

# Supplementary Information

## Missing Genes, Multiple ORFs and C-to-U Type RNA Editing in *Acrasis kona*

### (Heterolobosea, Excavata) Mitochondrial DNA

Cheng-Jie Fu<sup>\*1</sup>, Sanea Sheikh<sup>1</sup>, Wei Miao<sup>2</sup>, Siv G. E. Andersson<sup>3</sup>, Sandra L. Baldauf<sup>\*1</sup>

<sup>1</sup> Program in Systematic Biology, Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Uppsala 75236, Sweden

<sup>2</sup> Key Laboratory of Aquatic Biodiversity and Conservation, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

<sup>3</sup> Department of Molecular Evolution, Cell and Molecular Biology, Science for Life Laboratory, Biomedical Centre, Uppsala University, Uppsala 75105, Sweden

## Contents

<b>A. Supplementary Tables</b> .....	<b>3</b>
<b>B. Supplementary Figures</b> .....	<b>9</b>
<b>C. References</b> .....	<b>19</b>

## Supplementary Tables List

Table S1. Primers used in this study.

Table S2. Pair-wise sequence identity of mitochondrial protein-coding genes (OXPHOS pathway) in representative discobids.

Table S3. GC content of mtDNA encoded and 11 putative nuclear encoded mitochondrial protein (NcMt) genes in *Acrasis kona*.

Table S4. Posterior predictive tests of compositional heterogeneity in concatenated alignments of 19 mitochondrial proteins.

Table S5. Comparison of 11 putative nuclear DNA encoded mitochondrial proteins (NcMt) in *Acrasis kona* and their counterparts in the *Naegleria* spp. mtDNAs.

## Supplementary Figures List

Figure S1. A schematic map of gap-closing of *Acrasis kona* mtDNA based on long-rang PCR.

Figure S2. Tandemly repeated regions in *Acrasis kona* mtDNA.

Figure S3. Comparison of ribosomal protein operon gene order among representative  $\alpha$ -proteobacteria and jakobids.

Figure S4. A split *rps3* gene in *Acrasis kona* mtDNA.

Figure S5. Assembly coverage plots for the 11 putative nuclear encoded mitochondrial protein (NcMt) genes in the *Acrasis kona* nuclear contigs.

Figure S6. Phylogenies of 11 putative *Acrasis kona* nuclear encoded mitochondrial proteins.

Figure S7. Bayesian phylogeny of the E/E+/DYW domain of the DYW-type PPR proteins.

**Table S1. Primers used in this study.**

Name	Primer (5'-3')
<b>GAP_CLOSE(1-4)</b>	
GAP1_F1	TGGGTGCAGCATTTTTAGGACA
GAP1_R1	TCACATTCACCAGCTTGATCACAT
GAP1_nested_F2	CATCAATGGTATTGGGGTTATGG
GAP1_nested_R2	TTGTAGCACATGCAATTACTGGTTT
GAP2_F1	TGTGATCAAGCTGGTGAATGTGATTT
GAP2_R1	GCACCTGCACCTTTGGTTTTCC
GAP2_nested_F2	TGATCAAGCTGGTGAATGTGA
GAP2_nested_R2	TCCTCCATGCCACTTCCACTTT
GAP3_F1	GGAAAACCAAAGGTGCAGGTGCT
GAP3_R1	TGCCCTGTTGTATCTGTTGCTC
GAP3_nested_F2	TGATAGTCAAGCAGAAAGAGCA
GAP3_nested_R2	TCTAATATGTTGTTTAATTTTTCACT
GAP4_F1	GCAACAGATACAACAGGGGCAAGA
GAP4_R1	CGAACTCCCGACTTTTGGAGTGA
GAP4_nested_F2	AGAGTGTTAGGTCCAGTATGTACAGA
GAP4_nested_R2	TGCTTTCTATTATCTTCCAATTTACCT
<b>RT-PCR</b>	
nad1_est_F1	TCTGTTCAAAGAAGAAGAGGACCA
nad1_est_R1	AGCAAACCCATTCCACTAAA
nad1_est_F2	GCCGATAGCTGATGGTGTAATA
nad1_est_R2	AAAGCAAAACCCATTCCACT
atp6_est_F1	TTATTATTTGCACAAGGAG
atp6_est_R1	TTTGAGAAAATAGTAGATAAAA
atp6_est_F2	ATTGTTTGCTAATTTAGTGG
atp6_est_R2	ATCTTTTGAGAAAATAGTAG
cob_est_F1	TATTATGGTCAAGTGGTGTT
cob_est_R1	TCTGTATGTGATAAATATTCTGG
cob_est_F2	TGCATAGTAATGGTGCAAGT
cob_est_R2	TATTATAACCACCCCATACCCAA
cox3_est_F1	AGAGAGTGTTAAAGGTTGGTTT
cox3_est_R1	TGAAATCCATGAAATCCTGTT
cox3_est_F2	TGTTAGTAGCTTTAATAGAGAGTG
cox3_est_R2	ACAACCTAATCCCATTATCACAGA

**Table S2. Pair-wise sequence identity of mitochondrial protein-coding genes (OXPHOS pathway) in representative discobids.**

Gene	Sequence Identity (%)			
	Ak/Nf <sup>1</sup>	Ng/Nf	Ag/Jl	Ra/Ha
<i>atp1</i>	49.6	93.8	73.8	86.9
<i>atp3</i>	28.8	78.8	51.1	76.2
<i>atp6</i>	45.05	80.4	64.2	87.1
<i>atp8</i>	21.3	77.3	28.9	74.4
<i>atp9</i>	64.6	98.6	78.1	98.7
<i>cox1</i>	59.05	93	75	89.3
<i>cox2</i>	25.4	86.5	66.3	68.9
<i>cox3</i>	30.4	86.5	66	85
<i>cob</i>	44.65	74	70.7	80.8
<i>nad1</i>	43.9	85.3	70.4	90.1
<i>nad2</i>	30.2	59.6	47.8	80.1
<i>nad3</i>	38.8	86.7	68.1	92.4
<i>nad4</i>	51.15	74.7	68.9	83
<i>nad4L</i>	43.75	90.8	75.5	93.1
<i>nad5</i>	45.25	75.7	63.6	87.8
<i>nad7</i>	/	91.1	75.7	90.5
<i>nad6</i>	30.1	75.1	50.5	74.6
<i>nad8</i>	54.4	91.2	78.4	90.7
<i>nad9</i>	/	74.2	46.3	74.5
<i>nad11</i>	31.25	62.9	52.8	72.3

<sup>1</sup>: Ak: *Acrasis kona*; Ng: *Naegleria gruberi*; Nf: *Naegleria fowleri*; Ag: *Andalucia godoyi*; Ha: *Histiona aroides*; Jl: *Jakoba libera*; Ra: *Reclinomonas americana*.

