

Supplementary Information

First finding of free-living representatives of Prokinetoplastina and their nuclear and mitochondrial genomes

Denis V. Tikhonenkov^{1,2,*,#}, Ryan M. R. Gawryluk^{3,#}, Alexander P. Mylnikov^{1,†}, Patrick J. Keeling⁴

¹ Papanin Institute for Biology of Inland Waters, Russian Academy of Sciences, Borok, 152742, Russia

² AquaBioSafe laboratory, University of Tyumen, 625003, Tyumen, Russia

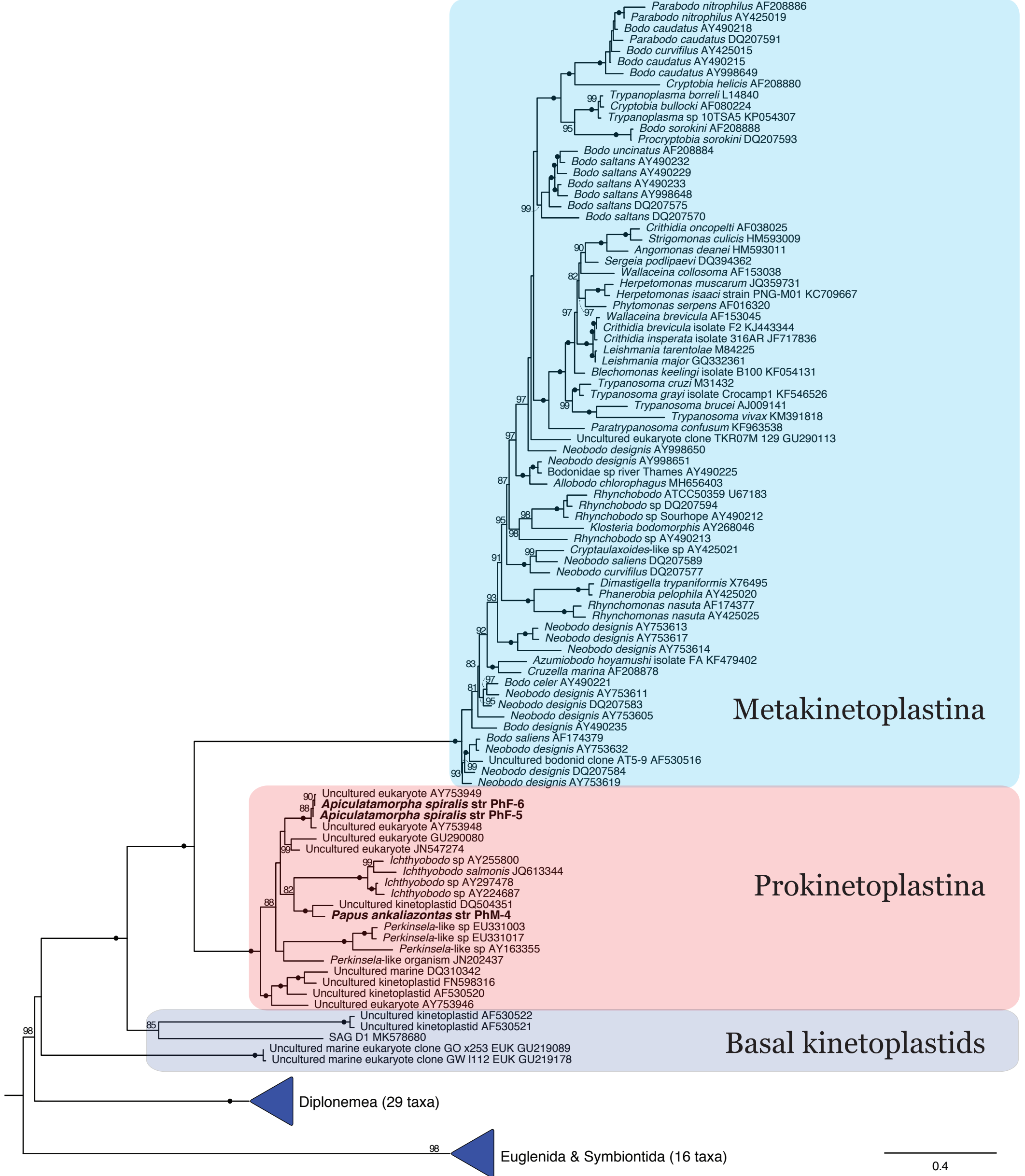
³ Department of Biology, University of Victoria, Victoria, V8W 2Y2, British Columbia, Canada.

