The cytochrome oxidase subunit 1 gene (*cox1*) from the dinoflagellate, *Crypthecodinium cohnii*

John E. Norman, Michael W. Gray*

Program in Evolutionary Biology, Canadian Institute for Advanced Research, Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

Received 11 July 1997

Abstract To date, no genes have been characterized from dinoflagellate mitochondrial DNA. Here we present the complete sequence of the gene (*cox1*) encoding subunit 1 of cytochrome *c* oxidase in the dinoflagellate, *Crypthecodinium cohnii*. Analysis of nucleotide and deduced amino acid sequences predicts a protein of 523 amino acids that is translated using universal initiation, stop and tryptophan codons. COX1 amino acid identity and phylogenetic tree analyses strongly support a close evolutionary relationship between dinoflagellates and apicomplexans; however, inclusion of the ciliates in this clade is less well supported, a result likely due to the highly derived nature of ciliate COX1 sequences.

Key words: Cytochrome oxidase; Alveolate; Codon usage; Molecular phylogeny; Evolution; (*Crypthecodinium cohnii*)

1. Introduction

Dinoflagellates comprise a large and structurally diverse group of unicellular eukaryotes (protists). Phylogenetic reconstructions, inferred from small subunit rRNA gene sequences, show that the dinoflagellates (Dinozoa) share a common ancestry with apicomplexans (Apicomplexa, a group of parasitic protists) and ciliates (Ciliophora) [1–4]. Within this assemblage, termed Alveolata (alveolates) [1,5], the dinoflagellates and apicomplexans cluster together, forming a sister group to the ciliates [1,4,6]. In addition to molecular data, the concept of an alveolate clade is supported by ultrastructural features such as tubular mitochondrial cristae, cortical alveoli [4,5] (outer membrane vesicles) and rows of microtubules just below the plasma membrane [5].

Most of what is known about mitochondrial genome structure and organization comes from examination of mitochondrial DNA (mtDNA) in the three most recently emerged eukaryotic lineages: animals, plants and fungi (reviewed in refs. [7,8]). Mitochondrial DNA has been characterized in only a fraction of the extant protist phyla; within the alveolate lineage, nothing is yet known about the dinoflagellate mitochondrial genome. In contrast, extensive sequence is available for mtDNA from four apicomplexan species (*Plasmodium falciparum* [9,10], *Plasmodium yoelii* [11], *Plasmodium gallinaceum* [12] and *Theileria parva* [13]) and two ciliates (*Paramecium aurelia* (reviewed in ref. [14]) and *Tetrahymena pyriformis* (reviewed in ref. [8]). The apicomplexan mitochondrial genomes (6–7 kbp) are the smallest known, containing only three protein-coding genes (*cox1*, *cox3* and *cob*) [9–13]. Among other interesting features of alveolate mtDNAs is the unusual arrangement of the mitochondrial rRNA genes: discontinuous in the ciliates and also rearranged in the case of the LSU rRNA gene of *T. pyriformis* (reviewed in refs. [8] and [14]); discontinuous and rearranged in the apicomplexans ([13]; reviewed in ref. [15,16]). From comparisons of the sequence of *cox1*, the gene specifying subunit 1 of cytochrome *c* oxidase (the most highly conserved mitochondrially encoded protein), the *T. pyriformis* COX1 protein appears to be one of the longest known (containing an insert of 108 aa that is present only in the ciliates) [17] whereas *P. falciparum* COX1 is thought to be the shortest. The mitochondrial translation system in both groups also appears to use non-universal initiation codons (reviewed in [14,15]), a situation similar to that described in mammals [18,19].

In this paper we report the first characterization of a mtDNA element from a dinoflagellate. Fragments of mtDNA, isolated from *Crypthecodinium cohnii*, have been found to contain an open reading frame (ORF) homologous to *cox1*. Analysis of *cox1* nucleotide and predicted amino acid sequence suggests that *C. cohnii* utilizes universal initiation, stop and tryptophan codons. We discuss these findings in the context of the proposed common ancestry of the dinoflagellates, apicomplexans and ciliates.

2. Materials and methods

2.1. Culturing methods and nucleic acid extraction

C. cohnii strain WH-d (kindly provided by Dr. Carl Beam (Brooklyn College, New York)) was grown axenically in MLH liquid medium [20] at 27°C in the dark with aeration. Mid-log phase cells were harvested by centrifugation (850×*g*, 10 m) and washed twice with resuspension buffer (25 mM Tris-HCl, 10 mM EDTA (pH 8.0)). Total nucleic acids were prepared by lysing cells in a French pressure cell (Aminco) (2000 lb/in^2) in resuspension buffer followed by standard phenol extraction of the resulting lysate [21]. Mitochondrial DNA was isolated by subcellular fractionation [22], except that a cell lysate was prepared by disrupting cells with glass beads [23]. Following sucrose gradient centrifugation of a crude mitochondrial fraction, purified mitochondria were found in the pellet.