**Table S3. GC content of mtDNA encoded and 11 putative nuclear encoded mitochondrial protein (NcMt) genes in *Acrasis kona*.**

mtDNA encoded proteins	Length (aa)	GC	GC1	GC2	GC3	ncDNA encoded proteins	Length (aa)	GC	GC1	GC2	GC3
<i>ORF149</i>	149	0.116	0.134	0.141	0.074	<i>RPS10</i>	227	0.398	0.445	0.33	0.419
<i>ORF603</i>	603	0.102	0.109	0.134	0.061	<i>RPL11</i>	200	0.395	0.485	0.400	0.300
<i>ORF246</i>	246	0.081	0.061	0.114	0.069	<i>RPL2</i>	256	0.422	0.488	0.461	0.316
<i>ORF802</i>	802	0.086	0.105	0.103	0.051	<i>RPS19</i>	99	0.377	0.354	0.364	0.414
<i>ORF1592</i>	1592	0.186	0.205	0.190	0.163	<i>RPL16</i>	194	0.400	0.412	0.371	0.418
<i>ORF376</i>	376	0.188	0.277	0.215	0.072	<i>RPS12</i>	172	0.403	0.430	0.448	0.331
<i>ORF61</i>	61	0.109	0.148	0.098	0.082	<i>RPS2</i>	254	0.419	0.425	0.331	0.500
<i>ORF399</i>	399	0.183	0.278	0.190	0.080	<i>NAD7</i>	432	0.464	0.475	0.373	0.544
<i>ORF253</i>	253	0.158	0.292	0.123	0.059	<i>NAD9</i>	195	0.400	0.451	0.297	0.451
<i>ORF69</i>	69	0.121	0.188	0.130	0.043	<i>SDH2</i>	318	0.437	0.406	0.374	0.531
<i>NAD1</i>	380	0.214	0.268	0.289	0.084	<i>COX11</i>	283	0.395	0.445	0.350	0.389
<i>NAD2</i>	518	0.133	0.166	0.195	0.039	Average	/	0.417	0.445	0.370	0.436
<i>NAD3</i>	124	0.161	0.218	0.202	0.065						
<i>NAD4</i>	465	0.164	0.196	0.237	0.060						
<i>NAD4L</i>	88	0.155	0.159	0.227	0.080						
<i>NAD5</i>	637	0.152	0.196	0.209	0.052						
<i>NAD6</i>	230	0.161	0.230	0.187	0.065						
<i>NAD8</i>	143	0.233	0.336	0.322	0.042						
<i>NAD11</i>	591	0.148	0.198	0.193	0.054						
<i>COB</i>	478	0.196	0.251	0.282	0.054						
<i>COX1</i>	663	0.248	0.324	0.326	0.095						
<i>COX2</i>	307	0.228	0.332	0.290	0.062						
<i>COX3</i>	316	0.190	0.282	0.222	0.066						
<i>ATP1</i>	557	0.266	0.357	0.343	0.097						
<i>ATP3</i>	333	0.172	0.228	0.207	0.081						
<i>ATP6</i>	261	0.143	0.188	0.199	0.042						
<i>ATP8</i>	126	0.124	0.127	0.167	0.079						
<i>ATP9</i>	72	0.296	0.361	0.431	0.097						
<i>RPS3_a</i>	966	0.124	0.130	0.115	0.125						
<i>RPS3_b</i>	1133	0.092	0.117	0.102	0.058						
<i>RPS7</i>	197	0.142	0.157	0.157	0.112						
<i>RPS8</i>	124	0.124	0.169	0.153	0.048						
<i>RPS13</i>	136	0.137	0.169	0.162	0.081						
<i>RPS14</i>	98	0.201	0.194	0.276	0.133						
<i>RPL5</i>	274	0.145	0.208	0.172	0.055						
<i>RPL6</i>	226	0.165	0.199	0.168	0.128						
<i>RPL14</i>	121	0.240	0.289	0.339	0.091						
Average	/	0.160	0.203	0.195	0.082						

**Table S4. Posterior predictive tests of compositional heterogeneity in concatenated alignments of 19 mitochondrial proteins.**

Taxa	Original dataset			Recoding using 6 Dayhoff groups			GenBank accession number
	z-score	p-value	Significance	z-score	p-value	Significance	
<i>Acrasis kona</i>	19.041	0	*	21.938	0	*	KJ679272
<i>Naegleria gruberi</i>	13.514	0	*	18.449	0	*	AF288092
<i>Acanthamoeba castellanii</i>	16.058	0	*	19.052	0	*	NC_001637
<i>Agrobacterium radiobacter</i>	8.507	0.004	*	4.673	0.01	*	NC_011985, NC_011983
<i>Allomyces macrogynus</i>	6.614	0	*	3.017	0.006	*	NC_001715
<i>Andalucia godoyi</i>	5.544	0.006	*	-1.126	0.896		KC353352
<i>Azospirillum brasilense</i>	8.385	0	*	4.941	0.004	*	NC_016617
<i>Bigelowiella natans</i>	7.994	0.008	*	6.811	0.004	*	HQ840955
<i>Bradyrhizobium japonicum</i>	7.958	0.008	*	6.435	0.004	*	NC_004463
<i>Capsaspora owczarzaki</i>	8.574	0	*	5.958	0.008	*	KC573038
<i>Chaetosphaeridium globosum</i>	2.876	0.024	*	1.479	0.084		NC_004118
<i>Chattonella marina</i>	10.989	0	*	0.896	0.158		NC_013837
<i>Chondrus crispus</i>	9.947	0	*	3.13	0.006	*	NC_001677
<i>Cyanidioschyzon merolae</i>	13.561	0	*	7.577	0	*	NC_000887
<i>Cyanophora paradoxa</i>	15.842	0	*	6.322	0.004	*	NC_017836
<i>Desmarestia viridis</i>	5.165	0.008	*	1.21	0.116		NC_007684
<i>Dictyostelium discoideum</i>	15.055	0	*	19.753	0	*	NC_000895
<i>Ehrlichia canis</i>	11.884	0	*	5.829	0.002	*	NC_007354
<i>Emiliana huxleyi</i>	6.772	0	*	1.073	0.132		NC_005332
<i>Geodia neptuni</i>	3.431	0	*	3.624	0.002	*	NC_006990
<i>Glaucocystis nostochinearum</i>	12.845	0	*	9.692	0	*	NC015117
<i>Hartmannella vermiformis</i>	10.794	0.002	*	11.958	0	*	NC_013986
<i>Hemiselmis andersenii</i>	4.129	0.004	*	0.124	0.368		NC_010637
<i>Histiona aroides</i>	3.225	0.006	*	0.08	0.432		KC353353
<i>Jakoba bahamiensis</i>	2.603	0.026	*	-0.994	0.842		KC353354
<i>Jakoba libera</i>	6.96	0.006	*	0.753	0.216		KC353355
<i>Magnetospirillum magnetotacticum</i>	7.81	0.006	*	7.676	0.004	*	NZ_AAAP0000 0000
<i>Magnetococcus</i> sp.	8.127	0.008	*	4.886	0.01	*	NC_008576
<i>Malawimonas californiana</i>	14.887	0	*	9.162	0	*	unpublished
<i>Malawimonas jakobiformis</i>	15.177	0	*	5.252	0	*	NC_002553
<i>Marchantia polymorpha</i>	1.434	0.082		0.887	0.174		NC_001660
<i>Mezorhizobium loti</i>	6.306	0.008	*	5.896	0.008	*	NC_002682, NC_002679, NC_002678
<i>Mesostigma viride</i>	9.064	0.002	*	1.211	0.116		NC_008240
<i>Micavibrio aeruginosavorus</i>	9.456	0.008 <sup>6</sup>	*	3.445	0.018	*	NC_016026

