INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Mitochondrial genome organization, expression and evolution in the dinoflagellate *Crypthecodinium cohnii*.

by

John E. Norman

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
February, 2001

DALHOUSIE UNIVERSITY

FACULTY OF GRADUATE STUDIES

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "Mitochondrial genome organization, expression and evolution in the dinoflagellate, *Crythecodinium cohnii*"

by John E. Norman

in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated: November 10, 2000

External Examiner

Research Supervisor

Examining Committee

DALHOUSIE UNIVERSITY

DATE: March 8, 20001

AUTHOR: John E. Norman

TITLE: Mitochondrial genome organization, expression and evolution in the dinoflagellate *Crypthecodinium cohnii*

DEPARTMENT(S): Biochemistry and Molecular Biology

DEGREE: Ph.D. CONVOCATION: May YEAR: 2001

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon request of individuals or institutions.

Signature of Author

Table of Contents

List of Figures

List of Tables

Abstract

To better understand how mitochondrial genomes evolve, I investigated the structure, organization, expression and evolutionary affiliations of mitochondrial genes in the non-photosynthetic dinoflagellate *Crypthecodinium cohnii*. In each of the two closest relatives of the dinoflagellates, apicomplexans (a group of intracellular parasites) and ciliates, mitochondrial DNA (mtDNA) has been characterized. Therefore, the specific question that this thesis addresses is: Is the mitochondrial genome of *C. cohnii* more similar to that of the apicomplexans or the ciliates? To answer this question, the DNA sequences of *cox1* and *cob* genes as well as flanking sequences were determined from *C. cohnii* mtDNA. The *cox1* gene 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* coding sequence 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* ORFs. Sequence analysis showed that the C-terminal end of one of the *cox1* ORFs can fold into a highly stable RNA structure, presumably allowing it to form a stable mRNA. A similar pattern is seen with *cob*, with the exception that *cob*-containing elements exhibit five different C-terminal flanking regions, generating three distinct ORFs. The arrangement of mitochondrial genes in *C. cohnii* appears to be the result of homologous recombination within the central repeat between different gene-containing 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.

Using a combination of DNA sequence determination and 3'-RACE analysis, two fragments (LSUG and LSUE) of the large subunit ribosomal RNA gene (*rnl*) were identified in *C. cohnii* mtDNA. Both coding elements exhibit a similar rRNA gene fragmentation pattern as seen in the mtDNA of *Plasmodium* species (apicomplexan). This observation is highly suggestive that the *rnl* gene is discontinuous and rearranged in *C. cohnii* mtDNA. Despite the similarities in rRNA fragmentation, the mtDNA organization of *C. cohnii* is very different from that described in members of the apicomplexans, representing the highest degrees of mtDNA structural divergence described to date between two closely related protist groups.

List of Abbreviations

A, adenosine
Å, angstrom = 10^{-10} metres
aa, amino acid
AMV, avian myeloblastosis virus
ATP, adenosine-5'-triphosphate
bp, base pair
BSA, bovine serum albumin
C, cytidine
°C, degree Celsius
CTP, cytidine-5'-triphosphate
d, day
dATP, 2'-deoxyadenosine-5'-triphosphate
dCTP, 2'-deoxycytidine-5'-triphosphate
dGTP, 2'-deoxyguanosine-5'-triphosphate
dNTP, 2'-deoxynucleoside-5'-triphosphate
dTTP, 2'-deoxythymidine-5'-triphosphate
ddATP, 2'3'-dideoxyadenosine-5'-triphosphate
ddCTP, 2'3'-dideoxycytidine-5'-triphosphate
ddGTP, 2'3'-dideoxyguanosine-5'-triphosphate
ddTTP, 2'3'-dideoxythymidine-5'-triphosphate
dH_2O, distilled water
DMSO, dimethyl sulphoxide
DNA, deoxyribonucleic acid
DNase, deoxyribonuclease
DTT, dithiothreitol
EDTA, ethylenediaminetetraacetate
g, gravity
G, guanosine
kb, kilobase
kbp, kilobase pairs
L, litre
LSU, large subunit
M, molar
mg, milligram
m, minute
mL, millilitre
µL, microlitre
mM, millimolar
µM, micromolar
MOPS, 3-(*N*- morpholino)-propanesulfonic acid
mRNA, messenger RNA
mt, mitochondrial
mtDNA, mitochondrial DNA
mtRNA, mitochondrial DNA
ng, nanogram = 10^{-9} grams
nm, nanometre = 10^{-9} metre
nt, nucleotide
OAc, acetate
ORF, open reading frame
pmol, picomole = 10^{-12} moles
RACE, rapid amplification of cDNA ends
RNA, ribonucleic acid
RNase, ribonuclease
rpm, revolutions per minute
rRNA, ribosomal RNA
RT, reverse transcriptase
S, Svedberg unit
s, second
SDS, sodium dodecyl sulfate
SSU, small subunit
T, thymidine
Tris, tris(hydroxymethyl)-aminomethane
tRNA, transfer RNA
UV, ultraviolet light
V, volts
vol, volume
v/v, volume by volume