2.2. PCR amplification, cloning and DNA sequencing

PCR amplifications were carried out using 50 μL reaction mixtures containing 25 mM glycine KOH (pH 9.3), 50 mM KCl, 2 mM $MgCl_2$, 1 mM DTT, 0.2 mM of each dNTP, 0.001% gelatin, 10 pmol of each primer, 0.5 units of *Taq* DNA polymerase (Gibco BRL) and 100 ng of total cellular DNA. Forward (5′-TTATTTTGRTTTTTTGGTCATCCTGARGT) and reverse (5′-TCTGGGTAGTCTGGTATTCKTCKTGGCA) primers, the design of which was based on conserved regions of COX1 sequence (LFWFFGHPEV and

*Corresponding author. Fax: (1) (902) 494-1355.
E-mail: wgray@is.dal.ca

Abbreviations: kbp, 10^3 base pairs; nt, nucleotide; bp, base pair; PCR, polymerase chain reaction; rRNA, ribosomal RNA; LSU, large subunit; SSU, small subunit; SDS, sodium dodecyl sulphate; EDTA, ethylenediaminetetraacetate; DTT, dithiothreitol; dNTP, deoxynucleoside triphosphate

PII S0014-5793(97)00938-1

MPRRIPDYPD, respectively), were a gift from Dr. B.F. Lang (Université de Montréal, Montréal, QC). Amplification was carried out using a 2 min denaturation period (94°C), followed by 30 cycles of 30 s denaturation at 94°C, 30 s annealing at 50°C and 30 s extension at 74°C. A 656-bp PCR product was visualized in a 1.0% agarose gel and cloned into pT7Blue T-Vector (Novagen) following manufacturer's specifications. Several constructs were sequenced using *fmol* cycle sequencing (Promega).

2.3. Southern hybridization analysis, mtDNA cloning and sequencing

A fraction enriched in mtDNA was hydrolyzed with either *Eco*RI or *Xba*I, and the products were electrophoresed in a 1.0% agarose gel containing 1×TAE (40 mM Tris-acetate, 1 mM EDTA) for 16 h. The resolved DNA fragments were transferred to a nylon membrane using conventional alkaline transfer protocols [21]. Hybridization was allowed to continue overnight with randomly labelled *cox1* PCR product [24] at 42°C in hybridization solution (5×SSPE = 180 mM NaCl, 10 mM NaH_2PO_4, 1 mM EDTA (pH 7.7)), 50% formamide and 1×BLOTTO (= 5% skim milk powder, 10% SDS) [25]. After hybridization, membranes were washed in 0.5×SSPE and 0.1% SDS followed by 0.1×SSPE and 0.1% SDS at 25°C, and finally for 30 min at 50°C in 0.1×SSPE and 1.0% SDS before being subjected to autoradiography. The sizes of *Eco*RI fragments that hybridized to the probe were determined by comparison with λ DNA/*Hin*dIII fragments. Additional mtDNA-enriched aliquots were cut with *Eco*RI and electrophoresed in 1.0% agarose gels (as described above), and regions corresponding to discrete hybridization bands were excised. Size-fractionated DNA was ligated into pBluescript KS+ (Stratagene) using T4 DNA ligase and the constructs were transformed into competent *E. coli* strain DH5α cells [26]. Positive clones were identified by hybridization of colony lifts [21] with randomly labelled *cox1* PCR product and used in the construction of a series of overlapping deletion clones, generated by digestion with exonuclease III/mung bean nuclease [27]. Inserts were sequenced on both strands using a *fmol* cycle sequencing kit (Promega).

2.4. Data and phylogenetic analyses

Sequence data were assembled using the GDE software package [28]) on a Sun SPARCstation 4. Amino acid alignments were performed using CLUSTAL W [29], manually optimized and examined using SeqVu 1.0.1 (J. Gardner, Garvan Institute of Medical Research, Sydney, Australia). COX1 phylogenetic trees were generated using the neighbor-joining method [30] (PHYLIP 3.5c [31], neighbor) based on distance matrices calculated (PHYLIP 3.5c [31], protdist) using a Dayhoff PAM amino acid substitution model [32]. Bootstrap resampling analyses (100 data sets) were performed (PHYLIP 3.5c [31], seqboot) to assess branch support [33].

