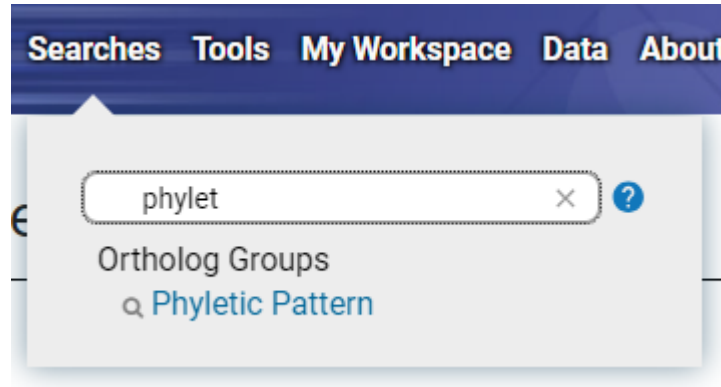# Exploring Phyletic Patterns in OrthoMCL

1. **Find ortholog groups with specific phyletic patterns.** Phyletic Pattern Expression is a flexible and powerful way to identify ortholog groups with a certain conservation pattern. The pattern is used to identify groups based on whether proteins from specific taxa are present or absent.

a. Go to the Phyletic Pattern search in OrthoMCL.org

b. Find ortholog groups that are Eukaryota specific. Arrange the taxonomic tree to include groups with proteins from Eukaryota but exclude proteins from Archaea and Bacteria. (Notice that using the tree creates an expression 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.)

Description

Find Ortholog Groups by phyletic pattern.

Phyletic Pattern Expression is a flexible and powerful way to identify ortholog groups with a certain conserva proteins from specific taxa are present or absent. Also, the pattern finds groups with a certain copy number (

Examples

These expressions find ortholog groups in which...

hsap>=5	there are five or more human sequences
hsap+ecol=2T	both human and E. coli are present.
hsap+ecol=1T	only one species of human or E. coli is present.

Expression: ARCH=0T AND BACT=0T

[Get Answer]

Key: ● = no constraints | ✔ (green) = must be in group | ✔ (yellow) = at least one subtaxon must be in group | ✖ = must not be in group | ✱ = mixture of constraints

expand all | collapse all

Type a taxonomic name

- ✱ **Root** (ALL)
  - ● **Eukaryota** (EUKA)
    - ● **Alveolates** (ALVE)
    - ● **Amoebozoa** (AMOE)
    - ● **Euglenozoa** (EUGL)
    - ● **Fungi** (FUNG)
    - ● **Metazoa** (META)
    - ● **Other Eukaryota** (OEUK)
    - ● **Viridiplantae** (VIRI)
  - → ✖ **Archaea** (ARCH)
    - ✖ Nitrosopumilus maritimus (strain SCM1) (nmar)
    - ✖ **Crenarchaeota** (CREN)
    - ✖ **Euryarchaeota** (EURY)
    - ✖ **Korarchaeota** (KORA)
    - ✖ **Nanoarchaeota** (NANO)
  - → ✖ **Bacteria** (BACT)
    - ✖ **Firmicutes** (FIRM)
    - ✖ **Other Bacteria** (OBAC)
    - ✖ **Proteobacteria** (PROT)

Phyletic
928,122 Ortholog Groups
Step 1

[+ Add a step]

Get Answer

2

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

d. Interpret your results. 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?

Ortholog Group	Total Number Proteins ⊗	Archaea ⊗	Bacteria ⊗	Alveolata ⊗	Amoeba ⊗	Euglenozoa ⊗
OG6_111091	243	0 / 27		140 / 141 (99%)	0 / 16 (0%)	0 / 74 (0%)
OG6_119698	146	0 / 27		138 / 141 (98%)	0 / 16 (0%)	0 / 74 (0%)
OG6_105220	677	0 / 27 (0%)	0 / 47 (0%)	132 / 141 (94%)	16 / 16 (100%)	73 / 74 (99%)
OG6_122551	154	0 / 27 (0%)	0 / 47 (0%)	114 / 141 (81%)	2 / 16 (13%)	0 / 74 (0%)
OG6_108921	420	0 / 27 (0%)	0 / 47 (0%)	84 / 141 (60%)	13 / 16 (81%)	0 / 74 (0%)
OG6_106907	492	0 / 27 (0%)	0 / 47 (0%)	136 / 141 (96%)	0 / 16 (0%)	0 / 74 (0%)
OG6_119330	171					

