Is the Dinoflagellate *Amoebophrya* Really Missing an mtDNA?

Ehsan Kayal [*,1] and David R. Smith[2]

[1]Fédération de Recherche 2424 Sorbonne Université & Centre National pour la Recherche Scientifique, Station Biologique de Roscoff, Roscoff, France

[2]Department of Biology, University of Western Ontario, London, ON, Canada

*Corresponding author: E-mail: ehsan.kayal@sb-roscoff.fr.

Associate editor: Aida Ouangraoua

Abstract

Mitochondrial DNA (mtDNA) is a universal hallmark of aerobic eukaryotes. That is why the recent suggestion by John et al. (2019. An aerobic eukaryotic parasite with functional mitochondria that likely lacks a mitochondrial genome. Sci Adv. 5(4):eaav1110.) that the aerobic dinoflagellate *Amoebophrya* sp. strain AT5 (Syndiniales) lacks mtDNA was so remarkable. Here, by reanalyzing recently published genomic and transcriptomic data from three *Amoebophrya* strains, we provide evidence of a cryptic, highly reduced mtDNA in this clade. More work is needed before one can definitively say if *Amoebophrya* has or does not have an mtDNA, but for now, the data are pointing toward the existence of one. Ultimately, we urge caution when basing supposedly absent genomic features on single line evidences.

Key words: anaerobe, genome loss, oxidative phosphorylation, *Perkinsus*, Syndiniales.

Mitochondria are, in many ways, inextricably tied to their genomes. Indeed, with one potential exception (discussed below), all aerobic eukaryotes have mitochondrial DNA (mtDNA). But why? Eukaryotic evolution has shown that nearly all of the genes originally housed on the proto-mitochondrial genome have been transferred to the nucleus or lost entirely (Roger et al. 2017). So, why leave behind a handful of mitochondrial genes, all dependent on a complex, metabolically costly gene expression system? Several hypotheses have addressed this question, including the colocation of gene and gene product for redox regulation of gene expression (CoRR) hypothesis (Allen 1993). Moreover, studies of anaerobic eukaryotes with mitochondrial-related organelles (MROs) have shown that it is possible to jettison a mitochondrial genome (Santos et al. 2018), and in extreme cases, the mitochondrion itself (Karnkowska et al. 2016). For example, the symbiotic MRO-containing anaerobes *Cryptosporidium* spp. (Apicomplexa), *Henneguya salminicola* (Myxozoa), and *Neocallimastix patriciarum* (Fungi) have all lost their mtDNAs (Rada and Tachezy 2019; Tsaousis and Keithly 2019; Yahalomi et al. 2020), as have the free-living anaerobic protists *Pygsuia biforma* (Obazoa), *Sawyeria marylandensis* (Heterolobosea), and *Mastigamoeba balamuthi* (Amoebozoa) (Gill et al. 2007; Barberà et al. 2010; Stairs et al. 2014). In each of these cases, however, the absence of oxidative phosphorylation has been a key catalyst for forfeiting the mtDNA. What of aerobic eukaryotes?

John et al. (2019) recently argued that the syndinialean parasite *Amoebophrya* sp. strain AT5 (Syndiniales, Dinoflagellates) lacks a mitochondrial genome despite having an otherwise conventional aerobic mitochondrion. If true, this would break the paradigm that mitochondrial oxidative phosphorylation is universally linked to the presence of mtDNA (Johnston and Williams 2016). Support for the loss of a mitogenome in *Amoebophrya* strain AT5 was based primarily on cell staining and whole-genome sequencing data. Using confocal laser scanning microscopy in conjunction with SYTOX staining, the authors did not observe DNA in *Amoebophrya* mitochondria but easily identified it in the nucleus. Similarly, their high-coverage total cellular DNA and RNA sequencing of strain AT5 revealed no obvious reads or contigs of mitochondrial origin, with the exception of two small (<55 nt) fragments of *cox1* (representing the C-terminal moiety of the protein) on scaffolds 46 and 1091, both of which had typical characteristics of nuclear DNA. No other pieces of *cox1* were identified.