3. Results

3.1. Isolation and sequence of the cox1 gene

An internal portion of *C. cohnii cox1* was initially identified by sequencing a 656-nt PCR product. In Southern hybridizations of a mtDNA-enriched fraction hydrolyzed with *Eco*RI or *Xba*I and probed with labelled PCR product, a smear ranging in size from <2.0 kbp to 23 kbp was visible in the lanes containing uncut control and the *Xba*I digest (Fig. 1, lanes 1 and 3), with the *Xba*I lane displaying a greater degree of homogeneity over the size range. In the *Eco*RI digest, four fragments of approximate size 3.2, 3.8, 5.0 and 5.5 kbp were observed (Fig. 1, lane 2), superimposed on a faint trailing smear starting at the 5.5-kbp band. The stoichiometries of the four *Eco*RI bands are unequal, with each of the two smaller bands displaying roughly twice the intensity of each of the two larger bands. Work currently in progress suggests that these size differences are the result of variability in sequence flanking the *cox1* gene.

Screening of a size-fractionated *Eco*RI library (as described in Section 2) yielded three clones (pCc15, pCc32 and pCc42), each containing a 5.0-kbp insert. Mapping and sequence analysis showed that each insert contained a single continuous ORF (1569 nt) homologous to *cox1*. The flanking regions of pCc15 and pCc42 are identical; all three constructs contain the same 124-nt sequence downstream of *cox1*, but sequence >400 nt upstream of *cox1* is different in the pCc15/pCc42 pair than in pCc32.

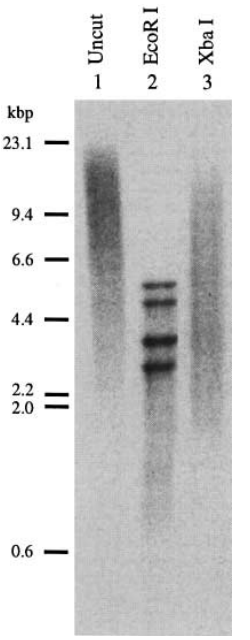

Fig. 1. Autoradiography showing the results of Southern hybridization analysis of a *C. cohnii* fraction enriched in mtDNA. The DNA was hydrolyzed with *Eco*RI (lane 2) or *Xba*I (lane 3) and probed with randomly labelled *cox1* PCR product. The undigested control fraction (lane 1) is labelled 'uncut'. Size markers (λ DNA hydrolyzed with *Hin*dIII) are denoted on the left side of blot.

Although we report the first case of a mtDNA element characterized from a dinoflagellate, a partial *cox1* sequence from *C. cohnii* already exists in GenBank (accession number L01984). Nucleotide and amino acid sequence comparisons clearly indicate that the genomic source of the two sequences is different and BLAST searches of the previously published partial COX1 sequence show higher alignment scores with bacterial than with mitochondrial COX1 homologs (data not shown); moreover, an oligonucleotide based on the L01984 sequence does not hybridize to *C. cohnii* DNA (W. Fischer and M. W. G. unpublished results). These observations suggest bacterial contamination as the most likely source of the L01984 sequence.

3.2. Comparison of COX1 amino acid sequences

The deduced amino acid sequence of *C. cohnii cox1* predicts a protein of 523 residues beginning with a universal initiation codon, ATG (Met), and terminating in a TAA stop codon. Continuity of the ORF and normal-length protein sequence rule out the presence of introns. Fig. 2 shows an alignment of COX1 sequences from *C. cohnii*, *P. falciparum*, *T. pyriformis*, *Trypanosoma brucei*, *Beta vulgaris*, *Neurospora crassa* and *Bos taurus*. Highest amino acid identity (59%) is between *C. cohnii* and *P. falciparum* COX1 sequences whereas the lowest value (33%) is with *T. pyriformis* COX1. Notably, COX1 from the more distantly related organisms *T. brucei* (kinetoplastid), *B. vulgaris* (plant), *N. crassa* (fungi) and *B. taurus* (animal)