⁴ Department of Botany, University of British Columbia, Vancouver, V6T 1Z4, British Columbia, Canada

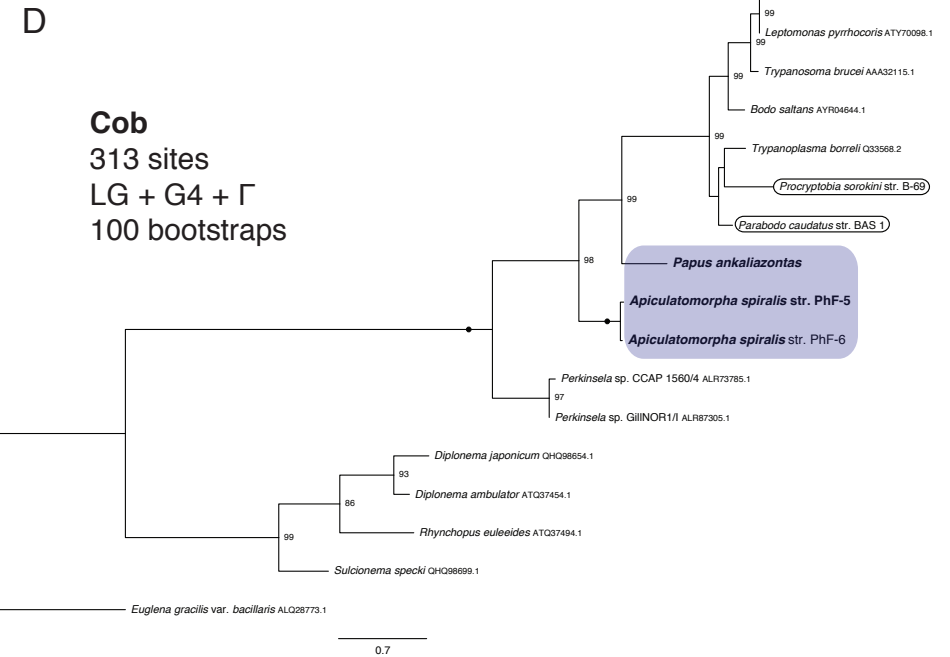
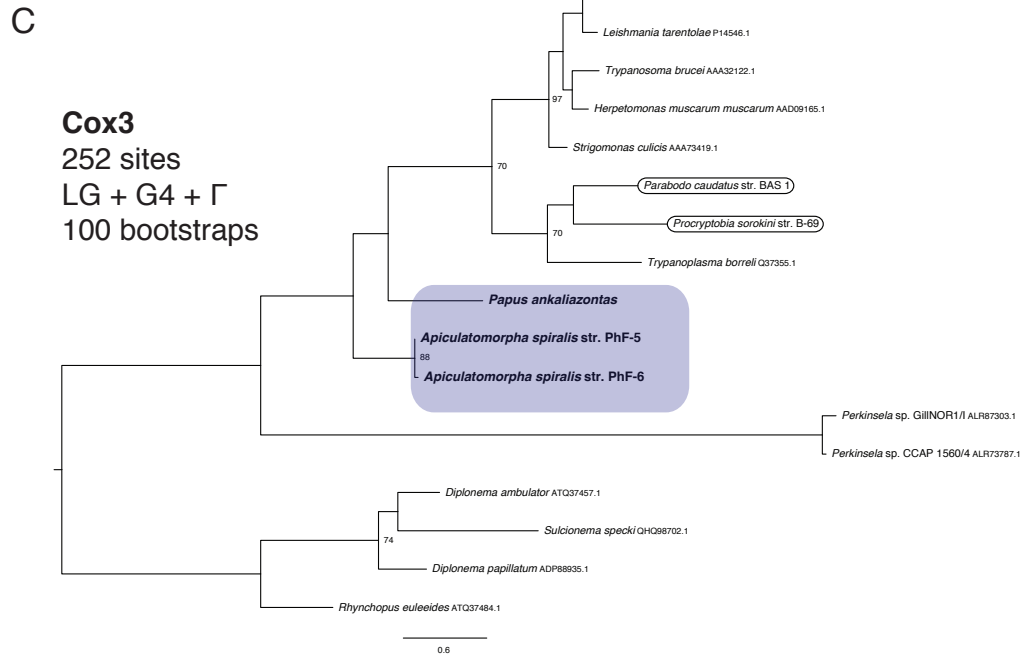
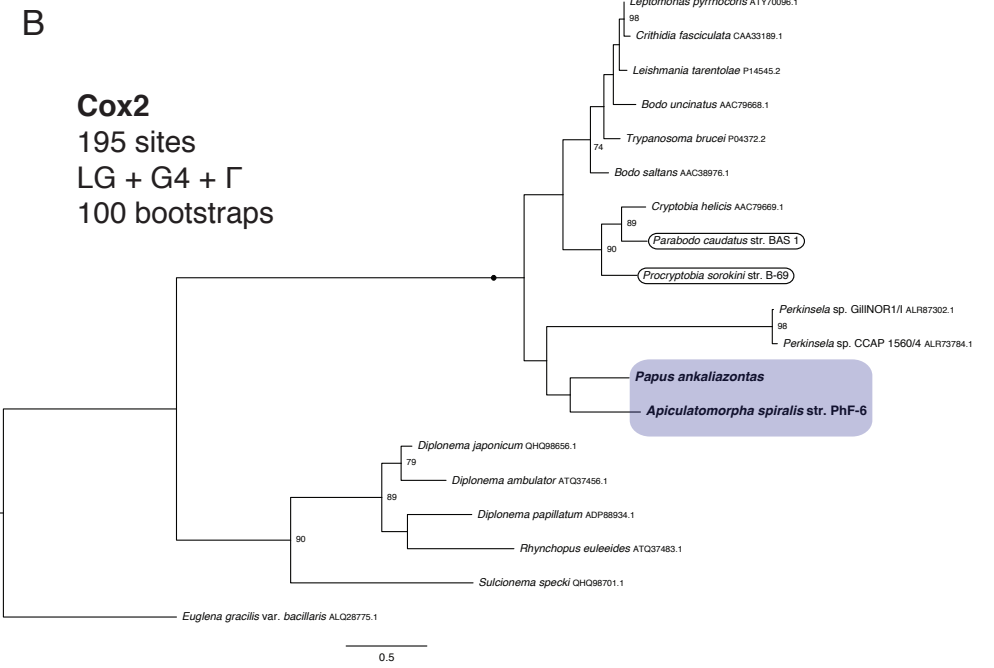
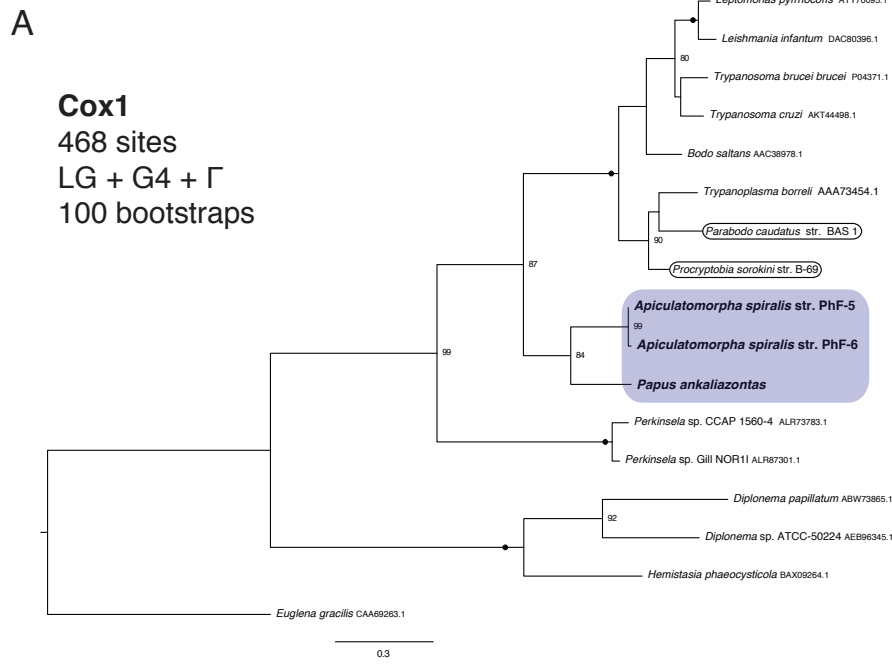
† Deceased May 30, 2019

