J Mol Evol (2001) 53:351–363
DOI: 10.1007/s002390010225

JOURNAL OF MOLECULAR EVOLUTION

A Complex Organization of the Gene Encoding Cytochrome Oxidase Subunit 1 in the Mitochondrial Genome of the Dinoflagellate, *Crypthecodinium cohnii*: Homologous Recombination Generates Two Different *cox1* Open Reading Frames

John E. Norman, Michael W. Gray

Program in Evolutionary Biology, Canadian Institute for Advanced Research, Department of Biochemistry and Molecular Biology, Dalhousie University, Room 8F-2, Tupper Medical Building, 5859 University Avenue, Halifax, Nova Scotia B3H 4H7, Canada

Received: 21 December 2000 / Accepted: 30 January 2001

Abstract. In the course of investigating mitochondrial genome organization in *Crypthecodinium cohnii*, a non-photosynthetic dinoflagellate, we identified four *Eco*RI fragments that hybridize to a probe specific for *cox1*, the gene that encodes subunit 1 of cytochrome oxidase. Cloning and sequence characterization of the four fragments (5.7, 5.1, 4.1, 3.5 kilobase pairs) revealed that *cox1* exists in four distinct but related contexts in *C. cohnii* mtDNA, with a central repeat unit flanked by one of two possible upstream (flanking domain 1 or 2) and downstream (flanking domain 3 or 4) regions. The majority of the *cox1* gene is located within the central repeat; however, the C-terminal portion of the open reading frame extends into flanking domains 3 and 4, thereby creating two distinct *cox1* coding sequences. The 3′-terminal region of one of the *cox1* reading frames can assume an elaborate secondary structure, which potentially could act to stabilize the mature mRNA against nucleolytic degradation. In addition, a high density of small inverted repeats (15–22 base pairs) has been identified at the 5′-end of *cox1*, further suggesting that hairpin structures could be important for gene regulation. The organization of *cox1* in *C. cohnii* mtDNA appears to reflect homologous recombination events within the central repeat between different *cox1* sequence contexts. Such recombining repeats are a characteristic feature of plant (angiosperm) mtDNA, but they have not previously been described in the mitochondrial genomes of protists.

Key words: Dinoflagellate — *Crypthecodinium cohnii* — Mitochondrial DNA — Recombination — *cox1* — Inverted repeat

Correspondence to: M.W. Gray; *email:* mwgray@is.dal.ca

Introduction

Mitochondrial genomes display a great deal of variability in size, structure, and organization (Gray et al. 1998), with the smallest mitochondrial DNAs (mtDNAs) occurring in members of the Apicomplexa, an exclusively parasitic phylum of unicellular eukaryotes (protists). The apicomplexan mitochondrial genome is only 6 kilobase pairs (kb) in length and contains just three open reading frames (ORFs) (Feagin 1994; Wilson and Williamson 1997) separated by small intergenic spacer regions. At the other extreme, land plants have the largest known mitochondrial genomes, which range in size from 180 to 2500 kb. Within the latter group, complete mtDNA sequences have been determined for the dicotyledonous angiosperms *Arabidopsis thaliana* (small mustard) and *Beta vulgaris* (sugar beet) and the bryophyte *Marchantia polymorpha* (a liverwort). The *M. polymorpha* mitochondrial genome (184 kb) encodes 94 identified genes (Oda et al. 1992) that comprise 33% of the mtDNA; in contrast, the twofold larger *A. thaliana* (367 kb) and *B. vul-*

garis (369 kb) mitochondrial genomes encode, respectively, 57 or 59 recognized genes that account for <15% of the entire genome, the remainder of which consists of introns, intergenic spacers, and repetitive sequence elements (Unseld et al. 1997; Kubo et al. 2000).

Dinoflagellates constitute a large group of diverse protists that usually possess two flagella and are often armored by a hard shell, or test. Phylogenetic reconstructions (Gajadhar et al. 1991; Sadler et al. 1992; Van de Peer et al. 1996) supported by morphological similarities (Siddall et al. 1995; Van de Peer et al. 1996) demonstrate that dinoflagellates share a common ancestry with apicomplexans, which together comprise a sister clade to ciliates. Although the mtDNA sequences of four apicomplexan species (reviewed in Feagin 1994; Wilson and Williamson 1997; see also McIntosh et al. 1998; Sharma et al. 1998) and two ciliates (reviewed in Cummings 1992; Gray et al. 1998) have been determined, surprisingly little is known about the mitochondrial genome of dinoflagellates.

Two recent studies have independently reported a *cox1* gene sequence from the nonphotosynthetic dinoflagellate, *Crypthecodinium cohnii* (Inagaki et al. 1997; Norman and Gray 1997). In both cases, Cox1 amino acid alignments were used as the basis for phylogenetic reconstructions that support a dinoflagellate/apicomplexan clade; however, inclusion of ciliates within this group is less well supported, because of the highly derived nature of ciliate Cox1 sequences.

Here we document our continuing investigation of *cox1* gene organization in the mtDNA of *C. cohnii*. Southern hybridization and sequence analyses reveal that in this organism, the *cox1* gene is found in four major contexts, with a *cox1* core region flanked by one of two different upstream and two different downstream sequences. These data provide the first insights into a novel and highly complex gene organization in a dinoflagellate mtDNA

Materials and Methods

Culturing of C. cohnii and Nucleic Acid Extraction. Procedures for culture of *C. cohnii* and nucleic acid extraction were as previously described (Norman and Gray 1997) with one exception: total and mitochondrial RNA fractions were prepared by incubating nucleic acids with RNase-free DNase I (Pharmacia) at 37°C for 30 min in a reaction mixture containing 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, and DNase I (1 unit per μg DNA). Following this treatment, samples were extracted with phenol according to standard protocols (Sambrook et al. 1989) and the recovered RNA was stored as a 50% ethanol solution at −20°C.

Southern and Northern Hybridization Analysis. A fraction enriched in mtDNA was hydrolyzed with *Eco*RI, *Eco*RV, or *Sst*I (Gibco BRL) following the supplier's protocols, and the products were electrophoresed at 1 V/cm in a 1.0% agarose gel containing 1X TAE (40 mM Tris-acetate, 1 mM Na_2EDTA) for 16 h. Resolved fragments were transferred to a Biotrans™(+) (ICN Biomedicals) nylon membrane using conventional alkaline transfer protocols (Ausubel et al. 1987; Sambrook et al. 1989). Southern hybridization and filter washing conditions were as previously outlined (Norman and Gray 1997).

In total, five different hybridization probes were used. The first comprised a 650-nt *cox1*-specific PCR product (labelled cox240-448 in Fig. 2) described previously (Norman and Gray 1997). Two other hybridization probes were prepared by agarose gel purification of restriction products. Hydrolysis of pCc15 with *Eco*RI and *Eco*RV produced a 997-nt fragment (P1 in Fig. 2) that covers the 5′-end of flanking region 1, whereas hydrolysis of pCc3 with *Eco*RI and *Dra*I gave a 236-nt fragment (P4, Fig. 2) that encompassed the terminal portion of flanking region 4. Two other probes were prepared using gel-purified PCR products as templates in labelling reactions. The first probe (P2, 371 nt in length) was amplified from clone pCc8 (Fig. 2) using the vector-based T7 and internal P37 (5′-ATTAATAGCTGGACATG-TAG-3′) primers. Similarly, a 1029-nt PCR product (P3, Fig. 2) was amplified from clone pCc8 using the two primers T3 and P6 (5′-TGTGGAGCTATAAACCATAAATC-3′). All PCR and restriction-fragment probes were radiolabelled using random hexamer extension (Ausubel et al. 1987) and freed of unincorporated isotope by chromatography on Sephacryl S-200 HR columns (Pharmacia) prior to hybridization.

Northern hybridization filters were generated by first electrophoresing RNA samples for 3 h at 80 V in a 1.2% agarose gel containing 3% formaldehyde and 1X MOPS buffer (200 mM 3-(*N*-morpholino) propanesulfonic acid, 50 mM NaOAc (pH 5.2), 10 mM EDTA). Following electrophoresis, RNA was visualized and photographed under UV light and transferred to a Biotrans™(+) nylon membrane (ICN Biomedicals) by capillary action. Transfers were carried out overnight at room temperature in 20X SSC (3 M NaCl, 0.3 M Na-citrate) followed by baking at 80°C for 2 h. Northern hybridizations were performed using cox240-448, P3, and P4 probes with Southern hybridization protocols. Following autoradiography, sizes of RNA transcripts were estimated by comparison with RNA size standards (Gibco BRL).