Acknowledgments

A very big thank you goes to Dr. Michael Gray for supervising my Ph.D. training. While at Dalhousie, Mike has provided me with much needed advice, patience and guidance as well as financial and moral support.

I thank my thesis advisory committee members, Drs. Susan Douglas, Bob Lee, Paul Liu and Hyo-Sung Ro, for their helpful advice and constructive criticism.

I also thank members of the "Gray Lab" past and present. Specifically, a sincere 'thanks' goes to Drs. Murray Schnare and David Spencer for their invaluable assistance, patience and stimulating debate. In addition, I acknowledge the help of Mike Charette, Tim Chipman, Tatsuya Ikeda, Amanda Lohan, David Price and Yoichi Watanabe.

I thank members of the Doolittle and Liu labs. In particular I thank John Archibald, Yan Boucher and Joel Dacks. I also thank Cynthia Chow and Quoc Quac for their excellent technical support.

I thank all members of the Biochemistry and Molecular Biology Department for their assistance. In particular, Dr. Ted Palmer for providing a quiet place for me to write.

Finally, I thank my wife, Joanne Norman, and my parents, Ed and Jessie Norman, for their encouragement as well as financial and moral support.

I. Introduction

A. Perspective.

1. Mitochondria

Mitochondria are cytoplasmic membrane-bound organelles that play an important role in cellular metabolism, disease and aging. In this capacity, they are the site of numerous biochemical functions such as oxidative phosphorylation and the production of ATP. Mitochondria contain their own genome (mitochondrial DNA, mtDNA) that is separate from the nuclear genome (nuclear DNA, nDNA). The mitochondrial genome encodes protein components of the five oxidative phosphorylation complexes as well as mitochondrial translation elements such as the ribosomal RNA (rRNA) and protein components of the mitochondrial ribosome as well as mitochondrial transfer RNAs (tRNAs). These gene products work in combination with nDNA-encoded proteins and RNAs that are targeted to the mitochondrion, allowing mitochondria to undertake DNA replication, transcription, transcript processing and translation.

It is thought that mitochondria arose by way of a single endosymbiotic event in which a pre-eukaryotic cell engulfed an α-proteobacterium. This event led to a symbiotic relationship wherein each member (host and prey) became dependent on the other (reviewed in Gray *et al.*, 1999; Lang *et al.*, 1999a). Subsequent migration of mtDNA genes to the nucleus, mtDNA rearrangement and loss of mitochondrial genes due to loss of a particular mitochondrial function have resulted in pronounced diversity of mitochondrial genome structure and organization among the eukaryotic lineages.

There are several reasons for studying mitochondrial genomes. First, it helps us to reconstruct eukaryotic evolution through comparison of mitochondrial gene phylogenies, gene content and gene organization. Second, it allows us to see how different eukaryotes have evolved in the presence of this organelle and how this evolution has in turn impacted on the evolution of mtDNA and the function of mitochondria. Finally, information gained about how mtDNAs evolve may be useful in understanding how larger genomes evolve.