<i>Midichloria mitochondrii</i>	4.18	0.008	*	0.146	0.34		NC_015722
<i>Monosiga brevicollis</i>	5.684	0.002	*	2.249	0.034	*	NC_004309
<i>Neorickettsia sennetsu</i>	4.089	0.004	*	3.545	0.008	*	NC_007798
<i>Nephroselmis olivacea</i>	1.882	0.042	*	1.205	0.11		NC_008239
<i>Novosphingobium aromaticivorans</i>	6.526	0.016	*	9.517	0.004	*	NC_007794
<i>Orientia tsutsugamushi</i> str. Ideda	17.564	0	*	8.216	0	*	NC_010793
<i>Pavlova lutheri</i>	6.215	0.002	*	-0.773	0.784		HQ908424
<i>Phytophthora infestans</i>	14.909	0	*	14.167	0	*	NC_002387
<i>Porphyra purpurea</i>	4.654	0	*	-0.552	0.716		NC_002007
<i>Prototheca wickerhamii</i>	1.884	0.04	*	0.96	0.168		NC_001613
<i>Reclinomonas americana</i>	1.288	0.106		-0.473	0.652		NC_001823
<i>Rhizopus oryzae</i>	3.762	0.014	*	4.7	0.008	*	NC_006836
<i>Rhodospirillum rubrum</i>	10.257	0.002	*	8.326	0	*	NC_007643
<i>Rhodospirillum centenum</i>	9.187	0.004	*	5.575	0.006	*	NC_011420
<i>Rhodomonas salina</i>	13.292	0	*	20.171	0	*	NC_002572
<i>Rickettsia prowazekii</i> str. Madrid	13.246	0	*	6.81	0	*	NC_000963
<i>Seculamonas ecuadoriensis</i>	0.507	0.258		0.053	0.432		KC353359
<i>Thalassiosira pseudonana</i>	9.779	0	*	2.922	0.014	*	NC_007405
<i>Tistrella mobilis</i>	8.045	0.004	*	6.099	0.008	*	NC_017966
<i>Alphaproteobacterium BAL199</i>	6.301	0.006	*	5.105	0.004	*	NZ_ABHC0100 0000
Global test	33.573	0		22.297	0		

**Table S5. Comparison of 11 putative nuclear encoded mitochondrial proteins (NcMt) in *Acrisis kona* and their counterparts in the *Naegleria* spp. mtDNAs.**

Protein	Species	Length (aa)	TargetP 1.1		MitoProt II		Predotar	Significance <sup>1</sup>
			P value	Cleavage length(aa)	P value	Cleavage length(aa)	P value	
RPS10	<i>A. kona</i>	227	0.647	17	0.900	13	0.76	+++
	<i>N. gruberi</i>	220	0.381	26	0.999	77	0.26	+
	<i>N. fowleri</i>	77	0.393	n.a.	0.150	n.a.	0.65	+
RPL11	<i>A. kona</i>	200	0.669	16	0.978	54	0.60	+++
	<i>N. gruberi</i>	148	0.272	n.a.	0.181	n.a.	0.32	—
	<i>N. fowleri</i>	145	0.390	82	0.625	n.a.	0.22	+
RPL2	<i>A. kona</i>	256	0.913	62	0.994	52	0.77	+++
	<i>N. gruberi</i>	268	0.310	n.a.	0.565	n.a.	0.10	+
	<i>N. fowleri</i>	266	0.182	n.a.	0.930	n.a.	0.35	+
RPS19	<i>A. kona</i>	99	0.708	14	0.981	n.a.	0.56	+++
	<i>N. gruberi</i>	88	0.633	54	0.879	n.a.	0.05	++
	<i>N. fowleri</i>	87	0.387	n.a.	0.675	n.a.	0.02	+
RPL16	<i>A. kona</i>	194	0.681	110	0.992	45	0.83	+++
	<i>N. gruberi</i>	141	0.412	n.a.	0.939	n.a.	0.47	+
	<i>N. fowleri</i>	141	0.648	16	0.989	n.a.	0.47	++
RPS12	<i>A. kona</i>	172	0.919	49	0.999	160	0.86	+++
	<i>N. gruberi</i>	128	0.294	n.a.	0.932	97	0.33	+
	<i>N. fowleri</i>	128	0.342	n.a.	0.364	15	0.47	—
RPS2	<i>A. kona</i>	254	0.341	n.a.	0.849	101	0.0	+
	<i>N. gruberi</i>	264	0.206	n.a.	0.950	n.a.	0.16	+
	<i>N. fowleri</i>	262	0.192	n.a.	0.566	27	0.03	+
NAD7	<i>A. kona</i>	432	0.674	30	0.586	n.a.	0.83	+++
	<i>N. gruberi</i>	395	0.094	n.a.	0.082	38	0.0	—
	<i>N. fowleri</i>	395	0.066	n.a.	0.155	n.a.	0.0	—
NAD9	<i>A. kona</i>	195	0.334	n.a.	0.817	n.a.	0.50	++
	<i>N. gruberi</i>	186	0.551	36	0.911	n.a.	0.01	++
	<i>N. fowleri</i>	185	0.294	23	0.636	23	0.0	+
SDH2	<i>A. kona</i>	318	0.838	26	0.933	40	0.65	+++
	<i>N. gruberi</i>	270	0.133	n.a.	0.113	n.a.	0.02	—
	<i>N. fowleri</i>	270	0.138	22	0.756	n.a.	0.04	+
COX11	<i>A. kona</i>	283	0.804	14	0.948	22	0.86	+++
	<i>N. gruberi</i>	222	0.015	21	0.327	n.a.	0.01	—
	<i>N. fowleri</i>	224	0.008	25	0.076	n.a.	0.01	—

<sup>1</sup>: (+): one of the predictions have P>0.50; (—): none of the predictions have P>0.50



