

**Deep origins of eukaryotic multicellularity revealed by the *Acrasis*
kona genome and developmental transcriptomes**

Sheikh et al.

Supplemental Data

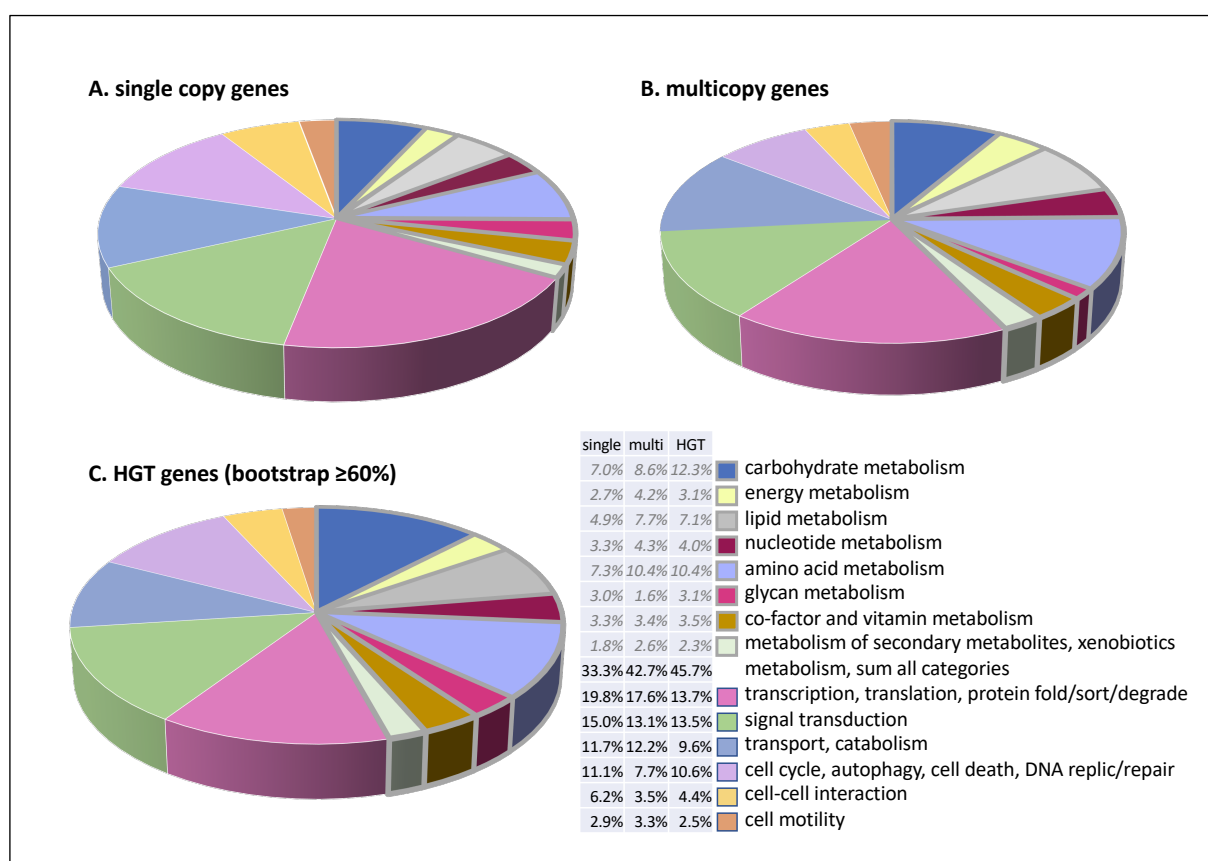


Figure S1: Predicted protein function for single-copy, multicopy and horizontal gene transfer (HGT) candidate accessions in *Acrasis kona*. Percentages of general functional categories are shown for *A. kona* predicted proteins for A) single copy, B) multicopy and 3) HGT candidate accessions. Functional categorization of predicted proteins was determined using GhostKOALA (Kanehisa et al. 2016).