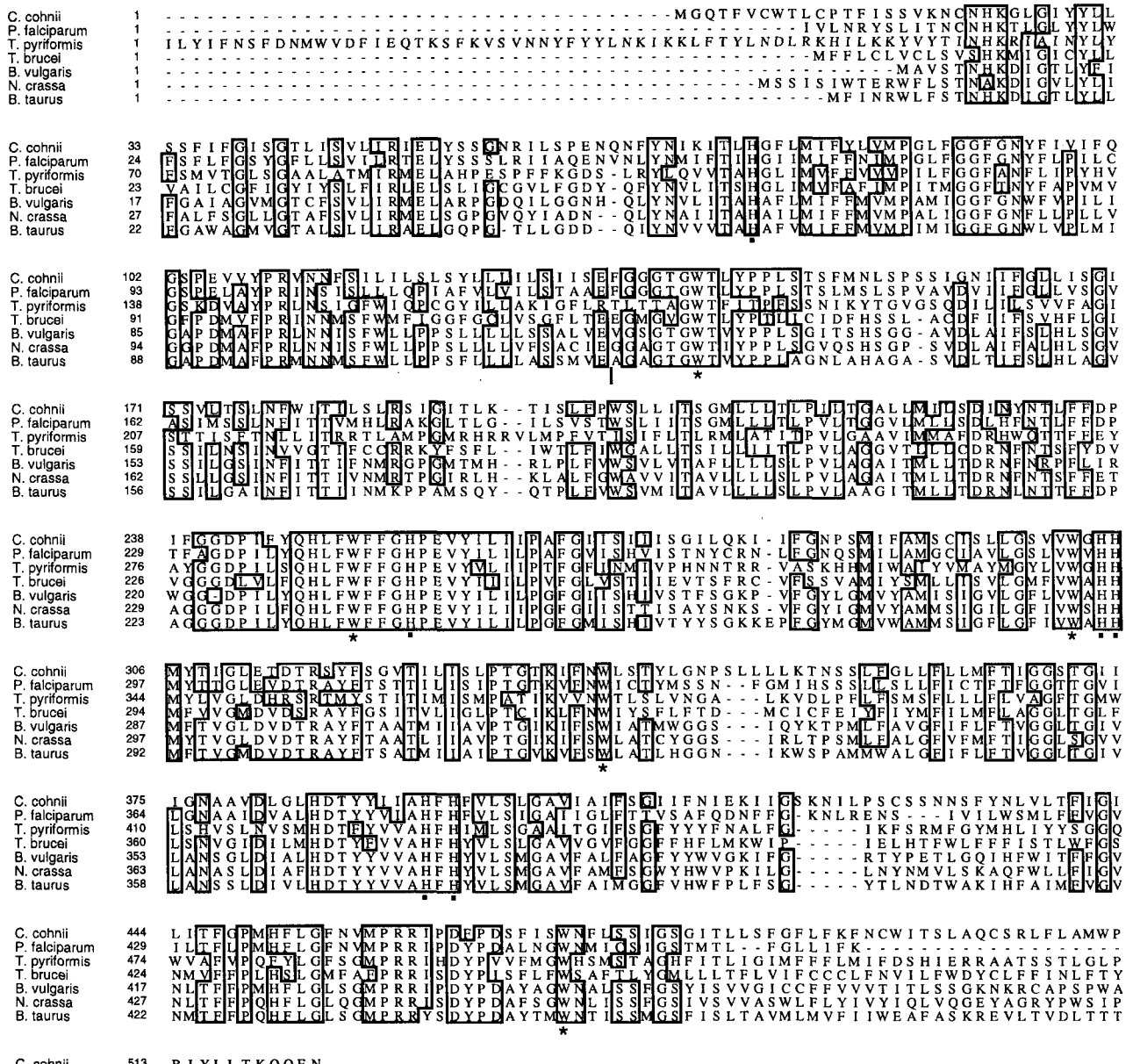

Fig. 2. Alignment of COX1 amino acid sequences. Boxed segments indicate regions that show >60% amino acid identity among all compared taxa. Dots (.) show the positions of the six conserved His residues. Asterisks (*) indicate the three conserved Trp residues that are encoded by TGA in *T. pyriformis* but by TGG in *C. cohnii* and apicomplexans. A stroke (|) denotes the position of a ciliate-specific insertion of 108 amino acids that has been omitted from the alignment. Inferred gaps are indicated by dashes (-). Taxa included in the alignment are: C. cohnii, *Crypthecodinium cohnii* (AF012554, this report); P. falciparum, *Plasmodium falciparum* (M76611); T. pyriformis, *Tetrahymena pyriformis* (X06133); T. brucei, *Trypanosoma brucei* (X01094); B. vulgaris, *Beta vulgaris* (X57693); N. crassa, *Neurospora crassa* (X14669); and B. taurus, *Bos taurus* (V00654). Numbers within parentheses refer to EMBL accession numbers.

show intermediate levels of identity (37%, 46%, 45% and 41%, respectively) to *C. cohnii* COX1.

