

## Orthology and Phyletic Patterns Exercise 8

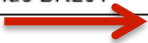
### 8.1 Getting to OrthoMCL from EuPathDB databases

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

- a. Go to the gene page for the *Encephalitozoon cuniculi* gene with the ID: ECU07\_0290.
- b. What does this gene do? It is annotated as hypothetical!
- c. Scroll down to the table labeled “Orthologs and Paralogs within MicrosporidiaDB”. Does this gene have orthologs in other *Encephalitozoon* species? What about other organisms?

Hint: click on the link below the table that takes you to OrthoMCL.

Orthologs and Paralogs within MicrosporidiaDB <a href="#">Hide</a>				
Gene	Organism	Product	is syntenic	has comments
<a href="#">ECU10_1730</a>	Encephalitozoon cuniculi GB-M1	hypothetical protein	no	no
<a href="#">EHEL_070250</a>	Encephalitozoon hellem ATCC 50504	hypothetical protein	yes	no
<a href="#">EHEL_101770</a>	Encephalitozoon hellem ATCC 50504	hypothetical protein	no	no
<a href="#">Ein07_0230</a>	Encephalitozoon intestinalis	hypothetical protein	yes	no
<a href="#">Ein10_1700</a>	Encephalitozoon intestinalis	hypothetical protein	no	no
<a href="#">EBI_25559</a>	Enterocytozoon bieneusi H348	hypothetical protein	yes	no
<a href="#">EBI_27325</a>	Enterocytozoon bieneusi H348	hypothetical protein	no	no
<a href="#">NCER_100684</a>	Nosema ceranae BRL01	hypothetical protein	no	no
<a href="#">NCER_101866</a>	Nosema ceranae BRL01	hypothetical protein	no	no


[Find the group containing ECU07\\_0290 in the OrthoMCL database](#)

- d. 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 pylum names – see image below.

**Group Summary (OG5\_127163)**

# Sequences	# Taxa	# Match Pairs (%)	Ave E-Value	Ave % Coverage	Ave % Identity
139	96	16552 (172.5%)	5.46e-08	88.5	32.5

Links to OG5\_127163

Pfam Domain Architecture

BioLayout Graph

No Multiple Seq. Alignment (too big)

Get Sequences (FASTA)

0 no ortholog;  group details

1 one ortholog;  phyletic patterns

n more than one ortholog;  pattern labels & counts

FIRM  PROT

OBAC  ARCH

EUGL  AMOE

VIRI  ALVE

FUNG  META

OEUK

saar	open	banl	mon	spne	cbol	bmal	tpse	rsol	yent	sent	cbur	vcho	ypes	sflc	flul	ecol	cej	wsuc	rpro	wend	bsui	atum	rtyp	gsul	cpne	mtub
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
drad	deth	ctep	tmar	mlep	syne	rball	tpal	aaeo	nmar	hbut	smar	ssol	msed	ihos	cmag	okor	nequ	halo	tvol	mmar	hwal	mjan	aful	mml	ibra	tbru
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
lmex	tviv	toon	tbrg	lmaj	linf	tcru	elnv	edis	ddis	ehis	gthe	room	atha	osat	micr	ppat	otau	crei	vcar	tpse	cmer	tthe	pviv	pfal	pber	pyoe
1	1	1	1	1	1	1	1	1	1	1	0	1	2	2	1	3	1	1	1	1	1	1	2	2	2	2
pkno	pcha	tpar	fann	bbov	cmur	hcan	cpaf	chom	aory	yilp	spom	psti	ncra	scer	egos	cimm	cpos	calb	mgri	klac	dhan	anid	afum	gzea	cqla	
2	2	1	1	1	2	2	2	2	1	1	1	1	1	1	1	1	1	1	2	0	1	1	1	1	1	1
ecun	sint	ebia	pchr	lbic	cneg	cnso	isca	dmel	aaeg	bmor	amel	cpip	phum	apis	agam	nvec	tadh	drer	trub	tnig	scint	oana	rnor	hsap	mmus	mdom
2	2	2	1	2	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3	3	3	2
mmu	clup	ptro	scab	ggal	cele	bmsa	cbri	sman	mbre	tvag	glae	glab	pram	glam												
4	3	3	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

- 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?

## 8.2 Using the phyletic pattern tool in OrthoMCL

**Note:** For this exercise use <http://www.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 Phyletic Pattern". The search can be accessed from multiple locations in OrthoMCL, see image below.

- 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.