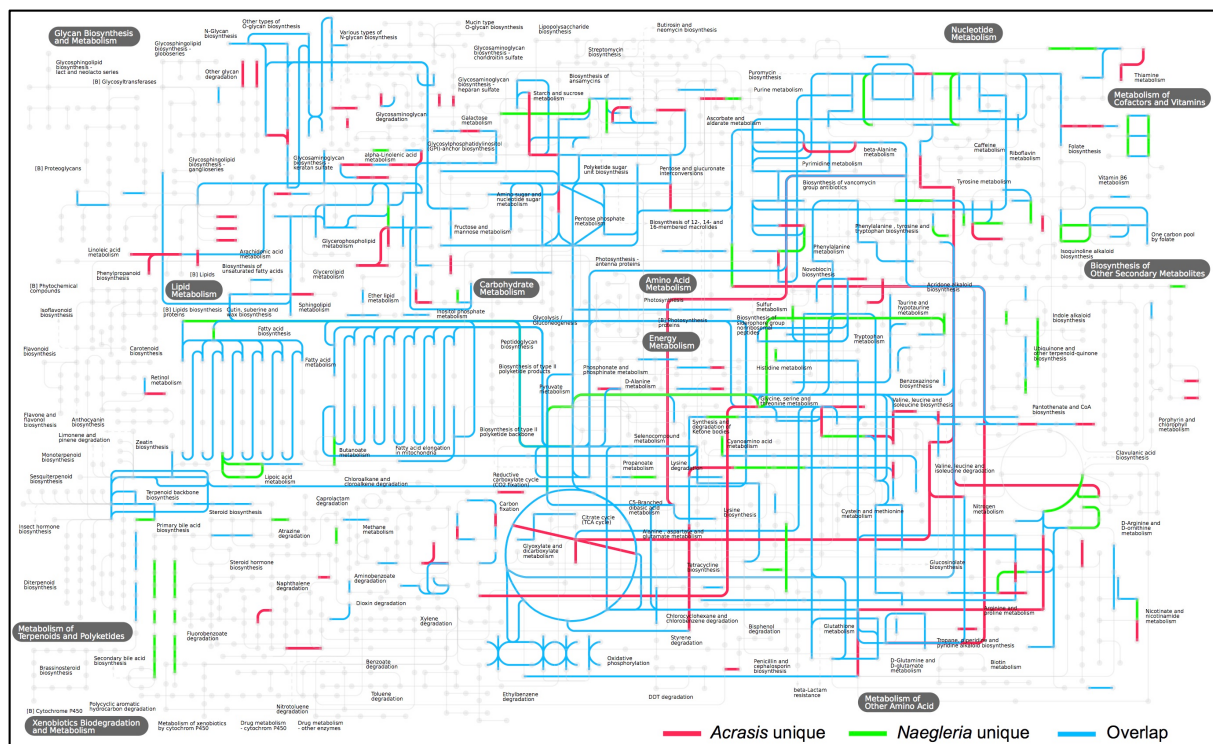
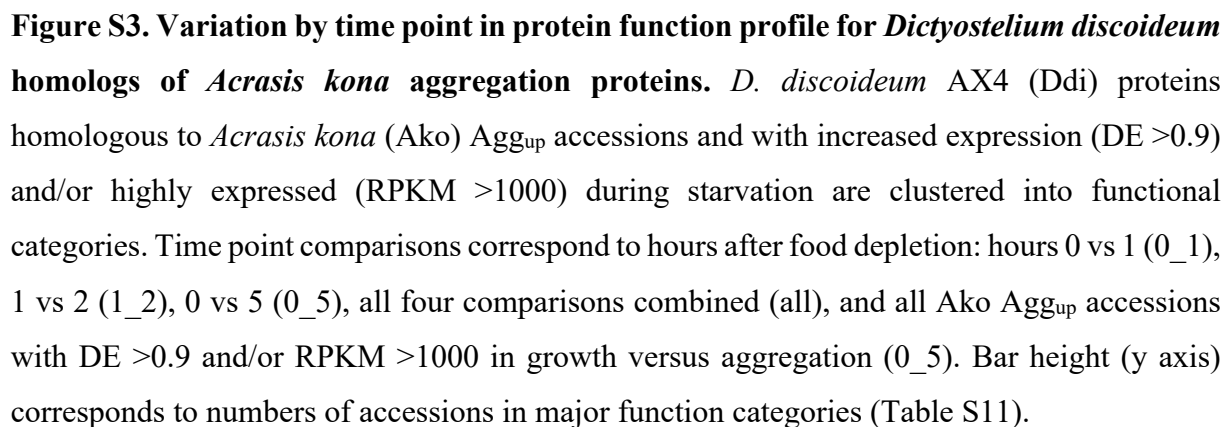


Figure S2. Metabolic capacity of *Acrisia kona* and *Naegleria gruberi*. Unique enzymatic pathways for *A. kona* and *N. gruberi* are shown in red and green, respectively, and the overlap between the two in blue. Pathways were predicted and graphically summarized using iPath (Letunic et al. 2008).