Tooltip over Alveolata column:

ALVEOLATA
Ciliates: 2 / 2
Apicomplexa
Haemosporida: 63 / 63
Coccidia: 50 / 51
Piroplasmida: 18 / 18
Other apicomplexa: 4 / 4
Other alveolata: 3 / 3

e. 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 to 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.

Expression: `ALVE=0T AND AMOE=0T AND ASCO=0T AND ARCH=0T AND BACT=0T`

Get Answer

Key: ● = no constraints | ✔ = must be in group | ✔ = at least one subtaxon must be in group | ✖ = must not be in group | ✱ = mixture of constraints

expand all | collapse all

Type a taxonomic name

▼ ✱ Root (ALL)
  ▼ ✱ Eukaryota (EUKA)
    ▶ ✖ Alveolates (ALVE)
    ▶ ✖ Amoebozoa (AMOE)
    ▶ ● Euglenozoa (EUGL)
    ▼ ✱ Fungi (FUNG)
      ● Allomyces macrogynus ATCC 38327 (amac)
      ● Catenaria anguillulae PL171 (cang)
      ● Conidiobolus coronatus (strain ATCC 28846 / CBS 209.66 / NRRL 28638
      ● Rozella allomycis (strain CSF55) (rall)
      ▶ ✖ Ascomycota (ASCO)
      ▶ ● Basidiomycota (BASI)
      ▶ ● Chytridiomycota (CHYT)
      ▶ ● Microsporidia (MICR)
      ▼ ● Mucoromycota (MUCO)
        ● Lichtheimia corymbifera JMRC:FSU:9682 (lcor)
        ● Mucor circinelloides 1006PhL (mcic)
        ● Mucor lusitanicus CBS 277.49 (mcir)
        ● Phycomyces blakesleeanus NRRL 1555(-) (pbla)
        ● Rhizophagus irregularis A1 (DAOM-664342) (rira)
        ● Rhizophagus irregularis C2 (rirc)
        ● Rhizophagus irregularis DAOM 181602=DAOM 197198 (rhiz)
        ● Rhizophagus irregularis DAOM 181602=DAOM 197198 (old build 201
        ● Rhizopus delemar RA 99-880 (rdel)
        ● Rhizopus delemar RA 99-880 (old build 2015-03-23) (rdel-old)
        ● Rhizopus microsporus var. microsporus ATCC 52814 (rmma)
    ▶ ● Metazoa (META)
    ▶ ● Other Eukaryota (OEUK)
    ▶ ● Viridiplantae (VIRI)
  ▶ ✖ Archaea (ARCH)
  ▶ ✖ Bacteria (BACT)

> This is as far as you can get using the tree since the button beside mcic and mcir do not accommodate the yellow checks. You must alter the Expression with specific instructions for mcic and mcir.
> mcic=1T AND mcir=1T

Phyletic
1,564 Ortholog Groups

+ Add a step

Step 1

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.

FungiDB
Release 62
9 Mar 2023
Fungal & Oomycete Informatics Resources

Search for...
phyl

Genes
Orthology and synteny
  Orthology Phylogenetic Profile
  Find genes that have an orthology-based phylogenetic profile that you specify.

5

OrthoMCL-7 (currently a beta release) has a new feature that directly identifies orthologs between two species with the ***Shared Orthologs by Organism*** 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 this search to find putative functions for genes hpil|HEP_00180900, hpil|HEP_00424400, and hpil|HEP_00057900.

Identify Proteins based on Shared Orthologs By Organisms

| Configure Search | Learn More |

🔄 Reset values to default

❓ **Query Organism**

| Hepatocystis sp. ex Piliocolobus tephrosceles |

❓ **Target Organism**

| Plasmodium falciparum 3D7 |

[Get Answer]

	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,