

Gene Ontology Enrichment

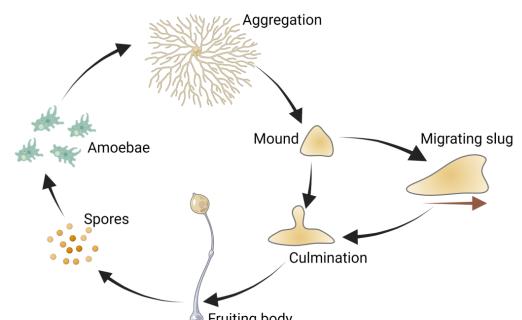
Learning objectives

- Identify a biologically meaningful set of regulated genes
- Analyze results using Gene Ontology (GO) enrichment
- Practice using tools to summarize big lists of GO terms
- Use Revigo to simplify and visualize GO results

Introduction



Dictyostelium (slime mold) is a model system for developmental biology. Its life cycle transitions from free-living single cells to a multicellular structure. Upon starvation, single cells (amoebae) begin to cooperate in a developmental process that leads to multicellular fruiting body



formation 24 hours later. These striking morphological changes are driven by large shifts in gene expression as different biological processes turn on and off.

In this exercise, we will focus on the **genes that become upregulated in the later stages of development**. By analyzing these genes, we can ask questions such as: What biological processes become active as *Dictyostelium* builds a multicellular structure?

To answer this, you will learn how to explore patterns in gene expression, use Gene Ontology (GO) enrichment to identify major biological themes, and visualize the results through tools like word clouds and Revigo.

1. Go to [AmoebaDB.org](#)¹ and search for RNA-Seq experiments.
2. In the list of datasets, look for the “**D. discoideum developmental program transcriptome**”.
3. Choose the Differential Expression (DE) search for this dataset.

¹ Note that this exercise uses AmoebaDB.org as an example database, but the same functionality is available on all VEuPathDB resources.

The screenshot shows the AmoebaDB search interface. On the left, there's a search bar with "rna" typed in, and a sidebar with sections for "Search for..." (Genes, Gene models, Transcriptomics) and "rna" (Gene Model Characteristics, Unannotated Intron Junctions, Microarray Evidence, RNA-Seq Evidence). The main area displays search results for "RNA-Seq Evidence". A legend at the top indicates filters for DE (Differential Expression), FC (Fold Change), P (Percentile), and SA (SenseAntisense). A search bar with "19 rows" is shown. The results table has columns for Organism, Data Set, and a "Choose a Search" section with buttons for P, DE, FC, P, and DE, FC, P, SA. The results list includes entries for *Acanthamoeba castellanii*, *Dictyostelium discoideum*, and *Dictyostelium purpureum*, each with a brief description and a "NEW" badge.

Organism	Data Set	Choose a Search
<i>Acanthamoeba castellanii</i> str. Neff	② Trophozoite transcriptome of <i>A. castellanii</i> (Wojtkawska et al.)	P
<i>Acanthamoeba castellanii</i> str. Neff	② Transcriptome during encystment of <i>A. castellanii</i> Neff (Bernard C. et al. 2022)	DE, FC, P
<i>Dictyostelium discoideum</i> AX4	② <i>D. discoideum</i> developmental program transcriptome (Parikh et al. 2010) NEW	DE, FC, P
<i>Dictyostelium discoideum</i> AX4	② Intracellular <i>M. marinum</i> infection of <i>D. discoideum</i> (Kjellin et al. 2019)	DE, FC, P
<i>Dictyostelium discoideum</i> AX4	② Intracellular <i>L. pneumophila</i> infection of <i>D. discoideum</i> (Kjellin et al. 2019)	DE, FC, P, SA
<i>Dictyostelium purpureum</i> QSDP1	② <i>D. purpureum</i> developmental program transcriptome (Parikh et al. 2010) NEW	DE, FC, P

4. This opens the Differential Expression search page which has data on the RNA sequence transcriptome of *Dictyostelium discoideum* during a 24-hour developmental program. *Hint: this information can be found in the “View Data Sets Used tab”*
 5. Run a search to find genes that were upregulated in late stages, i.e., at 24 hr (comparator sample) compared to 16 hr (reference sample). **How many genes do you find?**

Identify Genes based on *D. discoideum* AX4 *D. discoideum* developmental program transcriptome RNA-Seq (Differential Expression)

Configure Search [Learn More](#) [View Data Sets Used](#)

Experiment

D. discoideum developmental time course transcriptome unstranded

Reference Sample

0 hr
 4 hr
 8 hr
 12 hr
 16 hr
 20 hr
 24 hr
 NC4 prespore
 NC4 prestalk

Comparator Sample

0 hr
 4 hr
 8 hr
 12 hr
 16 hr
 20 hr
 24 hr
 NC4 prespore
 NC4 prestalk

Direction

fold difference >=

adjusted P value less than or equal to

6. Analyze the results for GO Enrichment

Analyze your Gene results with a tool below.

Gene Ontology Enrichment

Find Gene Ontology terms that are enriched in your gene result. [Read More](#)

Parameters

- Organism: Dictyostelium discoideum AX4
- Ontology: Biological Process
- Evidence: Computed, Curated
- Limit to GO Slim terms: No
- P-Value cutoff: 0.05 (0 - 1)

Submit

Choose the **Biological Process** ontology
Use default settings for the GO parameters

7. What functions are enriched in the gene set?

Analysis Results:

139 rows

		Genes in the	Genes in	Percent				
GO ID	GO Term	Genes in the	Genes in	Percent				
GO:0030198	extracellular matrix organization						9.67e-13	1.93e-12
GO:0043062	extracellular structure organization						9.67e-13	1.93e-12
GO:0045229	external encapsulating structure organization						7.39e-12	2.22e-11
GO:0072330	monocarboxylic acid biosynthetic process						1.69e-3	6.75e-3
GO:0006633	fatty acid biosynthetic process						4.08e-3	2.52e-2
GO:0046394	carboxylic acid biosynthetic process						4.08e-3	2.86e-2
GO:0016053	organic acid biosynthetic process	141	39	27.7	1.98	2.39	1.44e-5	4.08e-3
GO:0042060	wound healing	24	12	50.0	3.57	6.19	2.85e-5	6.29e-3
GO:0001778	plasma membrane repair	24	12	50.0	3.57	6.19	2.85e-5	6.29e-3
GO:0043436	oxoacid metabolic	390	83	21.3	1.52	1.71	4.04e-5	6.92e-3
								8.03e-2

8. Visualize related functions in Revigo. What are the major biological processes at this stage of *Dictyostelium* development?



9. What happens to the results if you increase the stringency of DE search to 4-fold upregulation and p-value of 0.01? How does this affect the GO enrichment?