

Supplementary Data for Hayes et al.

Consists of Tables S1 to S4, Figure S1, and supplementary references.

Table S1. Psi-BLAST results with BLOC-1 sub-units.

Blos1

NP_001478.1 (*H. sapiens*) – iteration 3 (132 hits, converged)

Posn	sequence ref	species	score	e-value
6.	NP_001478.1	<i>Homo sapiens</i>	160	2e-38
35.	NP_725401.2	<i>Drosophila melanogaster</i>	138	1e-31
43.	NP_499262.1	<i>Caenorhabditis elegans</i>	135	7e-31
54.	NP_180592.1	<i>Arabidopsis thaliana</i>	131	8e-30
130.	XP_500764.2	<i>Yarrowia lipolytica</i>	68.4	1e-10
141.	XP_451660.1	<i>Kluyveromyces lactis</i>	38.7	0.092

XP_451660.1 (*K. lactis*, see above) – iteration 2 (7 hits, converged)

Posn	sequence ref	species	score	e-value
1.	XP_451660.1	<i>Kluyveromyces lactis</i>	134	1e-30
2.	XP_002553358.1	<i>Lachancea thermotolerans</i>	129	4e-29
3.	NP_013512.1 (YLR408Cp)	<i>Saccharomyces cerevisiae</i>	118	7e-26
4.	XP_449775.1	<i>Candida glabrata</i>	117	1e-25
5.	XP_001646567.1	<i>Vanderwaltozyma polyspora</i>	112	4e-24
6.	XP_002495505.1	<i>Zygosaccharomyces rouxii</i>	112	4e-24
7.	NP_984939.1	<i>Ashbya gossypii</i>	103	2e-21
16.	ACO52029.1	<i>Rana catesbeiana</i>	36.2	0.47
34.	XP_500764.2	<i>Yarrowia lipolytica</i>	33.5	3.5

NP_013512.1 (YLR408Cp, *S cerevisiae*) – iteration 4 (61 hits)

N.B. converged at iteration 8, with 145 hits

Posn	sequence ref	species	score	e-value
1.	NP_013512.1 (YLR408Cp)	<i>Saccharomyces cerevisiae</i>	154	1e-36
24.	XP_451660.1	<i>Kluyveromyces lactis</i>	99.7	4e-20
44.	XP_500764.2	<i>Yarrowia lipolytica</i>	53.5	3e-06
54.	NP_499262.1	<i>Caenorhabditis elegans</i>	45.0	0.001
62.	XP_509120.1	<i>Pan troglodytes</i>	42.7	0.005
82.	NP_001478.1	<i>Homo sapiens</i>	39.7	0.054
140.	NP_725401.2	<i>Drosophila melanogaster</i>	35.4	0.79
202.	NP_180592.1	<i>Arabidopsis thaliana</i>	3.3	6.8

Snapi

YNL086Wp iteration 4 (12 hits, converged)

Posn	sequence ref	species	score	e-value
1.	NP_014313.1 (YNL086Wp)	<i>Saccharomyces cerevisiae</i>	122	4e-27
2.	XP_452820.1	<i>Kluyveromyces lactis</i>	104	2e-21
12.	XP_002494050.1	<i>Pichia pastoris</i>	77.0	3e-13

Cno

YDR357Cp – iterations 4 (51 hits, converged)

Posn	sequence ref	species	score	e-value
2.	EDN60686.1 (YDR357Cp)	<i>Saccharomyces cerevisiae</i>	129	4e-29
31.	XP_454923.1	<i>Kluyveromyces lactis</i>	105	1e-21
46.	XP_503339.2	<i>Yarrowia lipolytica</i>	92.0	8e-18
51.	XP_389585.1	<i>Gibberella zeae</i>	82.3	6e-15

Vab2 – iteration 6 (17 hits, converged)

Posn	sequence ref	species	score	e-value
1.	NP_010911.1 (Vab2p)	<i>Saccharomyces cerevisiae</i>	266	3e-69
5.	XP_504134.1	<i>Yarrowia lipolytica</i>	165	5e-39
17.	XP_001384198.2	<i>Scheffersomyces stipitis</i>	117	2e-24

YKL061Wp (Bli1p) – iteration 2 (6 hits, converged)