More recently, a genome study of two other *Amoebophrya* strains (A120 and A25) also reported nuclear-encoded *cox1* fragments, corresponding to the metal-binding sites located near the C-terminus of the protein and harboring mitochondrial transit peptides (fig. 1A) (Farhat et al. 2021). But, here again, the authors were unable to identify any regions encoding the highly conserved oxygen-binding domain of COX1, which is all the more intriguing given that *Amoebophrya* is an aerobic clade (John et al. 2019; Farhat et al. 2021). In the functional annotations of the genomes, Farhat et al. did uncover a slew of nuclear-encoded, mitochondrial-targeted proteins involved in mtDNA replication and gene expression, including a homolog of plant organellar DNA polymerases, a DNA-directed RNA polymerase, 31 mitochondrial ribosomal proteins, and a monomeric phenylalanine-tRNA (FARS2) ligase (fig. 1B; supplementary table S1, Supplementary Material online), all showing expression during at least one life-cycle stage. This has led to doubts about

Letter

whether *Amoebophrya* has truly lost its mitochondrial genome (Farhat et al. 2021).

Among the closest known relatives of *Amoebophrya* species for which significant mtDNA sequence data are available are members of the pathogenic dinoflagellate genus *Perkinsus*, whose mtDNAs are highly reduced, fragmented, and undergo induced frameshifts at AGG and CCC positions (Masuda et al. 2010; Zhang et al. 2011). Using *Perkinsus* mitochondrial gene sequences as BLAST queries against the *Amoebophrya* A120 and A25 genome assemblies, we identified in each strain an mtDNA-like contig encoding the missing oxygen-binding domain of COX1 (fig. 1A). These putative mtDNA contigs (supplementary file S1, Supplementary Material online) had lengths of 3,122 nt (A120) and 4,479 nt (A25), did not map to the nuclear genome scaffolds, and were AT-rich ($\geq$67%). This is consistent with the nucleotide composition of other dinoflagellates mitogenomes (Waller and Jackson 2009) but contrasts with the low AT content of the A120 and A25 nuclear genome assemblies ($< 53\%$), including the nuclear DNA-located *cox1* fragments, which are 38–51% AT. We also observed contiguous RNA-seq mapping across the full length of the mitochondrial contigs (max avg. coverage = 295X in A120 and 310X in A25), in line with the pervasive organelle transcription of other clades (Sanitá Lima and Smith 2017). No other mitochondrial-like genes were found on the mtDNA-like contigs, but each contained a terminal inverted repeat, which could be folded into a stem-loop secondary structure (fig. 1D; supplementary fig. S1, Supplementary Material online). Terminal inverted repeats are a common feature of mitochondrial genomes, including those of Alveolates (Waller and Jackson 2009; Smith and Keeling 2013), and are thought to increase the stability of linear DNA molecules (Tomáška et al. 2009).

The portions of the A120 and A25 mtDNA-like contigs that show sequence similarity to *cox1* are relatively small: 277 and 366 nt, respectively, including the final stop codon (supplementary file S2, Supplementary Material online). Similar to *Perkinsus*, the *cox1* segment of A120 requires a $+1$ frameshift at a GGC position to keep the coding region intact; this frameshift was not found in the corresponding region of the A25 *cox1* sequence (supplementary file S3, Supplementary Material online). To better understand how these segments might be functioning within *Amoebophrya* strains A120 and A25, we constructed an artificial COX1 protein sequence (arCOX1) based on the conserved regions of the nuclear and mitochondrial *cox1*-like fragments (fig. 1A and C; supplementary file S4, Supplementary Material online). The putatively mtDNA-encoded *cox1* pieces from A120 and A25 represent 72 and 45 amino acids, respectively, of arCOX1. The remaining 149–150 amino acids (for A25 and A120, respectively) of arCOX1 come from nuclear-encoded segments. Interestingly, only 39–48% of the nuclear *cox1*-like ORFs (when including the transit peptides) are represented in arCOX1. It is possible that the remaining peptides encoded by the nuclear *cox1*-like fragments are involved in the assembly of the COX1 subunit. Fragmentation and transfer of the C-terminal portion of *cox1* to the nucleus is ancient and widespread in eukaryotes (Gawryluk and Gray 2010),

suggesting that mechanisms already exist for the reassembly of a fragmented COX1 into a functional cytochrome *c* oxidase (CIII) complex.

