FungiDB: Orthology and Phyletic Patterns

Homology Early Globin Gene Gene Duplication Gene Duplication Gene Duplication Gene Duplication B mouse B chick B frog ORTHOLOGS ORTHOLOGS

1. Getting to OrthoMCL from EuPathDB databases

- a. Go to the gene record page for the Cryptococcus gattii gene CGB_L0350W
- **b.** What is the function of this gene? How can you infer its function?
- **c.** Click on the "Orthology and Synteny" link on the left. Does this gene have orthologs in other *Cryptococcus* species? What about other organisms <u>outside of fungi</u>? (hint: click on the *Ortholog Group OG5_127157*; this link will take you to OrthoMCL).

7 Orthology	and synteny				
Ortholog Group	OG5_127157				
 Orthologs and P 	aralogs within EuPa	athDB 🛢 Data sets			
To run Clustal On genes' button	nega, select genes from	the table below. Then choose the	sequence type and initiate the alignment	with the 'Run Clustal Om	ega for selected
Search this tab	le	Q Showing 142 rows	5		
⊥† Clustal Omega ঔ	↓ † Gene	↓† Organism	↓↑ Product	⊥† is syntenic	tî has ♂
0	AGR57_3474	Phanerochaete chrysosporium RP-78	Cation efflux protein	yes	no
	CC1G_14848	Coprinopsis cinerea okayama7#130	hypothetical protein	yes	no
	CKF44_05394	Cryptococcus neoformans var. grubii KN99	cation:cation antiporter	yes	no

d. Mouse over the colorful boxes in the table to reveal the full species and phylum names.