* Correspondence: tikho-denis@yandex.ru

equal contribution



Supplemental Figure S1. A maximum likelihood phylogeny of the 18S ribosomal RNA gene from diverse kinetoplastids, diplomonids, and euglenozoans (outgroup). Reconstructed with IQ-TREE v1.6.12 using the GTR+F+I+G4 model with 200 initial parsimony trees and 1000 ultrafast bootstrap replicates as a measure of support.



Supplementary Figure S2. Maximum likelihood phylogenies of mtDNA-encoded proteins demonstrate that they do not derive from prey bodonids. Phylogenies of (a) cytochrome c oxidase subunits I (Cox1), (b) II (Cox2), (c) III (Cox3), and (d) cytochrome b (Cob) are shown. The prokinetoplastids *P. ankaliazontas* and *A. spiralis* are indicated by bolded text inside of a coloured box; the names of their prey are enclosed in boxes. Non-parametric bootstrap values are shown as measures of statistical support; for clarity, values less than 70 are not shown and values of 100 are represented by black circles.

A

cob mRNA GCCGCTGTATGTTTCT**TTTGT**TTA**TTTTTTT**A**TTTTTTT**GTTTCG**TTT**GCG**TTT**GT**TTT**G**TTTTTTT**A**TTTTG**TTT**TTT**AGA-A-AT**GT**AG**TTG**-CG**ATT**
cob DNA GCCGCTGTATGTTTCT---G---A-----A-----G---CG---GCG---GT--G-----A---G---AGATATAT-G-A-TG**GC**GA--

cob mRNA **TTGTT**GT-C**TTG**TG**ATT**GACTCG**TTT**G**TTT**AG**TTT**AGG**TTT**A-G**TTT**GGCA**TTT**ACT-ATGGG**TAT**GC--AG**TTATT**AGTAGGAG**TAA**G**TTT**G**TTT**G
cob DNA --G--GT**C**--G-GA--GACTCG--G--AG--AGGA--AT**G**--GGCA--ACT**TA**-GGG-A-G**CT**TAG--A--AG-AGGAG-AAG---G-C--G

cob mRNA GAT**GTATT**G**TAA**T**GTTTTTTTTTTT**C**TTGG**TTAG**TGTTTT**ATG**TATGG**TAGAT**TTTT**GA-AGGA--GG**TTATAT**G**TTT**GCG**TTCT**TTGCAT**ATTG**TT**TT**G
cob DNA GA-G-A--G-AA-G-----C--GG---AG-G----A-G-ATGG-AGA----GATAGGAT**TT**GG--A-A-G--GCG--CT--GCA-A--GT---G