We did re-examine the genome assembly of *Amoebophrya* strain AT5 for putative mtDNA-like contigs, but none were found. Keep in mind, however, that despite the A120 and A25 genome assemblies being of better quality than that of AT5, no mtDNA-like contigs were identified in the genome assemblies. Nevertheless, we did uncover an mtDNA-like sequence from the RNA-seq reads downloaded from the NCBI BioProject PRJNA274490 (supplementary file S1, Supplementary Material online). This 812-nt sequence has an AT content of 65.4% and encodes a similar portion of COX1 as the mtDNA-like sequences of A120 and A25 (supplementary files S2–S4, Supplementary Material online), providing further support for the existence of mtDNA. Moreover, the *cox1* fragment from AT5 also requires a $+1$ frameshift at a GGC position similar to that found in A120 but absent in A25 (supplementary file S2, Supplementary Material online).

Much more work is needed before one can definitively say if members of the *Amoebophrya* genus have or do not have an mtDNA. But, for now, the data are pointing toward the existence of a mitochondrial genome, albeit one that might be among the most reduced of all eukaryotes, even more so than the highly reduced mtDNAs of other myzozoans (Gagat et al. 2017). Further detailed cataloguing and analysis of nuclear-encoded, mitochondrial-targeted proteins in *Amoebophrya*, especially those involved in mtDNA maintenance and in the electron transport chain of the oxidative phosphorylation pathway, will undoubtedly provide even more insights into mitochondrial function within this fascinating clade.