In recent years increased interest and effort has been directed at characterizing mitochondrial genomes from protists. These studies have been particularly fruitful in helping us to understand how mitochondria and organisms evolve. Despite this effort, a number of large and important protistan groups have yet to be explored. In particular, the dinoflagellates represent the largest uncharacterized group of eukaryotes with respect to mtDNA. This deficiency is particularly surprising because the two closest relatives of the dinoflagellates, apicomplexans (a group of intracellular parasites) and ciliates, were two of the first protist groups whose mtDNAs were characterized. Consequently, information on mitochondrial genome structure and organization in the dinoflagellates is required. Such information will not only further our understanding of how the mitochondrial genome evolved within a closely related group of protists (dinoflagellates, apicomplexans and ciliates), but will also contribute to our knowledge of how all eukaryotes that harbor a mitochondrion evolved.

2. Dinoflagellates

a. Dinoflagellates are a medically and ecologically important group.

Dinoflagellates are important for at least two reasons. First, a number of photosynthetic species produce potent neurotoxins that can lead to human sickness or death. Second, a number of symbiotic species such as members of the genus *Symbiodinium* are responsible for the photosynthetic activity of coral reefs.

b. *Crypthecodinium cohnii* is a model dinoflagellate. One of the best studied dinoflagellates is the "weedy" species *C. cohnii*. *C. cohnii* has long been a dinoflagellate species of choice for many reasons. First, its growth medium and growing conditions have been extensively studied and well defined (Gold & Baren, 1966; Tuttle & Loeblich, 1975). Second, *C. cohnii* is one of the few heterotrophic dinoflagellates that can be grown in a defined medium without adding a prey source. Third, axenic strains such as Whd are deposited in ATCC, making the organism readily available to many researchers. Fourth, the axenic nature of *C. cohnii* is particularly appealing when working with organellar genomes, in that contaminating bacteria would make it more difficult to isolate and characterize dinoflagellate mtDNA. Fifth, *C. cohnii* grows rapidly and can tolerate high culture densities. Finally, based on phylogenetic studies, *C. cohnii* is an early branching dinoflagellate (McNally *et al.*, 1994; Saunders *et al.* 1997); thus information gathered about this species can readily be compared with information about later diverging taxa. Such comparisons are especially useful in determining how members of this clade have evolved over time. Finally, this particular dinoflagellate species is non-toxic.

c. With respect to mtDNA, dinoflagellates are one of the last major uncharacterized groups of protists. Most of what is known about mitochondrial genome structure and organization comes from examination of mtDNA in the three most recently emerged eukaryotic lineages: animals, plants and fungi. Mitochondrial DNA has been characterized in only a fraction of the extant protist phyla. In particular, the dinoflagellates represent one of the last major protistan lineages for which a complete mtDNA sequence does not exist. In contrast, extensive mtDNA sequence is available from the apicomplexans and ciliates, two protistan lineages that share a recent common ancestor with the dinoflagellates. Characterization of mtDNA sequence from a dinoflagellate is required to complete the picture of mtDNA evolution within this clade.

Two recent studies have independently reported a *cox1* gene sequence from *C. cohnii* (Inagaki *et al.*, 1997; Norman & Gray, 1997). Based on *cox1* sequence divergence and gene organization (described in this thesis), *C. cohnii* appears to exhibit a derived type of mitochondrial genome.

The objective of this study was to increase our understanding of mitochondrial genome structure and gene content in dinoflagellates. This was accomplished by isolating mtDNA and determining mitochondrial genome sequence from the non-photosynthetic dinoflagellate *C. cohnii*. In particular, *cox1*, *cob*, *cox3* and *rnl* gene data were obtained using traditional molecular biology techniques such as DNA sequencing, Southern and northern analysis as well as phylogenetic reconstruction.

Another of the objectives of this study was to further our understanding of how mitochondrial genomes evolve. This study took a comparative approach in which

information gathered about *C. cohnii* mtDNA was compared with data available for the mitochondrial genomes of apicomplexans and ciliates as well as other organisms.

B. Dinoflagellates.

The dinoflagellates are small (0.01-2.0 mm) single-celled organisms that live in aqueous environments. The word "dino" has its origins in the Greek word for "whirling" and describes the type of swimming movement that is characteristic of this protist (Dodge, 1985). Most dinoflagellates are propelled by two flagella (Fig. 1.1). The first is called the transverse flagellum and encircles the middle of the cell, within an equatorial groove or girdle. The second flagellum (longitudinal) extends away from the posterior of the cell and is rooted within the sulcus groove. This groove forms a second channel in the lower portion of the cell, dividing it into right and left halves (Spector, 1984b).