								aro	up. o	as_	121	157											
								(1	39 se	quen	ces)												
								Add to Ba	ahee 🕥	Add to F	avorites												
equences & Statisti	tics PFam	domains (g	graphic)	PFa	am dom	ains (d	details)	MS/	4 C	luster	graph												
Phyletic Distribution	ion Hide																						
Legend: 0	no ortholog			W.	n a N	840	H D O	PROT	N C		521		ARCH										
1	one ortholog					01					8												
n	more than on	e ortholog		H	u a u	NG	= = =	META	= =		JK.	* 56 55											
show labels																							
saur oper bant	Imon spne cb	ot bmai bps	e rsol ye	nt sen 1 1	t cour	o voto	rpes ste 1 1	- Muli	ecol cje	N NSUC	orpro 0	wend b	sui atun 1 1	0	gsul 1	opne o	ntub d	rad d	teth cter	p tmar	mlep 1	o 1	tpal 0
aaeo mmar hbut	o 1	ed hos cma	o ckor ne	ngu hak	D NOL	o h	nwal mja 0 0	n atul n O	1 Dr	a Eru 2	imex	tviv to	ton torg	imaj 1	1	toru 2	einv e	ndis d	ldis ehi	gine o	ncom a	itha osa 2 1	O
ppat otau crei	vcar tose cm	er the put	v pfai pt	Mar Provide		and the second second	toar tare	h bboy a	mur tgo	n ncan	CEDBLE	chom a	ory yip	spom 1	psti	nora	scer e	gos ci	mm cpo	s calb	mgri	dac cha	amid
			0 0	0 0			0 0	0	0 0								_		1 1 1				1
1 0 0 atum gzea ogla 1 1 1 cele bmaa obri 1 2	0 0 1 ecun eint eb 1 1 0 sman mbre hu 3 1	ie pohr Bio 0 0 1 1 1 giae giai 0 0 0	0 0 0 0 0neg 0n 1 0 0 pram glu	0 0 1 80 1 2 1	dmei	0 84692 0	0 0 smor arm	O I I	0 0 hum api 1 1	s agar	TIVEC 3	tadh d	trub	trig 3	cint 1	oana 2	mor h	sap m	a nus mdo	en mm.u 4	0 Chup 3	atro ecal 3 3	0 00M
1 0 0 atum gzes ogis 0 000 Braa cbri 1 1 2	0 0 1 ecun eint et sman mbre tra 3 1	e pohr Ba o O 1 ag glae glai	b pram gli	0 0 1 1 2 3 3 3 3 3 3 3 3 3 3 3 3 3	O O O O O O O O O O O O O O O O O O O	O aneg b	0 0 mor arm	el coip 1	0 0 hum ap	s agan	3		ref trub	trig 3	T	oana	anor h	sap m	1 1 mus mdo 3 2	4	Chup 3	and a second	1 (00M) 2
atum gzes ogis oeie braa cbri 1 braa cbri 3 roup Statistics H Group	0 0 1 ecun eint et sman more he con 1 1 0 sman more he dide	s Average	0 cneg cn 1 gi	ty Ave	o dmol	Conr	0 0 mor any 1 1	y EC N	umbers		3	tadh d	ref Trub	ang 3	T	oana 2	mor h	aap m 3	1 1 mus mdo 3 2	4	Chup 3	pho 3 3	
atan pres opin 1 1 1 origina chi 1 2 Sroup Statistics H Group 065_127157	Contraction of the second of t	2 0 0 0 1 10 dae glat 0 0 0 11 0 10 dae glat 0 0 0 11 0 10	0 oneg on 1 pram gi	0 0 0 000 10 100 10 100 100	erage %	Conr	0 0 mor any	y EC N	umbers		Evec 3	tach d	rer Trub 2 2	3	T	2	nor h	aap m	1 1 mus máo 3 2	4	O Chap 3	3 3	
Troup Statistics H	C C C C C C C C C C C C C C C C C C C	s Average	0 original b pram gli 9% identi	ty Ave	erage %	Conr	0 0 mor and 1 1	y EC N	umbers		3	1	rer Trub 2 2 2	ang 3	T	2	mor h	3	1 1 mus más 3 2	4	O Chap 3	s s	
a de la de l	Contract of the second of	s Average	0 0 1 1 1 0 pram gin 1 1	0 0 0 000 000 000 000 000 000 000 000	erage %	Conr	0 0 mectivit	y EC N	umbers		3		rer Trub 2 2 2	3	T	2	a a a a a a a a a a a a a a a a a a a	3	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	4	3	ptro ecal	
Group Statistics H Group Statist	Contraction of the second	s Average	0 0 oran oran oran oran oran oran oran oran	0 0 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2	erage %	6 Conr	0 0 mor any 1 1	y EC N	umbers		3	1	rer Tub 2 2	3	T	2	mor h	3	3 2	4	3	ptro ecai	
Group Statistics H Group Statistics H Statistics H Statisti	Contraction of the second	s Average	v strategy	ty Ave	erage %	Conr	0 0 mor any 1 1	y EC N	umbers		3			3	1	2	and a second sec	3	1 1 1 mus mái 3 2	4	3	ptro ecation 3 3	
Geoup Group Good Datasetics H Group Good Datasetics H Group Good Datasetics H Good Dataset	0 0 1 0 0 1 1 1 3 1 3 1 3 1 4 1 4 1 4 1 4 1 4 1 4 1 4 1 4	s Average 34.7	v stralogy	ty Ave	erage %	is Conr	0 0 mor any 1 1	Y EC N	umbera		3		ner Indo	3	T	2	mor h	3	1 1 1 mus mds	4	3	ptro ocali	
Accession	V O O T V O	s Average 34.7	v strategy	ty Ave 38	erage %	6 Conr	nectivit	y EC N	umbern	ength	O P	techicus	Group:	3	1	©amai 2 ≎	EC Numi	bers	\$ Det	scriptic	on	S S	
Accession brur Dbru Tbru Tbru Tbru Tbru Tbru Tbru Tbru T	Comparison of the second	s Average 34.7	v stratogy	ty Ave 38	erage %	o nong b 1	0 0 0 mmor anno 1 1	y EC N	0 0 0 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	ength	0 P 0 Q 0 Q	revious 	Groups	3	54,	COMMA 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	EC Numt	bers	Cation	scriptic	on porter p	orotein,	putative

- e. Is this gene widely conserved? Does it have protein homologs in *Trypanosome brucei*, plants, or bacteria?
- **f.** Take a look at the PFAM domain architectures found under the *PFam domains* (graphic) tab. Do all the proteins in this group have similar domain architecture?

Accession Symbol Descrip	tion Count L	egend
PF01545 Cation_efflux Cation	efflux family 139	
ein Domain Architectures Hide		
Filter:		
 Accession 	Length	٥
aaeg AAEL000077-PA	371	
aaeo NP_213731	308	
afum Afu6g14170	878	
agam AGAP004461-PA	379	
amel XP_394569	352	
anid AN5347	838	
	574	

g. 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 phyletic pattern tool in OrthoMCL

All EuPathDB sites have a phyletic pattern search that uses OrthoMCL data under *Genes -> Evolution -> Orthology Phylogenetic Profile*. This search is very useful to identify genes in your organism of interest that are restricted in their phylogenetic profile. For example, you may be interested in identifying genes that are conserved among organisms in your genus but not present in the host as these genes may make good drug targets or vaccine candidates.