cob mRNA ACA--AG**TTGT****TTG**T**TTTTTT**G**TTG**ATTGACAT**ATTG**TAA**TG**T**TTTTTT**G**TTTTT**GA**TG**TTT**G**-AT**CT**GC**ATT**G**TATG**TT**GATT**GG**TAG**TT**TT**
cob DNA AC**ATT**AG--GT--G-A-----G--G-A--GACA-AT--G-AAA-G-A-----G-----GA-GTT**GT**TAT-C-GCA--G-A-G---GA--GG-AGG--

cob mRNA **TTGTTTT**G**TTTTT**G**TTT**G**TTTTT**G--ATAGCT**TTT**ATAGG**ATA**G**TTTT**GCC**ATT**GACCCAGATG**TCAT****TT**GAGGGCTTACTG**TG**TT**TT**AG**TAA**TATA
cob DNA --GT---G-A---GG---GA----G**TTA**-AGCTTT-A-AGGA-A-G---GCCA--GACCCAGATG-CA-A--GAGGGCTTACTG-G---AG-AA-A-A

cob mRNA **TTTT**GCT---AC**TTT**G-CC**CTGG**TTGGG**TTT**GAG**TTGG**TTT**ATT**--GGATTGAGG**TT**CAGAG**TTT**ATTCAGGAAGG**TA**-**CTT**GG**TAA**AGATACATT**CC**
cob DNA ---GCT**TTT**AC---G**CCCC**-GG--GGG**TT**-GAG**T**--GG--ATT**TT**GGATT**T**GAGG--CAGAG--ATTCAGGAAGG-**ATC**---GG-AAAGA-ACATT**CC**

cob mRNA **CTT**CATTT-G**TTG**TTGCC**ATT**GGG**TTT**G**TT**ATTT**TT**G**ATT**G**ATT**GCA**TTTTTTT**G**TTT**GCA**TTA****TTTTTT**GAG**TTCT**G**ATT**TT**GTT**-AGAT**CTG****TTTT**GT
cob DNA C--C**ATTTT**G--G--GCCA--GGG--G--ATTT**T**-GA--G-A--GCA-----G--GCA--A-----GAG--C-GA--GTT**T**AGA-CG----GT

cob mRNA **GTTTT**ATTTGGAGC--**GTTG****TTTTTT**G**TTT**AT**TG**TT**TTTT**A-GA--G**TTT**G**TTTTT**G**TTG**AT**TTT**G**TTTTT**G**ATTT**G**ATTTT**ATTT**TT**GT
cob DNA G----A---GGAG**CTT**G--G-----G---A--GA----A---ATG**ATT**TGA---G-----G--GA---G-----G-A---G-A-----A---GT

cob mRNA **TTT**ATTT**TT**GG**TATTTT**G**TTTTT**CA**TGA**GAAT**CTGG**TT**AA**T**AGTT**G**ATG**T**GA**T**GA**AGACAT**CTG**AGAAGG**TTAT**CCAG**A**---A**TGG****TTTTTTT**G
cob DNA TT--A---A--GG-A---G-----CA-GAAG**ATC**--GG--AA-AG--GA-G-GA-GAAGACAT**CTG**AGAAGG--A--CCAG**ATTTA**-GG-----G

cob mRNA AT**GTTTTT**GG**TG**T**GATTA**---AA**TCAAT**CTCG**ATAA**-**GTTTTT**G---G**TGTTTT**ATTT**TT**G**TGT**---**GTTTTT**GG**TTTTT**AT**GG**TT**TT**G**TTG**
cob DNA ATG-----GG-G-GA--**ATTTT**AA-CAAT-CCTGA-**AATG**-----G**TTTT**G-G---A-----G-G**TTTT**G-----GG-----A-GG---G--G-

cob mRNA **TGTT**G**TTTTT**G**TTT**AT**TCAT**TT**TT**G**ATTT**GCG**TTT**G-C**ATTTT**TG**TG**TT**TT**TT**TT**G**TTTTT**G**TTAG****TTTTTT**G**TATG**T**GT**TT**G**-GG**TTT**G
cob DNA -G--G-----G--A--CA--G--G-A--GCG--G**TC**A-----G-GT**GT**TTT---GT---G--AG-----G-A-GT**G**--G--G**TTG**---G

cob mRNA **TTG**CT**GTTT**A**TGTTTT**G**TTATTT**CC**TG**T**ATGG**TT**AG**TT**GC**TT**ACT****TTG****TTTT**G**TTT**G**TTGA**T**GAT****TTT**GG**TG**A**CT****TTTT**G**TTA**AT**TTA**
cob DNA --G-C-G---A-G---G--A--A---CC-G-A-GG--A-GG--GCAG**TTAC**--G---G---G-G**TTGA**-GAT---GG-GA-ACG---G---AA---A

cob mRNA **TTTTT**GTT**TTTT**G**TTTTTTT**A**TTTTTTT**TT**TTT**GGTTGTT**GGTA**
cob DNA -----GTT---G-----A-----GGTTGTT**GGTA**

B

Cob mRNA TTTTA**TTA**TT**TAA**TT**GTTT**CG-TA**TT**CG**TTTTTT**G**TTTTTTT**G**TTTTTTT**A-GA-AA**TTT**GA--G**TTG**CT**TTA****TTTT**GGC**TT**GA-GA-TTTGA-C**T**
Cob DNA TTTTA-TTA-TTAA--G---CG**TTA**--CG-----GG-----G---G---AT**G**ATAA--TG**ATT**G-GG-C--A---GGC--GAT**GAT**TT**TGAT**C-