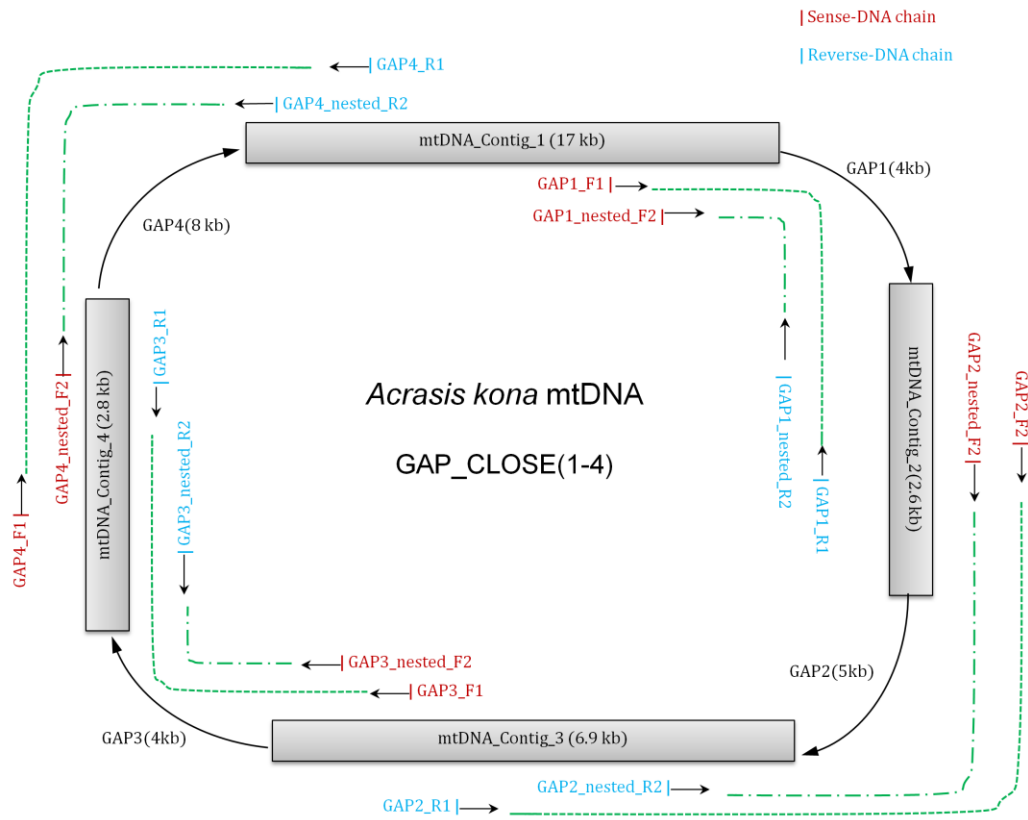


Fig S1. A schematic map of gap-closing of *Acrasis kona* mtDNA based on long-rang PCR.

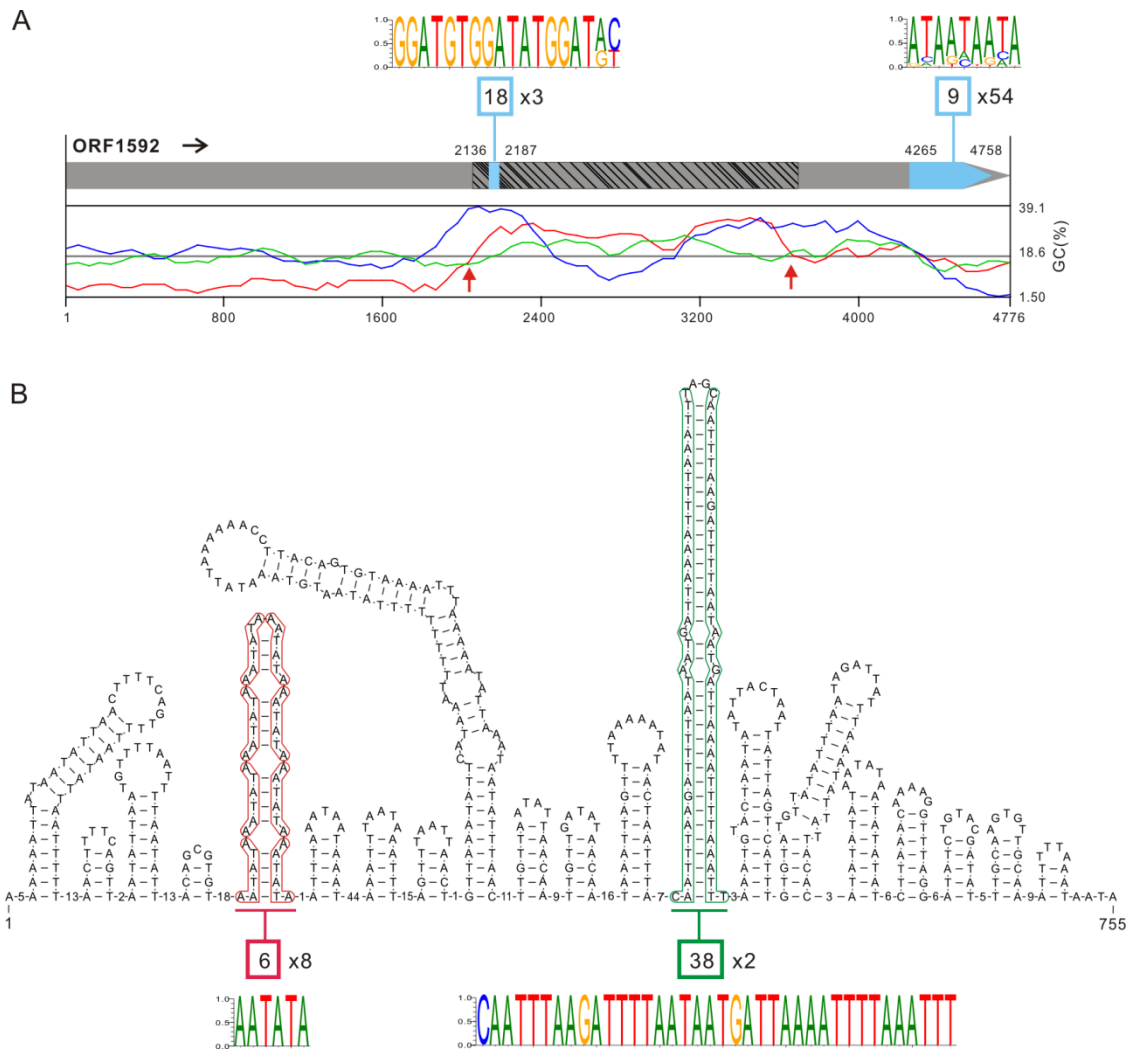
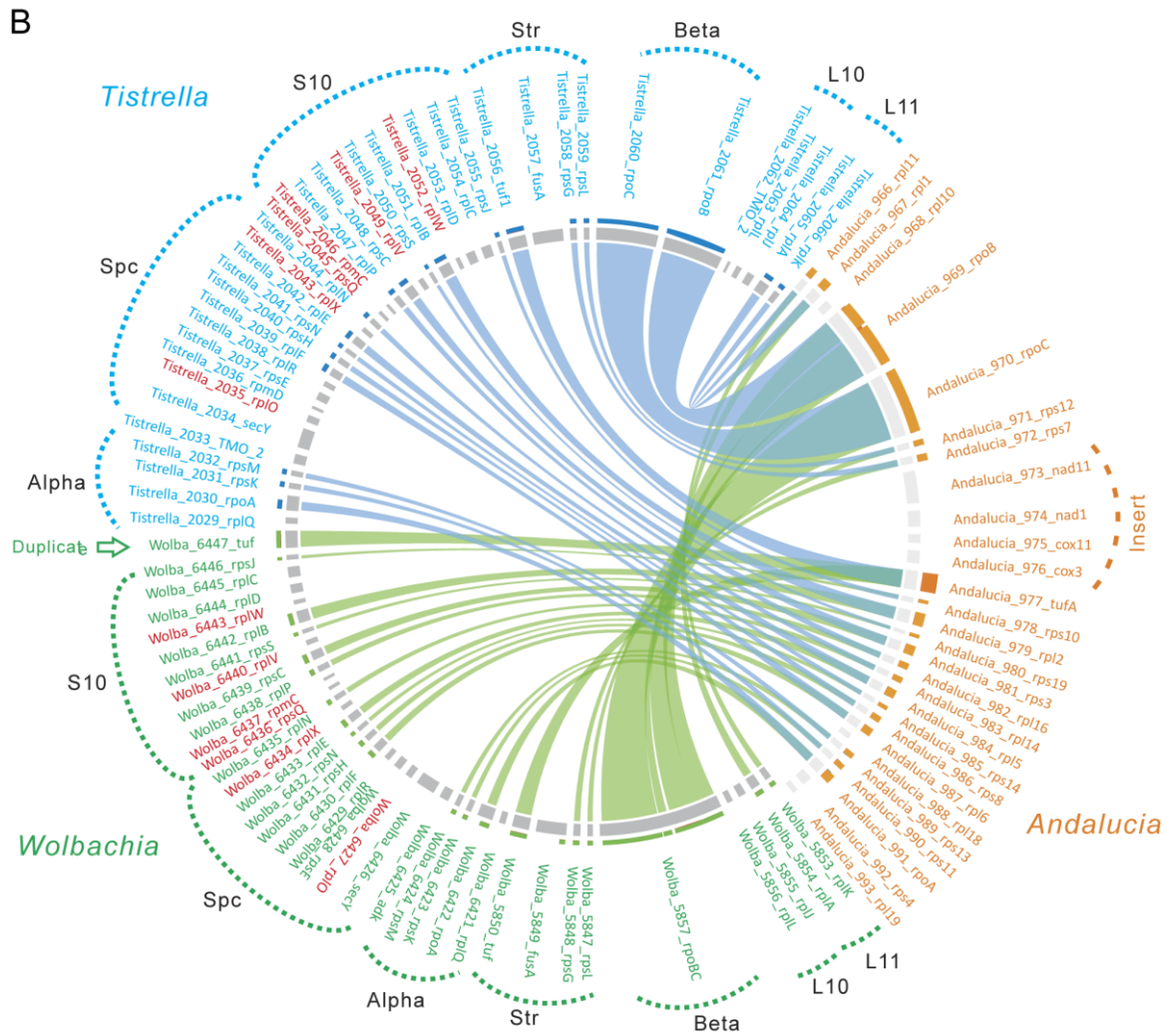