Amino acid identity is high throughout the conserved core region (between residues 59 and 520) and includes the six invariant histidine residues (dots) that bind heme *a*, Cu_B and heme a_3 (reviewed in ref. [34]). The C-terminal region displays the greatest degree of length variability, with *P. falciparum* COX1 being the shortest and *C. cohnii* COX1 displaying an intermediate length. The N-terminal region of COX1 tends to be more uniform in length than the C-terminal portion, a difference reflected in the level of sequence conservation in these two regions.

3.3. Codon usage

The most distinguishing feature of *cox1* in *C. cohnii* is the exclusive use of TGG to encode tryptophan (Trp) (Table 1); in contrast, many mitochondrial translation systems utilize both TGG and TGA (reviewed in [8] and [35]). TGA, a

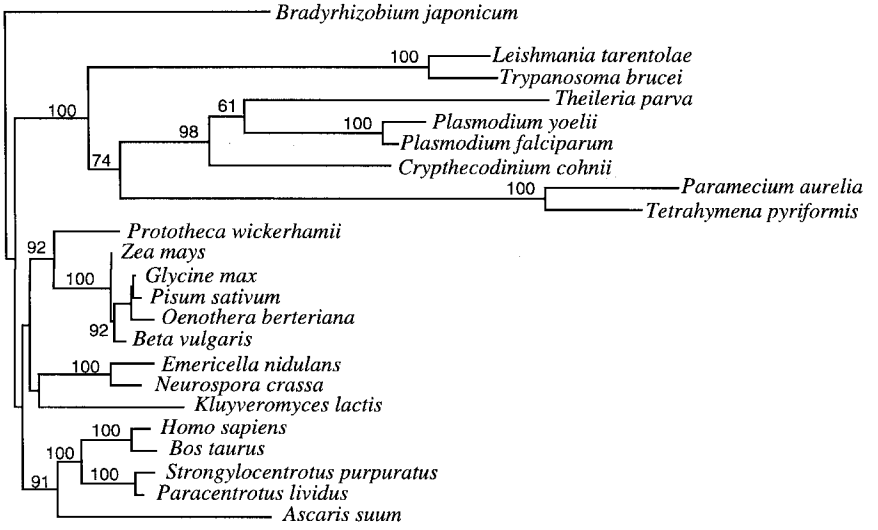

Fig. 3. Phylogenetic analysis of COX1 amino acid sequence by neighbor-joining. Numbers indicate the percentage of bootstrap resamplings (out of 100) that support the inferred topology; only values >90% are listed with the exception of the alveolate clade, where all values are shown. Complete amino acid alignment includes only the COX1 core region (amino acid residues 59–520, based on the alignment in Fig. 2). The scale bar indicates 10 substitutions per 100 amino acids. Taxa included in the analysis are those mentioned in Fig. 2 as well as *Bradyrhizobium japonicum* (X54800), *Leishmania tarentolae* (P14544), *Theileria parva* (S41689), *Plasmodium yoelii* (A38891), *Paramecium aurelia* (X15917), *Prototheca wickerhamii* (X68721), *Zea mays* (X02660), *Glycine max* (M16884), *Pisum sativum* (X14409), *Oenothera berteriana* (X05465), *Emericella nidulans* (X00790), *Kluyveromyces lactis* (X57546), *Homo sapiens* (V00662), *Strongylocentrotus purpuratus* (X12631), *Paracentrotus lividus* (J04815) and *Ascaris suum* (X54253). Numbers within parentheses refer to EMBL, GenBank or Swiss-Prot accession numbers.

stop codon in the universal system, is not found in *C. cohnii* COX1; in contrast, TGA is used exclusively in *T. pyriformis* COX1 to encode Trp, including five highly conserved Trp residues (Fig. 2). *C. cohnii cox1* (76% A+T) also exhibits a preference for A or T in the third (wobble) position of codons (87%) whereas first (excluding Trp and Met) and second position A+T biases are lower (69% and 62%, respectively). First-position preferences for A or T in Leu and Arg codons also contribute to the high A+T content of *C. cohnii cox1*. Of the nine Arg residues present, eight are encoded by AGA, one by AGG, and none by CGN. Similarly, of 78 Leu residues, 55 are specified by TTA, whereas CTC, CTA and CTG together account for only seven codons, and TTG is not used at all. Although codon usage in *cox1* might suggest an evolutionary drive toward increased A+T content in the mtDNA, it is noteworthy that flanking sequences have the same or lower A+T content as the *cox1* coding region.