Posn	sequence ref	species	score	e-value
1.	EEU07430.1 (YKL061Wp)	<i>Saccharomyces cerevisiae</i>	120	6e-26
6.	XP_448843.1	<i>Candida glabrata</i>	87.9	3e-16

YGL079Wp (Kxd1p) – iteration 5* (142 hits)

Posn	sequence ref	species	score	e-value
16.	NP_011436.1 (YGL079Wp)	<i>Saccharomyces cerevisiae</i>	142	1e-32
26.	AAQ13622.1	<i>Homo sapiens</i>	136	9e-31
64.	NP_648580.1	<i>Drosophila melanogaster</i>	125	1e-27
101.	NP_504831.1	<i>Caenorhabditis elegans</i>	101	4e-20
126.	NP_189557.2	<i>Arabidopsis thaliana</i>	84.2	4e-15

* from iteration 6, non-KxDL proteins align with $p < 0.001$

Table S2. HHpred and HHsenser results with BLOC-1 sub-units

Pair-wise comparison of profile hidden Markov models (HMMs) was carried out (HHpred, extended with HHsenser in some cases) with the indicated seed. Data shown for top hits: full length (Len), the calculated probability it shares the same structure as the seed, (Prob - which includes calculated secondary structure), the average number of false positives expected to score better than the hit (e-value - excludes secondary structure), the ranges of residues in the HMM that match for both the seed (Query) and the hit (Template), and the number of columns aligned between seed and hit HMMs (cols).

Blos1 HHpred

# Hit	Len	Prob	e-value	Query	Template	Cols
1 NP_001478.1 Blos1 (Hs)	125	100	<1e-50	1-125	1-125	125
2 NP_499262.1 Blos1 (Ce)	129	100	<1e-50	5-125	1-121	121
3 NP_180592.1 Blos1 (At)	152	100	<1e-50	2-120	34-152	119
4 NP_725401.1 Blos1 (Dm)	143	100	<1e-50	5-125	1-121	121
5 NP_013512.1 YLR408Cp (Sc)	122	96.9	0.0042	77-121	77-121	45

Blos1 → HHsenser

# Hit	Len	Prob	e-value	Query	Template	Cols
1 NP_001478.1 Blos1 (Hs)	125	100	<1e-50	1-125	1-125	125
2 NP_499262.1 Blos1 (Ce)	129	100	<1e-50	5-125	1-121	121
3 NP_180592.1 Blos1 (At)	152	100	2.8e-45	1-120	33-152	120
4 NP_725401.1 Blos1 (Dm)	143	100	3.8e-44	5-124	1-120	120
5 NP_013512.1 YLR408Cp (Sc)	122	97.5	0.00046	49-119	46-119	71

YLR408Cp HHpred

# Hit	Len	Prob	e-value	Query	Template	Cols
1 NP_013512.1 YLR408Cp (Sc)	122	100	<1e-50	1-122	1-122	122
2 NP_499262.1 Blos1 (Ce)	129	98.0	0.00013	35-121	38-117	80
3 NP_001478.1 Blos1 (Hs)	125	97.8	0.00021	48-120	51-120	70
4 NP_180592.1 Blos1 (At)	152	97.0	0.01	14-119	21-151	104
5 NP_725401.1 Blos1 (Dm)	143	96.9	0.012	36-119	39-115	77

YLR408Cp → HHsenser

# Hit	Len	Prob	e-value	Query	Template	Cols
1 NP_013512.1 YLR408Cp (Sc)	122	100	<1e-50	1-122	1-122	122
2 NP_499262.1 Blos1 (Ce)	129	100	6.8e-29	39-122	38-118	81
3 NP_001478.1 Blos1 (Hs)	125	100	1.6e-28	37-121	40-121	82
4 NP_180592.1 Blos1 (At)	152	100	4.7e-28	14-120	21-152	104
5 NP_725401.1 Blos1 (Dm)	143	99.9	2.4e-27	37-121	36-117	82

Snapin HHpred

# Hit	Len	Prob	e-value	Query	Template	Cols
1 NP_036569.1 snapin (Hs)	136	100	<1e-50	1-136	1-136	136
2 NP_722835.1 snapin (Dm)	134	100	8.7e-44	21-133	22-134	113
3 NP_500721.1 snapin (Ce)	122	100	4.1e-43	17-134	5-120	116
4 NP_178028.2 snapin (At)	138	99.9	4.7e-24	20-122	32-133	102
8 NP_014313.1 YNL086Wp (Sc)	102	95.8	0.46	30-110	21-101	81