Fig S2. Tandemly repeated regions in *Acrasis kona* mtDNA. (A) Two tandem repeats occur in the largest putative protein coding gene (ORF1592). The region of heterogeneous GC content is shaded. GC Frame plot of GC content at the 1<sup>st</sup>, 2<sup>ed</sup> and 3<sup>rd</sup> codon positions is shown below with a window size of 400. (B) The nucleotide repeat and the potential secondary structures in the largest noncoding region (755 bp) of *A. kona* mtDNA. The period size of nucleotide repeat is indicated in the box.





(Fig S3. Continued)

Fig S3. Comparison of ribosomal protein operon gene order among representative  $\alpha$ -proteobacteria and jakobids. (A) Sequences identified by BLASTp search of four *Rickettsiales* species against the free-living  $\alpha$ -proteobacteria *Tistrella mobilis* (NC\_017956) as the reference with a threshold of  $1e^{-10}$ . (B) Sequences identified by BLASTp search of *Rickettsiales* species *Wolbachia* and *Tistrella mobilis* against the jakobid species *Andalučia goyodi* (KC353352) as the reference with a threshold of  $1e^{-10}$ . The range of each protein operon is indicated with dotted line. Names of genes with discrepancy of their presence in the *Rickettsiales* species in comparison with Burger et al. 2013 are indicated in red. Genbank accession numbers of the genomes: NC\_002978 (*Wolbachia* endosymbiont of *Drosophila melanogaster*), NC\_010793 (*Orientia tsutsugamushi*), NC\_000963 (*Rickettsia prowazekii*) and NC\_015722 (*Candidatus Midichloria mitochondrii*).