Cob mRNA AGA-A**TTTT**AG**TTT**G---GG**TTTT**AA**TT**GGC**TTT**TG-----G--G**ATT**G**TT**ATGCAG----ATTT**TT**-G-TTGGTATGGTGGTAGCTTGAATGTACA
Cob DNA AGA**TA**-----AG---G**TTT**GG---AA-TTGGC--TT**GTTTTT**G**TTGA**-TG--ATGCAG**TTTT**ATTT**TT**T**TT**G**TT**GGTATGGTGGTAGCTTGAATGTACA

Cob mRNA ATAATATGTTTTTTTCTCTTTTGTAAAGTGT**TAT**TGTTGATGTCTTGATTATGAATTTGGATGGCTTGTTCGTTGTTTTCATATAGTATTGACTAGTGT
Cob DNA ATAATATGTTTTTTTCTCTTTTGTAAAGTGT**TAT**TGTTGATGTCTTGATTATGAATTTGGATGGCTTGTTCGTTGTTTTCATATAGTATTGACTAGTGT

Cob mRNA TATATATCTTTTATTTTGGCCATATTTGTAAAGT**TTTT**TATGTTTGT**TAA**TGTTT**GAT**AGTATTGCTATGGTTGATTTTCCGATTTATAGTTTTT
Cob DNA TATATATCTTTTATTTTGGCCATATTTGTAAAGT**TTTT**TATGTTTGT**TAA**TGTTT**GAT**AGTATTGCTATGGTTGATTTTCCGATTTATAGTTTTT

Cob mRNA ATTTT**TTA**AT**TGG**TAT**GCT**TTTAT**TGG**TTATGTTT**AC**CTCTTACTCAGATGTCATATTGAGGACTAACTGTTTTTAGTAATATTTTAGCTACTGTAC
Cob DNA ATTTT**TTA**AT**TGG**TAT**GCT**TTTAT**TGG**TTATGTTT**AC**CTCTTACTCAGATGTCATATTGAGGACTAACTGTTTTTAGTAATATTTTAGCTACTGTAC

Cob mRNA CTTTAGTGGGAGAA**TTTT**AGTTTTTTGATTTTGGGGTAGTGAATTTATTAGTGAAGTTACTTTAGTTAAGGTGCATTTATTACATATAGCTTTACCGTG
Cob DNA CTTTAGTGGGAGAA**TTTT**AGTTTTTTGATTTTGGGGTAGTGAATTTATTAGTGAAGTTACTTTAGTTAAGGTGCATTTATTACATATAGCTTTACCGTG

Cob mRNA GATAGCTTCGTTTTTTGTTTTAGCACATATGTTTTTTTACATTATTATTTGAGTTCCGATGCTATAGATCGGTATGTTTTTTATTTAGAAAGGTGTTTT
Cob DNA GATAGCTTCGTTTTTTGTTTTAGCACATATGTTTTTTTACATTATTATTTGAGTTCCGATGCTATAGATCGGTATGTTTTTTATTTAGAAAGGTGTTTT

Cob mRNA TTTGTTTAT**TGG**TTTTATTTTCGTGATTTAGTTT**TG**TTTTG**TAT**TTTTTTGTTTTTTCTTTGGGTATTTTGTTTTTTTGTTGGTGATTTTTGTTTTTCATG
Cob DNA TTTGTTTAT**TGG**TTTTATTTTCGTGATTTAGTTT**TG**TTTTG**TAT**TTTTTTGTTTTTTCTTTGGGTATTTTGTTTTTTTGTTGGTGATTTTTGTTTTTCATG

Cob mRNA AAGAATCTGGGTATTGTGAATGTTATGAAGACTTCTGAAAAAGTTATACCTGAGGAGTTT**GAT**TTGGTGATAAGTCAATCTCCGAAAGGGGGGGAA
Cob DNA AAGAATCTGGGTATTGTGAATGTTATGAAGACTTCTGAAAAAGTTATACCTGAGGAGTTT**GAT**TTGGTGATAAGTCAATCTCCGAAAGGGGGGGAA

Cob mRNA TG**GT**GAT**TTGG**TTTTT**GAC**GT**TGT**GGTTT**GTC**GATTTT**AG**GT**GG**TTT**GAG**TAGAGAGAAGCGGGAGTGT**TCC**GGGGGAGCATT**GTT**GAAATTT**G**TAA
Cob DNA TG**GT**GAT**TTGG**TTTTT**GAC**GT**TGT**GGTTT**GTC**GATTTT**AG**GT**GG**TTT**GAG**TAGAGAGAAGCGGGAGTGT**TCC**GGGGGAGCATT**GTT**GAAATTT**G**TAA

Cob mRNA GGGGTTTTT**GAC**GAAGTAGGTTTTT**GTT**GAGAGGTGTT**AA**TTTTTTATTTT**GTT**TAAGGTAGATTT**GAC**TTT**GTT**AGTTGGTTT**GTT**ATTTGGGGGA
Cob DNA GGGGTTTTT**GAC**GAAGTAGGTTTTT**GTT**GAGAGGTGTT**AA**TTTTTTATTTT**GTT**TAAGGTAGATTT**GAC**TTT**GTT**AGTTGGTTT**GTT**ATTTGGGGGA