Snapin → HHsenser

# Hit	Len	Prob	e-value	Query	Template	Cols
1 NP_036569.1 snapin (Hs)	136	100	2.2e-40	1-136	1-136	136
2 NP_722835.1 snapin (Dm)	134	100	3.9e-34	10-128	11-129	119
3 NP_500721.1 snapin (Ce)	122	100	1.4e-31	14-136	2-122	121
5 NP_178028.2 snapin (At)	138	99.8	1.1e-19	6-127	18-138	121
7 NP_014313.1 YNL086Wp (Sc)	102	97.7	0.0033	12-110	3-101	99

YLR086Wp HHpred

# Hit	Len	Prob	e-value	Query	Template	Cols
1 NP_014313.1 YNL086Wp (Sc)	102	100	<1e-50	1-102	1-102	102
2 NP_036569.1 snapin (Hs)	136	95.6	0.57	19-101	28-110	79
3 NP_500721.1 snapin (Ce)	122	95.3	1.9	20-101	17-97	81
4 NP_722835.1 snapin (Dm)	134	94.6	1.6	19-101	29-111	79

YLR086Wp → HHsenser

# Hit	Len	Prob	e-value	Query	Template	Cols
1 NP_014313.1 YNL086Wp (Sc)	102	99.9	1.6e-25	1-102	1-102	102
2 NP_036569.1 snapin (Hs)	136	97.7	0.0032	9-101	18-110	93
3 NP_722835.1 snapin (Dm)	134	97.1	0.027	21-101	31-111	81
4 NP_500721.1 snapin (Ce)	122	96.8	0.086	20-101	17-97	81

Cno (Dictyostelium discoideum) HHpred*

# Hit	Len	Prob	e-value	Query	Template	Cols
1 NP_060836.1 cno (Hs)	217	100	<1e-50	3-174	56-216	161
3 NP_648414.1 cno (Dm)	169	100	<1e-50	8-173	5-165	160
6 NP_495247.1 T24H7.4 (Ce)	106	96.5	0.18	12-106	8-103	94
7 NP_010644.1 YDR357Cp (Sc)	122	96.3	1.1	2-108	15-116	99

(*Breast Carcinoma Amplified Sequence-4 isoforms x3 excluded, as this human protein is so similar to cno)

Cno (Homo sapiens) HHpred*

# Hit	Len	Prob	e-value	Query	Template	Cols
1 NP_060836.1 cno (Hs)	217	100	<1e-50	1-217	1-217	217
3 NP_648414.1 cno (Dm)	169	100	<1e-50	61-217	5-167	156
6 NP_010644.1 YDR357Cp (Sc)	122	96.2	1.3	48-161	8-116	109
8 NP_495247.1 T24H7.4 (Ce)	106	95.9	0.64	82-159	27-103	77

T24H7.4 (Caenorhabditis elegans) HHpred

# Hit	Len	Prob	e-value	Query	Template	Cols
1 NP_495247.1 T24H7.4 (Ce)	106	100	<1e-50	1-106	1-106	106
2 NP_010644.1 YDR357Cp (Sc)	122	92.8	0.58	14-106	1-117	93
3 NP_648414.1 cno (Dm)	169	90.7	0.91	1-106	1-105	103
4 NP_060836.1 cno (Hs)	217	89.3	1.8	1-106	53-162	104

YDR357Cp HHpred

# Hit	Len	Prob	e-value	Query	Template	Cols
1 NP_010644.1 YDR357Cp (Sc)	122	100	<1e-50	1-122	1-122	122
2 NP_500377.1 DUF2365 (Ce)	157	97.8	0.013	6-111	42-152	104
3 NP_060092.1 DUF2365 (Hs)	223	96.7	0.23	17-109	123-220	91
4 NP_612037.1 DUF2365 (Dm)	302	95.8	0.44	16-109	202-300	92

DUF2365 (Caenorhabditis elegans, Y37E11B.3) HHpred

# Hit	Len	Prob	e-value	Query	Template	Cols
1 NP_500377.1 DUF2365 (Ce)	157	100	<1e-50	1-157	1-157	157
2 NP_060092.1 DUF2365 (Hs)	223	100	<1e-50	11-152	61-222	142
3 NP_612037.1 DUF2365 (Dm)	302	100	<1e-50	12-152	128-302	141
4 NP_194064.1 DUF2365 (At)	259	99.3	5.6e-11	55-127	43-115	73
5 NP_010644.1 YDR357Cp (Sc)	122	97.9	0.0062	41-151	5-110	104

Table S3. Sequences used for Blo1/snapin consensuses, and cno line-up & Tables S1/S2

The gene identifying (GI) number and species of sequences used to make consensuses and line-ups.