Many free-living dinoflagellate species possess a theca or shell that is composed of a layer of alveoli (membranous sacs) underneath the external cell membrane (Sleigh, 1989). In many dinoflagellate species, a cellulose-like material is deposited within the alveoli, thereby forming a protective covering around the cell (Morrill & Loeblich, 1983).

Dinoflagellates possess a number of subcellular structures that set them apart from many other protists. These include tricocysts (projectable spines), a cellulosic-like cell wall, chloroplasts, pustules (contractile vacuole-like organelles) (Sleigh, 1989) and an unusual nuclear organization (Spector, 1984a). The dinoflagellate nucleus lacks nucleosomes (Rizzo & Burghardt, 1982), possesses permanently condensed chromosomes with few histones (Rizzo & Nooden, 1974a; Rizzo & Nooden, 1974b) and

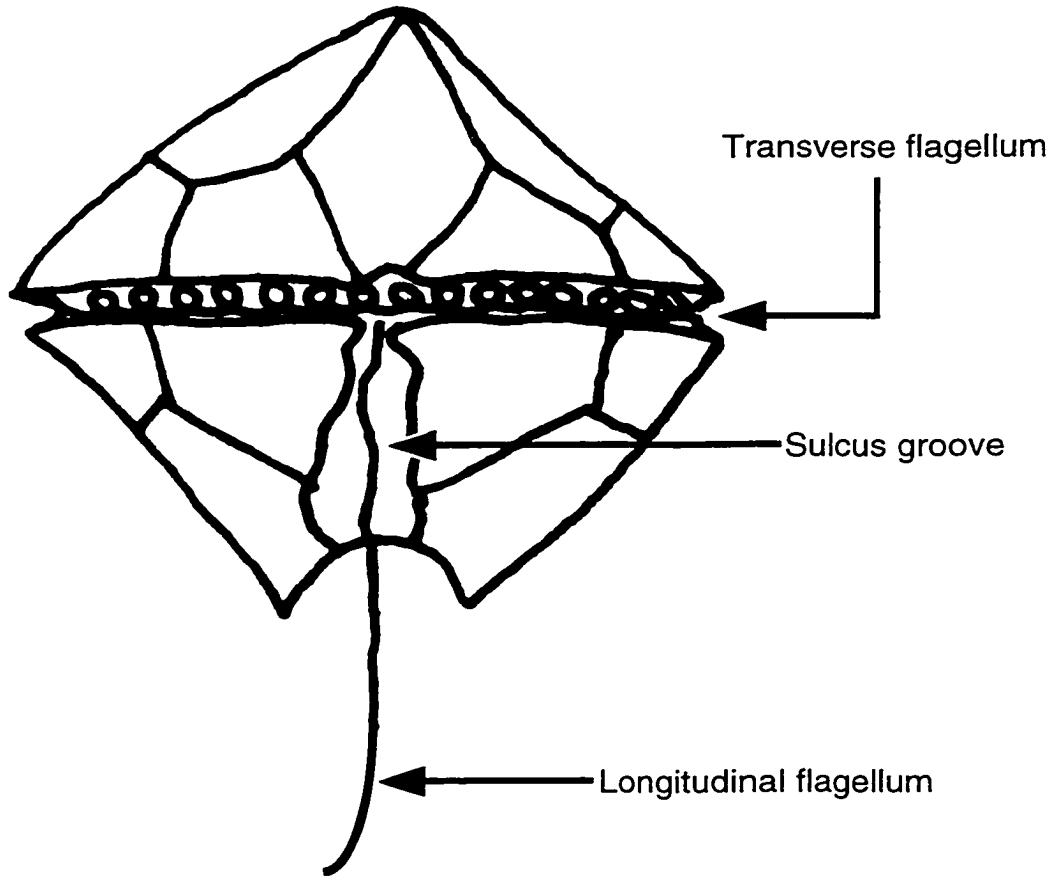


Figure 1.1. Line drawing of a 'typical' dinoflagellate. Modified from Spector (1984b).