• How many protein groups in OrthoMCL <u>do not</u> have any orthologs in bacteria or archaea? *Hint: go to the Phyletic Pattern search in the Evolution section of the Identify Ortholog Groups category. To specify a phyletic pattern, click on the icon next to the taxonomic group or species to include or exclude it.*

27 Martin Ma	Version 5 10 May 13			12 1		A EUP	athDB P
Ortholog Gro	UNICLE DD ups of Protein Sequences	Groups Quick Searc	ch: s	ynth" Sequ	ences Quic	k Search:	synth*
			19	bout OrthoMCL	Help Log	in Register	Contact Us
ome New Search 👻	My Strategies My Basket (0) Tools -	Data Summary 🕤 🛙	Downloads C	ommunity -	_	_	👷 My Fa
ata Summary	Identify Ortholog Group	s Identify I	Protein Seau	ences		Tools	_
Genomes: 150 Protein Sequences: 1,398,5	46 Text IDs	Text IDs	rotem sequ	ences	BLAS	T	
Ortholog Groups: 124,740	Group ID(s) Text Terms	Sequen	ce ID(s)		Assign	n your proteins t	o groups
ews and Tweets	Evolution	Text Ter	ms		Web	Services	on OrthoMCI
ommunity Resour	Phyletic Pattern Function	PFam II	D or Keyword	lanmont	• Fublic	auons menuonii	ING OTHIOMOL
Identify	Groups based on P	hyletic Pat	tern				
		nylotio i at					
Find Ortholog Gro	oups that have a particular phyletic pattern, i.e., that inc	lude or exclude taxa or specie	es that you specify				
_							
The search is con graphical tree dis	trolled by the Phyletic Pattern Expression (PPE) shown play is a friendly way to generate a pattern expression.	in the text box. Use either the You can always edit the expre	text box or the gr ssion directly. For	aphical tree display PPE help see the i	, or both, to s	specify your patte the bottom of thi	rn. The is page.
In the graphical tr	ee display:						
 Click on -/+ to Click on the ⁴ Refer to the let 	 show or hide subtaxa and species. icon to specify which taxa or species to include or exigend below to understand other icons. 	clude in the profile.					
						_	
	Expression: BACT=0T AND ARCH=0T				Get Answer	•	
Ke	Expression: BACT=0T AND ARCH=0T	ast one subtaxon must be in	group 洋 =must r	iot be in group *	Set Answer	onstraints	
ey:	Expression: BACT=0T AND ARCH=0T	ast one subtaxon must be in e	group X =must r	ot be in group ‡	Get Answer	constraints	
ey: *Root (AL	Expression: BACT=0T AND ARCH=0T ay: • =no constraints • =must be in group • =at le L): pria (BACT):	aast one subtaxon must be in s	group 🗙 =must r	ot be in group ‡	Get Answer	constraints	
ey: Ka *Root (AL [⊡] *Bact ×Fi	Expression: BACT=OT AND ARCH=OT wy: • =no constraints • =must be in group • =at le L): erria (BACT): minucutes (FIRM):	ast one subtaxon must be in state of the subtaxon must be in state of the subtaxon must be in subtaxon must be	group ¥ =must r \$cbot ¥cper	ot be in group *	=mixture of c	onstraints	
ey: K ≉Root (AL □ * #Bact × Fi □ * #Pi	Expression: BACT=OT AND ARCH=OT wy: • =no constraints • =must be in group • =at le .L): oria (BACT): immicutes (FIRM): roteobacteria (PROT):	ast one subtaxon must be in particular state of the subtaxon must	group ¥ =must r ≰cbot ¥cper	ot be in group ‡	Eet Answer =mixture of c	constraints	
ay: Ki ≉Root (AL ⊟ ¥Bacti ¥Fi ⊟ ¥Pi	Expression: BACT=OT AND ARCH=OT yr.	ast one subtaxon must be in s X bant 3 X aturn 3	group ¥ =must r &cbot ¥cper &bsui ¥rpro	ot be in group * Ximon X saur Xrtyp X wen	=mixture of c	ionstraints	
ey: Ki *Root (AL *XBact XFi *YPi 1	Expression: BACT=OT AND ARCH=OT yr. @ -no constraints # =must be in group # =at le L): eria (BACT): irmicutes (FIRM): roteobacteria (PROT): X sipha-Proteobacteria (PROA): X beta-Proteobacteria (PROB):	ast one subtaxon must be in s Xbant 3 Xaturn 3 Xbana 3	group ¥ =must r &cbot ¥ cper &bsui ¥ rpro &bpse ¥ rsol	iot be in group ≑ ≭lmon ≭saur ≭rtyp ≭wen	≕mixture of c	oonstraints	