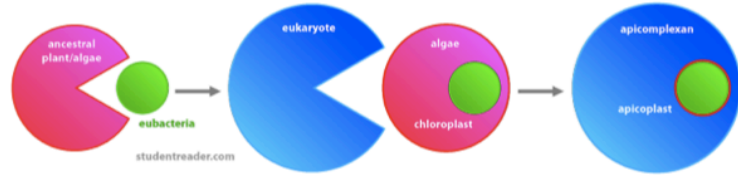
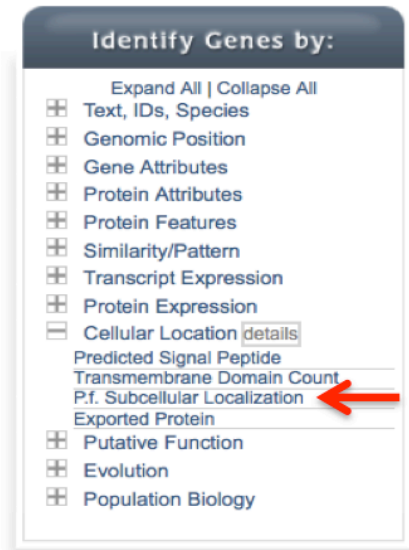
Key: =no constraints | =must be in group | =must not be in group | =at least one subtaxon must be in group | =mixture of constraints

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

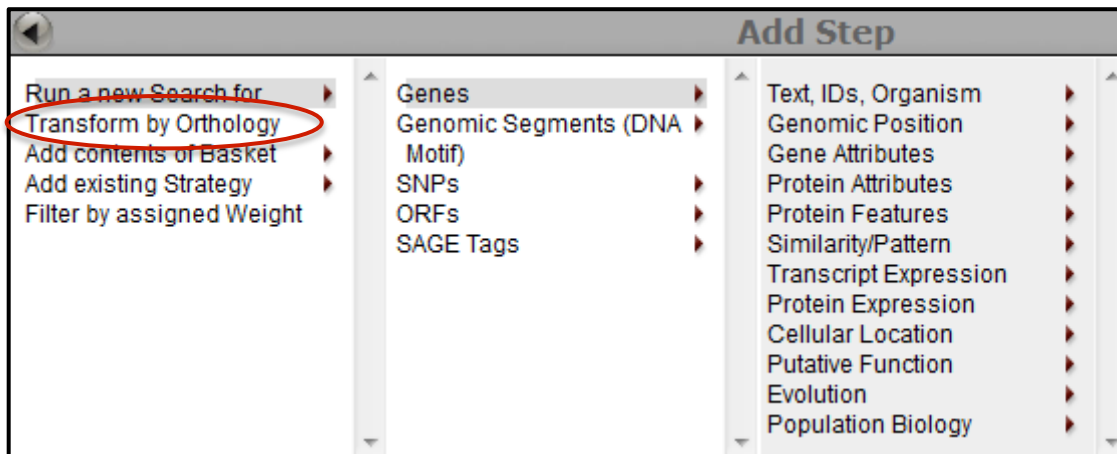
### 8.3 Using the orthology transform tool to identify apicoplast targeted genes in *Toxoplasma* and *Neospora*.

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

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

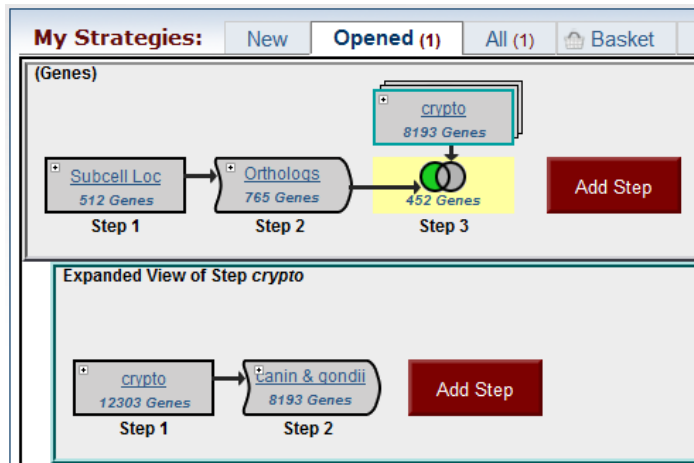


- b. Transform the results of the above search to their *Toxoplasma* orthologs.  
 Hint: add a step, then select “Transform by Orthology”. On the search page, select all *T. gondii* strains and *N. caninum* by holding the “control” key down while you select.



- 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.



**8.4(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.
- Generate the most comprehensive list of *P. falciparum* genes using PlasmoDB that may contain signal peptides. How many did you find?

## 8.4 (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, which should be a grey circle.
- Now you need to combine the above two searches with an intersect operation.  
Hint: from under the search menu item go to “Query History – Groups”, select the two results and click on intersect.
- How many groups did you return? If we have time, explore the multiple sequence alignments from some of these groups.