

Mining Proteomics Data

Protein expression in *P. falciparum* apicoplast vs. ER

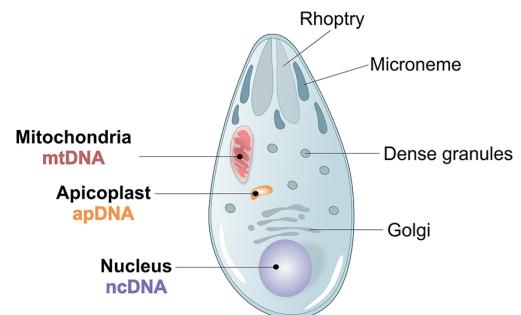
Learning objectives

- Explore proteomics data on VEuPathDB
- Perform a quantitative mass spec. evidence search

Introduction

The malaria parasite *Plasmodium falciparum* and other *Plasmodium* species contain a non-photosynthetic plastid organelle called the apicoplast that is crucial to the malaria parasite's survival. Due to the algal origin of the apicoplast (which contains its own DNA), many proteins and pathways are not shared by the human host, making it an attractive target for antimalarial drugs.

To identify genes that are unique to the apicoplast and not present in human hosts, we could ask the question: **What genes in *Plasmodium falciparum* have protein products that are present at a higher concentration in the apicoplast compared to the endoplasmic reticulum (ER)?** Furthermore, we can leverage protein subcellular localization data to limit the results to the apicoplast.



Overview of the search strategy: Use **Quantitative Mass Spec. Evidence** search in two steps-

- (1) identify genes that are upregulated in apicoplast vs. ER
- (2) intersect that search with an apicoplast localization search

1. Navigate to [PlasmoDB.org](https://plasmodb.org)

2. From the sidebar or header, search or scroll for “quantitative” and click on **Quantitative Mass Spec. Evidence**

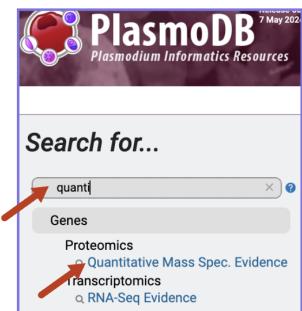
3. Run a search to find genes that are upregulated in the apicoplast sample compared to the ER sample.

a. Select the experiment called **Apicoplast and ER Proteomes**

(Quantitative)(Dd2) (Boucher et al)- it compares the samples we are interested in.

b. Choose the fold change (FC) search.

c. Configure this search to return all genes that are upregulated by 1.5 fold in the apicoplast sample compared to the ER sample.



How many genes did you get that have upregulated protein expression in apicoplast compared to ER?

Identify Genes based on Quantitative Mass Spec. Evidence

Legend: DCC Direct Confidence Comparison FC Fold Change

Filter Data Sets: 5 rows

Organism	Data Set	Choose a Search
Plasmodium berghei ANKA	Proteome of ApIP2 double vs single knockout (Modrzynska et al.)	DCC
Plasmodium falciparum 3D7	Long-lived merozoite proteome (Kumar et al.)	FC
Plasmodium falciparum 3D7	Proteome and phosphoproteome during intraerythrocytic development (Quantitative) (Pease et al.)	FC
Plasmodium falciparum 3D7	PfGRK4 regulated proteome at 29 and 37 hpi (quantitative) (Ganter et al.)	DCC
Plasmodium falciparum 3D7	Apicoplast and ER Proteomes (Quantitative)(Dd2) (Boucher et al.)	FC

Fold Change

Identify Genes based on P. falciparum 3D7 Apicoplast and ER Proteomes (Quantitative)(Dd2) Proteomics (fold change)

Configure Search Learn More View Data Sets Used

Reset values to default

For the Experiment: Apicoplast and ER Proteomes (Quantitative)(Dd2)

return protein coding Genes that are up-regulated with a Fold change >= 1.5 between each gene average expression value in the following Reference Samples: Apicoplast, ER

and its average expression value in the following Comparison Samples: Apicoplast, ER

Example showing one gene that would meet search criteria (dots represent this gene's expression values for selected samples)

Up-regulated

Expression Value Comparison

Expression Value Reference

Reference Comparison Samples Samples

For each gene, the search calculates:

$$\text{fold change} = \frac{\text{comparison expression value}}{\text{reference expression value}}$$

and returns genes when fold change >= 1.5

You are searching for genes that are up-regulated between one reference sample and one comparison sample.

Get Answer

Upregulated in apicoplast vs. ER *

PfDd2 Apico ER Prot (fc)
272 Genes

Step 1

+ Add a step

4. Limit the results by leveraging available subcellular localization data. PlasmoDB has a data set that returns genes with the transit peptides that mediate protein targeting to the apicoplast.
- Click on the **add step** button and find the **subcellular localization** search
 - Make sure Apicoplast localization is selected and click on the Run Step button

The screenshot shows the PlasmoDB search interface with three main panels:

- Step 1:** Shows a result for "PfDd2 Apico ER Prot (fc)" with 272 Genes. An arrow points from this panel to the "Add a step" button in the top right corner of the same panel.
- Step 2:** A modal window titled "Add a step to your search strategy". It contains three options: "Combine with other Genes", "Transform into related records", and "Use Genomic Colocation to combine with other features". The "Combine with other Genes" option is selected. A red arrow points to the "Choose how to combine with other Genes" section, which includes radio buttons for "1 INTERSECT 2", "1 UNION 2", "1 MINUS 2", and "2 MINUS 1".
- Search Panel:** On the right, a sidebar shows a list of categories under "Filter the searches below...". A red arrow points to the "Localization" section, where "Apicoplast" is selected from a dropdown menu. Below this is a "Run Step" button, which is circled in red.

5. How many genes did you identify? Explore your results.

Are you more confident that these genes are apicoplast genes? How would you use the PlasmoDB tools to boost your confidence in these so-called apicoplast genes?

The screenshot shows the search results page with the following details:

- Step 1:** Shows the initial result for "PfDd2 Apico ER Prot (fc)" with 272 Genes.
- Step 2:** Shows the result after adding the subcellular localization step, resulting in 139 Genes. This result is highlighted in yellow.
- Add a step:** A button in the bottom right corner of the search results panel.