ey: Ki *Root (AL = XBact XFi = XPi 1 1	Expression: [BACT=OT AND ARCH=OT yr.	ast one subtaxon must be in s Sbant 3 Saturn 3 Sband 3 Sband 3 Sband 3 Sband 3	group ¥ =must r &cbot ¥cper &bsui ¥rpro &bpse ¥rsol	tot be in group ★ Ximon X saut Xrtyp X wen	=mixture of c	constraints	
ey: Ki * Root (AL = X Bact X Fi = X Pi 1 1 1 1 1 1 1 1 1 1 1 1 1	Expression: [BACT=OT AND ARCH=OT yr. @ =no constraints # =must be in group # =at ie L); eria (BACT): irrnicutes (FIRM): roteobacteria (PROT): % alpha-Proteobacteria (PROA): % beta-Proteobacteria (PROB): % delta-Proteobacteria (PROD): % gama-Proteobacteria (PROG):	ast one subtaxon must be in s S bant 3 S atum 3 S bana 3 S gaul S cbur 3	group ¥ =must r &cbot ¥cper &bsui ¥rpro &bpse ¥rsol &ecol ¥ftul	Ximon Xsaur Ximon Xsaur Xrtyp Xwen Xsent Xsfle	=mixture of c =mixture of c & spne d	xonstraints Xyent Xype	25
ey: Ki *Root (AL Batt XFI B' XPI 1 1 1 1 1 1 1 1 1 1 1 1 1	Expression: [BACT=OT AND ARCH=OT yr.	ast one subtaxon must be in s S bant 3 S bant 3 S bana 3 S gaul S cbuy 3 S cjej 3	group ¥ =must r & cbot ¥ cper & bsui ¥ rpro & bpse ¥ rsol & ecol ¥ ftul & wsuc	ximon X saur Ximon X saur Xrtyp X wen Xsent X sfle	=mixture of c =mixture of c X spne d	xonstraints ≭yent ¥ype	25
ey: Ki * Root (AL = X Bact X Fi = X Pi 1 1 1 2 X O	Expression: BACT=OT AND ARCH=OT yr.	ast one subtaxon must be in s kbant 3 katum 3 kbana	group ¥ =must r &cbot ¥cper &bsui ¥rpro &bpse ¥rsol &ecol ¥ftul &wsuc icpne *ctep ¥	ximon Xsaur Xrtyp Xwen Xsent Xsfle deth Xdrad Xr	aet Answer =mixture of c %spne d %vcho	Xyent Xypr	es € ¥tmar ¥tp
ey: Ki * Root (AL * XBact XFI * XP 1 * XO * XArch	Expression: [BACT=OT AND ARCH=OT] yr. • =no constraints • =must be in group • =at ie L); rria (BACT): irriicutes (FIRM): roteobacteria (PROT): Kalpha-Proteobacteria (PROA): Koeta-Proteobacteria (PROB): Koeta-Proteobacteria (PROB): Kala-Proteobacteria (PROB): Kala-Proteobacteria (PROB): Kala-Proteobacteria (PROB): Kala-Proteobacteria (PROB): Kala-Proteobacteria (PROE): ther Bacteria (OBAC): aea (ARCH):	ast one subtaxon must be in t X bant 3 X atum 3 X bana 3 X gaul X cbur 3 X gaul X cbur 3 X gaul X cbur 3 X	group ¥ =must n \$cbot X cper \$bsul X rpro \$bspe X rsol \$wsuc \$cpne \$ctep \$	d ot be in group ♦ Ximon X sau Xinyo X wen Xsent Xstile deth Xdrad Xr	aet Answer =mixture of c %spne d %vcho	Xyent Xypr	≥s e ¥tmar ¥tp
ey: *Root (AL *Root (AL *Fi *Fi *Fi * * * * * * * * * * * * *	Expression: BACT-OT AND ARCH-OT ay: • -no constraints • = must be in group • =at le L): eria (BACT): Irmicutes (FIRM): rotoobacteria (PROT): Kalpha-Proteobacteria (PROB): K deita-Proteobacteria (PROB): K deita-Proteobacteria (PROB): K deita-Proteobacteria (PROD): K genion-Proteobacteria (PROG): K epsilon-Proteobacteria (PROG): ther Bacteria (OBAC): aea (ARCH): uryarchaeota (EURY):	ast one subtaxon must be in s X bant 3 X atum 3 X band 3 X gaul X cbur 3 X cbu	group X =must n 4cbot X cper 4bsul X rpro 4bspe X rsol 4wsuc 4cpne X ctep X 4halo X hwal 4balo X hwal	tot be in group ♦ Ximon Xsaur Xirtyp Xwen Xsent Xstie deth Xdrad Xr Xmsmi Xmjan	a Set Answer amixture of c am	Xyent Xypr b Xrbal Xsyn Xtvol	zs e ¥tmar ¥tµ
ey: • Root (Al • X Bach × FFI • X PFI • X PFI • X PFI • X Arch × EFI • X Arch • X EFI • X C	Expression: BACT-OT AND ARCH-OT ay: • -no constraints [* =must be in group] * =at le L): eria (BACT): Irmicutes (FIRM): rotoobacteria (PROT): Kalpha-Proteobacteria (PROA): Kdeita-Proteobacteria (PROB): Kdeita-Proteobacteria (PROB): Kdeita-Proteobacteria (PROB): Kdeita-Proteobacteria (PROB): Kapama-Proteobacteria (PROG): Kapama-Proteobacteria (PROG): Kapama-Proteobacteria (PROE): ther Bacteria (OBAC): aea (ARCH): uryarchasota (EURY): renarchaeota (CREN):	ast one subtaxon must be in r X bant 3 X band 3 X band 3 X band 3 X band 3 X band 3 X cjej 3 X cje	group X =must n 4cbot X cper 4bsul X rpro 4bsue X rsol 4csue X ful 4csue X f	ximon Xsau Ximon Xsau Xityo Xwen Xsent Xsile deth Xdrad Xr Xmsmi Xmjau Xmsed Xnma	 Set Answer mixture of c Sepne Xspne Xvcho xvcho xmiar xmmar xsmar 	Xyent Xypd Xyent Xypd Xtvol Xssol	zs e ¥tmar ¥tµ
