

Finding Ortholog Groups using Phyletic Patterns in OrthoMCL

Learning objectives

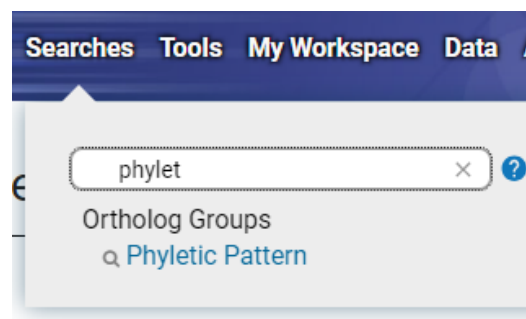
- Understand the utility of a phyletic patterns search
- Learn to do a phyletic pattern search on OrthoMCL

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](https://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.

- Find ortholog groups that are Eukaryota specific. Click selection icons in the taxonomic tree to include groups (✓) with proteins from Eukaryota but exclude (✗) 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.)

Expression: 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 Q ?

- 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

Configure Search Learn More

Description

Find Ortholog Groups by phyletic pattern.

Phyletic Pattern Expression is a flexible and powerful way to identify ortholog groups with a certain conservation pattern. Proteins from specific taxa are present or absent. Also, the pattern finds groups with a certain copy number (e.g., 1 or 2).

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.

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

Expression: 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) (Delacroixia coronata) (ccor)
 - Rozella allomycis (strain CSF55) (ra11)
 - Ascomycota (ASCO) ✓
 - Basidiomycota (BASI)
 - Chytridiomycota (CHYT)
 - Microsporidia (MICR)
 - Mucoromycota (MUCO)
 - 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
197,316 Ortholog Groups
Step 1 + Add a step

Get Answer

- 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?

Phyletic
197,316 Ortholog Groups
Step 1 + Add a step

197,316 Ortholog Groups Revise this search

Ortholog Group Results

1 2 3 ... 9,866 Rows per page: 20

Ortholog Group	Total Number Proteins	Archaea	Bacteria	Alveolata	Amoeba	Euglenozoa
OG6_111091	243	0 / 23 (0%)	0 / 47 (0%)	140 / 141 (99%)	0 / 16 (0%)	0 / 74 (0%)
OG6_119698	146	0 / 23 (0%)	0 / 47 (0%)	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_110220	171	0 / 27 (0%)	0 / 47 (0%)	136 / 141 (96%)	0 / 16 (0%)	0 / 74 (0%)

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

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 (✖) 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:

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)**
 - ✔ **Allozymetes macrocytus** ATCC 98327 (amac)
 - ✔ **Catenaria anguillulae** PL173 (cang)
 - ✔ **Conidiobolus coronatus** (strain ATCC 28846 / CBS 209.66 / NRRL 28638) (rall)
 - ✔ **Rozella alloomys** (strain CF55) (rall)
 - ✖ **Ascomycota (ASCO)**
 - ✔ **Basidiomycota (BASI)**
 - ✔ **Chytridiomycota (CHYT)**
 - ✔ **Microsporidia (MICR)**
 - ▼ **Macromycota (MACO)**
 - ✔ **Lichtheimia corymbifera** JMRCF-FSU/9682 (lcor)
 - ✔ **Mucor circinelloides** 10066P (**mcir**)
 - ✔ **Mucor lusitanicus** CBS 277.49 (**mcir**)
 - ✔ **Phycomyces blakesleeanae** NRRL 1555(-) (pb1a)
 - ✔ **Rhizoglyphus irregularis** A1 (DAXM-66-4342) (rira)
 - ✔ **Rhizoglyphus irregularis** C2 (rirc)
 - ✔ **Rhizoglyphus irregularis** DAXM 181602:DAXM 197198 (rhiz)
 - ✔ **Rhizoglyphus irregularis** DAXM 181602:DAXM 197198 (old build 20)
 - ✔ **Rhizopus delmar** RA 99-880 (rdel)
 - ✔ **Rhizopus delmar** RA 99-880 (old build 2015-03-23) (rdel-old)
 - ✔ **Rhizopus microsporus** var. **microsporus** ATCC 52814 (rma)
 - ✔ **Metazoa (META)**
 - ✔ **Other Eukaryota (OEUK)**
 - ✔ **Viridiplantae (VIRI)**
 - ✔ **Archaea (ARCH)**
 - ✔ **Bacteria (BACT)**

This is as
as you ca
get using
tree sinc
button b
mcir and
do not

This is as far as you can get using the tree since the button beside `mcic` and `mcir` do not

Phyletic
1,564 Ortholog Groups

Step 1

+ Add a step

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]	pvpcl YVG_00288,
hpil HEP_00398700	OG7_0001581	Zinc finger protein, putative [Source:UniProtKB/TrEMBL;Acc:A0A653HNS5]	pvpcl YVG_03170,
hpil HEP_00459500	OG7_0001581	FHA domain protein, putative [Source:UniProtKB/TrEMBL;Acc:A0A653H936]	pvpcl YVG_03170,
hpil HEP_00493000	OG7_0001581	FHA domain protein, putative [Source:UniProtKB/TrEMBL;Acc:A0A653HA98]	pvpcl YVG_03170,
hpil HEP_00406400	OG7_0001605	Atypical protein kinase, ABC-1 family, putative [Source:UniProtKB/TrEMBL;Acc:A0A653HNQ1]	pvpcl YVG_00827,
hpil HEP_00507600	OG7_0001605	Protein kinase, putative [Source:UniProtKB/TrEMBL;Acc:A0A653HRQ9]	pvpcl YVG_00827,
hpil HEP_00180800	OG7_0001608	G domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A653H180]	pvpcl YVG_00908,
hpil HEP_00217600	OG7_0001608	GN3L_Grn1 domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A653H298]	pvpcl YVG_00908,
hpil HEP_00219100	OG7_0001608	G domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A653HIB3]	pvpcl YVG_00908,
hpil HEP_00365300	OG7_0001608	G domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A653H6F9]	pvpcl YVG_00908,
hpil HEP_00382300	OG7_0001608	Nucleolar GTP-binding protein 2 [Source:UniProtKB/TrEMBL;Acc:A0A653HN10]	pvpcl YVG_00908,
hpil HEP_00239300	OG7_0001609	GOLD domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A653H2W2]	pvpcl YVG_03075,