Cloning and Sequence Analysis of mtDNA. Fractions enriched in mtDNA were hydrolyzed with either *Eco*RI or *Eco*RV and the products were ligated into pBluescript KS+ (Stratagene) using T4 DNA ligase, following which the constructs were transformed into competent *E. coli* strain DH5α cells (Hanahan 1983). Positive clones were identified by hybridization to colony lifts (Ausubel et al. 1987) using the cox240–448 probe. Four positive clones (pCc2.3, pCc15, pCc8, and pCc3) had inserts identical in size to bands visualized in the *Eco*RI lane by Southern analysis using the cox240–448 hybridization probe (Fig. 1). The sequences of the inserts in these clones were completely determined on both strands using a combination of automated and manual sequencing protocols. Sequence data have been submitted to GenBank under the following accession numbers: Cc2.3, AF182641; Cc15, AF012554; Cc8, AF182642; and Cc3, AF182643. Additional constructs were characterized and compared to the above four clones by *Eco*RI, *Eco*RV, *Sst*I, or *Xba*I mapping and partial sequence analysis.

Sequence data were assembled and scanned for ORFs using the Sequencher ver. 3.1 (Gene Codes) program. Additional protein-coding, rRNA, and tRNA gene searches were carried out by surveying NCBI databases using BLAST (Altschul et al. 1997). Analysis of repetitive sequences was performed using the MicroGenie (Queen and Korn, 1984) software package (SciSoft).

Results

Characteristics of Mitochondrial DNA Isolated from Whole Cells or Subcellular Fractions

A distinct banding pattern was not seen when DNA isolated from a crude mitochondrial fraction was hydro-

lyzed with various restriction endonucleases. However, specific and reproducible bands were revealed in different restriction digests by hybridization with a mitochondrial gene-specific probe (a *cox1*-specific PCR product) (Fig. 1); in particular, four bands appeared in the *Eco*RI digest, as previously noted (Norman and Gray 1997). As expected, the *cox1* probe produced a much weaker Southern hybridization signal with total DNA than with DNA isolated from a crude mitochondrial fraction (data not shown). A stronger hybridization signal was obtained with total DNA prepared by guanidine/proteinase K extraction than with DNA extracted with detergent/phenol-cresol, presumably because the former method allows a more efficient removal of DNA-bound proteins from mtDNA.

When total cellular DNA from *C. cohnii* was centrifuged in a CsCl density gradient, a faint diffuse zone was detected immediately above the main band. When individual gradient fractions were subjected to Southern hybridization, *cox1*-hybridizing sequences were distributed over a broad region of the gradient, extending from fractions 20–30; data not shown). The majority of the *cox1*-containing DNA was located just above the main DNA band (which comprised fractions 16–23), at the location of the diffuse satellite band (fractions 25–28) (data not shown).

Because we were not successful in isolating a *C. cohnii* mtDNA fraction that yielded a relatively simple, defined restriction pattern, our subsequent investigations have focused on cloned restriction fragments carrying typical mitochondrial genes.

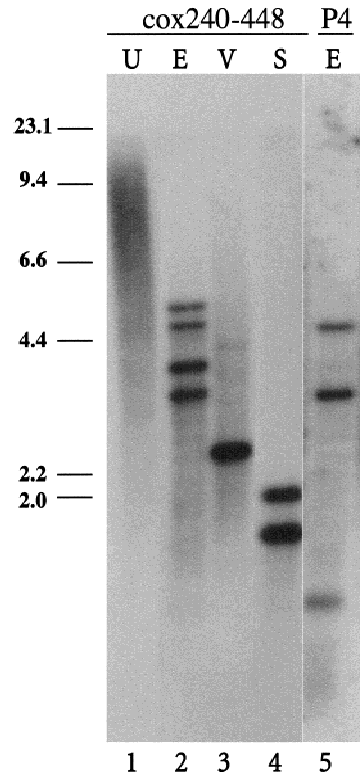

Fig. 1. Autoradiogram of Southern blot of *C. cohnii* mtDNA hydrolyzed with E, *Eco*RI (lanes 2 and 5); V, *Eco*RV (lane 3); or S, *Sst*I (lane 4). Untreated control lane is labelled U (lane 1). Blots were hybridized with radiolabelled cox240-448 PCR product (lanes 1–4) or a P4 restriction fragment (lane 5; see Fig. 2). The positions of size markers (kb; λ DNA hydrolyzed with *Hind*III) are indicated at the left of the figure.

Sequence Characterization of Four cox1-*Containing* Eco*RI Elements in* C. cohnii *mtDNA*

In Southern hybridizations, the cox240-448 probe hybridized to four *Eco*RI mtDNA fragments ranging in approximate size from 3.5 to 5.7 kb, superimposed on a trailing, low-level background smear beginning at about 5.7 kb. As shown in Fig. 1 (lane 2), the stoichiometries of the four bands are unequal, with each of the two smaller fragments exhibiting roughly twice the hybridization intensity of the two larger ones (Norman and Gray 1997). With the same probe, *Eco*RV produced a single 2.6-kb band overlaid on a faint trailing smear starting at approximately 10 kb and continuing to <2 kb in size (Fig. 1, lane 3). *Sst*I generated two hybridizing fragments, 1.5 and 2.0 kb in size (Fig. 1, lane 4), with the smaller band displaying twice the hybridization intensity of the larger band. Combined *cox1* DNA sequence and restriction data show that the region to which cox240-448 hybridizes does not contain *Eco*RI or *Sst*I sites; therefore, the multiple banding pattern seen in these lanes is not due to probe hybridization across internal restriction sites. Rather, the presence of multiple *cox1*-hybridizing *Eco*RI and *Sst*I bands was taken to indicate that the *cox1* gene in *C. cohnii* is present in different contexts that are stoichiometrically unequal.

Sequence analysis of the four cloned, *cox1*-hybridizing *Eco*RI fragments, whose organization is depicted in Fig. 2, revealed no genes (protein-coding, rRNA, or tRNA) other than *cox1*. Scattered throughout the regions flanking the *cox1* ORF are long stretches of noncoding DNA that contain numerous short (<27-nt) inverted and direct repetitive sequence elements. These *cox1* clones all contain a 2814-nucleotide (nt) central repeat (CR) unit that houses most of the *cox1* gene (Fig. 2) and is flanked by one of two different upstream and two different downstream domains. These four flanking domains are arrayed such that each of the two upstream regions, flank-1 (F1, 1882 nt) and flank-2 (F2, 302 nt), occurs in pairwise combination with either of the two downstream regions, flank-3 (F3, 963 nt) and flank-4 (F4, 382 nt). This allows a total of four possible combinations, all containing the *cox1* repeat, i.e., F1-CR-F3, F1-CR-F4, F2-CR-F3, F2-CR-F4. As shown in Fig. 2, the two largest fragments (5.7 and 5.1 kb) contain the same 5′ flanking sequence as do the two smallest fragments (4.1 and 3.5 kb); conversely, the 5.7 and 4.1 kb fragments have the same 3′ ends, as do the 5.1 and 3.5 kb fragments. To confirm that the four isolated *Eco*RI