Blo1		Snapin (x38)		Cno / cno-like (x24 each, + length)	
<u>GI</u>	<u>Species</u>	<u>GI</u>	<u>Species</u>	<u>GI</u>	<u>Species</u>
ANIMALS/PLANTS (x35)					
113195616	Danio rerio	148228503	Xenopus laevis	118086704	Gallus gallus 214
115692166	Strongylocentrotus purpuratus	148669506	Mus musculus	149244312	Lodderomyces elongisporus 161
116057994	Ostreococcus tauri	156365612	Nematostella vectensis	156350495	Nematostella vectensis 192
118398939	Tetrahymena thermophila	156843445	Vanderwaltozyma polyspora	156846164	Vanderwaltozyma polyspora 116
123449313	Trichomonas vaginalis	157104298	Aedes aegypti	157136111	Aedes aegypti 172
148235775	Xenopus laevis	158296992	Anopheles gambiae	158294340	Anopheles gambiae 184
15227736	Arabidopsis thaliana	167537388	Monosiga brevicollis	167538428	Monosiga brevicollis 265
156391143	Nematostella vectensis	168035684	Physcomitrella	170045619	Culex quinquefasciatus 193
157118066	Aedes aegypti	170585680	Brugia malayi	171684445	Podospora anserina 172
158301499	Anopheles gambiae	17538398	Caenorhabditis elegans	17536485	Caenorhabditis elegans 106 (T24H7.4)
170044487	Culex quinquefasciatus	189239210	Tribolium castaneum	189198760	Pyrenophora tritici 140
170590036	Brugia malayi	195999888	Trichoplax adhaerens	193582437	Acyrtosiphon pisum 164
196002611	Trichoplax adhaerens	198418879	Ciona intestinalis	19526908	Mus musculus 215
198421104	Ciona intestinalis	225462195	Vitis vinifera	196003246	Trichoplax adhaerens 177
218186731	Oryza sativa Indica	226502642	Zea mays	198429992	Ciona intestinalis 175
221101349	Hydra magnipapillata	238882874	Candida albicans	210075837	Yarrowia lipolytica 116
224124784	Populus trichocarpa	240848775	Acyrtosiphon pisum	21356449	Drosophila melanogaster 169
226459052	Micromonas pusilla	24581212	Drosophila melanogaster	221112508	Hydra magnipapillata 177
226484660	Schistosoma japonicum	254567451	Pichia pastoris	226458182	Micromonas pusilla 188
229367874	Anoplopora fimbria	254580643	Zygosaccharomyces rouxii	254573206	Pichia pastoris 125
240849438	Acyrtosiphon pisum	255633100	Glycine max	254584092	Zygosaccharomyces rouxii 108
255627133	Glycine max	255718239	Lachancea thermotolerans	255717240	Lachancea thermotolerans 131
260800377	Branchiostoma floridae	256086052	Schistosoma mansoni	255951050	Penicillium chrysogenum 168
281210911	Polysphondylium pallidum	260810887	Branchiostoma floridae	256727642	Nectria haematococca 158
281363363	Drosophila melanogaster	262098737	Phytophthora infestans	258569080	Uncinocarpus reesii 167
290997912	Naegleria gruberi	281204356	Polysphondylium pallidum	260944060	Clavispora lusitaniae 127
291228563	Saccoglossus kowalevskii	290985016	Naegleria gruberi	261353299	Verticillium albo-atrum 189
300169437	Selaginella moellendorffii	294659996	Debaryomyces hansenii	262112084	Phytophthora infestans 151
31980697	Mus musculus	297726007	Oryza sativa Japonica	270003414	Tribolium castaneum 125
4503955	Homo sapiens	300172237	Selaginella moellendorffii	281207674	Polysphondylium pallidum 192
47220703	Tetraodon nigroviridis	3152584	Arabidopsis thaliana	290987028	Naegleria gruberi 198
66814098	Dictyostelium discoideum	45191013	Ashbya gossypii	294659803	Debaryomyces hansenii 124
67479027	Entamoeba histolytica	50291031	Candida glabrata	295666097	Paracoccidioides brasiliensis 171
71416763	Trypanosoma cruzi	50305719	Kluyveromyces lactis	296425892	Tuber melanosporum 151
91090694	Tribolium castaneum	66811200	Dictyostelium discoideum	39971113	Magnaporthe oryzae 188
FUNGI (x27)					
116199867	Chaetomium globosum	6912674	Homo sapiens	45190845	Ashbya gossypii 129
119190001	Coccidioides immitis	71064114	Saccharomyces cerevisiae (YNL086W)	46135787	Gibberella zeae 180
119481547	Neosartorya fischeri			47208563	Tetraodon nigroviridis 231
145608262	Magnaporthe oryzae			50309819	Kluyveromyces lactis 124
146422819	Meyerozyma guilliermondii			51011035	Danio rerio 236
156043473	Sclerotinia sclerotiorum			6320564	Saccharomyces cerevisiae 122 (YDR357C)
156847365	Vanderwaltozyma polyspora			66813954	Dictyostelium discoideum 262
169624786	Phaeosphaeria nodorum			67903472	Aspergillus nidulans 167
171688892	Podospora anserina			68468893	Candida albicans 158
189198860	Pyrenophora tritici			76779596	Xenopus laevis 221
210075280	Yarrowia lipolytica			85113784	Neurospora crassa 181
242819386	Talaromyces stipitatus			8922952	Homo sapiens 217
255714152	Lachancea thermotolerans			isotig09812	Patiria miniata 197
255933021	Penicillium chrysogenum				
256735382	Nectria haematococca				
258565601	Uncinocarpus reesii 1704				
260949381	Clavispora lusitaniae				
261357041	Verticillium albo-atrum				
294656654	Debaryomyces hansenii				
295675039	Paracoccidioides brasiliensis				
296417511	Tuber melanosporum Mel28				
45190685	Ashbya gossypii				
46107490	Gibberella zeae				
50303437	Kluyveromyces lactis				
67540292	Aspergillus nidulans				
68482218	Candida albicans				
85079168	Neurospora crassa				