AK_13406

phosphoinositide phospholipase C (PI-PLC)
signaling, RTKc-RAS-PIP

Acrasis development				
	DE _{Agg}	DE _{Germ}	Reads _{Agg}	RPKM _{Agg}
AK_13406	2.79	-3.93	974	31.43
AK_04201	0.44	-1.61	373	12.76

Dictyostelium development			
DE: 1h	DE: 2h	DE: 5h	
-1.20	3.52	3.70	

Bacteria
Holozoa (animals)
Basidiomycota
Ascomycota
other Fungi
Discoba

RAxML x 416 sites

scale
0.2



Figure S4. An *Acrasis kona* aggregation-specific phosphoinositide phospholipase C derived by horizontal transfer (accession AK_13406). The tree shown was derived from a 415 amino acid alignment using RAxML and the LG+gamma model (Stamatakis 2006). Only bootstrap values over 60% are shown and branch lengths are drawn to scale according to the scale bar at the lower left (substitutions per site). Sequence names are colored according to higher order taxonomy shown in the lower left key. Some large groups of closely related sequences are collapsed to proportional triangles to increase readability of more relevant parts of the trees. The table to the immediate left shows gene expression values for growth vs aggregation (DE_{Agg}), aggregation vs germination. (DE_{Germ}) and numbers of raw reads (Read_{Agg}) and/or length-corrected reads (RPKM_{Agg}) for aggregating cells (Table S10). DE_{Log2} values for Ddi AX4 homologs are given for growth (0 hours) versus 1, 2 and 5 hours starvation (Santhanam et al. 2015).

location	structure, function	protein	Accession	Human vs:								Acrasis development					Dictyostelium development			
				Ako	Ngr	Cva	Ptr	Aan	Ddi	Sce	Spo	RPKM _{Gro}	RPKM _{Agg}	RPKM _{Gm}	DE _{Agg}	DE _{Gm}	RPKM ₀	RPKM ₅	DE _{0_1}	DE _{0_5}
universal	S1-KH cap	RPR6	AK_06097									121.57	111.68	124.00			10.49	8.21		
		CsI4	AK_08302									62.52	304.71	100.54	▲	▼	29.33	6.64	▼	▼
		RPR4	AK_08490									9.44	4.34	5.84			29.08	6.22	▼	▼
		RPR40	AK_02427									34.61	45.07	28.13			13.48	4.11		▼
			AK_09239									20.28	22.40	23.98						
	core (Exo9)	MTR3	AK_07336									20.08	16.56	6.92			31.71	5.83		▼
		Rpr41/ski6	AK_03026									10.25	11.82	6.92			28.15	6.24		▼
			AK_02663									18.21	18.09	12.78						
		RPR42	AK_15015									13.70	15.64	8.23			23.77	3.05		▼
			AK_03267									17.12	10.71	5.19						
nucleus	RNase	RPR43	AK_09379									6.11	3.16	4.36			14.50	1.37	▼	▼
		RPR45	AK_14128									43.30	49.90	37.56			13.47	5.26	▼	▼
		RPR46	AK_14924									88.50	88.63	203.14	▲		14.39	2.71	▼	▼
	helicase	Dis3/RPR44	AK_11571									35.11	41.70	35.33			13.04	2.71	▼	▼
		Mtr4	AK_07891									2.98	8.41	6.74	▲		16.26	2.93	▼	▼
			AK_01502									5.78	15.64	13.20	▲					
cytoplasm	RNase	Dis3L/RPR44	AK_09867									4837.59	4807.24	4852.46			1.90	2.30		
																	7.61	10.40	▲	
																	9.36	14.93		
	Ski complex	coupling	ski7/HBS1-like	AK_07586									28.73	19.22			46.50	96.25		▲
			AK_02870										11.81	0.77	▼					
			AK_09893										37.73	1.33	▼		50.22	51.15		
RNAi	dicer	dsRNA RNase	AK_13723									10.45	6.85	10.37			7.15	24.79		▲
		AK_08903										14.71	7.12	10.60						
	TAF11	TATA factor 11	AK_05317									17.76	13.89	10.26			8.25	14.29		
	RdRP	RNA-RNA pol	AK_05852									134.68	119.76	281.16	▲		0.53	3.98	▲	▲
		AK_07172										22.91	13.72	2.00	▼		1.21	2.99		
argonaut	RISC RNase																1.40	2.72	▲	▲
	piwi	RISC RNase	AK_08445									58.08	81.45	72.37			10.26	1.68	▼	▼
		AK_04633										144.98	134.38	137.92			21.15	101.79	▲	▲
																	3.27	0.83	▼	▼
																1.05	0.54	▼	▼	

