

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: Psura_72632.
- What does this gene do? It is annotated as unspecified product!
- 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.

HpaG806992	Hyaloperonospora arabidopsidis Emoy2	unspecified product	yes	no
HpaG806993	Hyaloperonospora arabidopsidis Emoy2	unspecified product	yes	no
MGG_09495	Magnaporthe oryzae 70-15	Arp2/3 complex subunit	no	no
MGL_2849	Malassezia globosa CBS 7966	hypothetical protein	no	no
MELLADRAFT_36782	Melampsora larici-populina 98AG31	hypothetical protein	no	no
NFIA_066590	Neosartorya fischeri NRRL 181	Arp2/3 complex subunit Arc16, putative	no	no
NCU03438	Neurospora crassa OR74A	Arp2/3 complex subunit	no	no
NEUDI_86726	Neurospora discreta FGSC 8579	unspecified product	no	no
NEUTE1DRAFT_117555	Neurospora tetrasperma FGSC 2508	hypothetical protein	no	no
PHYCA_507846	Phytophthora capsici LT1534	unspecified product	yes	no
PHYCA_548899	Phytophthora capsici LT1534	unspecified product	yes	no
PHYCA_577884	Phytophthora capsici LT1534	unspecified product	no	no
PHYCI_226786	Phytophthora cinnamomi var. cinnamomi	unspecified product	yes	no
PHYCI_93394	Phytophthora cinnamomi var. cinnamomi	unspecified product	yes	no
PITG_00292	Phytophthora infestans T30-4	hypothetical protein	yes	no
PITG_00293	Phytophthora infestans T30-4	conserved hypothetical protein	yes	no
PPTG_10680	Phytophthora parasitica INRA-310	hypothetical protein	yes	no
PPTG_10681	Phytophthora parasitica INRA-310	hypothetical protein	yes	no
PHYSO_288083	Phytophthora sojae	Actin-related protein Arp2/3 complex, subunit ARPC5	yes	no
PHYSO_547200	Phytophthora sojae	unspecified product	yes	no
PGTG_04942	Puccinia graminis f. sp. tritici CRL 75-36-700-3	hypothetical protein	no	no
PYU1_G004034	Pythium ultimum DAOM BR144	Similar to arcE: Actin-related protein 2/3 complex subunit 5 (Dictyostelium discoideum)	no	no
PYU1_G010688	Pythium ultimum DAOM BR144	Protein of unknown function	no	no
RO3G_04149	Rhizopus delemar RA 99-880	hypothetical protein	no	no
YIL062C	Saccharomyces cerevisiae S288c	Subunit of the ARP2/3 complex, which is required for the motility and integrity of cortical actin patches; has mRNA binding activity (ARC15)	no	no
SJAG_04098	Schizosaccharomyces japonicus yFS275	ARP2/3 actin-organizing complex subunit Arc5	no	no
SOCG_04276	Schizosaccharomyces octosporus yFS286	ARP2/3 actin-organizing complex subunit Arc5	no	no
SPAC1768.04c	Schizosaccharomyces pombe 972h-	ARP2/3 actin-organizing complex subunit Arc5	no	no
SS1G_05971	Sclerotinia sclerotiorum 1980 UF-70	hypothetical protein	no	no
SMAC_09493	Sordaria macrospora k-hell	unspecified product	no	no
SPPG_08730	Spizellomyces punctatus DAOM BR117	hypothetical protein	no	no
sr13589	Sporisorium reilianum SRZ2	related to subunit of the Arp2/3 complex	no	no
PMAA_023800	Talaromyces marneffei ATCC 18224	unspecified product	no	no
TSTA_032220	Talaromyces stipitatus ATCC 10500	Arp2/3 complex subunit Arc16, putative	no	no
TRIREDRAFT_2439	Trichoderma reesei QM6a	Arp2/3 complex protein	no	no
YALU02496g	Yarrowia lipolytica CLIB11zz	unspecified product	no	no

[View the group \(OGS_128798\) containing this gene \(PSURA_72632\) 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 pylum names – see image below.