3.4. Phylogenetic reconstruction based on COX1 sequence comparisons

Phylogenetic trees were constructed using several members of each major taxonomic group to accurately gauge COX1 global topology. A COX1 phylogenetic tree based on nearest-neighbor analyses (Fig. 3) shows that members of the kinetoplastids and alveolates branch together, forming the most basal mitochondrial clade with 100% bootstrap support. However, the long branch lengths within the two groups suggest than this affiliation is more likely indicative of a 'long branch attraction' artifact [36] than of a close evolutionary relationship. In fact, the phylogenetic distances between bac-

Table 1
Codon usage in *C. cohnii cox1* gene

aa	Codon	#	%	aa	Codon	#	%	aa	Codon	#	%	aa	Codon	#	%
Phe	TTT	42	85.7	Ser	TCT	24	46.1	Tyr	TAT	19	100	Cys	TGT	6	85.7
	TTC	7	14.3		TCC	2	3.9		TAC	0	0.0		TGC	1	14.3
Leu	TTA	55	70.5		TCA	26	50.0	***	TAA	1	100	***	TGA	0	0.0
	TTG	0	0.0		TCG	0	0.0	***	TAG	0	0.0	Trp	TGG	10	100
Leu	CTT	16	20.5	Pro	CCT	5	21.7	His	CAT	10	100	Arg	CGT	0	0.0
	CTC	2	2.6		CCC	0	0.0		CAC	0	0		CGC	0	0.0
	CTA	3	3.8		CCA	17	73.9	Gln	CAA	5	62.5		CGA	0	0.0
	CTG	2	2.6		CCG	1	4.4		CAG	3	37.5		CGG	0	0.0
Ile	ATT	26	35.1	Thr	ACT	9	28.1	Asn	AAT	25	89.3	Ser	AGT	5	71.4
	ATC	7	9.5		ACC	2	6.3		AAC	3	10.7		AGC	2	28.6
	ATA	41	55.4		ACA	21	65.6	Lys	AAA	8	72.7	Arg	AGA	8	88.9
Met	ATG	13	100		ACG	0	0.0		AAG	3	27.3		AGG	1	11.1
Val	GTT	10	55.6	Ala	GCT	5	50.0	Asp	GAT	8	100	Gly	GGT	21	42.9
	GTC	1	5.6		GCC	2	20.0		GAC	0	0.0		GGC	2	4.1
	GTA	7	38.8		GCA	3	30.0	Glu	GAA	6	75.0		GGA	24	4.1
	GTG	0	0.0		GCG	0	0.0		GAG	2	25.0		GGG	2	48.9

Percentage (%) indicates the proportion of codons used for each amino acid. Asterisks (***) denote stop codons.

terial COXA and non-kinetoplastid/alveolate COX1 sequences are shorter than distances found within the kinetoplastid/alveolate 'clade'. Within the alveolates, dinoflagellates and apicomplexans form a well-supported (98% bootstrap) monophyletic group, in agreement with nuclear rRNA trees, although overall support for the alveolate clade is low (bootstrap value 74%). Additional phylogenetic trees were also generated from much larger data sets that included > 60 taxa and eight different bacterial COX1 homologs (data not shown), but overall tree topology and bootstrap values did not vary significantly from those shown in Fig. 3.

4. Discussion

Amino acid sequence comparisons demonstrate that *C. cohnii* COX1 shares highest identity with *P. falciparum* COX1 and lowest with *T. pyriformis* COX1. The relatively low degree of amino acid identity between dinoflagellate and ciliate COX1 sequence is inconsistent with SSU rRNA phylogenies [1–4] but may be explained by the highly derived nature of many ciliate mitochondrial genes, including *cox1* [14,37,38].

Examination of nucleotide sequence in the vicinity of the predicted N-terminus of COX1 identifies a single potential ATG initiation codon. This is in contrast to the data available from apicomplexans (reviewed in ref. [15]) and ciliates (reviewed in ref. [14]) that suggest that alternate *cox1* initiation codons are used in these cases. Alternate initiation codons have been proposed for *P. falciparum* and *P. gallinaceum* (ATT) ([39], reviewed in ref. [15]), *T. parva* (AGT) [13] and *T. pyriformis* (ATA) [17] but the situations in *P. yoelii* [40] and *P. aurelia* [14] are unresolved. Why *C. cohnii* uses a conventional start codon when other members of the alveolates do not is unclear. However, considering that apicomplexans and ciliates also possess mitochondrial genes that have ATG start codons (e.g. *cob* in all apicomplexans [10,13,39,40] and *atp9* in *P. aurelia* [14,38]), all members of the alveolates, including dinoflagellates, may utilize both conventional and atypical start codons.