BLASTp e-value

- 0-100
- 99-70
- 69-40
- 39-25
- 24-10, cvg> 50%
- <10 / annotation

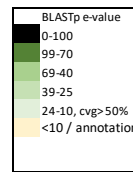


Figure S5. Sequence conservation and developmental expression of the *Acrasis kona* exosome. Homologs of exosomal proteins were identified by BLASTp using human queries. The strength of BLASTp hits (c-value) is indicated by color intensity (key at bottom right). BLASTp hits with high e-values (e^{-24} to e^{-10}) were scored as present only if their BLASTp alignments included >50% query coverage (cvg). Exosomal components annotated as present but with only very weak BLASTp scores ($> e^{-10}$) are indicated in pale yellow. RNAseq length-corrected read numbers (RPKM) are shown for actively growing (RPKM_{Gro}), aggregating (RPKM_{Agg}) and germinating (RPKM_{Germ}) *A. kona* cells and similarly for growth and aggregation in *Dictyostelium discoideum* AX4 (RPKM₀ and RPKM₅, respectively). Green and red arrows indicate substantially increased or decreased (DE >0.9 or DE <-0.9, respectively) gene expression for growth vs aggregation (DE_{Agg}) and aggregation vs germination (DE_{Germ}) in *A. kona* and similar time points for Ddi AX4, i.e., 0 vs 1 (DE_{0_1}) and 0 vs 5 (DE_{0_5}) hours of starvation. Accession numbers, BLASTp e-values and gene expression values are given in Table S14. Taxon names are abbreviated as follows: *A. kona* (Ako), *Aureococcus anophagefferens* (Aan), *Chlorella sp.* (Cva), *Ddi AX4* (Ddi), *Naegleria sp.* (Ngr), *Phaeodactylum tricornutum* (Ptr), *Saccharomyces cerevisiae* S288c (Sce) and *Schizosaccharomyces pombe* (Spo). Ddi AX4 gene expression values were derived from Santhanam et al. (2015).