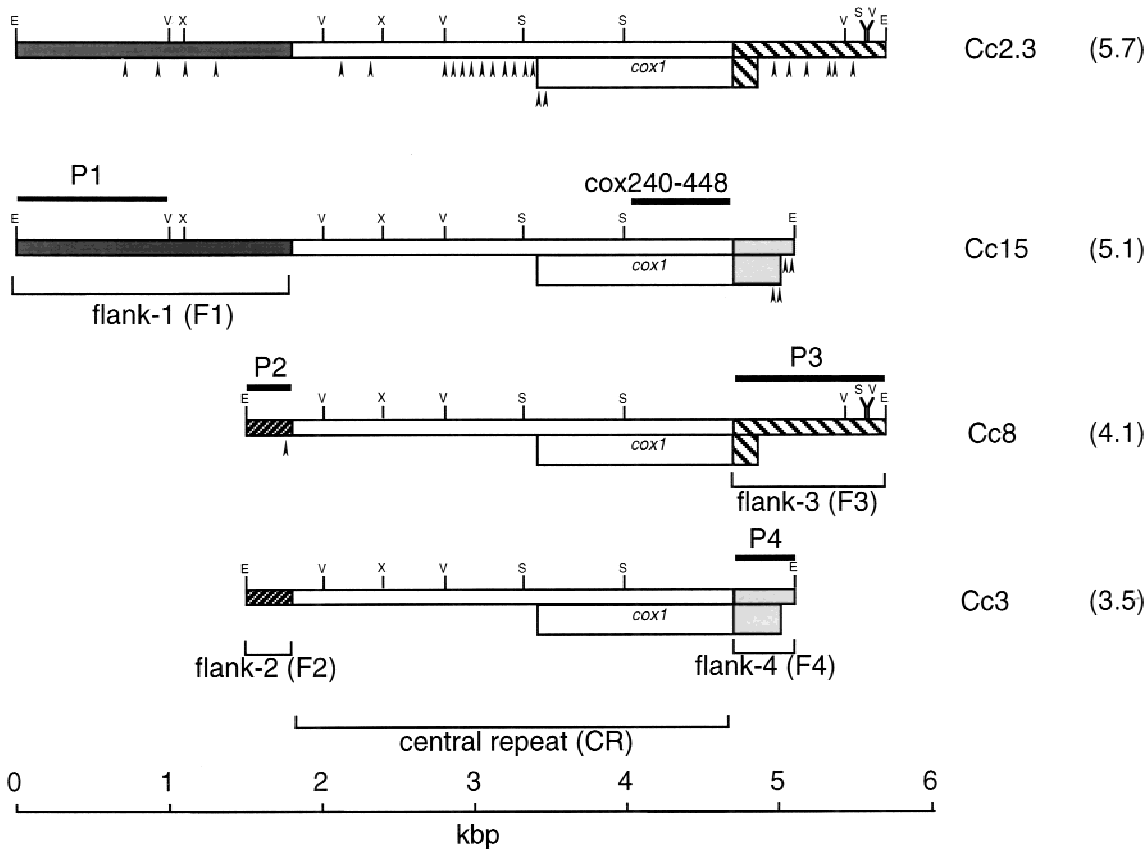

Fig. 2. Physical maps of the four major *cox1*-containing *Eco*RI fragments cloned from *C. cohnii* mtDNA. Names of inserts and their sizes (in parentheses; kb) are listed at the right of the figure. Open rectangles indicate identical regions in all four clones (central repeat, CR). Shaded rectangles denote flanking regions conserved in two of the four constructs (flank-1 to flank-4; F1 to F4). The open rectangles labelled '*cox1*' and adjoining shaded rectangles below each map delineate the length and position of the *cox1* ORF. Arrowheads denote approximate locations of IRs. Restriction sites: E, *Eco*RI; V, *Eco*RV; S, *Sst*I; X, *Xba*I. Solid black lines above each map indicate the location of five probes (cox240-448, P1, P2, P3, P4) used in Southern hybridizations. A size scale bar is located at the bottom of the figure.

clones are indeed representative of the four major *Eco*RI bands visualized by Southern analysis, hybridization was repeated using four flanking-region-specific probes (P1, P2, P3, and P4; see Fig. 2). Each probe hybridized to two of the four major *Eco*RI fragments, corresponding in size to that predicted on the basis of shared flanking sequence. For example, P1 detected two bands, 5.7 and 5.1 kb in size; P2 revealed bands of 4.1 and 3.5 kb; P3 hybridized to bands 5.7 and 4.1 kb (data not shown); and P4 identified two major bands of 5.1 and 3.5 kb, plus an additional minor band approximately 0.6 kb in size (Fig. 1, lane 5). These findings confirm that the four isolated *Eco*RI fragments are representative of the four major *Eco*RI bands visualized by Southern analysis.

Sequence alignments of the four *Eco*RI fragments revealed that distinct divergence points separate the CR from both the 5′- and 3′-flanking domains. As illustrated in Fig. 3A, the 5′-ends of the CR region align perfectly until the divergence point (labelled I) is reached, upstream of which flanking sequences 1 and 2 abruptly diverge. A similar pattern is evident at the 3′-end of the CR (II, Fig. 3B); however, in this case, the split into flanking domains 3 and 4 generates two *cox1* reading frames that differ within the C-terminal region of the encoded amino acid sequence. This feature, coupled with the presence of numerous small, directly oriented, repetitive sequence elements located upstream and downstream of the two divergence points (data not shown), suggests that the complex *cox1* organization observed in *C. cohnii* is likely a consequence of repeat-mediated recombination events. We surmise that illegitimate recombination within small repeats could initially have promoted large-scale genomic rearrangements, leading to the creation of duplicate *cox1*-containing elements (a two-membered repeat as described in Coulthart et al. 1990). Once generated, these *cox1* repeats would be capable of undergoing further rearrangement by recombination within the CR, generating the four different but related *cox1*-containing elements described here.

Analysis of Inverted Repeats

Analysis of *C. cohnii* mtDNA showed that the four *cox1* *Eco*RI clones contain 29 IR elements, ranging from 15 to 22 nt in length (Fig. 4). These repeats can be folded into

