

Data retrieval and download Exercise 9

9.1 Downloading a set of results and associated data.

For this exercise you can start with any gene list of results. Start with any result list you have generated, such as the DNA Motif search. Download all the genes:

The screenshot shows a workflow with four steps: Step 1 (DNA Motif 18800 Segments), Step 2 (Organism 18800 Genes), Step 3 (DNA Motif 18800 Segments), and Step 4 (Organism 4121 Genes). Step 4 is highlighted in yellow and shows 31 genes. A red circle highlights the 'Download 31 Genes' link in the top right of the results section.

All Results	Ortholog Groups	Ajellomyces				Aspergillus						Batrachochytrium		
		A. capsulatus (nr Genes: 0)		A. macrogynus	A. aculeatus	A. carbonarius	A. clavatus	A. flavus	A. fumigatus	A. nidulans	A. niger (nr Genes: 0)		A. terreus	B. dendrobatidis
		G186AR	NAm1	ATCC 38327	ATCC 16972	ITEM 5010	NRRL 1	NRRL3357	At293	FGSC A4	ATCC 1015	CBS 513.88	NH2624	JEL423
31	31	0	0	0	0	0	0	0	0	0	0	0	0	0

Gene Results | Genome View

First 1 2 Next Last | Advanced Paging | Add Columns

Gene ID	Genomic Location	Product Description
PYU1_G007170	PuIDAOMBR144_SC1242: 96,660 - 97,169 (-)	Similar to TPRXL: Putative protein TPRXL (Homo sapiens)

Download this list of results with the following associated data: Genomic Location, Product Description, Transcript Length and Predicted GO Function. Hint: click on the Download ## Genes link.

Hint: select the type of report to download and then click on the boxes to customize your report. The gene ID is automatically downloaded and so is not an option in the popup.

Download 31 Genes from the search:

Combine Gene results

Please select a format from the dropdown list to create the download report.

Tab delimited (Excel): choose from columns

Note: IDs will automatically be included in the report and the report will be sorted by ID.

Below select type and format for download

You may include additional columns in the report

clear all | expand all | collapse all
reset to current | reset to default

- Gene ID
- Text, IDs, Species
- Genomic Position
 - Chromosome
 - Genomic Location
 - Gene Strand
- Gene Attributes
 - Gene Type

9.2 Download the sequences of genes in a list of results.

Use the same list of results as in 9.1. Go to the download section and select “Configurable FASTA”. Download the ‘genomic’ sequences.

Now download the ‘transcript’ sequences. What is the difference?

Download 31 Genes from the search:
Combine Gene results

Please select a format from the dropdown list to create the download report.

FASTA (sequence retrieval, configurable) ▼

***Note: IDs will automatically be included in the report and the report will be sorted by ID.*

This reporter will retrieve the sequences of the genes in your result.

Choose the type of sequence: genomic protein CDS transcript

Choose the region of the sequence(s):

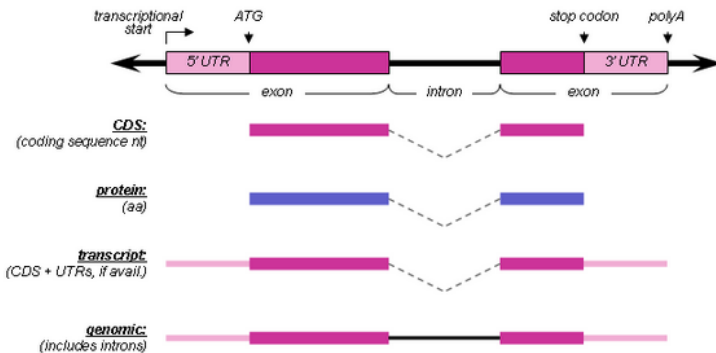
begin at: Transcription Start *** ▼ + ▼ 0 nucleotides

end at: Transcription Stop *** ▼ + ▼ 0 nucleotides

Download Type: Save to File Show in Browser

*** Note: If UTRs have not been annotated for a gene, then choosing "transcription start" may have the same effect as choosing "translation start".

Help



Note, that you can access and download sequence with the sequence retrieval tool (SRT) accessed from the tools menu on the home page:

- Retrieve Sequences By Gene IDs.
- Retrieve Sequences By Genomic Sequence IDs.
- Retrieve Multiple Sequence Alignments by Contig / Genomic Sequence IDs.
- Retrieve Sequences By Open Reading Frame IDs.

Tools:

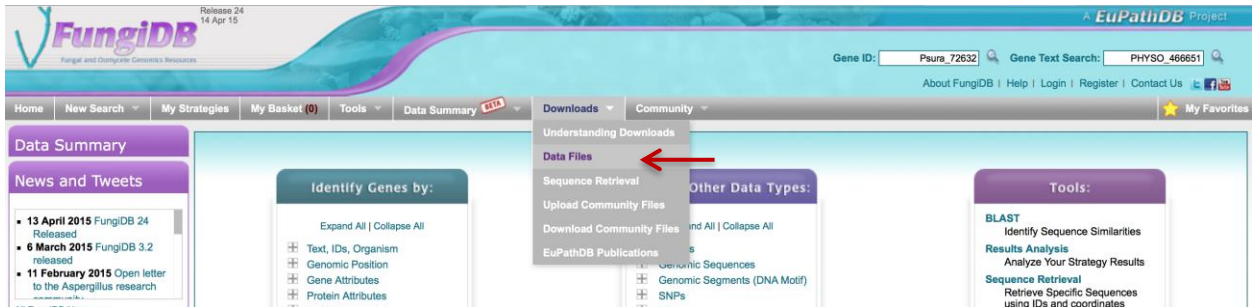
BLAST
Identify Sequence Similarities

Sequence Retrieval
Retrieve Specific Sequences using IDs and coordinates

PubMed and Entrez
View the Latest Pubmed and Entrez Results

9.3 Downloading large data files such as all coding sequences or all protein sequences for an entire genome.

Download files are available in the file download section of all EuPathDB sites
Hint: select “Data Files” under the “Download” menu in the grey tool bar.



Hint: navigate through the subfolders and find the files containing codon usage information for *P. capsici*. Folders without a strain designation contain species level data.

Name	Last modified	Size	Description
Parent Directory	-	-	-
Current_Release/	16-Apr-2013 17:44	-	-
release-2.0/	02-Nov-2012 19:37	-	-
release-2.1/	02-Nov-2012 16:46	-	-
release-2.3/			
Ndiscreta_FGSC_8579/	14-May-2013 14:06	-	-
Nfischeri/	14-Mar-2013 17:18	-	-
Nfischeri_NRI/			
Ntetrasperma/			
Ntetrasperma/			
Pcapsici/			
Pcapsici_LT1534/			
Pchrysochloris/			
Pchrysochloris/			
Paraminis/	14-Mar-2013 17:18	-	-
Parent Directory			
fasta/	14-Mar-2013 17:18	-	-
gff/	14-Mar-2013 17:19	-	-
transcriptExpression/	14-Mar-2013 17:18	-	-
txt/	14-Mar-2013 17:19	-	-
Parent Directory			
data/	08-Mar-2013 23:54	-	-
FungiDB-CURRENT_Phyca_LT1534_CodonUsage.txt	14-Mar-2013 17:19	1.1K	Codon usage table
FungiDB-CURRENT_Phyca_LT1534_InterproDomains.txt	14-Mar-2013 17:19	3.0M	Interpro features,