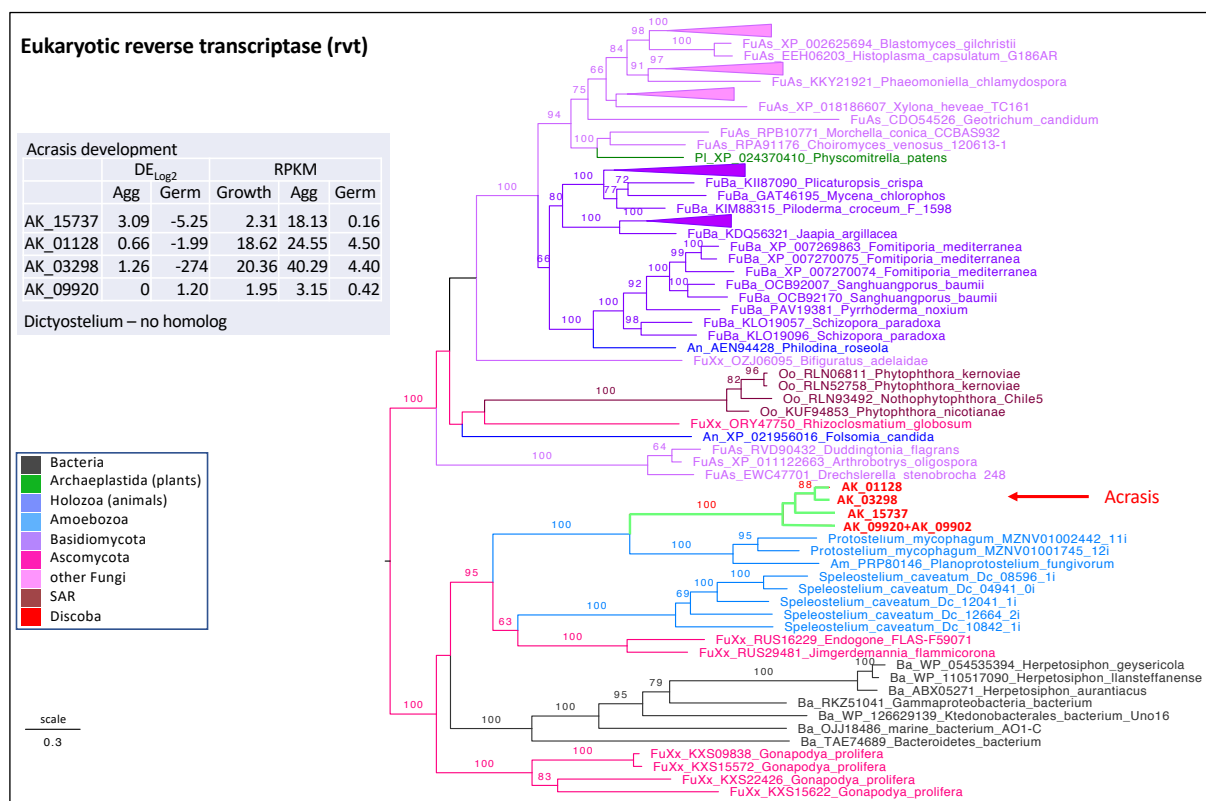


Figure S6. A family of reverse transcriptases in *Acrasis kona*. The tree shown was constructed from a 628 amino acid alignment using RAXML and the LG+gamma model (Stamatakis 2006). Only bootstrap values over 60% are shown and branch lengths are drawn to scale according to the scale bar at the bottom left (substitutions per site). Sequence names are colored according to higher order taxonomy according to the key to the lower left of the tree. Some large groups of closely related sequences are collapsed as triangles to increase readability of more relevant parts of the trees. The table to the upper left shows gene expression values for growth vs aggregation (DE_{Agg}), aggregation vs germination. (DE_{Germ}), and numbers of length corrected (RPKM_{Agg}) reads (Table S10). DE values for Ddi AX4 homologs are given for growth (0 hours) versus 1, 2 and 5 hours starvation (Santhanam et al. 2015).

Acrasis		Human vs:						Acrasis development			Dictyostelium development					
accession		Ako	Ngr	Aan	Cva	Ddi	Spo	Sce	RPKM ₀	RPKM ₅	DE _{0_5}	RPKM ₀	RPKM ₅	DE _{0_1}	DE _{2_5}	DE _{0_5}
Anaphase Promoting Complex																
Apc1	AK_09415								8.00	8.55		0.49	0.50			
Apc2	AK_01014								34.61	20.46		2.11	2.23			
Apc3	AK_05309								31.51	44.02		3.77	1.74	▲	▼	▼
Apc5	AK_09574								9.86	14.19		2.58	2.73			
Apc6	AK_13416								31.55	40.05		3.69	1.80	▲	▼	▼
Apc8	AK_14201								19.46	33.54	▲	19.43	19.55			
Apc10	AK_12692								0.00	0.00		3.58	1.40		▼	▼
Apc11	AK_13097						query		14.51	24.45		31.83	39.04	▲		
	AK_01639						query		8.45	1.12	▼	441.27	968.69	▲		▲
	AK_09363								24.80	7.04	▼					
Centromere																
Aurora	AK_01925								16.01	13.12		29.60	26.02	▲	▼	
	AK_03728								1.96	9.58	▲					
Incenp	AK_13745						query		0.89	9.40	▲	5.33	2.83	▲	▼	▼
CenpA	AK_00547						query		2.86	0.18	▼	5.60	3.65			
	AK_02433								1.48	0						
	AK_03408								6.52	9.11						
	AK_13776								3.16	1.01						
CenpE	AK_13279								3.73	4.43		2.03	1.29	▲	▼	
Inner Kinetochore																
Plk1	AK_05408								0	0		8.20	3.51		▼	▼
	AK_08166								6.63	43.68	▲					
Skp1	AK_05396								13.35	1.86	▼	722.01	1118.11			
	AK_09381								6.45	0.62	▼					
Outer Kinetochore																
Bub3	AK_09515								8.27	13.86		40.81	30.32		▼	
												44.15	130.97		▲	▲
Cdc20	AK_03968								3.60	8.68	▲	68.03	46.32		▼	
	AK_11901								6.75	1.57	▼					
Cdh1												6.63	21.26			▲
HORMA	AK_07587								14.13	3.43	▼					
Mad1	AK_10418								39.82	23.29		4.02	5.60		▲	
Mad2	AK_12843								12.18	28.79	▲	0	0			
MadBub	AK_04791								6.90	4.93		7.66	4.25	▲	▼	▼
	AK_10725								3.09	4.27						
Mps1	AK_15471								9.73	17.18	▲	4.29	2.08		▼	▼
RINT1	AK_06546								12.36	15.05		3.22	8.16		▲	▲
TRIP13	AK_04538								10.56	9.94		11.81	27.06		▲	▲
Nuclear Pore Proteins																
Nup107	AK_07456								0	0		10.38	6.80			
Nup160	AK_07444								102.45	101.56		3.09	2.47			