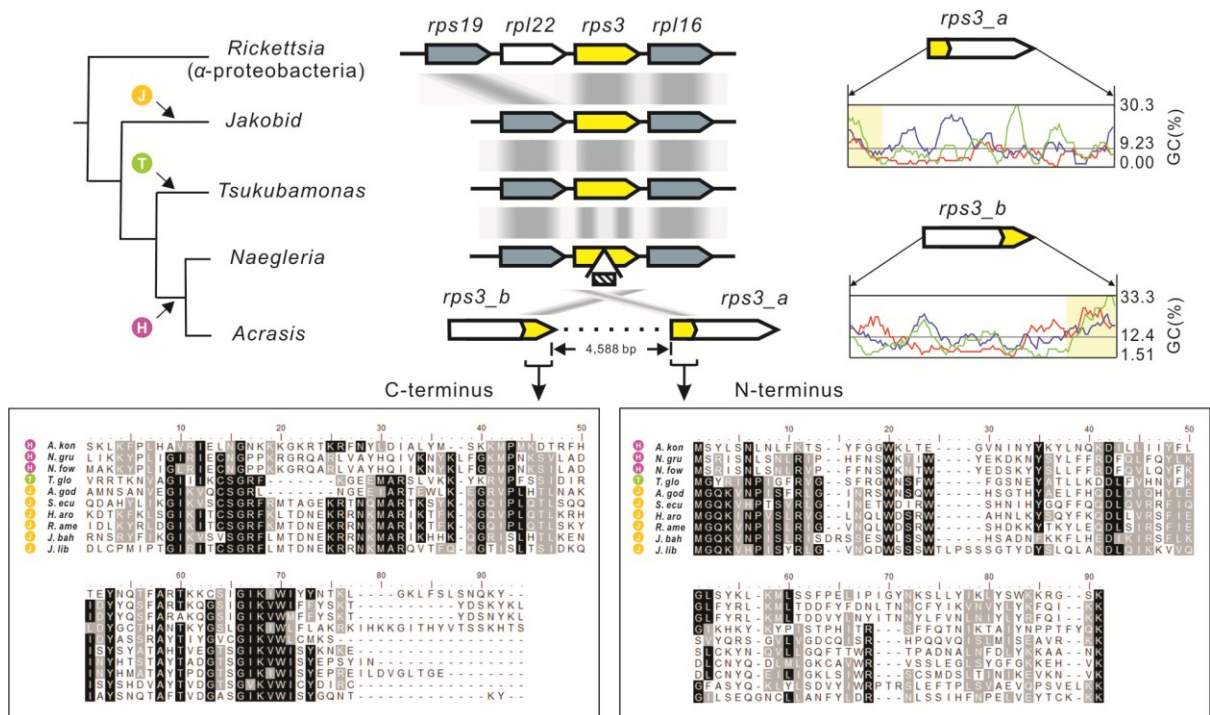


Fig S4. A split *rps3* gene in *Acrasis kona* mtDNA. The *rps3* gene organization in the α-proteobacteria (*Rickettsia*), Jakobida, Tsukubamonadida and Heterolobosea are shown next to a schematic species phylogeny. The *Naegleria rps3* gene has long insertions in the middle that are poorly aligned. The *Acrasis rps3* gene is split into two ORFs (*rps3\_a* and *rps3\_b*), both with terminal extensions. Grey intensity reflects the degree of similarity between the gene orthologs. Multiple sequence alignments for the regions corresponding to the two halves of *rps3* are shown below in boxes. Residues with 100% identity are shaded in black, with over 60% similarity shaded in grey. The GC content of *Acrasis rps3\_a* and *rps3\_b* at each codon position is shown as determined by GC Frame plot (window size 200).

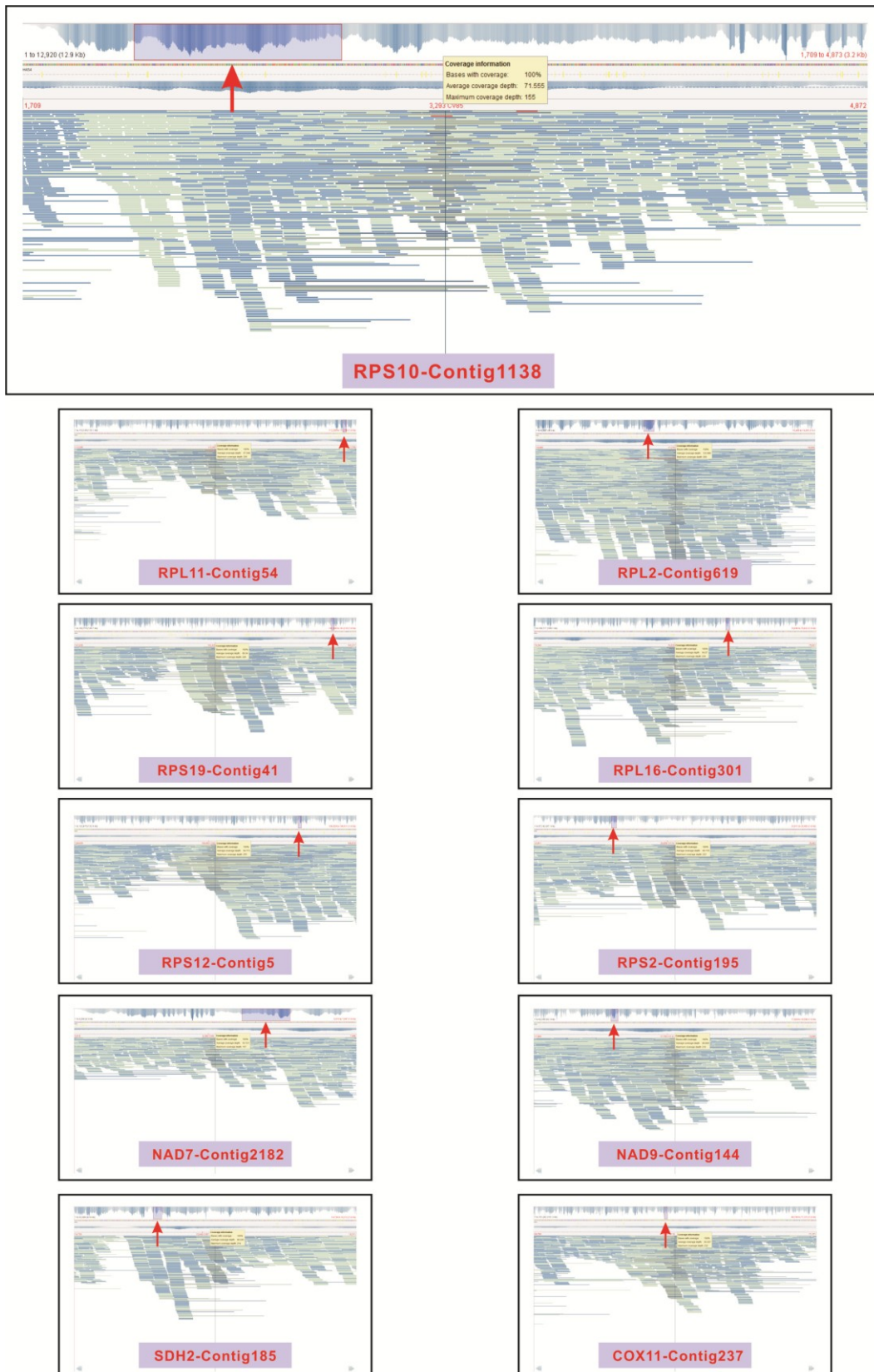
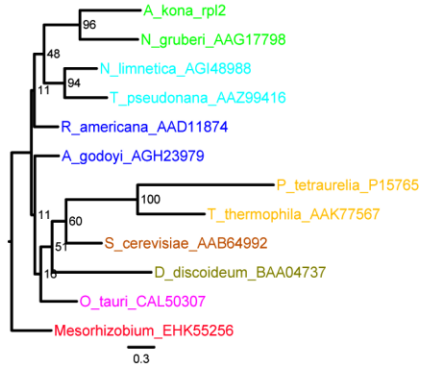
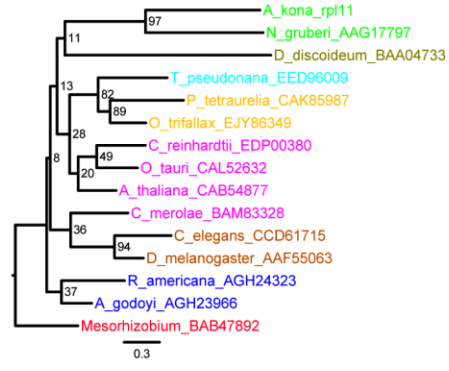


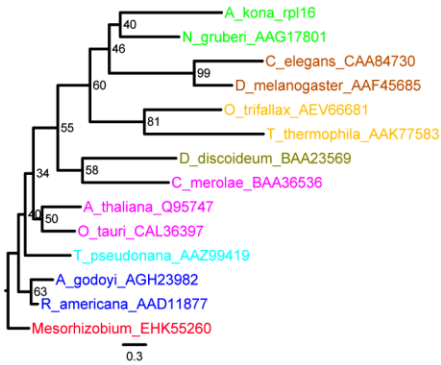
Fig S5. Assembly coverage plots for the 11 putative nuclear encoded mitochondrial protein (NcMt) genes in the *Acrasis kona* nuclear contigs. The corresponding loci of each NcMt gene candidate is indicated with red arrow.



