

Orthology and Phyletic Patterns Exercise 9


9.1 Getting to OrthoMCL from FungiDB databases

Note: For this exercise use <http://www.fungidb.org>

- Go to the gene page for the *Phytophthora ramorum* gene with the ID: Phyra_72568.
- What does this gene do? It is annotated as hypothetical!
- Scroll down to the table labeled “Orthologs and Paralogs within FungiDB”. Does this gene have orthologs in other *Oomycete* species? What about other organisms?
Hint: click on the link below the table that takes you to OrthoMCL. (if this link doesn't work look at the address bar, you may need to remove the last three characters: '/>)

Orthologs and Paralogs within FungiDB [Data Sets]

Gene	Organism	Product	is syntenic	has comments
Hyaar_801453	Hyaloperonospora arabidopsidis Emoy2	unspecified product	no	no
Hyaar_801460	Hyaloperonospora arabidopsidis Emoy2	hypothetical protein VOLCADRAFT_82641	no	no
Hyaar_809685	Hyaloperonospora arabidopsidis Emoy2	unspecified product	no	no
Hyaar_809686	Hyaloperonospora arabidopsidis Emoy2	unspecified product	no	no
Phyca_104576	Phytophthora capsici LT1534	Prolyl 4-hydroxylase alpha-2 subunit	no	no
Phyca_504760	Phytophthora capsici LT1534	Prolyl 4-hydroxylase alpha-2 subunit	no	no
Phyca_8777	Phytophthora capsici LT1534	predicted protein	yes	no
PITG_00069	Phytophthora infestans T30-4	conserved hypothetical protein	no	no
PITG_00070	Phytophthora infestans T30-4	conserved hypothetical protein	no	no
PITG_00206	Phytophthora infestans T30-4	conserved hypothetical protein	yes	no
PITG_00207	Phytophthora infestans T30-4	conserved hypothetical protein	yes	no
Phyra_73492	Phytophthora ramorum	Hypothetical protein	no	no
Phyra_73493	Phytophthora ramorum	Hypothetical protein	no	no
Phyra_93622	Phytophthora ramorum	Hypothetical protein	no	no
Phyra_93623	Phytophthora ramorum	Hypothetical protein	no	no
Physo_335951	Phytophthora sojae	E3 ubiquitin-protein ligase LINCR	yes	no
Physo_489528	Phytophthora sojae	Hypothetical protein VOLCADRAFT_80309	no	no
Physo_509415	Phytophthora sojae	prolyl 4-hydroxylase subunit alpha	yes	no
Physo_541336	Phytophthora sojae	predicted protein	no	no
Physo_541338	Phytophthora sojae	unnamed protein product	no	no
Physo_556371	Phytophthora sojae	Hypothetical protein CHLNCRAFT_143796	no	no
Pytul_G003945	Pythium ultimum DAOM BR144	Similar to RNF157: RING finger protein 157 (Homo sapiens);	no	no
Pytul_G003982	Pythium ultimum DAOM BR144	Similar to PH4: Putative HIF-prolyl hydroxylase PH-4 (Homo sapiens);	yes	no
Pytul_G012804	Pythium ultimum DAOM BR144	Similar to Ph4: Putative HIF-prolyl hydroxylase PH-4 (Mus musculus);	no	no


[View the group \(OG5_168148\) containing this gene \(Phyra_72568\) in the OrthoMCL database](#)

- Does this protein have orthologs in other organisms? Does it have any orthologs in bacteria or archaea?
Hint: mouse over the colorful boxes in the tables to reveal the full species and phylum names – see image below.