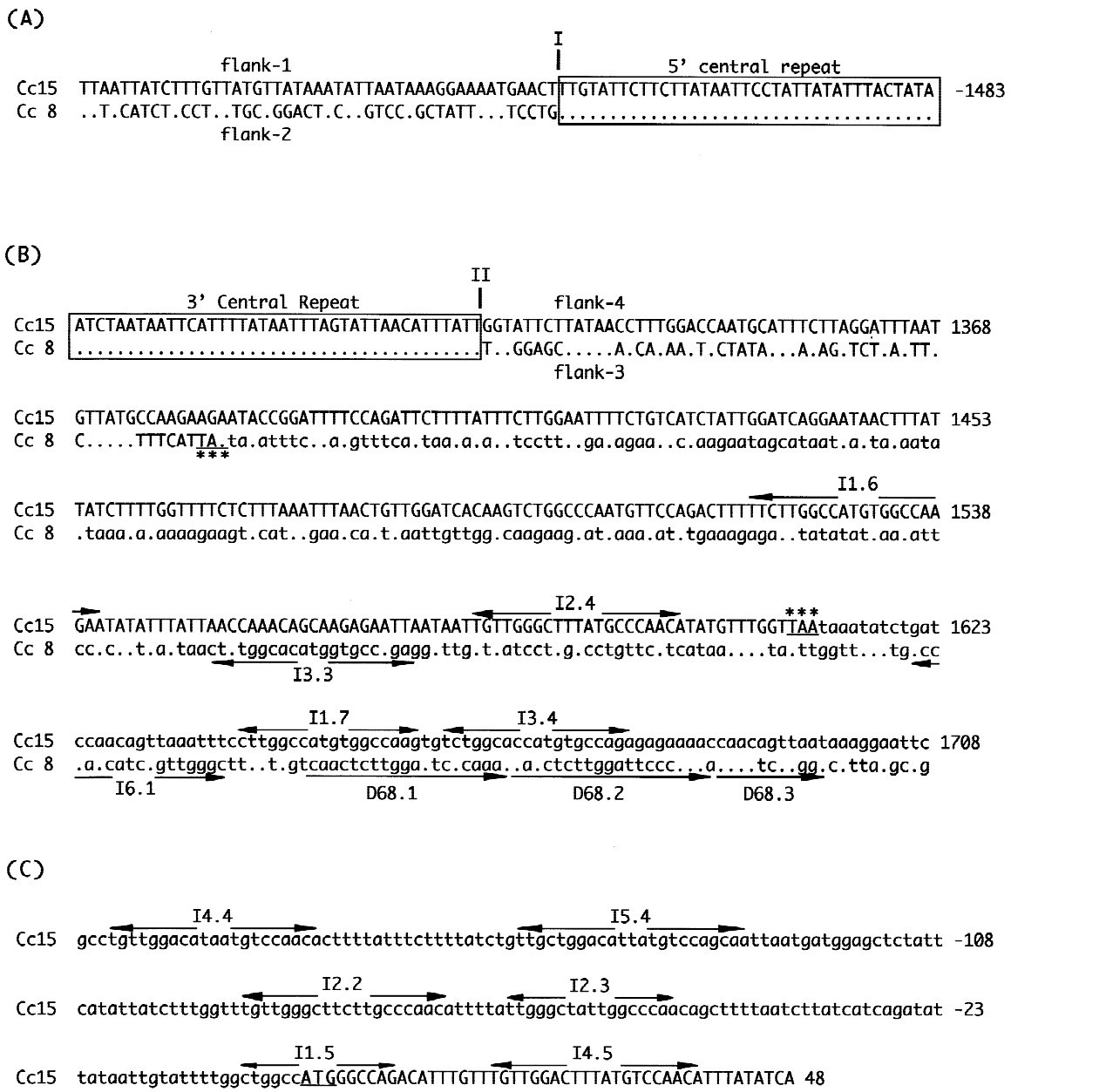

Fig. 3. Nucleotide sequence showing the location of divergence points and IRs within *cox1*-containing *Eco*RI fragments cloned from *C. cohnii* mtDNA. Nomenclature of flanking regions and insert designation are as described in Fig. 2. Sequence coordinates are relative to the inferred *cox1* initiation codon. (A) Sequence alignment showing the divergence point (I) between the 3′ ends of flank-1 and flank-2 leading into the 5′ end of the CR (enclosed by the solid line). Dots indicate positions of nucleotide identity in the bottom sequence relative to the top one. (B) Alignment showing the divergence point (II) between the 3′ end of the CR (enclosed by the solid line) leading into the 5′ ends of flank-3 and flank-4. Stop codons are underlined and highlighted by three asterisks (***). (C) Nucleotide sequence showing the location of IRs in the CR region close to the *cox1* initiation codon (which is shown in bold and underlined). In (B) and (C), uppercase letters denote sequence contained within the *cox1* ORF. Closely spaced horizontal arrows pointing in opposite directions located above (Cc15) and below (Cc8) the sequence delineate IRs; in (B), these repeats can potentially base pair to form the stem region of the repeat secondary structure depicted in Fig. 4. Names assigned to each IR follow the nomenclature of Fig. 4. Arrows labelled D.68.1, D68.2, and D68.3 in (B) denote tandem direct repeats. Note that a revision in the DNA sequence of insert Cc15 places the termination codon 39 nt further downstream than previously reported (Norman and Gray 1997).

stable, perfectly paired hairpin structures consisting of a 5- to 10-nt paired stem region closed by a 3- to 5-nt loop (Fig. 5). Paired regions tend to be GC-rich, whereas loops have a high AT content, as does *C. cohnii* mtDNA as a whole. As illustrated in Fig. 4, these repeats can be arrayed into one of six different groups based on stem-sequence similarity and loop length. Within each group, paired regions are highly conserved but loop sequences are more variable. Notably, base changes in the 5′-halves of palindromes are matched by compensating changes in the 3′-halves, thereby maintaining base pairing. For example, the group 4 consensus sequence is T*GT*TG-

IR#	Coordinates	5'-Stem	Loop	3'-Stem	Location
	Group 1	**--CTTGGCC**	**ATGT**	**GGCCAAG--**	
I1.1	1118-1135	--.....	....	.....--	F1
I1.2	1293-1310	--.....	.ATA	.....--	F1
I1.3	2299-2316	--.....	TAT.	.....--	CR
I1.4	2978-2995	--.....	TAT.	.....--	CR
I1.5	3397-3411	---C....	...-	.....G---	CR
I1.6	4922-4943	TT.....	....	.....AA	F4
I1.7	5042-5059	--.....	....	.....--	F4
	Group 2	**TGTTGGGC**	**TTTAT**	**GCCCAACA**	
I2.1	2898-2918	.....	.....	.....	CR
I2.2	3310-3332	.....	..CT.	.....	CR
I2.3	3338-3354	--.....	.A.TG	.....--	CR
I2.4	4980-5000	.....	.....	.....	F4
	Group 3	**-TCTGGCAC**	**CAT**	**GTGCCAGA-**	
I3.1	664-682	-.....	...	.....-	F1
I3.2	3007-3021	---....T.	TGA	.A...---	CR
I3.3	4954-4974	C.....	ATG	.....G	F3
I3.4	5062-5080	-.....	...	.....-	F4
	Group 4	**TGTTGGACC**	**-ACA-**	**GGTCCAACA**	
I4.1	2108-2128	.TG.....	-...-	.....CA.	CR
I4.2	2879-2895	--C.....	-...-	.....G--	CR
I4.3	3138-3158	.....-	TTT.T	-.....	CR
I4.4	3214-3234	.....A	-TA.-	T.....	CR
I4.5	3421-3441	.....-	TTT.T	-.....	CR
I4.6	5264-5284	.....-	C.A.A	-.....	F3
I4.7	5446-5464	-TC.....	-...-	.....GA-	F3
	Group 5	**TTGCTGGACA**	**-TTA-**	**TGTCCAGCAA**	
I5.1	272-290	--.....-	TAC.T	-.....--	F2
I5.2	893-913	-.....-...	-	.....-	F1
I5.3	2951-2973	.....-	A.G.A	-.....CR	
I5.4	3254-3277	.....-...	-	.....CR	
	Group 6	**GCCCAAC-**	**ATC-**	**-GTTGGGC**	
I6.1	5023-5040	.....-	...T	-.....	F3
I6.2	5119-5135	--...G.C	.CAA	G.C...--	F3
I6.3	5277-5295	.T.....A	.A.-	T.....A.	F3

Fig. 4. Nucleotide sequence comparison of IRs identified in the four major *cox1*-containing *Eco*RI fragments. Repeats are divided into three sections (5′-stem, loop, 3′-stem) based on the inferred ability of the 5′-stem to base pair with the 3′-stem. Repeats are classified into one of six groups on the basis of nucleotide sequence similarity. For each group a consensus sequence is shown in bold at the top. Exact nucleotide matches between individual repeats and the consensus sequences are shown as dots. Dashes denote alignment gaps. Names assigned to individual repeats are indicated at the left of the figure. The coordinates and locations of IRs are based on the nomenclature of Fig. 2. Underlined consensus nucleotides denote exact sequence matches among groups 1–5.

GACC//ACA//GGTCCA*CA*A and is identical to repeat I4.1 with the exception of two sets of compensating nucleotide substitutions (underlined) that maintain pairing within the base of the stem (i.e., T*TG*TGGACC//ACA//GGTCCA*AC*A; Figs. 4, 5). This conservation of secondary structure suggests that these hairpins have an as-yet-undefined functional role, thereby constraining their sequence divergence over time. Other intragroup differences include the number of nucleotides making up stem and loop regions as well as loop substitutions (Figs. 4, 5).

As shown in Fig. 4, several IRs have identical sequences but are found in different parts of the Cc15 and Cc3 *cox1* elements (i.e., I2.1 in CR and I2.4 in F4). These elements represent directly repeated sequences that have the potential to act as foci for recombination events. Intergroup sequence comparisons also indicate that groups 1–5 possess a common stem motif (-TGG-//-CCA-, underlined in Fig. 4) that is not shared by group 6. This shared similarity suggests that these elements may have originated from a common ancestral element.

Fig. 2 shows that most IRs occur within noncoding regions, clustered around the 5′- and 3′-ends of the *cox1* ORF, although two were identified close to the ends of the *cox1* ORF itself. In total, 14 IRs were identified within the central repeat region (CR), with the remainder occurring in the four flanking regions (Figs. 2, 4). This high density of repeats surrounding *cox1* suggests that these hairpins are functional, as does the presence of compensating substitutions in several base-paired palindromes.

Another noteworthy feature is that the atypical group 6 elements only occur within flank-3 (F3). This implies that flank-3 has a repeat composition different than that of the other flanking regions, information that may be useful in inferring the nature of the ancestral *cox1*-flanking sequences and establishing the origin of the more recently introduced regions.

Northern Analysis of Transcripts Expressed from the Two Different Versions of cox1 *in* C. cohnii *mtDNA*

As illustrated in Figs. 2 and 3B, *C. cohnii* mtDNA contains two variant *cox1* ORFs comprising an identical 1323-nt N-terminal portion followed by one of two different in-frame C-terminal regions. The longer ORF (536 amino acid residues) is found in Cc15 and Cc3, where

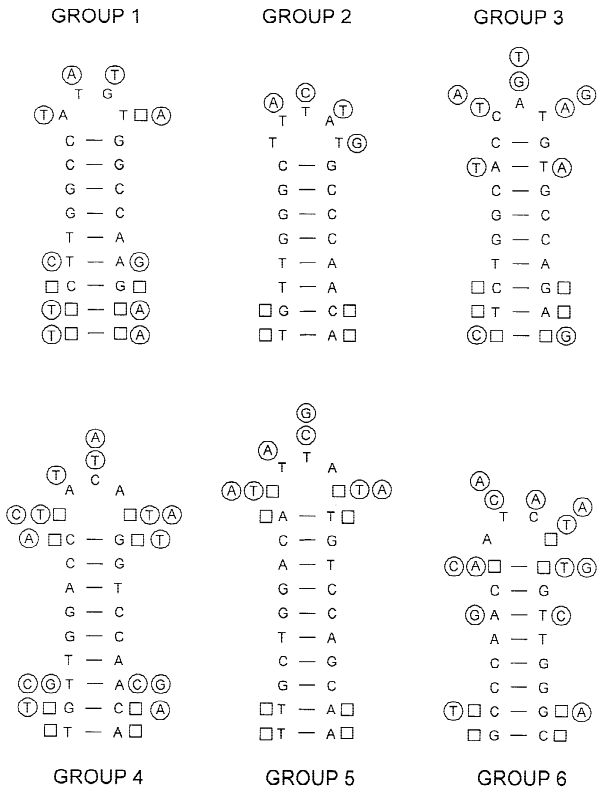

Fig. 5. Diagram depicting secondary structure models for IRs shown in Fig. 4. Each structure is based on the group consensus sequence (see Fig. 4). Differences from the consensus are shown circled around the outside of the consensus sequence. Squares indicate secondary structure positions for which nucleotides are missing in either the consensus or individual structures. RNA secondary structures were generated using the XRNA program (B. Weiser and H. Noller, University of California, Santa Cruz).