Cob mRNA TGGAT**TGT**AGGTTTT**GTT**GGGGTGGGATT**GTT**AAGT**GT**GGT**GAT**GTGTTTT**AG**GATTT**TG**GGGTTACGGTTT**GAT**TTTT**TG**GGTTT**GTT**ATTTAAAT
Cob DNA TGGAT**TGT**AGGTTTT**GTT**GGGGTGGGATT**GTT**AAGT**GT**GGT**GAT**GTGTTTT**AG**GATTT**TG**GGGTTACGGTTT**GAT**TTTT**TG**GGTTT**GTT**ATTTAAAT

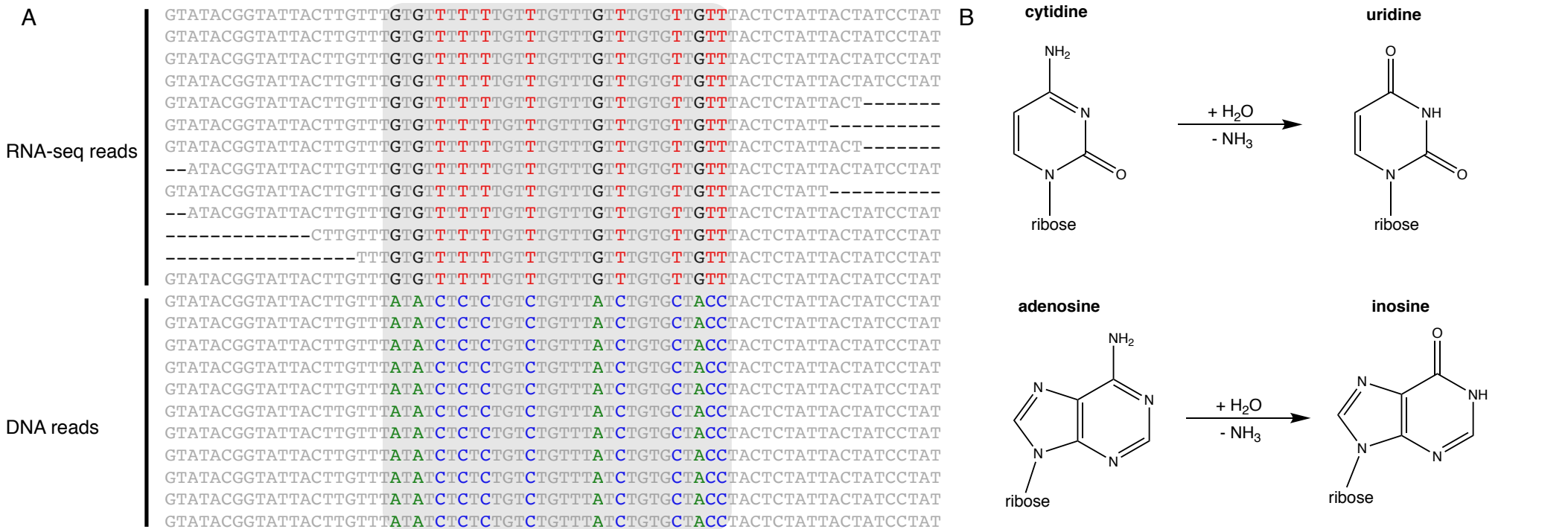
Cob mRNA TGGGCTTGGGTTGAAATTTTATGGACGGATGCTTTTGGGGTTAGAAATGGTTT**GAG**TTGGGAATGTTTGGTTT**GAT**TGAAATTT**GAG****TTT**G**TC****TTT**GGC
Cob DNA TGGGCTTGGGTTGAAATTTTATGGACGGATGCTTTTGGGGTTAGAAATGGTTT**GAG**TTGGGAATGTTTGGTTT**GAT**TGAAATTT**GAG**---G**TCG**---GGC

Cob mRNA **TG**ATTGAG---G**TTTTTT**GGTTTT**CC****TTT**GAA**TAA**GG**TTTTT**G**TTT**G**TTT**G-G**ATT**GTTTGG-GTT**TG**AT**AT**GAT**TTTT**G**TAT**GG**TG**-A**ATT**GTT
Cob DNA -GATTGAG**TTT**G-----GGTTT-CC---GAA-AAGG-----GT---G---G**TG**-A--GTTTGG**TG**TT-GA-A-GAT---G-A-GGG-G**TA**-A--GTT

Cob mRNA TGAGGTT**TTTTTT**G**TTT**G-G---A**TA**TGTT-G**TTA**TGATT**AA**T**AG**GT**TTTT**CG**TTT**G**CTTTTT**G**TTA**AT**TT**GGG-GA**TA**AT**TTTT**GTTTAT**TAA**
Cob DNA TGAGGTT-----G---G**TG****TTTTA**-A-GTT**TG**--A-GATT**AA**-A-GGT-----CG---G-C-----G--AAT--GGG**TGA**-AATT---GTTTTAT**TAA**