undergoes closed mitosis utilizing external spindles (Kubai & Ris, 1969; Leadbeater & Dodge, 1967; Spector & Triemer, 1981). Dinoflagellate nDNA is also unusual in that it contains a type of modified DNA base that has not been described in any other eukaryote (Rae, 1973; Rae, 1976). In fact, these unusual nuclear characteristics have prompted some authors to define dinoflagellates as "mesocaryotes", a term that implies that dinoflagellates occupy an intermediate position between eukaryotes and prokaryotes (Dodge, 1965). It is now accepted that these primitive characters are derived because dinoflagellates possess other morphological features, such as the presence of an elaborate cytoplasm with rough and smooth endoplasmic reticula, golgi bodies and mitochondria with tubular cristae, that clearly indicate that the Dinozoa are eukaryotes (Sleigh, 1989; Spector, 1984b).

1. Ecology of dinoflagellates.

While all dinoflagellates live in aquatic environments, the vast majority of them can be found swimming freely or floating in large bodies of salt water. They live in all latitudes extending from the Arctic and Antarctic seas to the warmer waters of the tropics. Many exist in marginal habitats, confined to the shallow water surrounding the shore where ocean salinity tends to be lower and food sources more plentiful. In fact, several species have been found that exist between grains of beach sand, or highly ephemeral intertidal pools and salt marshes. Despite the propensity for dinoflagellates to live in marine environments, a small percentage of dinoflagellates prefer fresh water, living in large lakes and ponds (Dodge, 1985).

2. Nutrient acquisition in dinoflagellates.

Dinoflagellates have adopted a wide range of feeding mechanisms and modes of acquiring nutrients. Traditionally members of the Dinozoa were characterized as either autotrophs or heterotrophs, with those species possessing plastids being viewed as strictly autotrophic. This view has changed, however, and it is now recognized that even primarily photosynthetic species supplement their nutrition by feeding on other organisms (Stoecker, 1999).

Approximately 50% of all dinoflagellates are strict heterotrophs. Only a few species such as *C. cohnii* can be grown in rich media containing a high concentration of salts and a carbon source (Gold & Baren, 1966; Tuttle & Loeblich, 1975). Most heterotrophs, however, feed by ingesting particulate matter, either living or dead (Hansen & Calado, 1999). Particle ingestion has been shown to occur in a number of different ways. For example, one group of dinoflagellates engulfs whole prey through the sulcus groove (Hansen & Calado, 1999). Initially it was thought that this feeding mechanism was restricted to naked early evolving species such as *Noctiluca* and *Oxyrrhis*; however, several late evolving armored dinoflagellates have now been shown to engulf their prey (Gaines & Elbrachter, 1987; Schnepf & Elbrachter, 1992). Thus it appears that this feeding mechanism occurs throughout the dinoflagellates.

A second group uses a feeding tube appendage to ingest particles. The most common type of feeding tube is the peduncle, which consists of a protoplasmic strand that protrudes from the mid-ventral area of the sulcus groove of the dinoflagellate and attaches to its food source. The ingested material flows into the peduncle and into one or

several vacuoles (Hansen, 1991). While whole organisms can be eaten using the feeding tube, if the prey is too large the peduncle may be used to pierce the plasma membrane of the prey organism and ingest its cytoplasmic contents. In fact, organelles from prey organisms can often be seen in food vacuoles (Hansen & Calado, 1999). This feeding mechanism is intriguing because it could provide a means by which dinoflagellates could secondarily acquire organelles without coming in contact with nDNA from the prey organism.

Heterotrophic dinoflagellates are abundant in marine environments, where they feed on a wide range of eukaryotes such as diatoms, other dinoflagellates, prymnesiophytes, cryptophytes, chlorophytes and injured metazoans or metazoan eggs. Dinoflagellates also feed on many different types of eubacteria such as cyanobacteria (Jeong, 1999). Heterotrophic dinoflagellates exhibit a wide range of feeding lifestyles including parasitic, symbiotic, saprophytic and holozoic (Jacobson, 1999; Jeong, 1999).