285 nt of the 1608-nt *cox1* reading frame extend into flank-4 (Fig. 3B). In contrast, the shorter *cox1* ORF, which occurs in Cc2.3 and Cc8, extends only 57 nt into flank-3, resulting in a reading frame of length 1380 nt (460 amino acids; Fig. 3B). Unlike the C-terminal 95 amino acids encoded by the longer *cox1* gene, the 19 amino acids of the shorter version align poorly with the C-terminal region of other Cox1 proteins; moreover, the shorter Cox1 sequence lacks the last two membrane-spanning domains that are present in all other characterized Cox1 sequences (Trumpower and Gennis 1994), including the longer *C. cohnii* Cox1.

Upstream of divergence point II, Cc2.3 and Cc8 are identical to inserts that possess the longer version of *cox1*. In addition, in Southern hybridization experiments, band intensity differences suggest a higher number of truncated than complete *cox1* genes in *C. cohnii* mtDNA. For example, the 1.5-kb *Sst*I band (Fig. 1, lane 4) is more intense than the larger version that encodes what appears to be the authentic *cox1*. Similarly, the 4.1-kb *Eco*RI fragment, which contains the truncated *cox1* version, hybridizes more intensely than the other smaller *Eco*RI band (3.5 kb).

To determine if the truncated *cox1* gene is functional

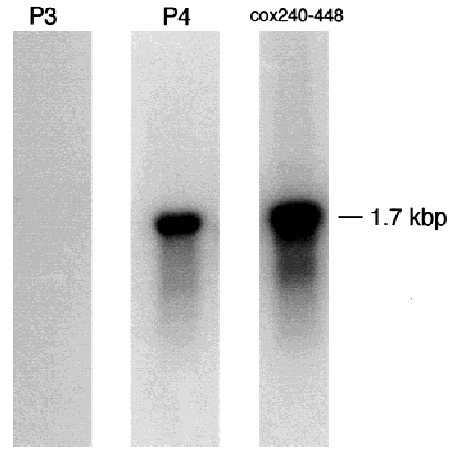

Fig. 6. Autoradiograms showing the results of northern hybridization analysis of RNA isolated from a *C. cohnii* mitochondrial fraction. Blots were hybridized with radiolabelled cox240-448 PCR product, or with P3 or P4 restriction fragments (Fig. 2), as indicated.

in *C. cohnii*, we searched for corresponding stable transcripts in RNA isolated from a subcellular fraction of mitochondria. A probe specific for the truncated version of *cox1* (flank-3) did not hybridize to northern blots (Fig. 6, lane P3), despite the fact that it produced a strong hybridization signal in Southern analysis performed at the same time (data not shown). This suggests that stable transcripts are not produced from this ORF. In contrast, reprobing of northern filters using P4, which is specific for the C-terminal portion of the longer *cox1* gene, revealed a 1.7-kb RNA (Fig. 6, lane P4). This transcript size is consistent with that predicted for the longer *cox1* ORF, and is the same size as a band detected using an internal *cox1* probe (Fig. 6, cox240–448). No detectable signal was obtained in northern experiments using *cox1* probes with total cellular or polyA$^+$ RNA fractions.

A further distinction between flank-3 and flank-4 pertains to the density of IRs located within the two regions (Fig. 3B). Approximately 118 nt downstream of divergence point II and within *cox1* of flank-4 are several stretches of sequence that have the potential to fold into an elaborate and highly stable stem-loop structure, with the UAA termination codon located within a helical region consisting of 18 consecutive base pairs (Fig. 7). In contrast, only a few IRs are located within the 5′ portion of flank-3 (Fig. 3B), and these occur well downstream (178 nt) of the termination codon. From these results we infer that flank-4 sequence, and potentially its folded RNA structure, may act to stabilize *cox1* transcripts.

The only region that exhibits a higher coverage of IRs is the segment of the CR that encompasses the 5′-end of the *cox1* gene (Fig. 2). As shown in Fig. 3C, two repeats occur within the ORF itself, and four others are located within the first 200 nt upstream of *cox1*. In fact, 41% (12) of identified IRs occur within the first 550 nt upstream of *cox1*, a region that constitutes only 8% of the total sequence within the *cox1*-containing elements. The placement of these IRs is intriguing and suggests that they

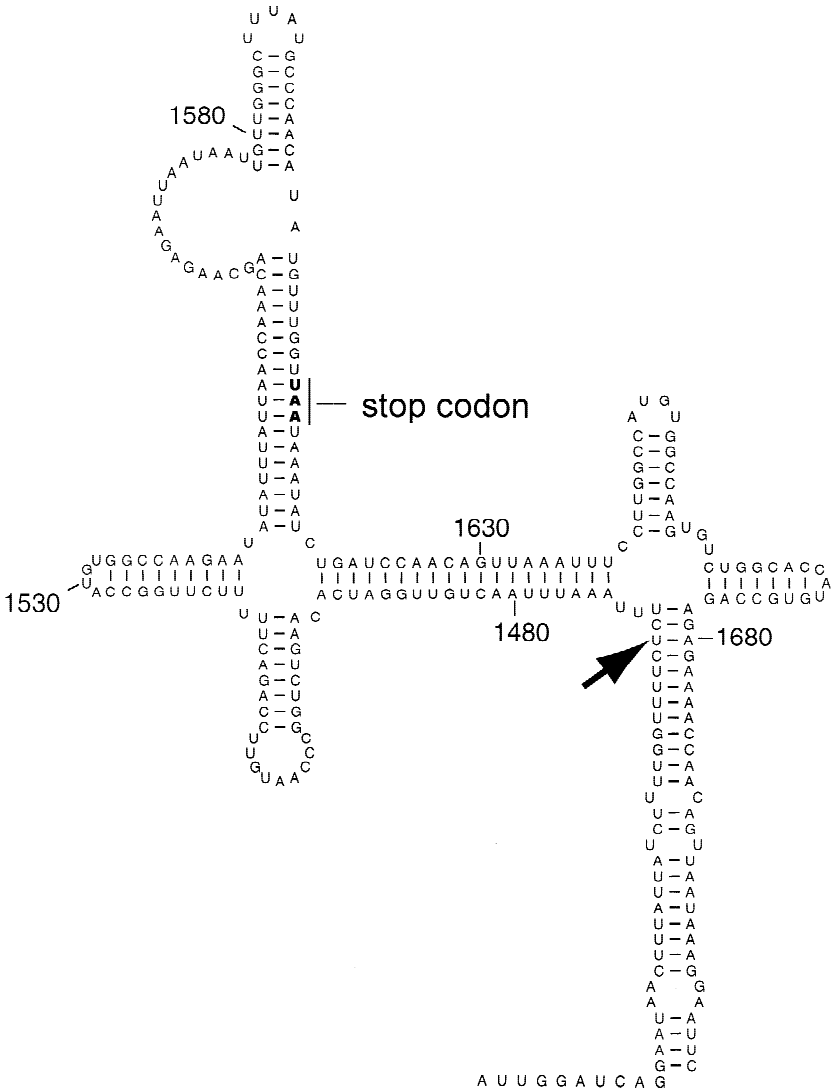

Fig. 7. Predicted secondary structure of the 3′ terminal portion and downstream flanking sequence of the RNA transcribed from the longer (presumably authentic) *cox1* gene in *C. cohnii* mitochondria. Sequence numbering is relative to the inferred *cox1* initiation codon. The UAA termination codon is indicated (positions 1609–1611). Arrow denotes predicted 3′ end of the *cox1* mRNA based on 3′-RACE analyses. The RNA secondary structure was generated using the XRNA program.

may be important in regulating gene expression and/or transcript stability near the 5′-end of the *cox1* coding region.