Materials and Methods

For each *Amoebophrya* strain, *cox1* sequences from *Perkinsus marinus* (HQ670240.1) and *Perkinsus chesapeaki* (HQ670238.1) were used as BLAST queries (v.2.9.0; e value 1e-03) against two libraries: one containing unique DNA-seq reads from Farhat et al. (2021) and another containing meta-transcriptomes assembled for each sampling time using data from Farhat et al. (2018). Bowtie2 v2.3.5 (–no-discordant –end-to-end) (Langmead and Salzberg 2012) and scripts from the SAMtools v.1.3.1 package (Li et al. 2009) were used to map positive hits to RNA-seq reads from the host (Farhat et al. 2018) in order to filter out contigs with read coverages. The integrity of the remaining contigs were checked by mapping DNA-seq reads and contigs were extended with MITObim v.1.8 (Hahn et al. 2013). The expression of the mtDNA-like contigs over a full infection cycle was monitored with Bowtie2. *cox1* fragment in AT5 was identified by BLAST queries (v.2.9.0; e value 1e-05) of A120 and A25 sequences against trimmed RNA-seq reads obtained from the NCBI BioProject PRJNA274490. COX1-like fragments were translated using the ExPASy Translate tool from SIB ExPASy Bioinformatics Resources Portal (http://web.expasy.org/translate/, last accessed February 2021) (Artimo et al. 2012) with the Standard Genetic Code and the $+1$ frameshift in A120
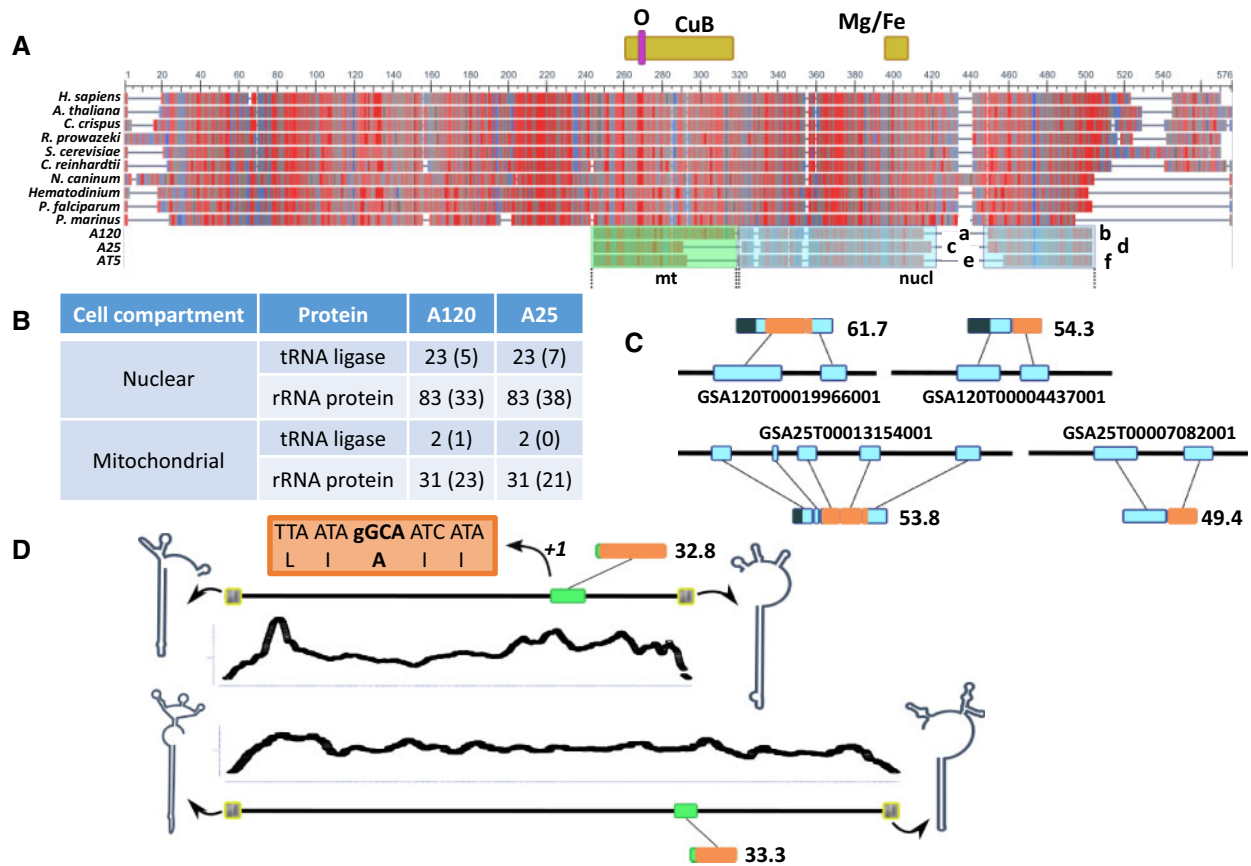
Fig. 1. Putative mitochondrial chromosome in three strains (A120, A25, and AT5) of the parasitoid *Amoebophrya* (Syndiniales, Dinoflagellates). (A) Alignment of concatenated COX1-like blocks into artificial COX1 protein sequences (arCOX1) identified in A25, A120, and AT5 to homologues from representative clades (sequence alignment can be found as supplementary file S4, Supplementary Material online). The nuclear-encoded *cox1*-like fragments (*a–f*) are derived from predicted proteins for each strain: *a* = GSA120T00019966001; *b* = GSA120T00004437001; *c* = GSA25T00013154001; *d* = GSA25T00007082001; *e* = g15932.t1; *f* = g833.t1. mt: mitochondrial-encoded fragment; nucl: nuclear-encoded fragments. Prediction of metal and oxygen-biding domains are based on the human COX1 (uniprot.org/uniprot/P00395). CuB: copper B-binding; Mg/Fe: magnesium and iron-biding; O: oxygen-biding. Colors in alignment are based on the hydropathy scale from NCBI Multiple Sequence Alignment Viewer 1.18.1 (https://www.ncbi.nlm.nih.gov/projects/msaviewer/, last accessed February 2021) where red and blue represent hydrophobic and hydrophilic residues, respectively. (B) Count of tRNA ligase and ribosomal proteins identified in A120 and A25. Numbers in brackets reflect the presence of mitochondrial transit peptide identified by TargetP v.2.0 and/or MitoProt II v1.101. (C) Predicted transcription of COX1-like fragments in the nucleus of A120 and A25. CDS are displayed as blue boxes transcribed in the same direction (left to right) and assembled into contiguous ORFs for nuclear genes, where dark regions correspond to mitochondrial transfer peptides. Orange-shaded regions of nuclear ORFs correspond to conserved blocks of arCOX1 in the alignment (A) and numbers display the GC content. (D) Predicted transcription of COX1-like ORFs from the mtDNA-like contig found in A120 and A25. All genomic features are displayed proportionally to the length of the contigs. DNA-read coverage is displayed for each contig. CDS are displayed as green boxes transcribed in the same direction (left to right). Shaded regions of nuclear ORFs correspond to conserved blocks of arCOX1 in the alignment (A) and numbers display the GC content. Schematic representations of secondary structures of terminal repeats identified in the mtDNA-like contigs (gray boxes at the end of the contigs) were predicted with the RNA structure Web Servers; for larger images of these structures, see supplementary files S5 and S6, Supplementary Material online.