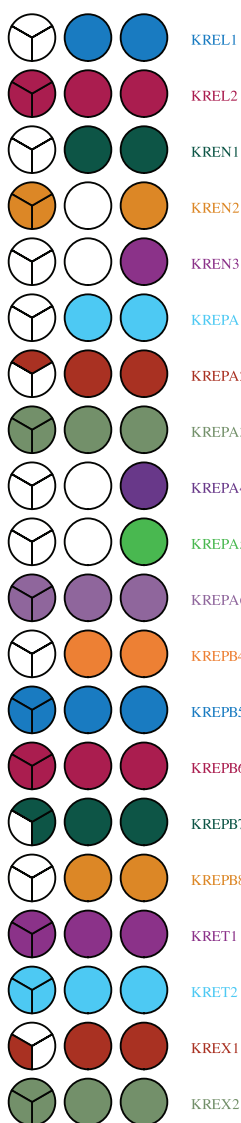
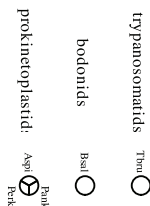
Cob mRNA TGTT**TT**ATAAAAAAAAA---AAA
Cob DNA TGTT--ATAAAAAAAAA**TTT**AAA

Supplementary Figure S3. Representative U indel editing in *Papus ankaliazontas* and *Apiculatamorpha spiralis*. Nucleotide alignments of the cytochrome b gene and edited mRNA sequences are presented for (a) *P. ankaliazontas* and (b) *A. spiralis*. Editing of Cob illustrates the broadscale patterns of RNA-editing observed in these two new species: *P. ankaliazontas* mitochondrial mRNA transcripts are extensively edited along their entire length, whereas *A. spiralis* counterparts were typically edited at their termini. Bold red nucleotides denote U insertions and blue nucleotides denote U deletions (although represented by T). Cob is the 6th most heavily edited mRNA in each of *P. ankaliazontas* and *A. spiralis*.

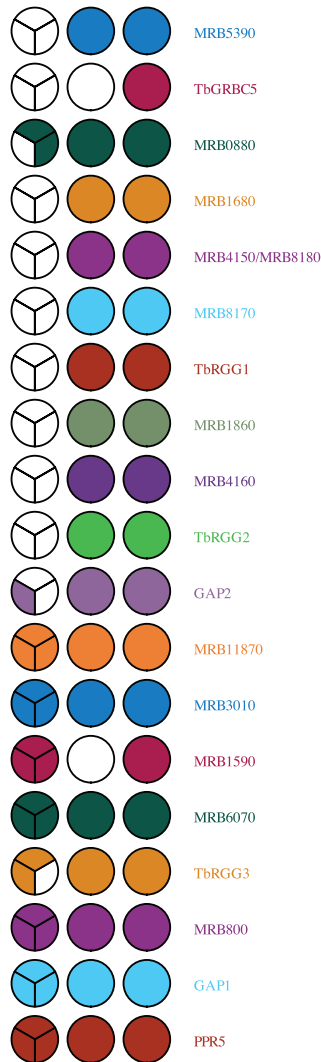
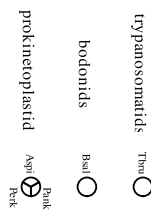


Supplementary Figure S4. Evidence for C-to-U and A-to-I RNA-editing in the putative *Papua ankaliazontas* mitochondrial 12S rRNA. (a) Multiple alignment of RNA-seq and genomic DNA reads from *P. ankaliazontas*. Deaminative editing of cytosine to uracil and adenosine to inosine is observed as C-to-T and A-to-G changes, respectively. A cluster of 12 deamination edits were observed over a short span of the 12S rRNA, representing the first description of deaminative editing in kinetoplastids; no incompletely edited sites were observed. (b) Chemical reactions underlying C-to-U and A-to-I RNA-editing.

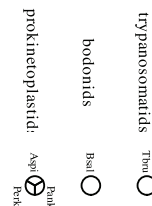
RECC20



MRB1



Others



Supplementary Figure S5. RNA-editing machinery in prokinetoplastids. Coulsen plot highlighting the presence and absence of *Trypanosoma brucei* RNA-editing components in *Bodo saltans*, and the prokinetoplastids, *Perkinsela*, *Papus ankaliazontas*, and *Ankaliazontas spiralis* PhF-6. To accommodate all genes, the plot shows components from the RECC20 and MRB1 and ‘others’ separately. Filled in wedges indicate the presence of a subunit, and empty wedges indicate absence.

Supplemental Table S1. Statistics for *Papus ankaliazontas* and *Apiculatomorpha spiralis* genome assemblies. Genomes were assembled for each species with SPAdes, and decontaminated of prey kinetoplastids and co-cultured bacteria using a combination of Automet (for prokaryotes), and blastn using transcriptome assemblies from cultures that contained the eukaryotic prey. Most statistics are derived from Quast analysis; however, transcript mapping rate and number of introns are derived from BLAT analyses, paired read mapping rate – referring to the percentage of genomic DNA reads that map to the decontaminated genome assemblies – were derived using bbmap, and transcript number refers to number of transcriptome coding sequences predicted by Transdecoder. Predicted genome size corresponds to Bayesian kmer-based estimates generated by GCE. Intergenic distances are based on manually inspected gene mapping on several of the longest contigs.