3. Photosynthetic dinoflagellates.

a. Plastid types in dinoflagellates. Photosynthetic dinoflagellate species appear to have acquired plastids secondarily in at least three different events. Most dinoflagellate plastids are pigmented with chlorophylls *a* and c_2 and the accessory pigment peridinin. Plastids in these species also tend to be surrounded by three membranes (Dodge, 1989; Palmer & Delwiche, 1998; Schnepf, 1993). The peridinin-type plastid appears to have been acquired from the red algal lineage (Delwiche, 1999; Durnford *et al.*, 1999). Interestingly, phylogenetic trees based on 18S rRNA genes show that this plastid type is not monophyletic within the dinoflagellates. In fact, peridinin-type plastid species appear

to be randomly scattered throughout the dinoflagellates. In stark contrast, phylogenetic reconstruction using two plastid-encoded genes (23S rRNA and *psbA)* suggest that the peridinin-containing dinoflagellates form a monophyletic group, branching as a sister group to the apicomplexans (Zhang *et al.*, 2000).

A second major group are the fucoxanthin-containing species. These dinoflagellates contain plastids that are pigmented with chlorophylls *a* and *c* and the accessory pigment fucoxanthin. This plastid type is similar to that found in the haptophytes (a group of marine algae) and heterokonts (diatoms, kelps and water moulds). It has been suggested that this plastid type arose when a dinoflagellate engulfed a haptophyte alga (Delwiche, 1999; Tengs *et al.*, 2000) or a diatom (Chesnick *et al.*, 1996; Palmer & Delwiche, 1998). Because both the haptophytes and heterokonts secondarily acquired their plastids from a red alga, the dinoflagellate host would be a tertiary endosymbiont. A third group of dinoflagellates contains plastids that are surrounded by four membranes and contain chlorophylls *a* and *b*, features suggestive of a green algal origin for these plastids (Palmer & Delwiche, 1998; Watanabe *et al.*, 1990; Watanabe *et al.*, 1987).

b. Plastid genome sequence and organization in dinoflagellates. Recently, the first dinoflagellate plastid genes were isolated and sequenced from two different peridinin-containing species, *Heterocapsa triquetra* (Zhang *et al.*, 1999) and *Amphidinium operculatum* (Barbrook & Howe, 2000). The plastid DNA (ptDNA) of these two dinoflagellates is unlike any plastid genome previously characterized. It is composed of numerous mini-circles (2.1-3.1 kbp), each containing a single gene, a

putative control region and spacer DNA. In *H. triquetra* nine different genes have been characterized including the chloroplast protein-coding genes *psaA*, *psaB*, *psbA*, *psbB*, *psbC*, *atpA* and *petB*. In addition, minicircles containing plastid 23S and 16S rRNA genes were isolated and sequenced (Zhang *et al.*, 1999). In *Amphidinium operculatum*, only five chloroplast protein-coding genes have been identified so far. Three (*psaA*, *psbA*, *psbB)* of these genes were also found in *H. triquetra*, whereas the remaining two genes (*atpB* and *petD*) are so far unique to *A. operculatum*.

Besides coding sequence, each mini-circle contains a tripartite non-coding region that is highly conserved among mini-circles containing different genes. These regions, however, are not conserved among different species (Barbrook & Howe, 2000; Zhang *et al.*, 1999). The purpose of these tripartite regions is still under investigation, but their repetitive nature suggests that they may be involved in plastid DNA replication (for example, by acting as origins of replication). Alternatively, they may provide sites through which the mini-circles anchor to the plastid membrane during plastid division. This would ensure that genes are not lost by unequal segregation (Zhang *et al.*, 1999). In *H. triquetra*, the tripartite region contains a central core domain (188 nt) that is flanked by two shorter non-identical repeat regions (135 nt) that appear to be related to one another. There are also short direct and inverted repeat regions located throughout the non-coding region (Zhang *et al.*, 1999). The mini-circles of *A. operculatum* also displays a tripartite region but the sequence is different from that of *H. triquetra*, with the central core being 49 nt in length (Barbrook & Howe, 2000).

One point of interest is that circles that contain the same gene are not identical in

length. For example, in *H. triquetra* three mini-circles were identified that harbored truncated genes. Two of these contained a truncated *psbC* gene and one possessed a truncated 16S rRNA gene. In *Heterocapsa pygmara* four *psbA* bands (1.2-3.0 kbp) were also visualized during Southern hybridization. Preliminary sequence analysis suggests that the difference in size is due to length variability in the non-coding region (Zhang *et al.*, 1999).