Group: OG5_128798

(82 sequences)

Add to Basket Add to Favorites

Sequences & Statistics PFam domains (graphic) PFam domains (details) ISA Cluster graph

Phyletic Distribution Hide

Legend: 0 no ortholog 1 one ortholog n more than one ortholog

FRM PROT OBAC ARCH
 EUGL AMOE VIRI ALVE
 FUNG META GEUK

show labels

saur	oper	bant	lmon	spne	cbot	bmal	bpse	rsol	yent	sent	obur	vcho	ypes	sfile	ftul	ecol	cje	wsuc	rpro	wend	bsul	atum	rtyp	gsul	cpne	mtub	drad	deih	steg	lmar	mlep	syne	
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	0	0	0
rbal	tpal	aseo	nmar	fbut	smar	ssol	msed	ihos	cmaq	ckor	nequ	halo	tvol	mmar	hwal	mjan	aful	mami	lbra	tbru	lmex	tviv	toon	tbrg	lma	lnt	trcu	ehv	adis	ddia	eha	gthe	
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	0	0	0
room	atha	osat	mioc	ppat	otau	orei	vcar	tpse	cmel	ltthe	pviv	pfal	pber	pyoe	pkno	pcha	tpar	tann	bbov	cmur	tgon	ncan	cpar	chom	soy	yip	spom	psti	ncra	soer	eges	cimm	
1	2	1	0	1	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	
cpes	calb	mgri	klac	dhan	lanid	afum	gzea	cqla	scun	ieit	ebie	pchr	fbic	cneg	cnco	isca	dml	aaag	bmor	amel	cpip	phum	apis	agam	nvec	tadh	drer	trub	tnig	oint	loana	rnor	
1	2	0	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
haap	mmus	mdom	mmul	clup	ptro	ecab	ggal	cele	bmaa	cbri	smar	mbre	tvag	glae	glab	pram	glam	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	2	2	1	2	2	2	2	2	2	1	1	4	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Group Statistics Hide

Group	Average E-value	# Sequences	Average % Connectivity	% Similar Pairs	Average % Identity	Avg % Match Length	EC Numbers	Keywords	Domains
OG5_128798	2.4E-7	82	50	38	43.1	92.1		complex; arp2/3; arp2/3 complex	ARP2/3 complex 16 kDa subunit (p16-Arc)

EC Number none

List of Sequences Hide

Get Sequences: As Fasta file As new strategy

Filter:

Accession	Taxon	Length	Previous Groups	EC Numbers	Description
aaeg AAEL009059-PA	Aedes aegypti	153	OG2_73438, OG3_12298, OG4_12680	N/A	N/A
afum Afu3g10700	Aspergillus fumigatus Af293	194	OG3_12298, OG4_12680	N/A	Arp2/3 complex
agam AGAP008895-PA	Anopheles gambiae str. PEST	156	OG3_12298, OG4_12680	N/A	N/A

- Take a look at the PFAM domain architectures. 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?

9.2 Using the phyletic pattern tool in OrthoMCL

UNDOABLE: *P. sojae* was removed from ORTHOMCL

Use <http://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?

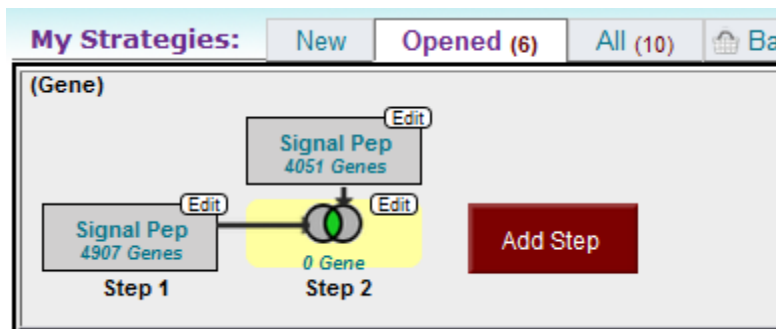
Hint: click on the icon to specify which taxa or species to include or exclude in the profile.