	<i>P. ankaliazontas</i>	<i>A. spiralis</i>
Contigs	5,118	9,927
N50 (bp)	5,464	2,438
Assembly length (bp)	22,454,015	24,339,364
Longest contig (bp)	31,864	25,801
GC (%)	55.8	48.02
Coverage (SPAdes)	6	2
N's per 100 kb	86.59	0.82
Transcript number	21,181	32,854
Transcript mapping rate (%)	73	37
Number of introns	0	0
Intergenic distance (bp)	~78	~1200
Paired read mapping rate (%)	44	6.5

Supplemental Table S2. Table of RNA editing statistics for *Papus ankaliazontas* and *Apiculatomorpha spiralis* strain PhF-6, based on comparisons between edited mitochondrial mRNAs identified in transcriptome datasets and genomic contigs.

Species	Gene	mRNA completeness	Gene completeness	Editing Type	Pre-edited length	Edited ORF Length	% Size increase	Insertions	Deletions	Observations
<i>P. ankaliazontas</i>	Nad1	complete	Uncertain 5'	Pervasive central and 3' editing, diminished at 5' end.	394	684	73,60406091	329	39	Stop codon formed by 3 U deletion.
<i>P. ankaliazontas</i>	Nad4	incomplete	internal problems	Pervasive central and 5' editing, diminished at 3' end.	711	1028	44,58509142	349	32	Cannot assemble a full-length mRNA lacking internal stop codons. Dominant version has stop codons, but is highly supported by the reads. Possible pseudogene? Stop codons exist in a high GC area, suggestive of loss of editing.
<i>P. ankaliazontas</i>	Nad5	complete	complete	pan	805	1517	88,44720497	885	49	Stop codon generated by U insertion. Required some manual curation.
<i>P. ankaliazontas</i>	Nad7	complete	complete	pan	723	1218	68,46473029	542	47	Stop codon generated by U insertion.
<i>P. ankaliazontas</i>	Nad8	complete	complete	pan	274	456	66,42335766	207	25	
<i>P. ankaliazontas</i>	Nad9	incomplete	incomplete	Pervasive central and 3' editing, diminished at 5' end.	265	467	76,22641509	218	16	Cannot assemble full-length mRNA. Truncated at 5' end, without a good candidate start codon. Like Nad4, there is a GC-rich region, supported by reads, that looks like it should be edited, but isn't.
<i>P. ankaliazontas</i>	Cob	complete	complete	pan	642	1195	86,13707165	575	49	2 possible G deletions. Stop codon generated by U insertion.
<i>P. ankaliazontas</i>	Cox1	complete	complete	5' and 3' editing	1392	1665	19,61206897	282	10	U insertion generates start codon. U insertion generates stop codon.
<i>P. ankaliazontas</i>	Cox2	complete	complete	pan	340	651	91,47058824	322	11	Stop codon generated by U insertion.
<i>P. ankaliazontas</i>	Cox3	possibly complete	likely complete	pan	377	771	104,5092838	411	17	Stop codon generated by U insertion. Possibly truncated at 5' end.
<i>P. ankaliazontas</i>	Atp6	incomplete	unknown	pan	349	693	98,56733524	383	39	
<i>P. ankaliazontas</i>	rps12	likely complete	complete	pan	162	315	94,44444444	166	13	Stop codon generated by U insertion. No ATG start identified, but likely complete based on MSA.
<i>P. ankaliazontas</i>	Murf1	incomplete	complete	unknown (likely pan)	162	310	91,35802469	150	1	Cannot resolve 3' end of transcript
<i>P. ankaliazontas</i>	putative 9S	possibly complete	complete	unedited	676	676	0	0	0	
<i>P. ankaliazontas</i>	putative 12S	possibly complete	complete	minimal	1050	1050	0	0	0	C-to-U, A-to-I editing
<i>A. spiralis</i>	Nad1	complete	complete	pan	481	732	52,18295218	331	80	
<i>A. spiralis</i>	Nad4	incomplete	complete	5' editing	1874	1899	1,334044824	42	17	Possible pseudogene (N-terminal part of protein in transcript but broken in all contigs)
<i>A. spiralis</i>	Nad5	possibly complete	complete	5' editing	1196	1347	12,62541806	198	47	
<i>A. spiralis</i>	Nad7	complete	complete	5' and 3' editing	1128	1192	5,673758865	96	32	

<i>A. spiralis</i>	Nad8	complete	complete	pan	278	462	66,18705036	236	52	
<i>A. spiralis</i>	Nad9	incomplete	complete	pan	322	525	63,04347826	248	45	Termini incomplete in cDNA, in frame stop
<i>A. spiralis</i>	Cob	incomplete	complete	5' and 3' editing	1373	1479	7,720320466	148	42	Likely more editing at 5' and 3' ends but not helped by transcripts
<i>A. spiralis</i>	Cox1	possibly complete	complete (but two contigs)	5' editing	1476	1476	0	19	19	
<i>A. spiralis</i>	Cox2	complete	complete	5' and 3' editing	519	594	14,45086705	128	51	
<i>A. spiralis</i>	Cox3	complete	complete	5' and 3' editing	762	795	4,330708661	79	46	
<i>A. spiralis</i>	Atp6	possibly complete	complete	pan	206	357	73,30097087	190	39	
<i>A. spiralis</i>	putative 9S	possibly complete	complete	unedited	334	334	0	0	0	
<i>A. spiralis</i>	putative 12S	possibly complete	complete	unedited	1108	1108	0	0	0	