CHOOSE DATABASE](#) and select [Ensembl Fungi Genes 58](#) from the drop-down menu. Click on [CHOOSE DATASET](#) and select [Schizosaccharomyces pombe genes \(ASM294v2\)](#) from the drop-down menu.

The screenshot shows the Ensembl Fungi BioMart interface. At the top, there's a navigation bar with links for BLAST, BioMart, FTP, and Docs & FAQs. Below the navigation bar, there are buttons for New, Count, and Results. On the right side, there are links for URL, XML, Perl, and Help. The main area has three dropdown menus: 'Dataset' set to 'Ensembl Fungi Genes 58', 'Dataset' set to 'Schizosaccharomyces pombe genes (ASM294v2)', and 'Filters' set to '[None selected]'. On the left, there's a sidebar with sections for 'Dataset' (set to 'Schizosaccharomyces pombe genes (ASM294v2)'), 'Filters' (set to '[None selected]'), and 'Attributes' (listing 'Gene stable ID' and 'Transcript stable ID'). At the bottom, it says 'Ensembl Genomes release 58 - January 2024 (c) EBI'.

2. Filters: Open the **GENE** tab and paste your list of IDs into the text box under **Input external references ID list**. Select **UniProtKB/Swiss-Prot ID(s) [e.g. A0ZWU1]** from the drop-down menu above to specify the type of IDs you are giving. Click on the **Count** button in the top left-hand corner. Your filter should apply to 36/7,268 genes.

**Dataset 36 / 7268 Genes**  
Schizosaccharomyces pombe genes (ASM294v2)

**Filters**  
UniProtKB/Swiss-Prot ID(s) [e.g. A0ZWU1]: [ID-list specified]

**Attributes**  
Gene stable ID  
Transcript stable ID

**Dataset**  
[None Selected]

**REGION:**

**GENE:**

- Limit to genes (external references)...  Only  Excluded
- Input external references ID list [Max 500 advised]   
P31411  
013742  
060159  
094287
- Transcript count >=
- Transcript count <=
- Gene type

Please restrict your query using criteria below  
(If filter values are truncated in any lists, hover over the list item to see the full text)

Choose file No file chosen

Ensembl Genomes release 58 - January 2024 (c) EBI

3. Attributes: Select **Features** from the options on the top. Open the **GENE** tab, unselect **Transcript stable ID** and select **Gene name** and **Gene description**. Open the **EXTERNAL** tab, scroll down to External References and select **PomBase ID**.

**Dataset 36 / 7268 Genes**  
Schizosaccharomyces pombe genes (ASM294v2)

**Filters**  
UniProtKB/Swiss-Prot ID(s) [e.g. A0ZWU1]: [ID-list specified]

**Attributes**  
Gene stable ID  
Gene name  
Gene description  
PomBase ID

**Dataset**  
[None Selected]

**External References (max 3)**

- ChEMBL ID
- Enzyme EC Number ID
- European Nucleotide Archive ID
- Fission Yeast Phenotype Ontology ID
- INSDC protein ID
- KEGG ID
- MEROPS - the Peptidase Database ID
- NCBI gene (formerly Entrezgene) description
- NCBI gene (formerly Entrezgene) accession
- NCBI gene (formerly Entrezgene) ID
- Orthologous Gene ID
- PDB ID
- PomBase ID
- PomBase (peptide) ID
- PomBase Ontology ID
- PomBase PubMed ID
- PSI-MOD, Post Translational Modification Ontology ID
- RefSeq DNA ID
- RefSeq mRNA ID
- RefSeq mRNA predicted ID
- RefSeq peptide ID
- RefSeq peptide predicted ID
- RFAM ID
- Sequence Ontology ID
- Sequence Publications ID
- SPD ID
- STRING ID
- tRNAscan-SE ID
- UniParc ID
- UniProtKB/SpliceVariant ID
- UniProtKB/Swiss-Prot ID
- UniProtKB/TrEMBL ID
- WikiGene description
- WikiGene name
- WikiGene ID

Ensembl Genomes release 58 - January 2024 (c) EBI

4. Results: Click on the **Results** button in the top left-hand corner to view your output table. Select **All** from the drop-down menu to open the full table in a new tab.