ey: • Root (Al • Root (Al • X Bact × Fri • X Pri • X	Expression: BACT-OT AND ARCH-OT ay: • =no constraints • =must be in group • =at ie L): eria (BACT): irmicutes (FIRM): roteobacteria (PROT): * lipha-Proteobacteria (PROA): * beta-Proteobacteria (PROB): * delta-Proteobacteria (PROD): * delta-Proteobacteria (PROD): * gama-Proteobacteria (PROD): * gama-Proteobacteria (PROE): ther Bacteria (OBAC): aes (ARCH): uryarchaeota (EURY): renarchasota (CREN): anoarchaeota (NANO):	ast one subtaxon must be in r X bant 3 X atum 3 X gaul X cipe 3 X cipe 4 X	sproup X =must n schot X cper sbsu X rpro sbsu X rpro sbsu X rpro sbsu X rsol schot X thut status schot X thus schot X	ximon X-saut Ximon X-saut Xrtyp Xwen Xsent Xsfle deth Xdrad Xr Xmsmi Xmjai Xmsed Xnma	aet Answer =mixture of c x spne d x vcho ntub x miej n x mmar r x smar	Xyent Xypt o Xrbal Xsyn Xtvol Xssol	os e ¥tmar ¥tp
ey: * Root (Al * Soot (Al *	Expression: BACT-OT AND ARCH-OT ay: • =no constraints • =must be in group • =at ie L): eria (BACT): irmicutes (FIRM): roteobacteria (PROT): * lpha-Proteobacteria (PROA): * beta-Proteobacteria (PROB): * delta-Proteobacteria (PROB): * delta-Proteobacteria (PROB): * delta-Proteobacteria (PROE): * delta-Proteobacteria (PROE): * delta-Proteobacteria (PROE): * delta-Proteobacteria (PROE): * delta-Proteobacteria (PROE): * uryarchaeota (OBAC): aea (ARCH): uryarchaeota (EURY): renarchaeota (CREN): anoarchaeota (KANO): orarchaeota (KORA):	ast one subtaxon must be in a X bant 3 X atum 3 X bmal 3 X bmal 3 X bmal 3 X cjej 3 X atua 3 X cjej 3 X atua 3 X catua 3	group X =must r Kobot X oper Kobot X rpro Kobot X rol Kasoc X full Kasoc Kobot X hwal Kabot X hwal	ximon Xsaul Ximon Xsaul Xrtyp Xwen Xsent Xsfie deth Xdrad Xr Xmsmi Xmjar Xmsed Xnma	aet Answer =mixture of c xspne d xvcho ntub xmlej n xmmar r xsmar	Xyent Xypr o Xrbal Xsyn Xtvol Xssol	25 e ¥tmar ⊁tp
ey: ● Root (Al ● X Bact X Fi ● X Pri =	Expression: BACT-OT AND ARCH-OT ay: • =no constraints • =must be in group • =at ie L): eria (BACT): irmicutes (FIRM): roteobacteria (PROT): * lipha-Proteobacteria (PROA): * beta-Proteobacteria (PROB): * delta-Proteobacteria (PROB): * delta-Proteobacteria (PROG): * gama-Proteobacteria (PROG): * gama-Proteobacteria (PROG): * gama-Proteobacteria (PROG): * uryarchaeota (CREN): anoarchaeota (CREN): anoarchaeota (KORA): : ryota (EUKA):	ast one subtaxon must be in a statum 3 Katum 3	group X =must r Kobol X oper Kobol X rpro Kobol X rpro Kobol X flui Kobol X flui Kobol X hival	tt be in group ♦ Ximon Xsaut Xrtyp Xwen Xsent Xsfle deth Xdrad Xr Xmsmi Xmjan Xmsmi Xmjan	aet Answer =mixture of c %spne d d x vcho ntub X miej n X mmar r X smar	Xyent Xypr o Xrbal Xsyn Xtvol Xssot	ss e ¥tmar ≭tp
ey: *Root (Al *XBact XF, *XP, *XP, *XF, *XF, *XF, *XF, *XC, *XArch *ExArch *ExArch *ExArch *ExArch *ExArch *ExArch *ExArch *ExArch *ExArch	Expression: BACT-OT AND ARCH-OT ayr • =no constraints	ast one subtaxon must be in Kbant 3 Katum 3 Kbmal 3	group X =must r Kobol X oper Kobol X rpro Kopos X rsol Kecol X ftul Kecol X ftu	sten group ♦ Ximon Xsaur Xrtyp Xwen Xsent Xsfie deth Xdrad Xr Xmami Xmjan Xmsed Xnma	aet Answer =mixture of c % spne d % vcho ntub % miej n % mmar r % smar	Xyent Xype o Xrbal Xsyn Xtvol Xssol	e Xtmar Xtp
ey: K © Root (Al © XBact XF © XPA : : : : : : : : : : : : :	Expression: BACT-OT AND ARCH-OT ayr	ast one subtaxon must be in a Katum 3 Koban 3 Koban 3 Koban 3 Koban 3 Koban 3 Koban 3 Koban 3 Koban 3 Koban 3 Katu 3 Koban 3 K	group X =must r Kobol X oper Kobol X rpro Kopes X rsol Kecol X ftul Kecol X ftu	ximon Xsaur Ximon Xsaur Xrtyp Xwen Xsent Xsfie deth Xdrad Xr Xmsmi Xmjan Xmsed Xnma	d Xvcho Mubb Xmley Mubb Xmley Mubb Xmley Mubb Xmley	Xyent Xype o Xrbal Xsyn Xtvol Xssol	e ¥tmar ¥tp