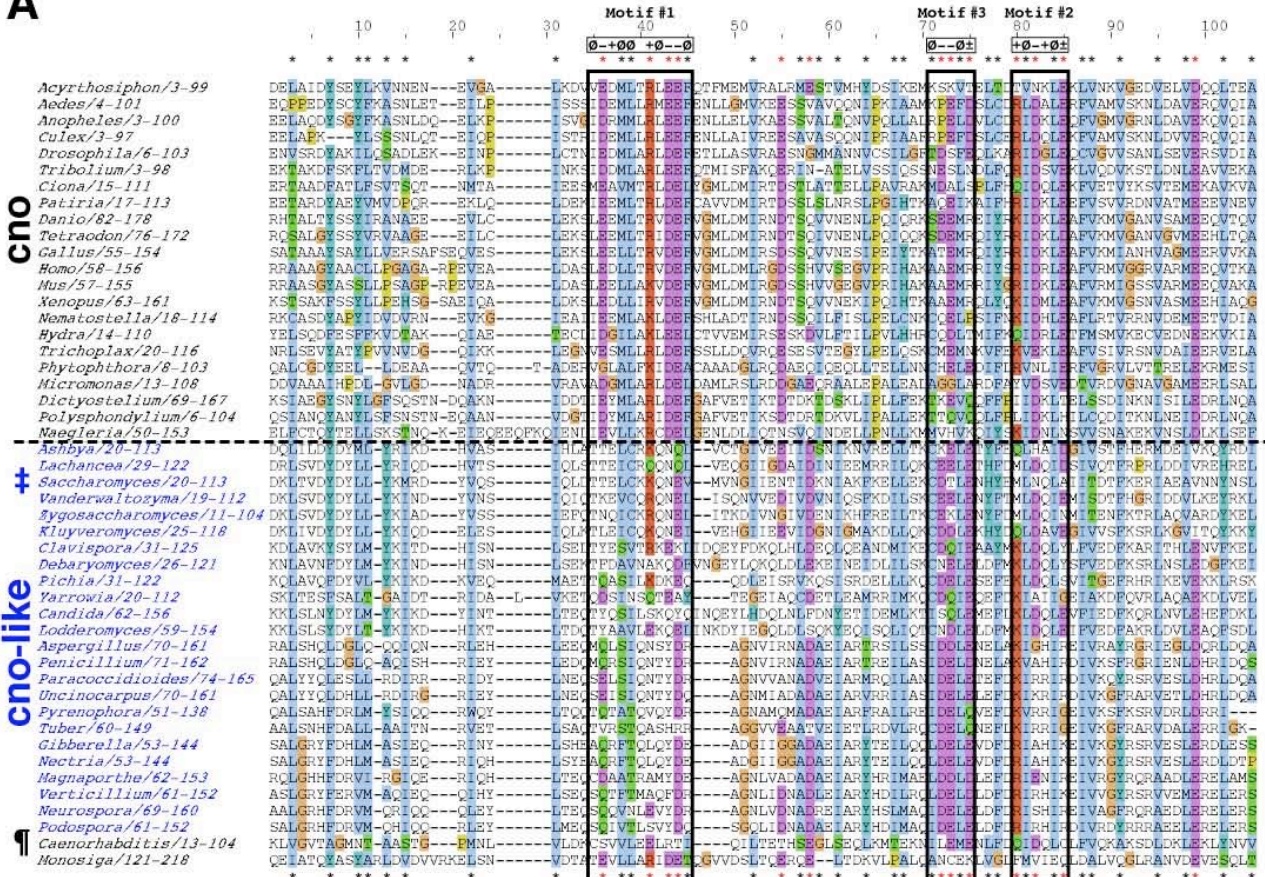
Table S4. Documented interactions within Vab2 complex

