

Orthology and Phyletic Patterns

1. Getting to OrthoMCL from EuPathDB databases

Note: For this exercise use <http://cryptodb.org> and <http://orthomcl.org>

- Go to the gene page for the *Cryptosporidium parvum* gene with the ID: cgd7_2290
- What does this gene do? It is annotated as a hypothetical protein!
- Scroll down to the table labeled “Orthologs and Paralogs within CryptoDB”. Does this gene have orthologs in other *Cryptosporidium* species? What about other

Orthologs and Paralogs within CryptoDB Hide

Gene	Organism	Product	is syntenic	has comments
Chro.70261	<i>Cryptosporidium hominis</i> TU502	hypothetical protein	yes	no
CMU_034340	<i>Cryptosporidium muris</i> RN66	hypothetical protein, conserved	yes	no

View the group (OG5_127679) containing this gene (cgd7_2290) in the OrthoMCL database



organisms? (hint: click on the link below the table that takes you to OrthoMCL).

- 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 table to reveal the full species and phylum names – see image below).

Group: OG5_127679
(110 sequences)

Add to Basket Add to Favorites

Sequences & Statistics | Pfam domains (graphic) | Pfam domains (details) | MSA | Cluster graph

Phyletic Distribution Hide

Legend:

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

show labels

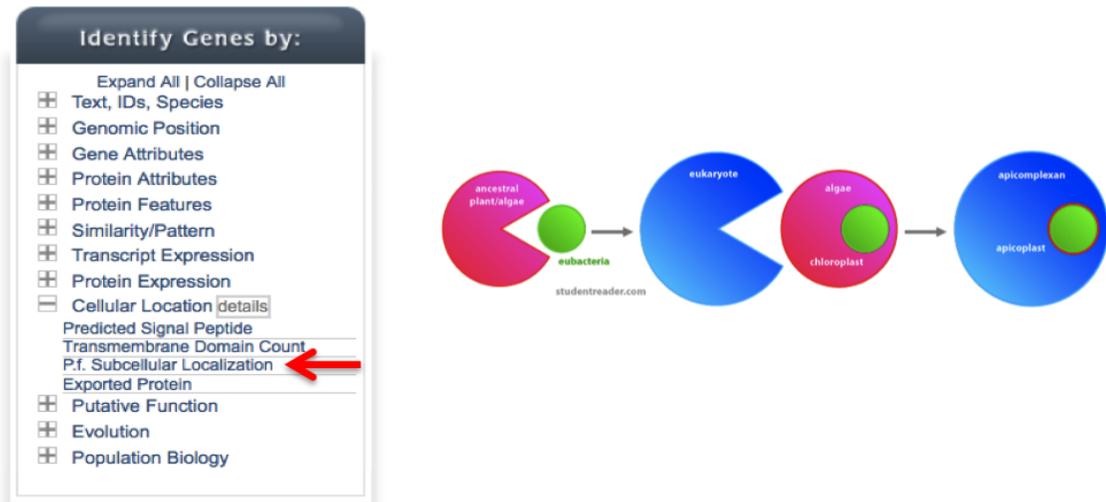
saur	cpar	bant	lmon	spne	cbol	bmali	bpose	rsoi	yent	sent	cbur	vcho	ypes	sfe	ftul	ecol	cje	wsuc	rpro	wend	bsui	atum	rtyp	gsul	spne	mtub	drad	deth	clep	tmar
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mlep	syne	rba1	tpal	aaeg	rmar	hbut	smar	ssol	msed	lhos	cmag	ckor	nequ	halo	tvol	mmar	hwai	mjan	afui	msmi	lbra	lbru	lmex	tviv	lcon	lbrg	lmaj	linf	lcru	eniv
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	2	1	1	1	1	1	1	2	1
edis	odis	ehis	gthe	room	atha	osat	micr	ppat	otau	crei	vcar	tpse	cmcr	lthe	pviv	pfai	pber	pyoe	pkno	pcha	lpar	lann	bbov	cmur	lgon	ncan	cpar	chom	aory	yip
1	1	1	1	2	3	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
spom	psli	ncra	scer	egos	dimm	cpas	calb	mgra	kiac	chan	anid	afum	gzee	cgia	ecun	eint	ebie	pchr	lbic	cneg	cned	isca	dme1	aaeg	bmor	amel	cpip	phum	apis	agam
1	1	1	1	1	1	1	1	2	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
nvec	tach	drer	trub	tniq	cint	oana	rnor	hsap	mmus	mdom	mmul	clup	ptro	ecab	ggal	cele	bmaz	cbri	smar	mbre	tvag	glae	glab	pram	glam					
1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	2	1	3	1	1	1	1	1	1					

- Take a look at the PFAM domain architectures found under the PFM domains (graphic) tab. Do all the proteins in this group have similar domain architecture?
- 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?