and AT5 were identified manually. Conserved COX1-like blocks for each strain were concatenated to create an artificial COX1 "protein" based on their alignment to reference orthologues using the online version of MAFFT v7 aligner (Katoh et al. 2019). The artificial COX1 proteins were realigned to orthologues and the prediction of functional domains were inferred from the human COX1 (https://uniprot.org/uniprot/P00395, last accessed February 2021). Mitochondrial transfer peptides for nuclear fragments were predicted with TargetP v.2.0 (Armenteros et al. 2019) from the DTU Health Tech online server (https://services.healthtech.dtu.dk/, last accessed February 2021) and MitoProt II v1.101 (Claros and Vincens 1996). Terminal repeats were identified by BLASTn

from the NCBI Web BLAST server (https://blast.ncbi.nlm.nih.gov/Blast.cgi, last accessed February 2021) of putative mtDNA contigs against themselves and the secondary structure were predicted using the RNAstructure Web Servers (https://rna.urmc.rochester.edu/RNAstructureWeb/index.html, last accessed February 2021) with default parameter choosing DNA as Nucleic Acid Type (Bellaousov et al. 2013).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Data Availability

No new data were generated in support of this research. The data underlying this article can be found on EMBL-EBI BioProject PRJEB39972 and available at http://application.sb-roscoff.fr/blast/hapar/download.html (last accessed February 2021), as well as on NCBI BioProject PRJNA274490.

References

Allen JF. 1993. Control of gene expression by redox potential and the requirement for chloroplast and mitochondrial genomes. *J Theor Biol*. 165(4):609–631.

Armenteros JJA, Salvatore M, Emanuelsson O, Winther O, Von Heijne G, Elofsson A, Nielsen H. 2019. Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance*. 2(5):e201900429.

Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, et al. 2012. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res*. 40(W1):W597–W603.

Barberà MJ, Ruiz-Trillo I, Tufts JY, Bery A, Silberman JD, Roger AJ. 2010. *Sawyeria marylandensis* (heterolobosea) has a hydrogenosome with novel metabolic properties. *Eukaryot Cell*. 9(12):1913–1924.

Bellaousov S, Reuter JS, Seetin MG, Mathews DH. 2013. RNAstructure: web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res*. 41:471–474.

Claros MG, Vincens P. 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem*. 241(3):779–786.

Farhat S, Florent I, Noel B, Kayal E, Da Silva C, Bigeard E, Alberti A, Labadie K, Corre E, Aury J-M, et al. 2018. Comparative time-scale gene expression analysis highlights the infection processes of two Amoebophrya strains. *Front Microbiol*. 9:2251.

Farhat S, Le P, Kayal E, Noel B, Bigeard E, Corre E, Maumus F, Florent I, Alberti A, Aury J-M, et al. 2021. Rapid protein evolution, organellar reductions, and invasive intronic elements in the marine aerobic parasite dinoflagellate Amoebophrya spp. *BMC Biol*. 19(1):1.