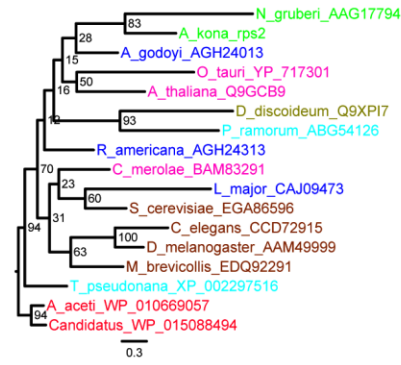
A. rpl12 (223 aligned positions)



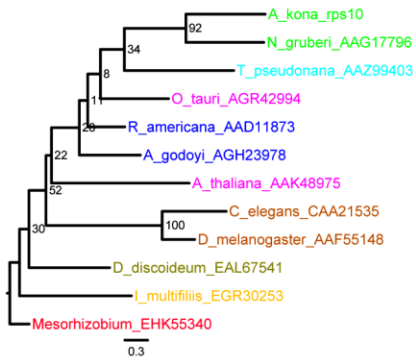
B. rpl11 (126 aligned positions)



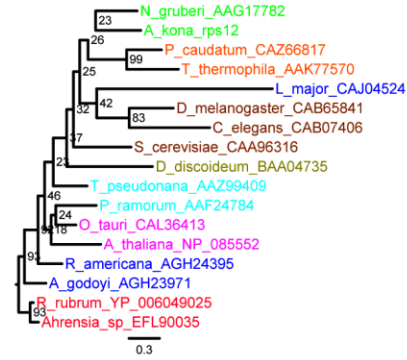
C. rpl16 (118 aligned positions)



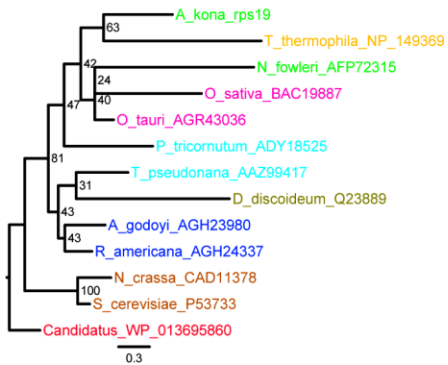
D. rps2 (171 aligned positions)



E. rps10 (99 aligned positions)

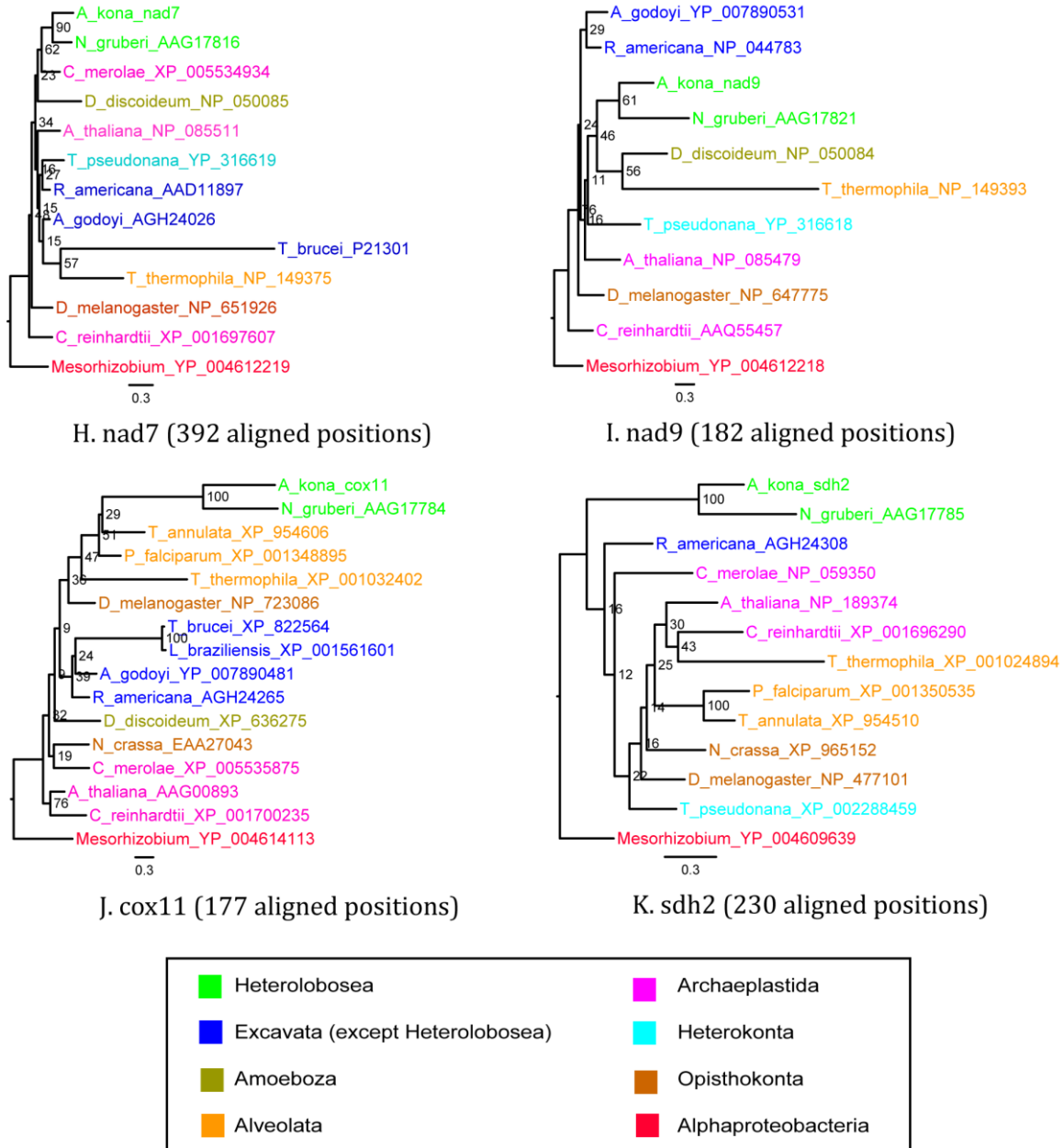


F. rps12 (121 aligned positions)



G. rps19 (80 aligned positions)

(Fig S6.)