- Modify the above search to identify protein groups that do not contain orthologs from eukaryotes? How many did you find?
- Find all groups that contain orthologs from at least one species of *Cryptococcus* (cneo, cneg) and *Aspergillus* (anid, afum) but not from any bacteria or archaea. To run this search, you will have to use the expression panel at the top of the search. (Remember this question is requiring that the groups must contain at least one *Cryptococcus* and one *Aspergillus* species). See the description and help at the bottom of the search page.

In the graphical tree display:	
 Click on -/+ to show or hide subtaxa and species. Click on the [®] icon to specify which taxa or species to include or exclude in the profile. Refer to the legend below to understand other icons. 	
Expression:	Get Answer
Key: • =no constraints 🖌 =must be in group 🥓 =at least one subtaxon must be in group 🗱 =must not be in group	

The correct expression to use is: BACT=0T AND ARCH=0T AND cneo+cneg>=1T AND afum+anid>=1T



- What happens if you revise the strategy to increase the stringency where *Cryptococcus bacillisporus* (cneo), *Cryptococcus neoformans var grubii* H99 (cneg) *Aspergillus nidulans* (anid), and *Aspergillus fumigatus* (afum) *must be* in the group?
- Find all groups that contain at least one subtaxon from *Basidiomycota* and *Ascomycota* but not from bacteria or archaea, or microsporidia

Key: 🔍 =no constraints 🖋 =must be in group	≪ =at leas	t one sul	btaxor	n must be i	n group	X =must i	not be in gr	oup 🌞 =r	nixture of o	constraints	;		
∻Root (ALL) :													
[⊡] ≍ Bacteria (BACT):													
X Firmicutes (FIRM):				×bant	≍ cbot	×cper	XImon	≍saur	×spne				
[⊟] ≭ Proteobacteria (PROT):													
¥alpha-Proteobacteria (PROA):				Xatum	Xbsui	🗙 rpro	🗙 rtyp	×wend					
🗯 beta-Proteobacteria (PROB):				≍bmal	×bpse	≭rsol							
🗙 delta-Proteobacteria (PROD):				≭gsul									
🗯 gama-Proteobacteria (PROG):				×cbur	×ecol	🗯 ftul	≍sent	≭sfle	×vcho	×yent	×ypes		
≭epsilon-Proteobacteria (PROE):				🗙 cjej	×wsuc								
X Other Bacteria (OBAC):				×aaeo	×cpne	×ctep ×	deth Xd	rad Xmt	ub Xmle	p Xrbal	×syne	≭tmar	×
^{⊡⊷} ¥Archaea (ARCH):													
Euryarchaeota (EURY):				Xaful	X halo	X hwal	×msmi	×mjan	×mmar	≭tvol			
Crenarchaeota (CREN):				×cmaq	≍ hbut	🗯 ihos	≭msed	≭nmar	≍smar	×ssol			
×Nanoarchaeota (NANO):				×nequ									
Korarchaeota (KORA):				X ckor									
[∼] ¥Fungi (FUNG):													
XMicrosporidia (MICR):	×ecun	🗙 ebie	×ei	nt									
✓Basidiomycota (BASI):	● cneo	cneg	© Ib	ic © pch	r								
✓Ascomvcota (ASCO):	afum	anid	aory	●cgla ●	cimm 🏼 o	pos © dha	n ©egos (∮gzea ©k	ac 🔍 mgri	● ncra ●	psti • sce	r © calb	•