Gagat P, Mackiewicz D, Mackiewicz P. 2017. Peculiarities within peculiarities? Dinoflagellates and their mitochondrial genomes. *Mitochondrial DNA Part B*. 2(1):191–195.

Gawryluk RM, Gray MW. 2010. An ancient fission of mitochondrial cox1. *Mol Biol Evol*. 27(1):7–10.

Gill EE, Diaz-Triviño S, Barberà MJ, Silberman JD, Stechmann A, Gaston D, Tamas I, Roger AJ. 2007. Novel mitochondrion-related organelles in the anaerobic amoeba *Mastigamoeba balamuthi*. *Mol Microbiol*. 66(6):1306–1320.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. *Nucleic Acids Res*. 41(13):e129–e129.

John U, Lu Y, Wohlrab S, Groth M, Janouškovec J, Kohli GS, Mark FC, Bickmeyer U, Farhat S, Felder M, et al. 2019. An aerobic eukaryotic parasite with functional mitochondria that likely lacks a mitochondrial genome. *Sci Adv*. 5(4):eaav1110.

Johnston IG, Williams BP. 2016. Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retention. *Cell Syst*. 2(2):101–111.

Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, Novák L, Žárský V, Barlow LD, Herman EK, et al. 2016. A eukaryote without a mitochondrial organelle. *Curr Biol*. 26(10):1274–1284.

Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. 20(4):1160–1166.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9(4):357–359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

Masuda I, Matsuzaki M, Kita K. 2010. Extensive frameshift at all AGG and CCC codons in the mitochondrial cytochrome c oxidase subunit 1 gene of *Perkinsus marinus* (Alveolata; Dinoflagellata). *Nucleic Acids Res*. 38(18):6186–6194.

Rada P, Tachezy J. 2019. The proteome of *T. vaginalis* hydrogenosomes. In: Tachezy J, editor. Hydrogenosomes and mitosomes: mitochondria of anaerobic eukaryotes. Berlin (Germany): Springer. p. 177–204.

Roger AJ, Muñoz-Gómez SA, Kamikawa R. 2017. The origin and diversification of mitochondria. *Curr Biol*. 27(21):R1177–R1192.

Sanitá Lima M, Smith DR. 2017. Pervasive, genome-wide transcription in the organelle genomes of diverse plastid-bearing protists. *G3* 7:3789–3796.

Santos HJ, Makiuchi T, Nozaki T. 2018. Reinventing an organelle: the reduced mitochondrion in parasitic protists. *Trends Parasitol*. 34(12):1038–1055.

Smith DR, Keeling PJ. 2013. Gene conversion shapes linear mitochondrial genome architecture. *Genome Biol Evol*. 5(5):905–912.

Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E, Dellaire G, Soanes DM, Van Der Giezen M, Roger AJ. 2014. A SUF Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist *Pygsuia*. *Curr Biol*. 24(11):1176–1186.

Tomáška L, Nosek J, Kramara J, Griffith JD. 2009. Telomeric circles: universal players in telomere maintenance? *Nat Struct Mol Biol*. 16(10):1010–1015.

Tsaousis AD, Keithly JS. 2019. The mitochondrion-related organelles of Cryptosporidium species. In: Tachezy J, editor. Hydrogenosomes and mitosomes: mitochondria of anaerobic eukaryotes. Berlin (Germany): Springer. p. 243–266.

Waller RF, Jackson CJ. 2009. Dinoflagellate mitochondrial genomes: stretching the rules of molecular biology. *Bioessays* 31(2):237–245.

Yahalomi D, Atkinson SD, Neuhof M, Sally Chang E, Philippe H, Cartwright P, Bartholomew JL, Huchon D. 2020. A cnidarian parasite of salmon (Myxozoa: Henneguya) lacks a mitochondrial genome. *Proc Natl Acad Sci USA*. 117(10):5358–5363.

Zhang H, Campbell DA, Sturm NR, Dungan CF, Lin S. 2011. Spliced leader RNAs, mitochondrial gene frameshifts and multi-protein phylogeny expand support for the genus perkinsus as a unique group of alveolates. *PLoS One* 6(5):e19933.