Documented interactions shown in Fig. 4 between the six components of the Vab2 complex identified by Krogan *et al.* (2006). In that paper, raw mass spectrometry data were analysed to produce a score of the probability that the interaction is genuine (= "prob"). ~7,100 core interactions scored over a threshold of $p = 0.273$. One interaction within this complex (Vab2p-Bli1p, indicated by ‡) was one of 7,000 further reactions that scored lower than the threshold – strength 0.261. Interactions that repeat one shown higher up the table are not numbered, with cross-referencing ("Xref") to the first occurrence indicated. References as indicated.

#	Bait	interactor	prob	Xref	ref
1.	Vab2p	Cnl1p	0.896		(1)
2.	Vab2p	Kxd1p	0.767		(1)
3.	Vab2p	Bls1p	0.683		(1)
4.	Vab2p	Snn1p	0.702		(1)
5.	Vab2p	Bli1	0.261‡		(1)
6.	Bls1p	Kxd1p	0.818		(1)
7.	Bls1p	Snn1p	0.647		(1)
8.	Cnl1p	Bli1p	0.748		(1)
9.	Bli1p	Kxd1p	0.404		(1)
10.	Bli1p	Bls1p	0.286		(1)
11.	Kxd1p	Snn1p			(2)
12.	Snn1p	Bli1p			(3)
	Snn1p	Bli1p		12	(4)
13.	Cnl1p	Kxd1p			(4)
	Vab2p	Kxd1p		2	(4)
	Vab2p	Snn1p		4	(4)
	Bli1p	Snn1p		12	(5)
	Cnl1p	Kxd1p		13	(5)

Supplementary Figures: Fig S1 (legend overleaf)