Group Summary (OG5_128292)

# Sequences	# Taxa	# Match Pairs (%)	Ave E-Value	Ave % Coverage	Ave % Identity
94	89	8742 (200%)	1.35e-14	92.3	36

Links to OG5_128292

- Pfam Domain Architecture**
- Layout Graph
- Multiple Sequence Alignment
- Get Sequences (FASTA)

0 no ortholog;
 1 one ortholog;
 n more than one ortholog;

group details
 phyletic patterns
 pattern labels & counts

bant	cbok	cper	lmon	saur	spna	bmal	bpse	rsol	cbur	ecol	ftul	sent	sife	vcho	yent	ypes	cjei	wsuc	atum	bsui	rpro	rtyp	wend	gsul	aaeo	cpne	ctep	deth	drad	mtub	mlep
rbal	syne	smar	tpal	cmad	nbud	ihos	msed	hmar	smar	ssol	ckor	nequ	aful	halo	hwai	rsmi	mjan	mma	tvol	lbra	linf	lmaj	lmex	tbru	tbrg	tcon	tgru	tviv	ddis	ehis	edis
envv	gthe	atha	osat	ppat	rcom	micr	crei	olau	ycar	tpse	cmer	lthe	pber	pcha	pfal	pkne	pviv	pyoe	bbow	tam	tpar	chom	emur	cpar	nicar	lgon	afum	anid	aory	cgla	clmm
cpso	dhan	egos	grea	klac	mgn	ncra	ptli	scer	calb	spom	yilp	acur	ebie	eint	creo	cneg	lbic	pchr	apis	aate	agan	amel	bmor	cpip	dmei	phum	isca	nvac	tadh	drer	trub
tnig	clnt	clup	hsap	mmu	scab	mdom	mmus	oana	ptro	rnor	ggal	bmae	cbri	cele	sman	glam	glab	glae	mbre	pram	tvag										

List of Protein Sequences

Number	Accession	Taxon	Length	Previous Groups	Description
1	ddis ddx56	Dictyostelium discoideum AX4	686	OG1_2009, OG3_11952, OG4_11690	
2	dmei FBpp0076917	Drosophila melanogaster	560	OG1_2009, OG3_11952, OG4_11690	
3	anid AN6374	Emericella nidulans	611		Uncharacterized ORF; S. cerevisiae ortholog DBP9 has role in ribosomal large subunit assembly, rRNA processing
4	edis EDI_277890	Entamoeba dispar SAW760	552	OG4_11690	ATP-dependent RNA helicase dbp9, putative
5	ehis EHI_078560	Entamoeba histolytica HM-1:IMSS	551	OG1_2009, OG2_72614, OG3_11952, OG4_11690	DEAD/DEAH box helicase, putative
6	envv EIN_037600	Entamoeba invadens IP1	534	OG4_11690	DEAD box ATP-dependent RNA helicase, putative

- e. Take a look at the PFAM domain architectures. Do all the proteins in this group have similar domain architecture?
- f. Based on the orthologs, what do you think this protein might be doing? If you had to give this gene a name, what would you call it?

9.2 Using the phyletic pattern tool in OrthoMCL

Note: For this exercise use <http://www.beta.orthomcl.org>

- a. How many protein groups in OrthoMCL do not have any orthologs in bacteria or archaea?
 Hint: go to “Search for Groups by Evolution...Phyletic Pattern”.



- b. How many protein groups do not contain orthologs from eukaryotes?
 Hint: click on the icon to specify which taxa or species to include or exclude in the profile.

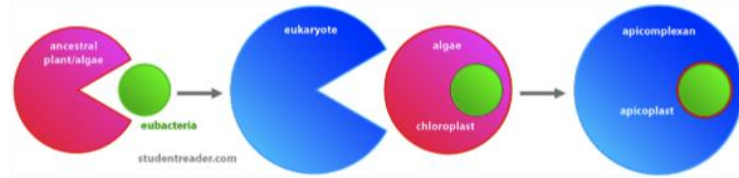
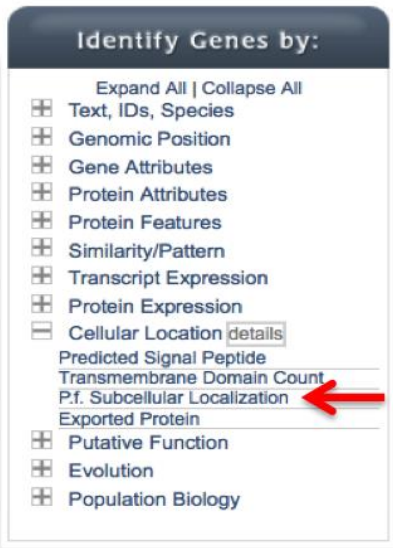
Key: ○ =no constraints | ✓ =must be in group | ✗ =must not be in group | ✓ =at least one subtaxon must be in group | ✳ =mixture of constraints