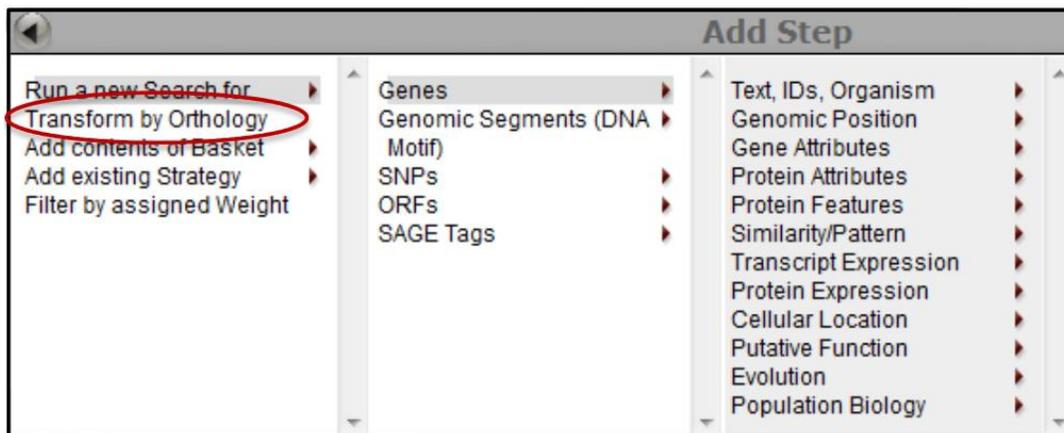
2. Using the orthology transform tool to identify apicoplast targeted genes in *Toxoplasma* and *Neospora*.

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.

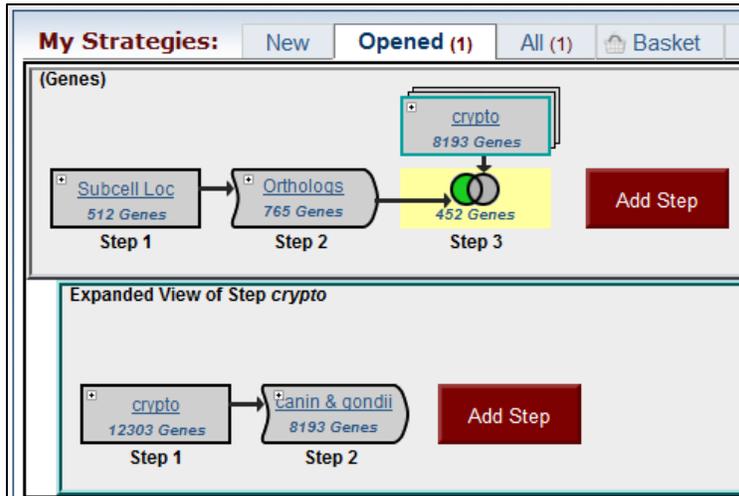


- b. Transform the results of the above search to their *Toxoplasma* orthologs. Hint: add a step, then select “Transform by Orthology”. On the search page, select all *Toxoplasma* and *Neospora*.



- 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.



d. How many of these genes do not have orthologs in mammals (are not conserved in mammals)? (hint: to find this out you will have to add a step and use the “orthology phylogenetic search” under the evolution category).

- To find ortholog groups that do not contain any mammalian proteins, navigate to mammals in the phyletic tree (expand metazoan, then expand chordata). Click on the circle next to “Mammalia” twice (red x).

The screenshot shows the 'Add Step' dialog box in the EuPathDB interface. The 'Orthology Phylogenetic Profile' section is active, displaying a list of organisms. The 'Mammalia (MAMM)' entry is selected with a red 'x' in the 'must not be in group' column. The 'All Organisms' list includes Bacteria (BACT), Firmicutes (FIRM), Proteobacteria (PROT), Other Bacteria (OBAC), Archaea (ARCH), Euryarchaeota (EURY), Crenarchaeota (CREN), Nanoarchaeota (NANO), Korarchaeota (KORA), Eukaryota (EUKA), Alveolates (ALVE), Amoebozoa (AMOE), Euglenozoa (EUGL), Viridiplantae (VIRI), Fungi (FUNG), Metazoa (META), Platyhelminthes (PLAT), Nematodes (NEMA), Arthropoda (ARTH), Chordata (CHOR), Actinopterygii (ACTI), Aves (AVES), Mammalia (MAMM), Tunicates (TUNI), Other Metazoa (OMET), and Other Eukaryota (OEUK).