The existence of the truncated *cox1* gene raises questions about its origin. One possibility is that it arose when flank-3, present elsewhere in the genome, replaced flank-4 through a spurious recombination event (e.g., between short repeated sequences). In such a situation, flank-4 could have been deleted or moved to another position in the mtDNA. In the latter situation, it should be possible to locate the 'missing' C-terminal portion of *cox1*. As seen in Fig. 1 (lane 5), probe P4, which targets only the C-terminal region of the longer *cox1* gene, hybridized to three bands, the two largest of which (5.1 and 3.5 kb) correspond to the *Eco*RI inserts contained in plasmid clones pCc15 and pCc3; however, a third (0.8 kb) band is also visible. This 0.8-kb fragment may represent the portion of *cox1* that is missing from the truncated version of the gene. Though this observation is provocative, numerous attempts to clone this particular *Eco*RI fragment have so far been unsuccessful.

Minor cox1 *Arrangements in* C. cohnii *mtDNA*

In addition to the four major *Eco*RI fragments, a *cox1* probe hybridized to other *Eco*RI products comprising the faint background smear seen in Fig. 1 (lane 2). To investigate the nature of these low-abundance fragments, 17 additional *Eco*RI clones were isolated by library screening and characterized. Restriction mapping and partial sequence analysis revealed that 12 of these 17 inserts corresponded to one of those in pCc15, pCc8, or pCc3 (Fig. 8); however, a small number contained unique inserts that appeared to be variants of the four major types. Two of the variants (Cc2.2 and Cc7) seemed to be terminally truncated versions of the 3.5-/5.1- and 5.7-kb fragments, respectively, whereas two other inserts (Cc11 and Cc26.1) possessed internally deleted regions (Fig. 8). Notably, the region that is missing in Cc11 is 20 nt in length and corresponds to the second (D68.2) of three tandemly duplicated direct repeats (Fig. 3B). Finally, two constructs (pCc32 and pCc11) contained unique 5′ sequence flanking the *cox1* gene. Com-

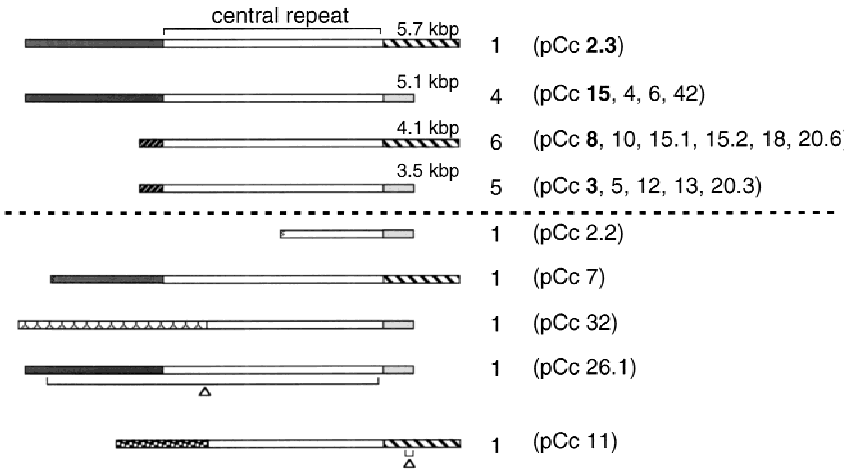

Fig. 8. Schematic representation of all *cox1 Eco*RI fragments that were cloned and characterized from *C. cohnii* mtDNA in this study. To the right of each map are the number and (in parentheses) the names of individual clones that were found to contain the particular insert depicted. Rectangles that have the same shading denote flanking regions that are shared by different constructs. The CR is delineated by open rectangles. Internally deleted regions are indicated by triangles. The dotted line separates the four major *Eco*RI fragments characterized in detail here (above) from the minor variants identified by cloning (below).

parisons of the points of divergence in the latter constructs and in flank-1 and flank-2 revealed no identity among them. In fact, the start of the unique region occurred within the CR (Fig. 8); also, these unique *cox1*-flanking regions contained repetitive sequence elements similar to those seen in all other *cox1*-flanking regions.

Like the major *cox1*-containing elements, these minor elements appear to have arisen by recombination events; however, their low copy number indicates that, individually, they comprise a small proportion of *C. cohnii* mtDNA. These results suggest that whereas there are four major *cox1*-containing elements in *C. cohnii*, the mitochondrial genome also contains a sub-population of less-abundant elements that appear to be derived from the four major *Eco*RI versions. In this regard we note that one of the previously published versions of *C. cohnii cox1* (Inagaki et al. 1997) does not resemble any of our constructs beyond conserved protein-coding domains, implying that this copy of *cox1,* which was cloned using a polymerase chain reaction approach, may represent a minor component of *C. cohnii* mtDNA.

Investigation of cox1 *Organization Downstream of 3′* Eco*RI Termini*

In *T. parva* (Kairo et al. 1994), *cox1* is localized in a terminal portion of the mitochondrial genome. To further investigate the genomic context of the *cox1* gene in *C. cohnii* mtDNA we isolated, mapped, and partially sequenced two *Eco*RV *cox1* clones (pCcV14 and pCcV22) that extended beyond the 3′ *Eco*RI site present in pCc15 and pCc3. Sequence analysis of CcV14 and CcV22 revealed that both were devoid of recognizable genes, contained numerous small repeats, and had the same sequence to a position approximately 1.5 kb downstream of the flank-4 *Eco*RI site. Beyond this point, the two sequences displayed very little similarity (data not shown). These results indicate that there are no identifiable mitochondrial genes immediately downstream of *cox1* in *C. cohnii* mtDNA and that the C-terminal flanking sequences appear to be heterogeneous in nature.

Discussion

Subcellular Localization of cox1 *Coding Sequences*

In this study, we found that the *cox1*-containing *Eco*RI fragments were enriched in a mitochondrial subcellular fraction and concentrated in a minor component of whole cell DNA located just above the main band DNA in a CsCl isopycnic gradient. Corresponding transcripts were readily detected in RNA isolated from a subcellular fraction of mitochondria, but not in total cellular RNA or in polyA$^+$ RNA. These observations argue that the *cox1* sequences described here are contained in the mitochondrial genome of *C. cohnii,* and do not represent transferred mitochondrial sequences now resident in (and perhaps expressed from) nuclear DNA.

A further arguments in this regard is that base composition and codon usage are very different in the *C. cohnii cox1* gene compared with recently published nuclear gene sequences from this organism. For example, the *cox1* coding region has a G+C content of only 27%, compared with 58%–66% in the nuclear *dip1* (AF255444), *dip5* (AF255661), and *dapC* (AF255446) genes of *C. cohnii.* In the *cox1* gene, codon usage is heavily biased toward codons ending in T or A, whereas this bias is not evident in available nuclear protein-coding sequences from *C. cohnii.* Thus, in *cox1,* codons ending in T and A account for >95% of codons used in the TCN (Ser), CCN (Pro), and ACN (Thr) families; by contrast, in the three nuclear genes cited above, codons ending in T and A represent, on average, only 39% (Ser), 59% (Pro), and 49% (Thr) of the total codons used in these same families. On the strength of these observations and arguments we infer that the *cox1* sequences cloned and characterized here are encoded by the mitochondrial genome.

Mitochondrial Gene and Genome Organization in the Apicomplexa

It is now well established that dinoflagellates and apicomplexans share a common ancestry, and that the dinoflagellate-apicomplexan clade is specifically allied with the ciliates (Gajadhar et al. 1991; Sadler et al. 1992; Van de Peer et al. 1996). Despite this phylogenetic relationship, the results reported here show that *cox1* gene arrangement in *C. cohnii* mtDNA is very different from that found in apicomplexans and ciliates. In *C. cohnii*, *cox1* exists in four major sequence contexts, in all of which the *cox1* ORF is flanked upstream and downstream by long stretches of noncoding DNA. In contrast, in both apicomplexan and ciliate mtDNAs, *cox1* constitutes a unique ORF that is separated from flanking genes by relatively short intergenic spacer regions. In fact, the entire apicomplexan mitochondrial genome is only slightly larger (6–7 kb) than the two largest *cox1* elements described here in *C. cohnii.* This pronounced variation in the length of noncoding sequences implies that constraints on mtDNA size are quite different in the two organismal groups.