BLAST e-value

0-100

99-70

69-40

39-25

24-10 + cvg>50%

<10 + annotation

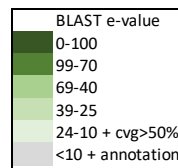


Figure S7. The *Acrasis kona* kinetochore and life cycle-stage specific expression patterns. Homologs of kinetochore proteins were identified by BLASTp using *Homo sapiens* sequences as queries. The strength of BLASTp hits (e-value) is indicated by color intensity according to the key at the bottom right. BLASTp hits with high e-values (e^{-24} to e^{-10}) were scored as present only if their BLASTp alignments included >50% query coverage (cvg). RNAseq length-corrected read numbers (RPKM) are shown for *A. kona* for actively growing (RPKM₀) and aggregating (RPKM₅) cells and for similar time points for *Dictyostelium discoideum* AX4 (growth: RPKM₀, aggregation: RPKM₅). Green and red arrows indicate substantially increased or decreased gene expression (DE >0.9, DE <-0.9, respectively) for both *A. kona* and *D. discoideum* for growth vs aggregation (DE_{0_5}). Accession numbers, BLASTp e-values and gene expression values are given in Table S15. Taxon names are abbreviated as follows: *A. kona* (Ako), *Aureococcus anophagefferens* (Aan), *Chlorella variabilis* (Cva), *D. discoideum* AX4 (Ddi), *Naegleria gruberi* (Ngr), and *Saccharomyces cerevisiae* S288c (Sce). *D. discoideum* AX4 gene expression values were derived from Santhanam et al. (2015).

