Finding Ortholog Groups using Phyletic Patterns in OrthoMCL

Learning objectives
- Understand the utility of a phyletic patterns search
- Learn to do a phyletic pattern search on OrthoMCL

Introduction

Phyletic Patterns are a flexible and powerful way to identify ortholog groups with a certain **pattern of conservation** across taxonomic groups. Finding ortholog groups with specific phyletic patterns is useful because it lets you answer important biological and evolutionary questions by looking at **which organisms have (or lack) a gene or protein family**. For instance:

- Pathogen-specific genes (present in pathogens, absent in hosts) can be identified via phyletic patterns, helping to find drug or vaccine targets.
- Phyletic patterns can guide genome annotation by highlighting likely functional orthologs and genes with incorrect or missing annotations.
- Phyletic patterns can identify genes linked to specific biological functions or traits. If a gene is present in organisms showing a certain trait (e.g., photosynthesis, motility, pathogenicity) and absent in those without it, that pattern can highlight candidate genes responsible for the trait and molecular mechanisms underlying the phenotype.

1. Go to the **Phyletic Pattern** search in OrthoMCL.org[1]

---
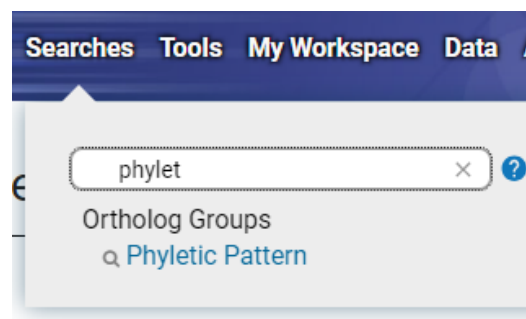
[1] VEuPathDB is updated regularly, so the numbers you see while completing the exercise may differ slightly from those shown in the screenshots.

2. **Find ortholog groups that are Eukaryota specific.** Click selection icons in the taxonomic tree to include groups (✓) with proteins from Eukaryota but exclude (x) proteins from Archaea and Bacteria. Note that the yellow check mark requires at least one member of a group, while the green check requires ALL members.

(Notice that using the tree creates an expression in the text box above. It's also possible to ignore the tree and write an expression for your phyletic pattern. See more information in the Learn More tab.)

3. Find all groups that contain orthologs from at least one species of Ascomycota fungi (1T) but not from bacteria, archaea, or metazoan (0T).

4. **Interpret your results.** Examine the results table. What information do the columns provide? Some of the columns provide data for representation of the orthogroup in different taxonomic groups. Sort the result table by the Alveolata column (descending) and hover over the Alveolata cell in the first row. Can you tell the distribution of Alveolata in the group?

5. Revise your search to find groups that: do not contain orthologs from Alveolates, Amoebozoa, Archaea, Bacteria and Ascomycota but contain at least one protein from *Mucor circinelloides f. circinelloides* 1006PhL (mcic) OR *Mucor circinelloides f. lusitanicus* CBS 277.49 (mcir).

If you are getting frustrated trying to figure this one out, you have a right to be! If your results look different, hover over the search step and click to revise the parameter search. OrthoMCL also allows you to enter a text expression to precisely define the phyletic pattern. Try to figure out what expression to use before looking at the next page. (Hint: start by assigning the "do not contain" parameter (x) using check boxes to Alveolates, Amebozoa, Archaea, Bacteria and Ascomycetes. Next, use the expression window to add "AND" followed by specific criteria for Mucor spp. Use the *Learn More* tab for more information on phyletic pattern expressions.

ALVE=0T AND AMOE=0T AND ASCO=0T AND ARCH=0T AND BACT=0T AND mcic+mcir=1T

https://orthomcl.org/orthomcl/app/workspace/strategies/import/c1883ab75f86053d

Useful information

All VEuPathDB genomics sites (e.g., FungiDB) have an integrated phyletic pattern search that uses OrthoMCL to return lists of genes. For example, you use the "**Orthology Phylogenetic Profile**" search to identify genes in your organism of interest that are restricted in their profile. For example, you frequently want to identify genes that are conserved among organisms in your genus of interest but not present in the host as these genes may make good drug targets or vaccine candidates.

OrthoMCL-7 has a new feature that directly identifies orthologs between two species with the **Shared Orthologs** search. This is particularly useful when an experiment has been done on one species, but the results must be extrapolated to another species; or the gene annotations of a less studied organism can be inferred from a model species. For example, *Hepatocystis sp.* has very few functional gene annotations, however most of its genes have orthologs in *Plasmodium falciparum* 3D7, which has extensive annotation. Use the Shared Ortholog From List search to find putative functions for genes hpil|HEP_00180900, hpil|HEP_00424400, and hpil|HEP_00057900.

The Shared Orthologs by Organism search allows you to find ALL orthologs between two species.

Identify Proteins based on Shared Orthologs By Organisms

Accession	Group ID	Description	Target ID
hpil|HEP_00189600	OG7_0001577	Cullin-1, putative [Source:UniProtKB/TrEMBL;Acc:A0A653H1F5]	pvpc|YYG_00288,
hpil|HEP_00398700	OG7_0001581	Zinc finger protein, putative [Source:UniProtKB/TrEMBL;Acc:A0A653HNS5]	pvpc|YYG_03170,
hpil|HEP_00459500	OG7_0001581	FHA domain protein, putative [Source:UniProtKB/TrEMBL;Acc:A0A653H936]	pvpc|YYG_03170,
hpil|HEP_00493000	OG7_0001581	FHA domain protein, putative [Source:UniProtKB/TrEMBL;Acc:A0A653HA98]	pvpc|YYG_03170,
hpil|HEP_00406400	OG7_0001605	Atypical protein kinase, ABC-1 family, putative [Source:UniProtKB/TrEMBL;Acc:A0A653HNQ1]	pvpc|YYG_00827,
hpil|HEP_00507600	OG7_0001605	Protein kinase, putative [Source:UniProtKB/TrEMBL;Acc:A0A653HRQ9]	pvpc|YYG_00827,
hpil|HEP_00180800	OG7_0001608	G domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A653H180]	pvpc|YYG_00908,
hpil|HEP_00217600	OG7_0001608	GN3L_Grn1 domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A653H298]	pvpc|YYG_00908,
hpil|HEP_00219100	OG7_0001608	G domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A653HIB3]	pvpc|YYG_00908,
hpil|HEP_00365300	OG7_0001608	G domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A653H6F9]	pvpc|YYG_00908,
hpil|HEP_00382300	OG7_0001608	Nucleolar GTP-binding protein 2 [Source:UniProtKB/TrEMBL;Acc:A0A653HN10]	pvpc|YYG_00908,
hpil|HEP_00239300	OG7_0001609	GOLD domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A653H2W2]	pvpc|YYG_03075,