Despite differences in *cox1* organization, the mitochondrial genomes of *C. cohnii* and the apicomplexans share a number of similarities centered around repetitive sequence elements. An example of this is seen in *T. parva* mtDNA (Kairo et al. 1994), where *cox1* is located immediately downstream of one of two large terminal IR regions. Although the mtDNA repeat arrangements in *C. cohnii* and *P. falciparum* differ in detail, *cox1* hybridization did reveal a number of similarities of a more general nature. For example, uncut mtDNA from the two organisms migrates in agarose gels as a smear ranging in apparent size from <23 kb to >6.0 kb (see Fig. 1, lane 1). Endonucleases having a single restriction site in the *Plasmodium* 6.0-kb element produce a band of this size (Preiser et al. 1996), superimposed on a background smear starting at 6 kb. A similar smear is also visible following *Eco*RI hydrolysis of *C. cohnii* mtDNA; however, this treatment produces four bands superimposed on a trailing smear, unlike the single-band pattern seen in *P. falciparum.* In both cases the trailing smear is thought to result largely from terminal heterogeneity; however, in *C. cohnii* some of this heterogeneity also appears to be due to truncations occurring upstream of *cox1*. This is similar to what has been suggested in the case of *P. falciparum,* where a random assortment of mtDNA termini is thought to arise from variously truncated versions of the 6-kb element (Preiser et al. 1996). In *C. cohnii* we infer the presence of mtDNA molecules that have different sequences flanking *cox1* and thus different *Eco*RI sites and fragment sizes. In particular, this appears to be the case for those fragments that map downstream of the 3′-terminal *Eco*RI site in *cox1*-containing clones. The terminal heterogeneity present in *P. falciparum* has been attributed to DNA degradation or nonspecific initiation of mtDNA replication that coincides with homologous recombination. In this case, it is thought that incompletely replicated mtDNA molecules that have overhanging 3′-ends are able to invade homologous regions of other DNA duplexes (Fig. 9A; Y-branched, three-way DNA junction), thus creating a DNA replication fork (Preiser et al. 1996). As yet, we have no evidence for such a mechanism in *C. cohnii*; however, if such a recombination-mediated replication system operates in dinoflagellates, it might explain the apparent heterogeneity of *cox1* element ends and the occurrence of unique *cox1*-containing elements.

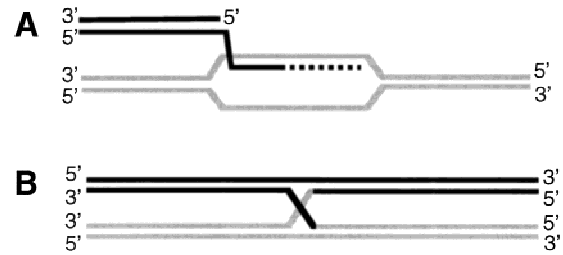

Fig. 9. Models showing two types of recombination-generated DNA junctions. (A) A Y-branched, three-way DNA junction is created when one segment of DNA (black) that has a 3′-overhang invades an homologous region present in a second DNA duplex (grey), thereby creating a replication fork that allows the invading strand to be extended (black dashed line) using the second DNA duplex as a template. (B) Depiction of an X-branched, four-way DNA junction whereby recombination results in strand exchange between two DNA duplexes (black and grey) starting at the crossover point.

The arrangement of *cox1*-containing *Eco*RI fragments as a two-membered family of repeats is consistent with homologous recombination via Holliday junction formation (Fig. 9B; X-branched, four-way DNA junction). We assume that the four *cox1*-hybridizing bands that we observe in *C. cohnii* are derived in evolution from a single *cox1* element. We imagine a process in which duplication of the *cox1* gene and dispersion of the resulting copies to different parts of the genome would initially create two distinct *cox1*-containing repeat elements, each having unique flanking domains. Homologous recombination within the CR of two *cox1* regions would then generate four different contexts, with the CR flanked by one of two upstream and one of two downstream flanking domains. The occurrence of *cox1* in multiple distinct contexts, as documented here, strongly suggests that *C. cohnii* mtDNA does undergo repeat-mediated homologous recombination.

Large and Small Repetitive Elements in Mitochondrial DNA

One of the most surprising findings of this study is that *cox1* gene organization in *C. cohnii* is strikingly similar to gene arrangements seen in angiosperm (flowering plant) mtDNAs. First, the presence of four gene-specific restriction fragments, each the result of pairwise combi-

nations of a CR repeat unit flanked by two sets of upstream and downstream flanking regions, is quite common in plant mtDNAs (Hanson and Folkerts 1992; Wolstenholme and Fauron 1995). Second, comparisons show that *cox1* flanking domains in *C. cohnii* share very little sequence similarity, and that the transition between the CR and the adjacent flanks is abrupt. This has also been documented in plants, where discrete divergence points are the result of large-scale genomic rearrangements that shuffle once-adjacent segments of DNA to other parts of the genome (Bailey-Serres et al. 1986; Coulthart et al. 1990). Third, within *C. cohnii* mtDNA we have detected many small direct and inverted repetitive sequences, <27 nt in length, scattered throughout the CR and flanking regions. Similar small repeats have also been described in the mitochondrial genome of numerous plants (reviewed in André et al. 1992). Finally, in *C. cohnii* we identified a low-abundance subpopulation of *cox1*-containing elements that appear to be recombination-generated versions of the more abundant elements. Similar substoichiometric copies of repeat-containing elements (sublimons), often the product of illegitimate recombination between small repeats, have been described in angiosperm mtDNAs (Small et al. 1987). Because the mitochondrial genomes of plants and dinoflagellates are not specifically related, these similar recombining repeat arrangements must have arisen independently in the dinoflagellate and flowering plant lineages.

Small IRs are found in the mtDNAs of many organisms, including dinoflagellates (this report), green algae (Boer and Gray 1991; Nedelcu and Lee 1998) and land plants (André et al. 1992), apicomplexans (Kairo et al. 1994), fungi (Paquin et al. 1997), animals (Watanabe et al. 1999), and red algae (Ohta et al. 1998). In most cases, however, these repeats are limited to noncoding regions such as intergenic spacers or introns. A notable exception is a red algal species, *Cyanidioschyzon merolae* (Ohta et al. 1998), in which two IRs occur within the C-terminal region of *cox1*, similar to what we describe here.

Repetitive Elements and Mitochondrial Gene Expression

In *C. cohnii*, we detect a single discrete *cox1* transcript, suggesting that *cox1* has its own promoter and transcription initiation site. Any posttranscriptional processing would have to occur rapidly, as larger *cox1* transcripts are not detected in the steady-state RNA population. In contrast, the 6-kb element in *P. falciparum* is polycistronically transcribed, with individual mature RNAs being generated by processing (Ji et al. 1996). *C. cohnii* mtDNA contains two major but slightly different versions of the *cox1* coding element, only the longer of which (present in Cc15 and Cc3) specifies all 12 expected membrane-spanning domains and is stably expressed. The fact that the truncated version of *cox1* (present in Cc2.3 and Cc8) does not support production of a stable *cox1* transcript is somewhat surprising, considering that the same sequence is present for more than 1.5 kb upstream of the translation initiation site in each *cox1* variant and extends more than 1.3 kb into the coding region. Therefore, we would expect that both genes should have the same transcription initiation and promoter sites and would be transcribed. Our results therefore suggest that *cox1* transcript abundance may be regulated by sequences occurring at or downstream of the C-terminus. In *C. cohnii*, IRs that can be folded into an elaborate secondary structure are present within the C-terminal region of the longer *cox1* gene but are absent in the truncated version. Accordingly, we infer that both forms of *cox1* may be expressed but that the truncated *cox1* transcript may be rapidly degraded because it does not contain the correct stabilizing RNA sequence and/or higher order structure. A stability control mechanism involving secondary structure elements has been described for pea (Dombrowski et al. 1997) and *Brassica* (Bellaoui et al. 1997) mitochondrial mRNAs; in these cases, IRs in 3′-untranslated regions fold into stable stem-loop structures, thereby providing signals for mRNA processing factors as well as enhancing the stability of upstream sequences. Observations in *P. falciparum* (Ji et al. 1996) also suggest that differences in mitochondrial transcript abundance may be controlled by RNA stability elements. An alternative possibility is that in *C. cohnii* mitochondria, *cox1* transcription is under the control of a sequence element that is located within the 3′-end of the *cox1* ORF. So far, however, such a mechanism has not been reported in mitochondria.