Gene stable ID	Gene name	Gene description	PomBase ID
<a href="#">SPBC29A3.14c</a>	<a href="#">trt1</a>	telomerase reverse transcriptase 1 protein Trt1 [Source:PomBase;Acc:SPBC29A3.14c]	<a href="#">SPBC29A3.14c.1</a>
<a href="#">SPAC15A10.08</a>	<a href="#">ain1</a>	alpha-actinin [Source:PomBase;Acc:SPAC15A10.08]	<a href="#">SPAC15A10.08.1</a>
<a href="#">SPAC16E8.07c</a>	<a href="#">vph1</a>	V-type ATPase V0 subunit a (predicted) [Source:PomBase;Acc:SPAC16E8.07c]	<a href="#">SPAC16E8.07c.1</a>
<a href="#">SPAC29B12.02c</a>	<a href="#">set2</a>	histone lysine methyltransferase Set2 [Source:PomBase;Acc:SPAC29B12.02c]	<a href="#">SPAC29B12.02c.1</a>
<a href="#">SPAC2C4.07c</a>	<a href="#">dis32</a>	3'-5'-exoribonuclease activity Dis3L2 [Source:PomBase;Acc:SPAC2C4.07c]	<a href="#">SPAC2C4.07c.1</a>
<a href="#">SPACUNK4.10</a>		glyoxylate reductase (predicted) [Source:PomBase;Acc:SPACUNK4.10]	<a href="#">SPACUNK4.10.1</a>
<a href="#">SPBC16E9.11c</a>	<a href="#">pub3</a>	HECT-type ubiquitin-protein ligase E3 Pub3 (predicted) [Source:PomBase;Acc:SPBC16E9.11c]	<a href="#">SPBC16E9.11c.1</a>
<a href="#">SPBC30D10.10c</a>	<a href="#">tor1</a>	phosphatidylinositol kinase Tor1 [Source:PomBase;Acc:SPBC30D10.10c]	<a href="#">SPBC30D10.10c.1</a>
<a href="#">SPBC19C7.11</a>		CIC chloride channel (predicted) [Source:PomBase;Acc:SPBC19C7.11]	<a href="#">SPBC19C7.11.1</a>
<a href="#">SPBC17F3.01c</a>	<a href="#">rga5</a>	Rho-type GTPase activating protein Rga5 [Source:PomBase;Acc:SPBC17F3.01c]	<a href="#">SPBC17F3.01c.1</a>
<a href="#">SPCC23B6.03c</a>	<a href="#">tel1</a>	ATM checkpoint kinase [Source:PomBase;Acc:SPCC23B6.03c]	<a href="#">SPCC23B6.03c.1</a>
<a href="#">SPBC24C6.08c</a>	<a href="#">bhd1</a>	folliculin/Birt-Hogg-Dube syndrome ortholog Bhd1 [Source:PomBase;Acc:SPBC24C6.08c]	<a href="#">SPBC24C6.08c.1</a>
<a href="#">SPBC4B4.03</a>	<a href="#">rsc1</a>	RSC complex subunit Rsc1 [Source:PomBase;Acc:SPBC4B4.03]	<a href="#">SPBC4B4.03.1</a>
<a href="#">SPBC887.02</a>		CIC chloride channel (predicted) [Source:PomBase;Acc:SPBC887.02]	<a href="#">SPBC887.02.1</a>
<a href="#">SPBC1604.15</a>	<a href="#">gpi16</a>	pig-T, Gpi16 (predicted) [Source:PomBase;Acc:SPBC1604.15]	<a href="#">SPBC1604.15.1</a>
<a href="#">SPCC1620.11</a>	<a href="#">nup97</a>	nucleoporin Nic96 homolog [Source:PomBase;Acc:SPCC1620.11]	<a href="#">SPCC1620.11.1</a>
<a href="#">SPBC609.02</a>	<a href="#">ptn1</a>	phosphatidylinositol-3,4,5-trisphosphate3-phosphatase Ptn1 [Source:PomBase;Acc:SPBC609.02]	<a href="#">SPBC609.02.1</a>
<a href="#">SPCC18.18c</a>	<a href="#">fum1</a>	fumarate hydratase (predicted) [Source:PomBase;Acc:SPCC18.18c]	<a href="#">SPCC18.18c.1</a>