Changes in mitochondrial codon frequencies can occur as a result of AT selection pressure. Such pressure is manifested by an increase in the A+T content of spacer regions or codon third positions relative to first (with the exception of Leu or Arg codons) and second positions [35,41]. *C. cohnii* COX1 does exhibit a bias for A or T in codon third positions (Table 1); however, the A+T content of spacer regions is equal to or less than that of coding regions, making it difficult to argue that AT selection is at work. Table 1 also shows that all 10 Trp residues are encoded using the universal TGG codon; however, in many mitochondrial systems, TGA (a universal stop codon) also codes for Trp (reviewed in ref. [35]). Based on the COX1 results, it seems possible that TGA in *C. cohnii* is either an unassigned or stop codon. As in *C. cohnii*, *Plasmodium* mitochondrial protein genes do not appear to use TGA [9,15]; on the other hand *T. pyriformis* mtDNA uses TGA almost exclusively ([17]; G. Burger, M.W.G. et al., unpublished data) whereas the *P. aurelia* mitochondrial genome uses both TGA and TGG equally [14,42].

It has been suggested [35,43] that virtually all non-plant mitochondria (those of the oomycete, *Phytophthora infestaus*, being an exception [41]) use TGA to encode Trp, and that this switch in TGA coding occurred after the split of green plants from other eukaryotes. Recently, several non-plant species have been characterized whose mtDNAs do not use TGA to encode Trp, but which affiliate with species whose mtDNA uses TGA to specify Trp. Examples are the chytrid, *Allomyces*, contrasted with other fungi [44] and *Euglena* contrasted with kinetoplastids [45] (see also data compiled by the Organelle Genome Megasequencing Program (OGMP); http://megasun.bch.umontreal.ca/ogmp). This variant assignment of TGA is similar to what is found within the alveolates and further illustrates the point that codon usage is highly variable and basically uninformative with respect to understanding global phylogenetic relationships [44,46].

The COX1 phylogenetic tree (in agreement with amino acid identity values) shows strong support for a dinoflagellate and apicomplexan clade (98% bootstrap) but displays poor resolution within the alveolate clade. This low degree of support in part may reflect artifactual grouping of unrelated taxa (e.g. alveolates and kinetoplastids) as a result of long branch attraction [36] due to an accelerated rate of sequence divergence. Agreement of SSU rRNA trees [1–4] and morphological data [4,5] with COX1 trees implies that members of the alveolates share a common ancestry; however their *cox1* genes are obviously evolving at an accelerated and unequal rate. This is particularly evident in ciliate COX1 protein sequences, which are even more derived than COX1 homologs in kinetoplastids [37], a separate and earlier branching eukaryote lineage in nuclear rRNA trees [47]. This extreme divergence complicates phylogenetic reconstruction and assessment of relationships. It also highlights the importance of isolating and characterizing other dinoflagellate mitochondrial genes, which should provide additional information into codon usage and insights about the evolution of dinoflagellates, apicomplexans and ciliates.

Acknowledgements: We thank D.F. Spencer and M.N. Schnare for assistance and helpful comments. This work was supported by an operating grant (MT-4124) from the Medical Research Council of Canada to M.W.G. The authors also gratefully acknowledge salary support in the form of an MRC Studentship to J.E.N. and a fellowship from the Canadian Institute for Advanced Research (Program in Evolutionary Biology) to M.W.G.

References

[1] Gajadhar, A.A., Marquardt, W.C., Hall, R., Gunderson, J., Ariztia-Carmona, E.V. and Sogin, M.L. (1991) Mol. Biochem. Parasitol. 45, 147–154.
[2] Schlegel, M. (1991) Eur. J. Protistol. 27, 207–219.
[3] Cavalier-Smith, T. (1993) Microbiol. Rev. 57, 953–994.
[4] Van de Peer, Y., Van der Auwera, G. and De Wachter, R. (1996) J. Mol. Evol. 42, 201–210.
[5] Siddall, M.E., Stokes, N.A. and Burreson, E.M. (1995) Mol. Biol. Evol. 12, 573–581.
[6] Sadler, L.A., McNally, K.L., Govind, N.S., Brunk, C.F. and Trench, R.K. (1992) Curr. Genet. 21, 409–416.
[7] Gray, M.W. (1989) Annu. Rev. Cell Biol. 5, 25–50.
[8] Gray, M.W. (1992) Int. Rev. Cytol. 141, 233–357.
[9] Feagin, J.E., Werner, E., Gardner, M.J., Williamson, D.H. and Wilson, R.J.M. (1992) Nucleic Acids Res. 20, 879–887.
[10] Vaidya, A.B., Lashgari, M.S., Pologe, L.G. and Morrisey, J. (1993) Mol. Biochem. Parasitol. 58, 33–42.
[11] Vaidya, A.B., Akella, R. and Suplick, K. (1989) Mol. Biochem. Parasitol. 35, 97–108.
[12] Aldritt, S.M., Joseph, J.T. and Wirth, D.F. (1989) Mol. Cell. Biol. 9, 3614–3620.
[13] Kairo, A., Fairlamb, A.H., Gobright, E. and Nene, V. (1994) EMBO J. 13, 898–905.
[14] Cummings, D.J. (1992) Int. Rev. Cytol. 141, 1–64.