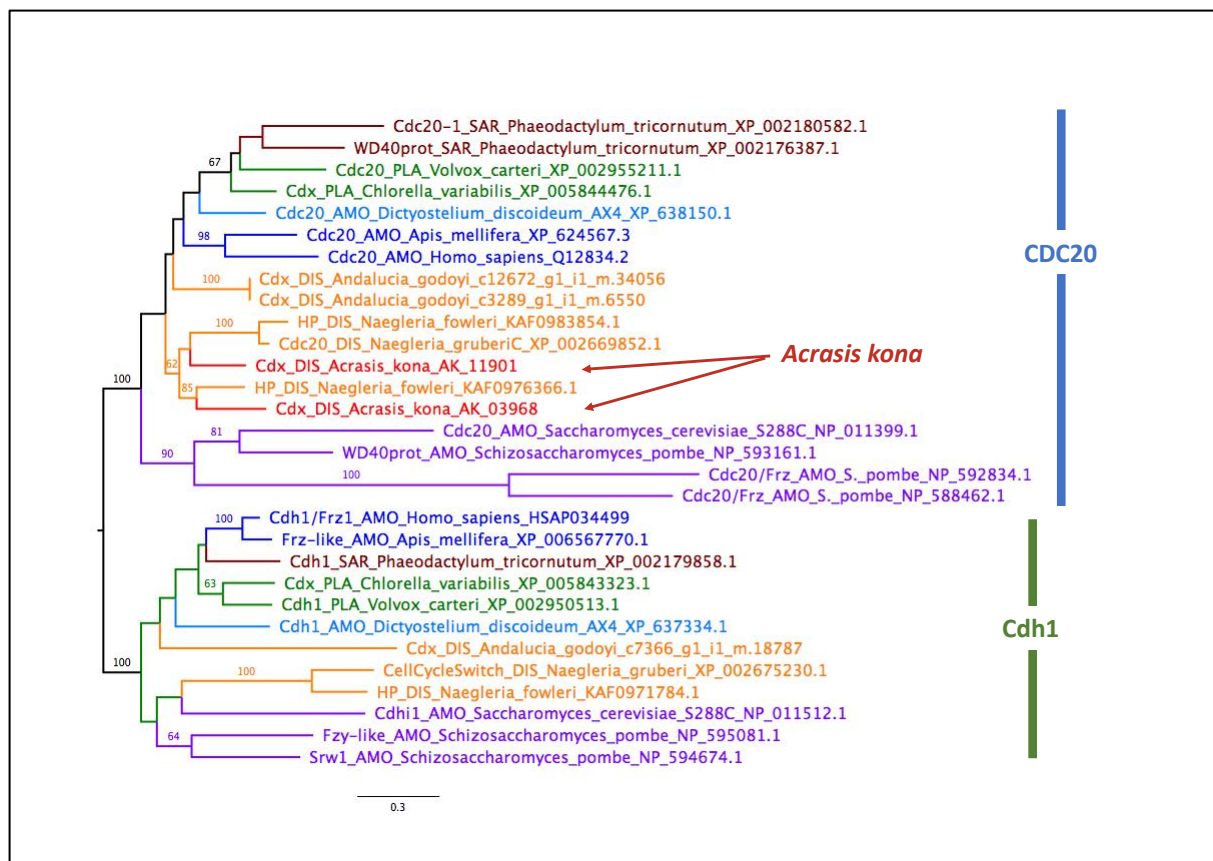


Figure S8. Both of the autophagy control proteins of *Acrasis kona* are of the CDC20-type. Protein sequences were identified by taxon-specific BLASTp using human protein sequence queries, and all hit sequences with e -values $< e^{-35}$ were retrieved. Sequences were aligned and trimmed using the Auto-1 algorithm of Trim-A1 (Capella-Gutierrez et al. 2009). The tree was constructed from the resulting 306 aligned positions using RAxML with the LG+gamma amino acid substitution model and bootstrapping with automated stopping (Stamatakis 2006). Sequences are labeled according to their primary annotation (otherwise denoted as hypothetical proteins or HP), and major taxonomy as follows: SAR (Stramenopila, Alveolata and Rhizaria), PLA (Archaeplastida), DIS (Discoba), Amo (Amorphea). Numbers on branches show bootstrap values over 60%. The tree is drawn to scale according to the scale at the bottom (substitutions per site).