<a href="#">SPBC1773.17c</a>		glyoxylate reductase (predicted) [Source:PomBase;Acc:SPBC1773.17c]	<a href="#">SPBC1773.17c.1</a>
<a href="#">SPAC17H9.09c</a>	<a href="#">ras1</a>	GTPase Ras1 [Source:PomBase;Acc:SPAC17H9.09c]	<a href="#">SPAC17H9.09c.1</a>
<a href="#">SPAC637.05c</a>	<a href="#">vma2</a>	V-type ATPase V1 subunit B [Source:PomBase;Acc:SPAC637.05c]	<a href="#">SPAC637.05c.1</a>
<a href="#">SPAC17A2.13c</a>	<a href="#">rad25</a>	14-3-3 protein Rad25 [Source:PomBase;Acc:SPAC17A2.13c]	<a href="#">SPAC17A2.13c.1</a>
<a href="#">SPCC4G3.02</a>	<a href="#">aph1</a>	bis(5'-nucleosidyl)-tetraphosphatase [Source:PomBase;Acc:SPCC4G3.02]	<a href="#">SPCC4G3.02.1</a>
<a href="#">SPCC290.03c</a>	<a href="#">nup186</a>	nucleoporin Nup186 [Source:PomBase;Acc:SPCC290.03c]	<a href="#">SPCC290.03c.1</a>
<a href="#">SPBC3D6.07</a>	<a href="#">gpi3</a>	pig-A, phosphatidylinositol N-acetylglucosaminyltransferase subunit Gpi3 (predicted) [Source:PomBase;Acc:SPBC3D6.07]	<a href="#">SPBC3D6.07.1</a>
<a href="#">SPCC18B5.11c</a>	<a href="#">cds1</a>	replication checkpoint kinase Cds1 [Source:PomBase;Acc:SPCC18B5.11c]	<a href="#">SPCC18B5.11c.1</a>
<a href="#">SPBC428.01c</a>	<a href="#">nup107</a>	nucleoporin Nup107 [Source:PomBase;Acc:SPBC428.01c]	<a href="#">SPBC428.01c.1</a>
<a href="#">SPBC2D10.18</a>	<a href="#">abc1</a>	ABC1 kinase family ubiquinone biosynthesis protein Abc1/Coq8 [Source:PomBase;Acc:SPBC2D10.18]	<a href="#">SPBC2D10.18.1</a>
<a href="#">SPAPYUG7.03c</a>	<a href="#">mid2</a>	medial ring protein Mid2 [Source:PomBase;Acc:SPAPYUG7.03c]	<a href="#">SPAPYUG7.03c.1</a>
<a href="#">SPAC869.10c</a>	<a href="#">put4</a>	proline specific plasma membrane permease Put4 (predicted) [Source:PomBase;Acc:SPAC869.10c]	<a href="#">SPAC869.10c.1</a>
<a href="#">SPAC1002.03c</a>	<a href="#">gls2</a>	glucosidase II alpha subunit Gls2 [Source:PomBase;Acc:SPAC1002.03c]	<a href="#">SPAC1002.03c.1</a>
<a href="#">SPCC4B3.14</a>	<a href="#">cwf20</a>	complexed with Cdc5 protein Cwf20 [Source:PomBase;Acc:SPCC4B3.14]	<a href="#">SPCC4B3.14.1</a>
<a href="#">SPCC11E10.02c</a>	<a href="#">gpi8</a>	pig-K [Source:PomBase;Acc:SPCC11E10.02c]	<a href="#">SPCC11E10.02c.1</a>
<a href="#">SPAC1805.15c</a>	<a href="#">pub2</a>	HECT-type ubiquitin-protein ligase E3 Pub2 [Source:PomBase;Acc:SPAC1805.15c]	<a href="#">SPAC1805.15c.1</a>
<a href="#">SPBC146.13c</a>	<a href="#">myo1</a>	myosin type I [Source:PomBase;Acc:SPBC146.13c]	<a href="#">SPBC146.13c.1</a>
<a href="#">SPBC146.06c</a>	<a href="#">fan1</a>	Fanconi-associated nuclease Fan1 [Source:PomBase;Acc:SPBC146.06c]	<a href="#">SPBC146.06c.1</a>

- (a) Yes, the 36 UniProt IDs correspond to 36 genes. However, not all of them have a gene name assigned to them (e.g. SPACUNK4.10).