(Fig S6. Continued)

Fig S6. Phylogenies of 11 *Acrasis kona* putative nuclear-encoded mitochondrial proteins (NcMt). Trees were derived by maximum likelihood analysis of amino acid sequences with the number of aligned positions for individual data sets indicated below the respective trees. Numbers indicate bootstrap values from 1000 replicates of rapid maximum likelihood bootstrapping with RAXML v7.3.3 (Stamatakis 2006). Taxa were selected to give a broad sampling of eukaryotes based on (Desmond et al. 2011) but with emphasis on Discoba. Taxon labels are color coded as indicate at the bottom right. Trees are drawn to scale as indicated by the scale bars.



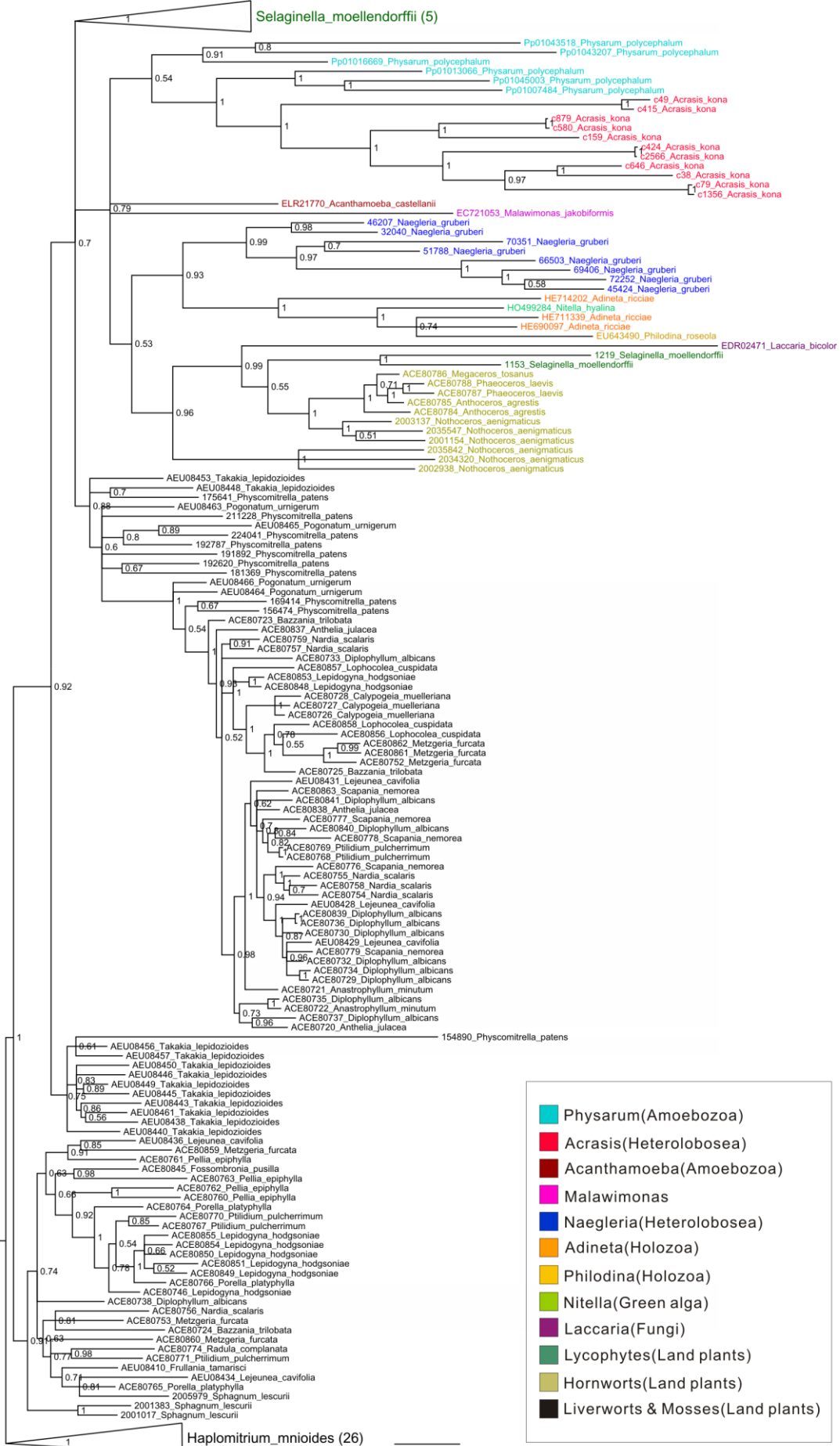


Fig S7. Bayesian phylogeny of the E/E+/DYW domain of the DYW-type PPR proteins. Taxa include a broad sampling of early branching land plants [Liverworts (Haplomitriopsida and Jungermanniopsida), Mosses (Bryopsida, Polytrichopsida, Sphagnopsida and Takakiopsida), Hornworts (Anthocerotopsida) and Lycophytes (Isoetopsida)], and those outside land plants [Heterolobosea (*Acrasis kona*; *Naegleria gruberi*), Amoebozoa (*Acanthamoeba castellanii*; *Physarum polycephalum*), Metazoa (*Adineta ricciae*; *Philodina roseola*), Fungi (*Laccaria bicolor*), Charophyta (*Nitella hyaline*), *Malawimonas jakobiformis*]. The candidate protein sequences from species outside of land plants were retrieved based on (Iyer et al. 2011; Schallenberg-Rudinger et al. 2013). The tree shown was derived using the program PhyloBayes v1.4 (Lartillot et al, 2009) using an alignment of 201 universally-aligned amino acid positions. Values correspond to posterior probabilities; all nodes with posterior probability less than 0.5 are collapsed. Taxon labels include Genbank accession numbers and are color coded according to the key at the bottom right. The tree is drawn to scale as indicated by the scale bar.

## References

1. Burger G, Gray MW, Forget L, Lang BF. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. *Genome Biol Evol* 5: 418-438.
2. Desmond E, Brochier-Armanet C, Forterre P, Gribaldo S. 2011. On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes. *Res Microbiol* 162: 53-70.
3. Iyer LM, Zhang D, Rogozin IB, Aravind L. 2011. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res* 39: 9473-9497.
4. Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25: 2286-2288.
5. Schallenberg-Rudinger M, Lenz H, Polsakiewicz M, Gott JM, Knoop V. 2013. A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. *RNA Biol* 10: 1549-1556.
6. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688-2690.