3. Combining 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 OrthoMCL groups that contain the word • *phosphatase*. Use wild cards to search for any word combinations.

D	A O	athoM(C	LDB ^{Release} 23 Jul 2	5015	2	120		A EuPat	hDB Proje	ect
¥	Ortho	olog Groups of Protei	n Sequences	Groups Qu	ick Search:	*phosphatase*	Sequences Quick	Search:	synth*	۹
						About OrthoMCI	_ Help Login Re	gister Contact	Us 🗾 f	
Но	me New Sea	rch 👻 My Strategies	My Basket (0) Tools	- Data Summary -	Downloads	Community -			숚 My Fav	orites
(Gro	ups) Text 1560 Groups Step 1	Add Step					Stra	ategy: Phyle	etic(2) * Renam Duplicat Save A Shar Delet	te te te
					1					
1560	Groups from	Step 1 Revise								
Strat Gro	egy: Phyletic	(2) Phyletic pattern								
Strat Gro F	up Results	Phyletic pattern Next Last Advance	ed Paging			Do	wnload Add t	to Basket	Add Colum	nns
Gro F	egy: Phyletic	(2) Phyletic pattern Next Last Advance Sequences 3	ed Paging	◆ Average % ▼ Connectivity	Q 💼 🗘 EC	Do Numbers Q	wnload Add t	co Basket	Add Colun Found in	nns
Gro F	egy: Phyletic up Results irst 1 2 3 4 5 Group 0G5_126538	(2) Phyletic pattern Next Last Advance Sequences I ili	Average % 3 1	Average % Connectivity 33.8	3	Do Numbers Q	wwnload Add t	Score 3	Add Colun Found in Sequences	nns Q

Add a step and run a phyletic pattern search for groups that contain at least • one subtaxon from Ascomycota and Basiodiomycota. *Hint:* make sure everything has a red x on it except for Fungi, which should be yellow check).

^{B−} ¥ Viridiplantae (VIRI):																
Streptophyta (STRE):	Xatha	≭osat	X ppa	* ×rcor	n 🗙 m	nicr										
Chlorophyta (CHLO):	Xcrei	≭otau	×vca													
*Rhodophyta (RHOD):	×cmer															
Cryptophyta (CRYP):	≭gthe															
🗱 Bacillariophyta (BACI):	≭tpse															
^{──} ≑Fungi (FUNG) :																
XMicrosporidia (MICR):	Xecun	×ebie	¥ eint													
✓Basidiomycota (BASI):	● cneo	● cneg	Ibic	© pch	r											
Ascomycota (ASCO):	afum	anid	aory	©cgla ●	cimm	cpos	O dhan	egos	© gzea	klac	mgri	ncra	ø psti	scer	● calb	spom
── X Metazoa (META):																
Platyhelminthes (PLAT):	≍sman															
XNematodes (NEMA):	X bmaa	🗙 cbri	×cele													
XArthropoda (ARTH):	Xapis	×aaeg	Xaga	m 🗙 ame	el 🗙 br	mor 🕽	Срір	×dmel	X phum	≭isca						

How many groups did you return? •



• Explore the phyletic pattern.

Group Results	Phyletic pattern
Legend: 0 no	
1 or	ne ortholog Brita el EUGL, Brita el MAMBE Brita el MAMBE
n m	one than one ortholog BF Ca N FIGHER BF Ca N FIGHER AF Ca N FIGHER
Pohyletic pattern	labels
Advanced	Paging
	Keywords: synthase
	Plam Domains: divocosyttransferase family 20
OG5_135116	
	Keywords: n/s
	Pfam Domains: Fungal protein of unknown function (DUF1752)
OG5_135798	
	Kaywords: v/u
	Pfam Domains: 6-phosphofnucto-2-kinase
OG5_136577	
	naymetas no
000 100000	
035_136969	
	Keywords: min
	Pfam Domains: v/a
OG5_137024	
	Museuka sharakakara

• Click on the group OG5_137024. Explore the multiple sequence alignments and Cluster graphs.