The intriguing observation that IRs also surround the region encoding the N-terminus of the *C. cohnii cox1* ORF implies that such repeats may be important in regulating gene expression at the 5′- as well as the 3′-end of *cox1*. As in *C. cohnii*, the *P. yoelli cox1* gene also contains an IR within its 3′ end, less than 20 nt upstream of the 5′ end of *cob*; the latter is probably co-transcribed with *cox1* (Suplick et al. 1990), with the two coding regions then being separated by RNA processing. By contrast, *cox1* appears to be independently transcribed in *C. cohnii*, suggesting that if processing of mitochondrial mRNA transcripts occurs in this organism, it plays a different role than it does in members of the Apicomplexa.

The large number and wide distribution of small IRs seen in *C. cohnii* is very similar to that described for several *Chlamydomonas* species (Boer and Gray 1991; Nedelcu and Lee 1998). In both groups, repeats display a common GC-rich palindromic motif (-TGG-//N_n//-CCA- in *C. cohnii* and TRCTCGG//N_n//CCGAGYA in *Chlamydomonas reinhardtii*) with loop sequences being poorly conserved. Another notable similarity is the compensating base changes in paired regions that maintain secondary structure, as also described for several fungi

(Paquin et al. 1997). The potential to maintain pairing strongly suggests that these secondary structures are functionally important, and further supports the hypothesis that turnover of *cox1* mRNA may be controlled by C-terminal IRs in *C. cohnii*. However, we cannot rule out the possibility that these sequence elements are involved in some other process, such as replication or transcription. Another possibility is that they may be mobile genetic elements, capable of spreading throughout the mtDNA of *C. cohnii*, as has been suggested in the case of several fungal species (Butow et al. 1985; Paquin et al. 1997).

Acknowledgments. We thank Drs. M.N. Schnare and D.F. Spencer for assistance, and these individuals as well as Dr. A. Lohan for critical review of the manuscript. 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

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

André C, Levy A, Walbot V (1992) Small repeated sequences and the structure of plant mitochondrial genomes. Trends Genet 8:128–132

Ausubel F, Brent R, Kingston R, Moore D, Seidman J, Smith J, Struhl K (1987) Current protocols in molecular biology, 1st edn. John Wiley and Sons, New York

Bailey-Serres J, Hanson DK, Fox TD, Leaver CJ (1986) Mitochondrial genome rearrangement leads to extension and relocation of the cytochrome *c* oxidase subunit I gene in sorghum. Cell 47:567–576

Bellaoui M, Pelletier G, Budar F (1997) The steady-state level of mRNA from the Ogura cytoplasmic male sterility locus in *Brassica* cybrids is determined post-transcriptionally by its 3′ region. EMBO J 16:5057–5068

Boer PH, Gray MW (1991) Short dispersed repeats localized in spacer regions of *Chlamydomonas reinhardtii* mitochondrial DNA. Curr Genet 19:309–312

Butow RA, Perlman PS, Grossman LI (1985) The unusual *var1* gene of yeast mitochondrial DNA. Science 228:1496–1501

Coulthart MB, Huh GS, Gray MW (1990) Physical organization of the 18S and 5S ribosomal RNA genes in the mitochondrial genome of rye (*Secale cereale* L.) Curr Genet 17:339–346

Cummings DJ (1992) Mitochondrial genomes of the ciliates. Int Rev Cytol 141:1–64

Dombrowski S, Brennicke A, Binder S (1997) 3′-Inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators. EMBO J 16:5069–5076

Feagin JE (1994) The extrachromosomal DNAs of apicomplexan parasites. Annu Rev Microbiol 48:81–104

Gajadhar AA, Marquardt WC, Hall R, Gunderson J, Ariztia-Carmona EV, Sogin ML (1991) Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. Mol Biochem Parasitol 45:147–154

Gray MW, Lang BF, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Brossard N, Delage E, Littlejohn TG, Plante I, Rioux P, Saint-Louis D, Zhu Y, Burger G (1998) Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Res 26:865–878

Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol 166:557–580

Hanson MR, Folkerts O (1992) Structure and function of the higher plant mitochondrial genome. Int Rev Cytol 141:129–172

Inagaki Y, Hayashi-Ishimaru Y, Ehara M, Igarashi I, Ohama T (1997) Algae or protozoa: phylogenetic position of euglenophytes and dinoflagellates as inferred from mitochondrial sequences. J Mol Evol 45:295–300

Ji Y-e, Mericle BL, Rehkopf DH, Anderson JD, Feagin JE (1996) The *Plasmodium falciparum* 6 kb element is polycistronically transcribed. Mol Biochem Parasitol 81:211–223

Kairo A, Fairlamb AH, Gobright E, Nene V (1994) A 7.1 kb linear DNA molecule of *Theileria parva* has scrambled rDNA sequences and open reading frames for mitochondrially encoded proteins. EMBO J 13:898–905

Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for $tRNA^{Cys}$(GCA). Nucleic Acids Res 28:2571–2576

McIntosh MT, Srivastava R, Vaidya AB (1998) Divergent evolutionary constraints on mitochondrial and nuclear genomes of malaria parasites. Mol Biochem Parasitol 95:69–80

Nedelcu AM, Lee RW (1998) Short repetitive sequences in green algal mitochondrial genomes: potential roles in mitochondrial genome evolution. Mol Biol Evol 15:690–701

Norman JE, Gray MW (1997) The cytochrome oxidase subunit 1 gene (*cox1*) from the dinoflagellate, *Crypthecodinium cohnii*. FEBS Lett 413:333–338

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. J Mol Biol 223:1–7

Ohta N, Sato N, Kuroiwa T (1998) Structure and organization of the mitochondrial genome of the unicellular red alga *Cyanidioschyzon merolae* deduced from the complete nucleotide sequence. Nucleic Acids Res 26:5190–5198

Paquin B, Laforest M-J, Forget L, Roewer I, Wang Z, Longcore J, Lang BF (1997) The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. Curr Genet 31:380–395

Preiser PR, Wilson RJM, Moore PW, McCready S, Hajibagheri MAN, Blight KJ, Strath M, Williamson DH (1996) Recombination associated with replication of malarial mitochondrial DNA. EMBO J 15:684–693

Queen C, Korn LJ (1984) A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Res 12:581–599

Sadler LA, McNally KL, Govind NS, Brunk CF, Trench RK (1992) The nucleotide sequence of the small subunit ribosomal RNA gene from *Symbiodinium pilosum*, a symbiotic dinoflagellate. Curr Genet 21:409–416

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning. A laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York

Sharma I, Pasha ST, Sharma YD (1998) Complete nucleotide sequence of the *Plasmodium vivax* 6 kb element. Mol Biochem Parasitol 97:259–263

Siddall ME, Stokes NA, Burreson EM (1995) Molecular phylogenetic evidence that the phylum Haplosporidia has an alveolate ancestry. Mol Biol Evol 12:573–581

Small ID, Isaac PG, Leaver CJ (1987) Stoichiometric differences in DNA molecules containing the *atpA* gene suggest mechanisms for the generation of mitochondrial genome diversity in maize. EMBO J 6:865–869

Suplick K, Morrisey J, Vaidya AB (1990) Complex transcription from

the extrachromosomal DNA encoding mitochondrial functions of *Plasmodium yoelii*. Mol Cell Biol 10:6381–6388

Trumpower BL, Gennis RB (1994) Energy transduction by cytochrome complexes in mitochondrial and bacterial respiration: the enzymology of coupling electron transfer reactions to transmembrane proton translocation. Annu Rev Biochem 63:675–716

Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nature Genet 15:57–61

Van de Peer Y, Van der Auwera G, De Wachter R (1996) The evolution of stramenopiles and alveolates as derived by "substitution rate calibration" of small ribosomal subunit RNA. J Mol Evol 42:201–210

Watanabe KI, Bessho Y, Kawasaki M, Hori H (1999) Mitochondrial genes are found on minicircle DNA molecules in the mesozoan animal *Dicyema*. J Mol Biol 286:645–650

Wilson RJM, Williamson DH (1997) Extrachromosomal DNA in the Apicomplexa. Microbiol Mol Biol Rev 61:1–16

Wolstenholme DR, Fauron CM-R (1995) Mitochondrial genome organization. In: Levings CS III, Vasil IK (eds) The molecular biology of plant mitochondria. Kluwer Academic Publ, Dordrecht, The Netherlands, pp 1–59