A



B

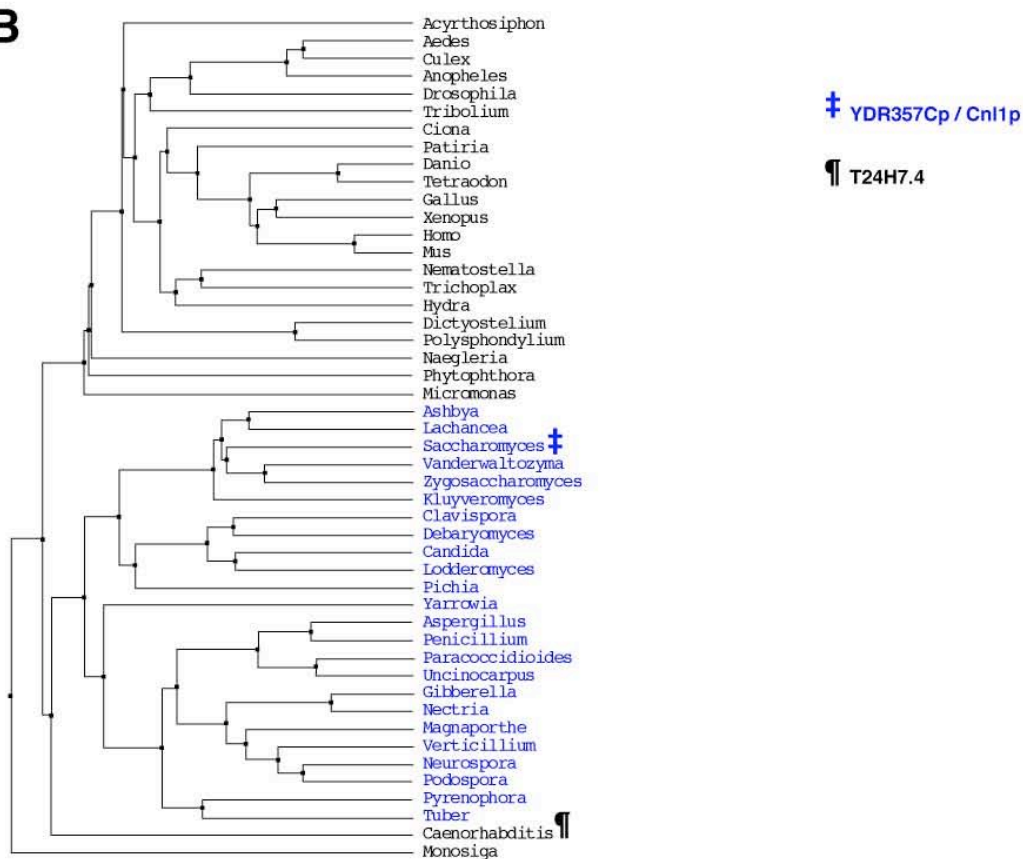


Figure S1. Relationship between cno homologues and fungal cno-like proteins.

(A). 24 cno homologues (in black, including T24H7.4[¶]) and 24 fungal cno-like homologues (in blue, including YDR357Cp/Cnl1p[‡]) were sorted in Kalign by average distance using BLOSUM62. The figure shows the region of maximal alignment across a central block of ~100 amino acids. The dotted line indicates that the two families separated almost completely (but see B). Residues at conserved positions are coloured: hydrophobic (blue), negatively charged (purple), positively charged (red), hydrophilic (T/S/N/Q: green), and tyrosine and histidine (turquoise). Also coloured are all prolines (yellow) and glycines (pink). Highly conserved positions are also indicated by asterisks (above and below, black if hydrophobic, red if charged). In addition, each of the three proposed motifs corresponding to those in Fig. 3 (Motifs #1 and #2 identified in cno⁽⁶⁾, and Motif #3 identified in this study) are boxed, and summarized above, where “Ø” is hydrophobic, and “+” “-“ & “±” are charged positive, negative or both. Motif #1 is only poorly preserved from cno (top half) into cno-like sequences (bottom half), but Motifs #2 and #3 are well conserved across both groups of sequences.

(B). Tree of the regions shown in (A), showing species. Cno homologues in nematode[¶] and the unicellular flagellate *Monosiga* are intermediate in character between the other cno sequences and the fungal cno-like family, indicating that the two main groups of sequences might have derived from a single ancestor. Sequences are listed in Table S3.

Supplementary References:

1. Krogan, NJ, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 2006; 440:637-43.
2. Chin, CH, SH Chen, CW Ho, MT Ko, and CY Lin. A hub-attachment based method to detect functional modules from confidence-scored protein interactions and expression profiles. *BMC Bioinformatics* 2010; 11 Suppl 1:S25.
3. Uetz, P, et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 2000; 403:623-7.
4. Ito, T, T Chiba, R Ozawa, M Yoshida, M Hattori, and Y Sakaki. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 2001; 98:4569-74.
5. Yu, H, et al. High-quality binary protein interaction map of the yeast interactome network. *Science* 2008; 322:104-10.
6. Cheli, VT and EC Dell'Angelica. Early origin of genes encoding subunits of biogenesis of lysosome-related organelles complex-1, -2 and -3. *Traffic* 2010; 11:579-86.