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

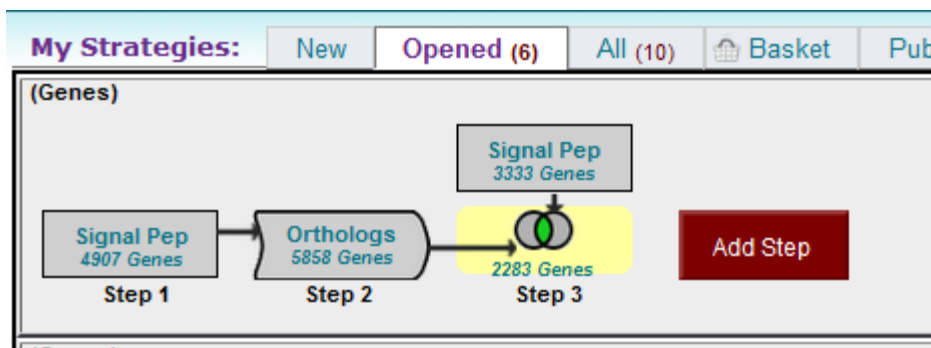
NOTE: All EuPathDB sites including FungiDB also have a phyletic pattern search that uses OrthoMCL data under Genes -> Evolution -> Orthology Phylogenetic Profile.

9.3 Use the orthology transform tool to identify *P. sojae* genes containing signal peptides also found in *P. ramorum*.

- Go to Fungi DB.
- How many *P. sojae* genes are annotated with signal peptides (just use the default settings)?
- Use intersection to see the shared *P. ramorum* signal peptide genes. How did that work?



- We'll have to use a different method. First transform the *P. sojae* results into their *P. ramorum* orthologs, use these in the intersection.



- How many of the *P. ramorum* orthologs of *P. sojae* 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.

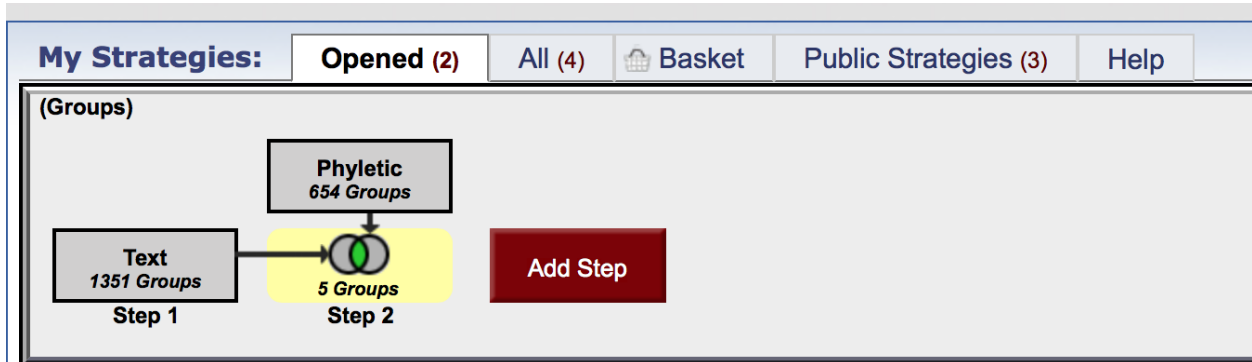
9.5 (optional) Integrated searches in OrthoMCL

Find all oomycete proteins that are likely phosphatases that do not have orthologs outside of oomycetes.

Use OrthoMCL.org

- Use the text search to find groups that contain the word “phosphatase”.
- Run a orthology phylogenetic profile search for groups that contain any oomycete protein but do not contain any other organism outside oomycetes.



Hint: make sure everything has a red x on it except for oomycetes, 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.

Group: OG5_152554
(9 sequences)

Add to Basket  Add to Favorites 

Sequences & Statistics PFam domains (graphic) PFam domains (details) **MSA** Cluster graph

MUSCLE (3.7) multiple sequence alignment

```

pram| fgeneshl_pg.C_scaffold_1400      MATTRTNSTRSDNRTVLFDAEDNVPATTRTNRTASSRSNTASTRSRAATASSRSRSRR
pram| fgeneshl_pg.C_scaffold_4800      -----
pram| fgeneshl_pg.C_scaffold_1400      --MATSSSTAHSGLAPLRHALLSNGSLPEGFVASIAVFLAKTRHLQRDRSRSGASGFEHE
pram| fgeneshl_pg.C_scaffold_1400      -----
pram| fgeneshl_pg.C_scaffold_1400      -----
pram| fgeneshl_pg.C_scaffold_1400      -----
pram| fgeneshl_pg.C_scaffold_1100      -----
pram| fgeneshl_pg.C_scaffold_1300      -----
pram| fgeneshl_pg.C_scaffold_1300      -----

```