NOTE: All EuPathDB sites including FungiDB also have a phyletic pattern search that uses OrthoMCL data under Genes -> Evolution -> Orthology Phylogenetic Profile.

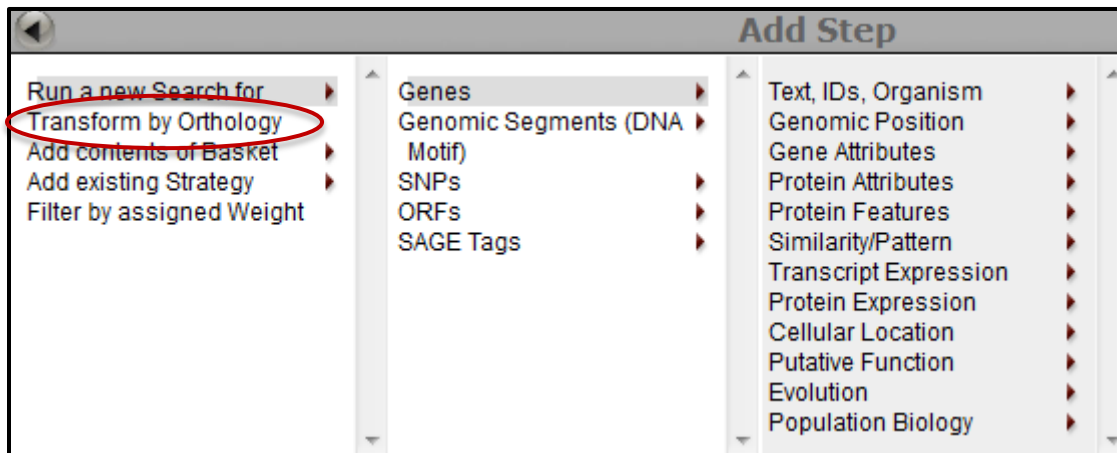
9.3 Using the orthology transform tool to identify apicomplast targeted genes in *Toxoplasma* and *Neospora*.

Note: For this exercise use <http://eupathdb.org>

- a. Start by finding genes in *Plasmodium* that are predicted to target to the apicomplast.
 Hint: click on “Cellular Location” then on “P.f. Subcellular Localization”; see image below.

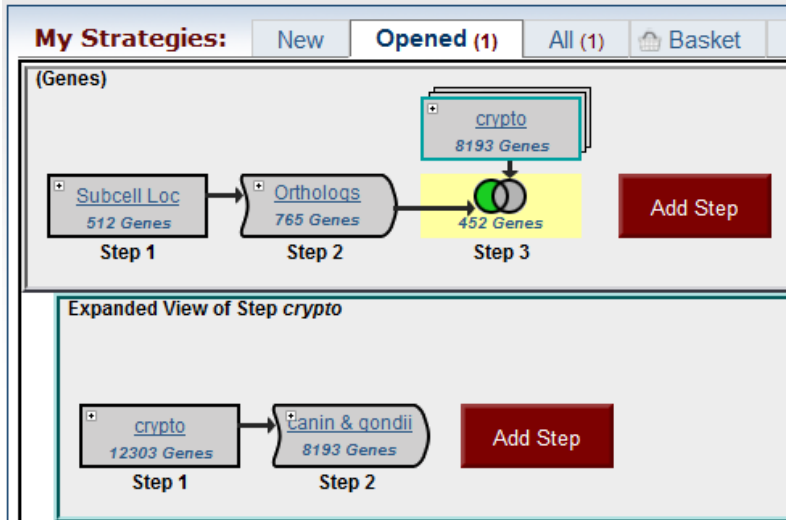


- b. Transform the results of the above search to their *Toxoplasma* orthologs.
 Hint: add a step, then select “Transform by Orthology”. On the search page, select all *Toxoplasma* and *Neospora*.