equences & Statistics PFam domain	s (graphic) PFam domains (details)	MSA Cluster	graph		
yletic Distribution Hide					
Legend: no ortholog n more than one ortholog	第 도 의 [FRM] 第 도 의 [PR	11 н ц о СВ н ц о VI а н ц о СВ	AC If a s Arch If a s Arch		
Chow labels Immunol (seen) Abolt (seen) Seen (seen) Abolt (seen) Seen (seen) Abolt (seen) Seen (seen) Abolt (seen) <td>Space Table Space <th< td=""><td>ecol cjej wsuc 0 0 0 1 inf 0 0 0 0 0 0 0 0 0 0 0 0 0 1 2 0 1 0 0 0 0 0</td><td>Top: Norm State S</td><td>ctep Imax miep syme rb 0 0 0 0 0 0 0</td><td>III beam abaro max max<</td></th<></td>	Space Table Space Space <th< td=""><td>ecol cjej wsuc 0 0 0 1 inf 0 0 0 0 0 0 0 0 0 0 0 0 0 1 2 0 1 0 0 0 0 0</td><td>Top: Norm State S</td><td>ctep Imax miep syme rb 0 0 0 0 0 0 0</td><td>III beam abaro max max<</td></th<>	ecol cjej wsuc 0 0 0 1 inf 0 0 0 0 0 0 0 0 0 0 0 0 0 1 2 0 1 0 0 0 0 0	Top: Norm State S	ctep Imax miep syme rb 0 0 0 0 0 0 0	III beam abaro max max<
roup Statistics Hide	ge % Identity Average % Connectivity EC	Numbers			
OG5_137024 20 36.5	96.3				
Number none					
st of Sequences Hide					
Get Sequences: As Fasta file As no Filter:	aw strategy				
* Accession	Taxon	Characteristic Length	Previous Groups	C Numbers	Description
afum Afu8g04180	Aspergillus fumigatus Af293	813	OG3_20656, OG4_20833	N/A	C2H2 zinc finger protein
anid AN1420	Emericella nidulans	822	N/A	N/A	Uncharacterized ORF

• Explore the multiple sequence alignments from some of these groups and nodes for selected species. (Hint: click on a *group ID* and then open the *MSA* and *Cluster graph* tabs).

	Group: OG5_137024											
		(20 se	equences)									
	Add to Basket add to Favorites 🖒											
Sequences & Statistics PFam	omains (graphic) PFam domains (detail	s) MSA	Cluster graph									
MUSCLE (3.7) mult	ple sequence alignment											
cneg [CNAG_03689T0 cnee [CNBG_0091T0 spom NP_596071 ylip XP_506071 cimm [CTMG_04429T0 cpos [CPAG_06227 anid]AN1420 aory A0090103000033 afum]Afu8g04180 nora NCU05589T0 gzea [FGST_10087]bic [eu2.Lbscf0006g06320 calb [CANT_00816 calb]calb_gc5314.orf19.2450 peti XP_001387834 dhan XP_459608 egos [SIP5 klac XP_454949	MGNAPSSHATPPPSSSTNTASISERER MGNTJSHATPPSSSTSNAPTSERER MGNTIGKEQQKNAASVSEDYHREGOV- MGNAATKESRSSQEHSARHSRGGSTAS MGNAATKESRSSQEHSARHSRGGSTAS MGNAQTKETRPHOJSSRNRMGRGGS MGNSQTKETRPPISSSNRSHOWGSGS MGNNGTKERAPPTPTNRSHGGGHGRSI MGNSGTKEFRPISSSNRSHOWGSSS MGNNATKESRDPGDASRRYSQPLDPSIF MGNSTKESRPEGDASRRYSQPLDPSIF MGNSTKESRPEGDASRRYSQVLDPSIF MGNVPAKETASRSSSV5DVRNSTSST 	SSSAQPVNTS SSSAQPMNTS AAYDHYRPSI HGRSPYGDRH HGRSPYGDRH SSTOSGRESS DHLRERQDNI HGSSSSTSSB HGSSSSTSSB SSSTSSS SSSTSSS SSGSFKLN-	STAGRIQRPS TAGRTQRPS RSKRPF OGIHASSS-RALRGG OGIHASSS-RALRGG NSEGSRBHRSS RSTRG RSTRG 	SVSFHSGLT SVSFHSGLT 								
scer scer_s288c:YMR140W cgla XP_447290	MGNVPGKIDQEDSFNDVRPDSSYNTTSS MGNVPAKLDQDASSYTGRSTYT-DSSS#	NSVIKQYDEE SGVGAFRGAC	5ASSRVRTRRT1 GGTMDETRARGRRT5	ISLVNNILN SSLVGNILN								

- *Examine the Cluster Graph tab:* Modify the E-value cutoff slider.
 - What happens when you increase the E-value?
 - What happens when you decrease the E-value?
 - Can you identify subclusters?