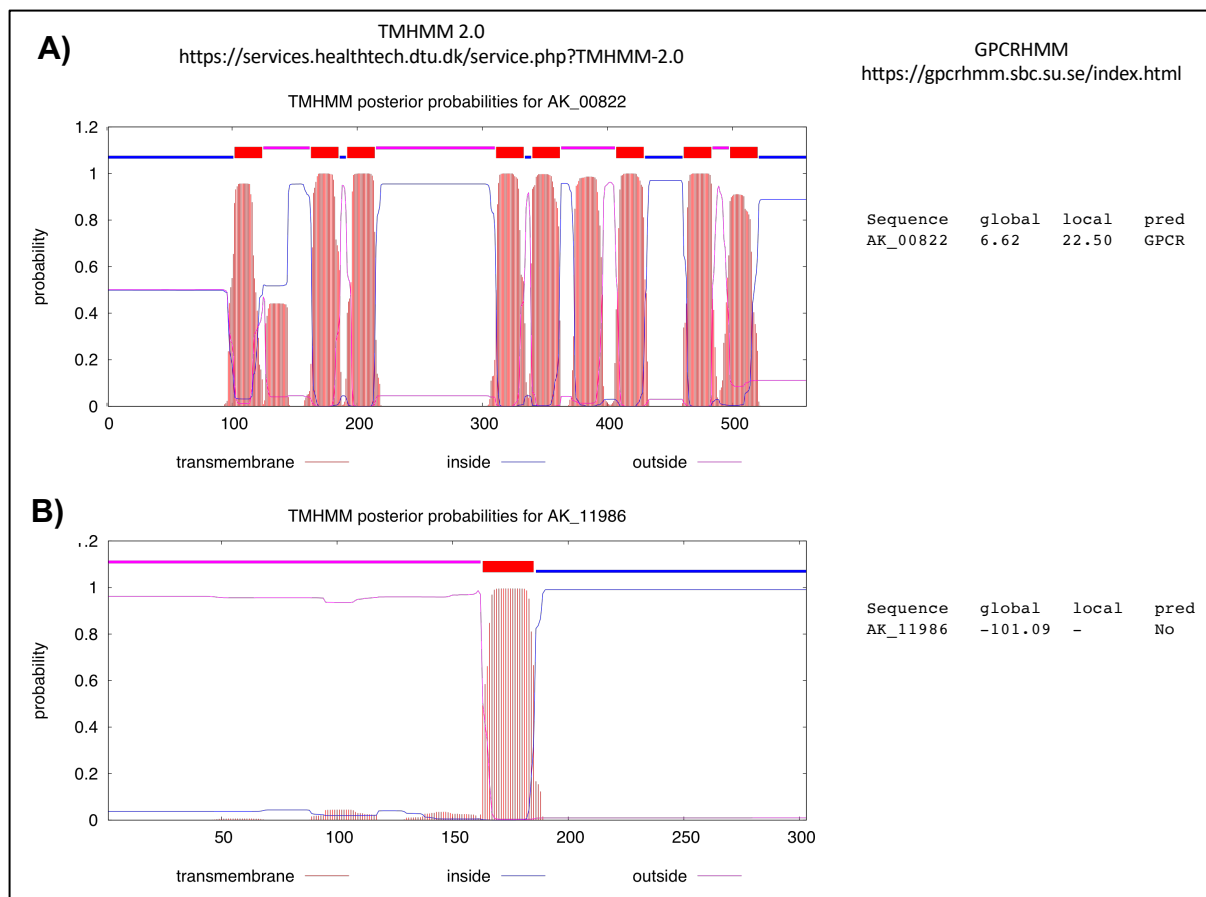


Figure S9. Two extracellular signaling proteins predicted from novel *Acrasis kona* aggregation accessions.. All unannotated *A. kona* aggregation proteins with predicted transmembrane domains (Table S3) were analyzed using the Hidden Markov Model based GPCR predictor TMHMM 2.0 (services.healthtech.dtu.dk/service.php?TMHMM-2.0; Wistrand et al. 2006). A) AK_00822 is strongly predicted to encode a classic GPCR with 8-9 transmembrane domains and large loops on both sides of the plasma membrane. B) AK_11986 is predicted to encode a membrane signaling or receptor protein with a single membrane spanning domain and large domains on both external and internal membrane surfaces.

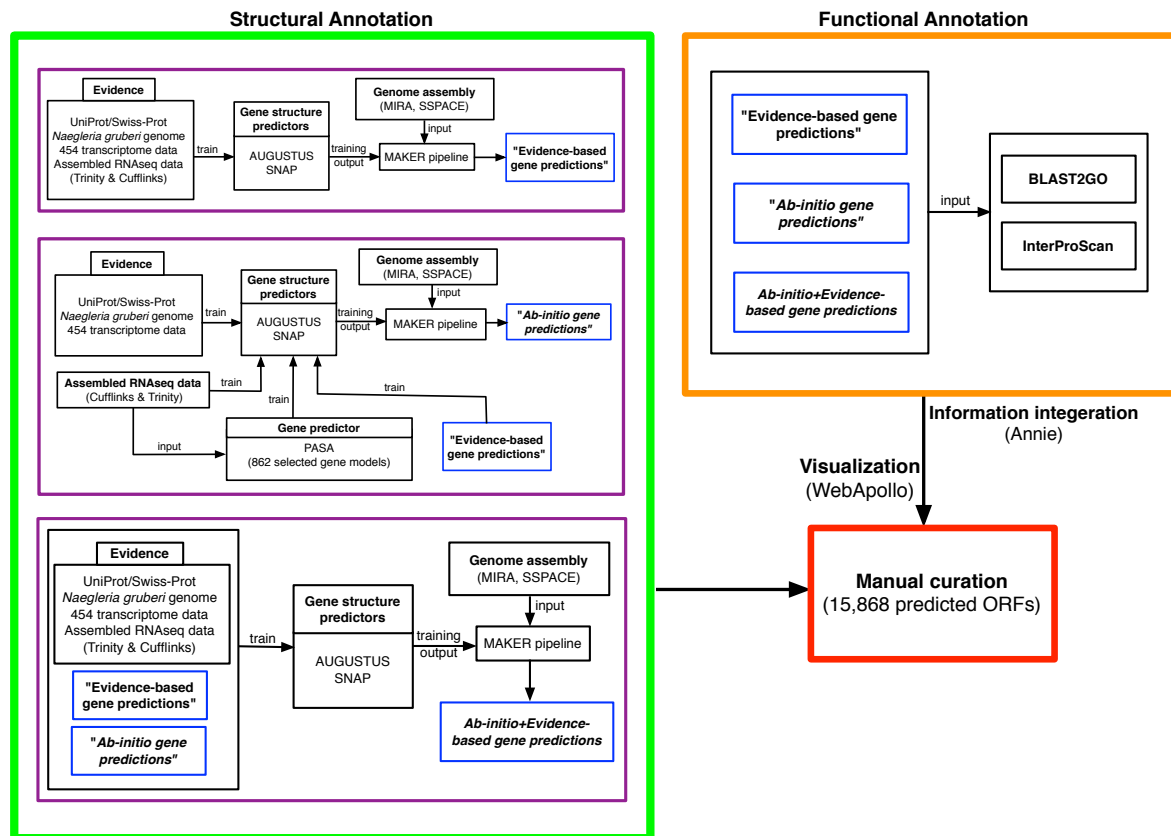


Figure S10. *Acrasis kona* genome annotation pipeline. The pipeline shown was used to perform structural (left) and functional (right) annotation followed by manual curation of the *A. kona* nuclear genome. Three different strategies were used to train gene predictors for structural annotation to produce evidence-based gene predictions, *ab-initio* gene predictions and evidence+*ab-initio* gene predictions.