[15] Feagin, J.E. (1994) Annu. Rev. Microbiol. 48, 81–104.
[16] Feagin, J.E., Mericle, B.L., Werner, E. and Morris, M. (1997) Nucleic Acids Res. 25, 438–446.
[17] Ziaie, Z. and Suyama, Y. (1987) Curr. Genet. 12, 357–368.
[18] Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R. and Young, I.G. (1981) Nature 290, 457–465.
[19] Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A. (1981) Cell 26, 167–180.
[20] Tuttle, R.C. and Loeblich, A.R. (1975) J. Phycol. 14, 1–8.
[21] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[22] Spencer, D.F., Gray, M.W. and Schnare, M.N. (1992) in: Modern Methods of Plant Analysis (Linskens, H.F. and Jackson, J.F., Eds.) vol. 14, pp. 347–360, Springer-Verlag, New York.
[23] Lang, B., Burger, G., Doxiadis, I., Thomas, D.Y., Bandlow, W. and Kaudewitz, F. (1977) Anal. Biochem. 77, 110–121.
[24] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) Current Protocols In Molecular Biology, vol. 1, John Wiley and Sons, New York.
[25] Reed, K.C. and Mann, D.A. (1985) Nucleic Acids Res. 13, 7207–7221.
[26] Hanahan, D. (1983) J. Mol. Biol. 166, 557–580.
[27] Henikoff, S. (1984) Gene 28, 351–359.
[28] Smith, S.W., Overbeek, R., Woese, C.R., Gilbert, W. and Gillevet, P.M. (1994) Comp. Appl. Biosci. 10, 671–675.
[29] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[30] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[31] Felsenstein, J. (1993) Phylip (Phylogeny Inference Package) Version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle.
[32] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1979) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O., Ed.) vol. 5, suppl. 3, pp. 345–352.
[33] Felsenstein, J. (1985) Evolution 39, 783–791.
[34] Trumpower, B.L. and Gennis, R.B. (1994) Annu. Rev. Biochem. 63, 675–716.
[35] Osawa, S., Jukes, T.H., Watanabe, K. and Muto, A. (1992) Microbiol. Rev. 56, 229–264.
[36] Felsenstein, J. (1978) Syst. Zool. 27, 401–410.
[37] Pritchard, A.E., Seilhamer, J.J. and Cummings, D.J. (1986) Gene 44, 243–253.
[38] Pritchard, A.E., Sable, C.L., Venuti, S.E. and Cummings, D.J. (1990) Nucleic Acids Res. 18, 163–171.
[39] Feagin, J.E. (1992) Mol. Biochem. Parasitol. 52, 145–148.
[40] Suplick, K., Morrisey, J. and Vaidya, A.B. (1990) Mol. Cell. Biol. 10, 6381–6388.
[41] Karlovcky, P. and Fartmann, B. (1992) J. Mol. Evol. 34, 254–258.
[42] Pritchard, A.E., Seilhamer, J.J., Mahalingam, R., Sable, C.L., Venuti, S.E. and Cummings, D.J. (1990) Nucleic Acids Res. 18, 173–180.
[43] Jukes, T.H. and Osawa, S. (1990) Experientia 46, 1117–1126.
[44] Paquin, B. and Lang, B.F. (1996) J. Mol. Biol. 255, 688–701.
[45] Yasuhira, S. and Simpson, L. (1997) J. Mol. Evol. 44, 341–347.
[46] Lang, B.F., Cedergren, R. and Gray, M.W. (1987) Eur. J. Biochem. 169, 527–537.
[47] Sogin, M.L., Gunderson, J.H., Elwood, H.J., Alonso, R.A. and Peattie, D.A. (1989) Science 243, 75–77.