- c. Although *Cryptosporidium* is an apicomplexan parasite it has actually lost its apicoplast! Can you use this fact to refine your results from the above search?

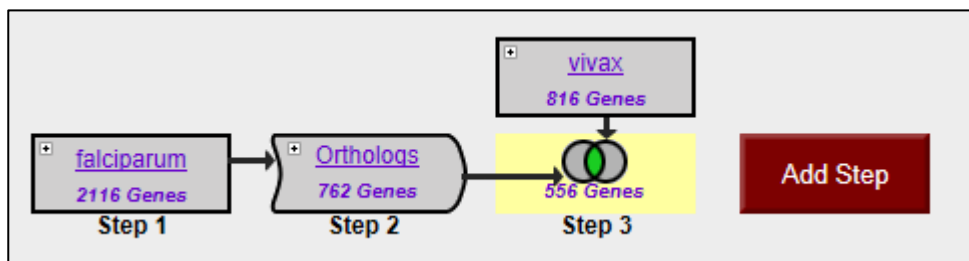
Hint: try subtracting out any orthologs present in *Cryptosporidium*. You will need to use a nested strategy.



9.4(Optional) Use the orthology transform tool to identify as many *P. falciparum* genes containing signal peptides as possible.

Note: For this exercise use <http://www.plasmodb.org>

- How many genes in *P.falciparum* are annotated with signal peptides (use default settings)?
- How many *P. vivax* genes are annotated with signal peptides (use default settings)?
- How many genes on these two lists are in common?
Hint, use the ortholog query to transform between organisms.

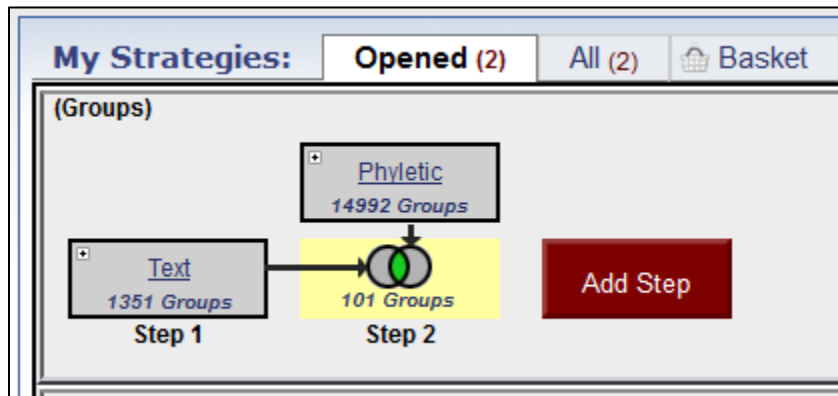


- How many *P. falciparum* orthologs of vivax genes with signal peptides do not themselves contain signal peptides? Why might this be the case? Look at a couple of these using the synteny viewer to generate some hypotheses.
- Using PlasmoDB, generate the most comprehensive list of *P. falciparum* genes that may contain signal peptides. How many did you find?

9.5 (optional) Integrated searches in OrthoMCL

Find all plant proteins that are likely phosphatases that do not have orthologs outside of plants.

- Use the text search to find groups that contain the word “phosphatase”.
- Run a phyletic pattern search for groups that contain any plant protein but do not contain any other organism outside plants.
Hint: make sure everything has a red x on it except for plants (Viridiplantae (VIRI)), which should be a grey circle.



- How many groups did you return? Explore the multiple sequence alignments from some of these groups.
Hint: click on a group ID and open the MSA tab.