Recently, plastid small subunit (SSU) rRNA genes were sequenced from three fucoxanthin-containing dinoflagellates. Phylogenetic reconstruction places these dinoflagellates within the haptophyte plastid clade. This result suggests that fucoxanthin-containing dinoflagellates acquired their plastid from a haptophyte alga and therefore are of tertiary endosymbiotic origin (Tengs *et al.*, 2000). In contrast, plastid gene phylogenies suggest that peridinin-containing dinoflagellates are more closely related to chromist plastids (Zhang *et al.*, 1999). These results support the hypothesis that dinoflagellates acquired their plastids independently from two separate lineages.

Surprisingly, the plastid genome organization described for dinoflagellates is somewhat similar to the bizarre mtDNA organization recently described in *Dicyema* (mesozoan). In this organism, mitochondrial electron transport genes (*cox1*, *cox3* and *cob*) are encoded by separate small circles (1600-1700 nt) (Watanabe *et al.*, 1999). Similar to the situation in *H. triquetra* and *A. operculatum*, each circle contains one gene, a possible control region and spacer DNA.

4. *Crypthecodinium cohnii*.

C. cohnii is a typical theca-containing dinoflagellate that possesses a transverse and posterior flagellum. Although it can be grown axenically on defined medium this dinoflagellate is typically heterotrophic and feeds by way of a peduncle feeding tube (Hansen & Calado, 1999). This species is not known to be photosynthetic, but it produces 4-5 times more carotenoids (a photosynthetic accessory pigment) in the presence of light than in its absence (Tuttle & Loeblich, 1975). This result has occasionally been interpreted to mean that *C. cohnii* (and possibly all dinoflagellates) are ancestrally photosynthetic (Cavalier-Smith, 1999). In addition, Kubai & Ris (1969) have identified a membrane structure that they assert is a degenerate plastid. While their micrograph clearly shows what appears to be a plastid-like organelle, this type of structure has not been reported by any other investigators and was not seen in electron micrographs prepared for the current work.

C. cohnii cells contain large quantities of nDNA, with estimates ranging from 3.8 to 7.3 pg per nucleus (Allen *et al.*, 1975). This amount is approximately double that estimated for haploid human cells, which contain 2.7 pg DNA per nucleus (Holm-Hansen, 1969). In *C. cohnii* the nDNA is packed into 99 (Kubai & Ris, 1969) or 100 (Haapala & Soyer, 1974) chromosomes. Denaturation kinetic profiles suggest that *C. cohnii* nDNA contains 55-60% repetitive sequence interspersed with non-repetitive DNA (Allen *et al.*, 1975). This value is typical for most eukaryotes.

Dinoflagellates undergo typical mitosis (G1-S-G2-M) yet they exhibit some unusual nuclear characteristics. For example, the chromosomes of *C. cohnii*, like those of

many other dinoflagellates, are permanently condensed throughout the cell cycle. Overall, chromosomes look like stubby arches. These structures are composed of tightly packed orderly whorls of filaments that are between 25 and 50Å in diameter (Rae & Steele, 1978; Spector, 1984a). This filament diameter is much narrower than that determined for chromatin in other eukaryotes. These chromosomes also lack histones and nucleosomal structure, and have a 10:1 DNA to protein ratio, which in most other eukaryotes is more typically 1:1 (Bhaud *et al.*, 2000). In *C. cohnii* six basic, low-molecular-weight proteins have been found to associate with condensed DNA. These proteins are lysine-rich but they have a lower affinity for DNA than typical histone proteins. Another unusual feature is that they bind more tightly to single-stranded than to double-stranded DNA (Vernet *et al.*, 1990).

