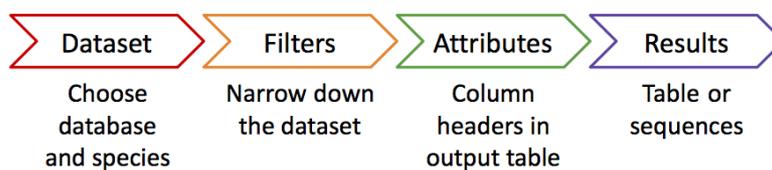


Demo: Ensembl Fungi BioMart

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

1. How many genes are found in *Fusarium solani* that do not have an orthologue in *Fusarium oxysporum*?
2. How many of these are associated with a pathogenic phenotype of ‘reduced virulence’?
3. Export the gene name, locations and GO terms associated with these genes
4. Export their cDNA sequences

BioMart queries: The 4 step process



Click on [BioMart](#) in the top header of a [fungi.ensembl.org](#) page to go to: [fungi.ensembl.org/biomart/martview](#)

NOTE: These answers were determined using BioMart Ensembl Fungi 42

Step 1: Choose database and dataset

The screenshot shows the Ensembl Fungi BioMart interface. The 'Dataset' dropdown menu is open, showing three options: '- CHOOSE DATABASE -', 'Ensembl Fungi Genes 42', and 'Ensembl Fungi Variations 42'. A callout box points to the first option with the text 'Step 1: Choose Genes database'.

The screenshot shows the Ensembl Fungi BioMart interface. The 'Dataset' dropdown menu is open, showing two options: 'Ensembl Fungi Genes 42' and 'Fusarium solani genes (v2.0)'. A callout box points to the second option with the text 'Step 2: Choose Fusarium solani genes (v2.0) as the dataset'.

Step 2: Choose appropriate filters

We want to find all genes in *Fusarium solani* that **do not** have an orthologue with *Fusarium oxysporum*. We need to filter the dataset to look only at these genes.

Dataset
Fusarium solani genes (v2.0)

Filters
[None Selected]

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:

MULTI SPECIES COMPARISONS:

Homologue filters

Paralogous Fusarium solani Genes

Only

Excluded

PROTEIN DOMAINS AND FAMILIES:

Top tip: Click **Count** to check if your filters work

Dataset 6033 / 16464 Genes
Fusarium solani genes (v2.0)

Filters
Orthologous Fusarium oxysporum 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:

MULTI SPECIES COMPARISONS:

Homologue filters

Orthologous Fusarium oxysporum Genes

Only

Excluded

PROTEIN DOMAINS AND FAMILIES:

Using the count function we can see that there are about 6,000 *F. solani* genes (out of a total of 16,464) that do not have an orthologue in *F. oxysporum*. We also want to find out how many of these are associated with the PHI-base Pathogen phenotype ‘reduced virulence’.

Update the count

Step 2e: Expand the PATHOGEN PHENOTYPES (PHI-BASE) section

Step 2f: Choose the Reduced virulence option

We can now see from the count information that we have 3 genes associated with this phenotype that do not have any orthologues with *F. oxysporum*.

Step 3: Select Attributes

Attributes are defined by what we would like to learn about the data. We want to find out more information about the genes:

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

We can answer points 1-3 in a single query, we will need to do a second query to answer point 4.

Step 3a: Click on Attributes

Step 3b: In the 'Features' section. Expand the GENE section

Make sure that features is selected at the top of the page.

Expand the **GENE** section, **deselect** Transcript stable ID, and **select** Chromosome/scaffold name, Gene start and Gene end, and Gene name.

Step 3c: Expand the EXTERNAL section

Step 3d: Choose GO term accession and name

EXTERNAL:

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

GOSlim GOA

- GOSlim GOA Accession(s)
- GOSlim GOA Description

Pathogen Phenotypes (source: PHI-base)

- PHI-base ID
- Plant host
- Pathogenic phenotype
- Experimental condition

External References (max 3)

- European Nucleotide Archive ID
- INSDC protein ID
- KEGG Pathway and Enzyme ID
- MEROPS - the Peptidase Database ID
- NCBI gene ID
- Rfam 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 can download the data if you desire. The table presented shows a sub-sample of 10 results to enable you to check you have the correct attributes.

Step 4a: Click Results

Step 4b: Change the number of rows to 20 to view all results

Export all results to: File TSV Unique results only **Go**

Email notification to: []

View: 20 rows as

Gene stable ID	Gene name	Chromosome/scaffold name	Gene start (bp)	Gene end (bp)	GO term accession	GO term name
NechaG64937	PDA1	14	1131753	1133840	GO:0005506	iron ion binding
NechaG64937	PDA1	14	1131753	1133840	GO:0020037	heme binding
NechaG64937	PDA1	14	1131753	1133840	GO:0004497	monooxygenase activity
NechaG64937	PDA1	14	1131753	1133840	GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
NechaG64937	PDA1	14	1131753	1133840	GO:0016020	membrane
NechaG64937	PDA1	14	1131753	1133840	GO:0016021	integral component of membrane
NechaG64937	PDA1	14	1131753	1133840	GO:0016491	oxidoreductase activity
NechaG64937	PDA1	14	1131753	1133840	GO:0046872	metal ion binding
NechaG64937	PDA1	14	1131753	1133840	GO:0055114	oxidation-reduction process
NechaG73962	PEP2	14	1141191	1142037		
NechaG73960	PEP5	14	1129115	1131280	GO:0016021	integral component of membrane
NechaG73960	PEP5	14	1129115	1131280	GO:0055085	transmembrane transport
NechaG73960	PEP5	14	1129115	1131280	GO:0016020	membrane

Each attribute becomes a column in the table

We can see that all of these genes are located in the same region. Perhaps this has something to do with why they are not found in *F. oxysporum*...? You can click on the location links and explore the synteny between the two species.

What about the fourth point? ‘Export their cDNA sequences?’

In the Attributes section there are some ‘radio buttons’. You can only choose attributes from one of these at a time. If we want Sequence data we have to do a separate query.

Step 3.2: Select attributes to answer Question 3

From the results page, click back to [Attributes](#) in the left-hand navigation panel.

Also expand the **HEADER INFORMATION** section and **deselect** Transcript stable ID and **select** Gene name.

Step 4.2: View results for the sequences

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, R.J. *et al*

Ensembl BioMart: a hub for data retrieval across taxonomic space.

Database (Oxford) 2011:bar030

Exercises: BioMart

BioMart Exercise 1 – Convert IDs

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

Q92338	O13728
P49776	O74769
Q09170	Q9USK4
O14040	Q9Y804
O94552	Q9US55
O14075	O94574
O94380	P87172
Q9USP5	Q9P7Y8
Q9Y7Z8	Q10331
P78847	O94418
O94526	Q9UTG2
O14326	Q9URZ3
P42657	P08647
O74335	O74964
O14026	O74630
O14356	O13339
P31411	O13742
O60159	O94287

BioMart Exercise 2 – Export PHI-base data and Orthologues

Use Ensembl Fungi BioMart to retrieve all *Zygomycota tritici* genes that have been linked to a loss of pathogenicity. Export the gene IDs, homology type and confidence of their orthologues in *Blumeria graminis*, *Botrytis cinerea*, *Cryptococcus neoformans* and *Saccharomyces cerevisiae*.

Q1: Do all of these genes in *Z. tritici* have an orthologue in these species?

Q2: Find the only orthologue in *Cryptococcus neoformans*. Is this a high confidence orthologue?

We will explore more about this orthologue in the exercise section for Comparative Genomics.

BioMart Exercise 3 – Finding genes by protein domain

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