Dinoflagellates also exhibit a number of atypical nuclear features that are only apparent during mitosis. During this stage of the cell cycle, the nuclear envelope remains intact and dividing chromosomes are attached to it. Large cytoplasmic invaginations as well as cytoplasmic channels traverse the nucleus, with each nucleus containing microtubules (mitotic spindle) that do not come in contact with chromosomes (Bhaud *et al.*, 2000; Kubai & Ris, 1969). Permanent nuclear envelopes have been described in other eukaryotes such as yeast, euglenids and diatoms, but in these organisms the mitotic spindle is still intranuclear. In the dinoflagellates, however, the mitotic spindle does not come in contact with dividing chromosomes but instead perforates the nucleus by way of cytoplasmic conduits. Exactly how this type of chromosome separation evolved is unknown. Initially it was thought that the dinoflagellates were more similar to

prokaryotes, using membrane separation to segregate chromosomes. However, this method of chromosome separation in bacteria is thought to be incorrect (Gordon & Wright, 1998). Therefore, the question remains whether or not nuclear membrane separation causes chromosomes to be pulled apart in dinoflagellates and, if so, whether this is an ancestral or derived character state (Bhaud *et al.*, 2000).

C. Phylogeny.

1. Phylogenetic placement of dinoflagellates within the eukaryotes.

Dinoflagellates comprise a large and structurally diverse group of unicellular eukaryotes (protists). Yet early phylogenetic studies were mostly focused on establishing the relationship between dinoflagellates and other early-branching eukaryotes (see Saunders *et al.*, 1997 for review). This search was mostly stimulated by the question: Are dinoflagellates mesokaryotic (intermediate between eukaryotes and prokaryotes) as proposed by Dodge (1965)? Other ancestral relationships that had been proposed included a dinoflagellate-chromophtye clade (Cavalier-Smith, 1975; Ragan, 1978) and a dinoflagellate-ciliate alliance (Taylor, 1978; Taylor, 1976). Analysis of 5S rRNA sequence data from *C. cohnii* (Hinnebusch *et al.*, 1981) suggested that the dinoflagellates have a stronger phylogenetic affinity with later diverging eukaryotic lineages than with early evolving ones. This finding challenged the conclusion that the apparently 'primitive' nature of the dinoflagellate nucleus was an ancestral characteristic. Instead the conclusion was drawn that these 'primitive' characters evolved by reduction. This inference was further bolstered by U1-U5 small nuclear RNA primary and secondary

structure data that firmly placed the dinoflagellates within the eukaryotes (Liu *et al.*, 1984; Reddy *et al.*, 1983).

The first small subunit (SSU) rRNA data obtained for a dinoflagellate was from the photosynthetic species *Prorocentrum micans*. Distance phylogenetic reconstructions and secondary structure analysis were interpreted to support the basal separation of the dinoflagellates (mesokaryotic hypothesis) relative to later evolving eukaryotes (Herzog & Maroteaux, 1986). Subsequent phylogenetic reconstructions using additional taxa showed that the dinoflagellates (Dinozoa) share a common ancestry with apicomplexans (Apicomplexa) and ciliates (Ciliophora) (Fig. 1.2) (Cavalier-Smith, 1993; Gajadhar *et al.*, 1991; Schlegel, 1991; Van de Peer *et al.*, 1996). Within this assemblage, termed Alveolata (alveolates) (Gajadhar *et al.*, 1991; Siddall *et al.*, 1995), the dinoflagellates and apicomplexans cluster together, forming a sister group to the ciliates (Cavalier-Smith, 1993; Gajadhar *et al.*, 1991; Sadler *et al.*, 1992). According to this analysis, the alveolates branch within the eukaryotic crown and are only distantly related to putatively early-branching eukaryotes such as *Giardia*.

In addition to molecular data, the concept of an alveolate clade is supported by ultrastructural features such as tubular mitochondrial cristae, cortical alveoli (Siddall *et al.*, 1995; Van de Peer *et al.*, 1996) and rows of microtubules just below the plasma membrane (Siddall *et al.*, 1995).

Most members of the alveolates fit neatly within one of the major phyla (Dinozoa, Apicomplexa or Ciliophora). However, the placement of the parasitic genus *Perkinsus* is still uncertain. The apical structure of the zoospore of *Perkinsus marinus* caused Perkins

Fig. 1.2. Nuclear SSU rRNA phylogenetic tree showing inferred relationships among the major eukaryotic lineages. The tree was generated using the neighbour joining algorithm (NEIGHBOR; Felsenstein, 1993).The scale bar indicates 10 substitutions per nucleotide. (David F. Spencer, unpublished data).

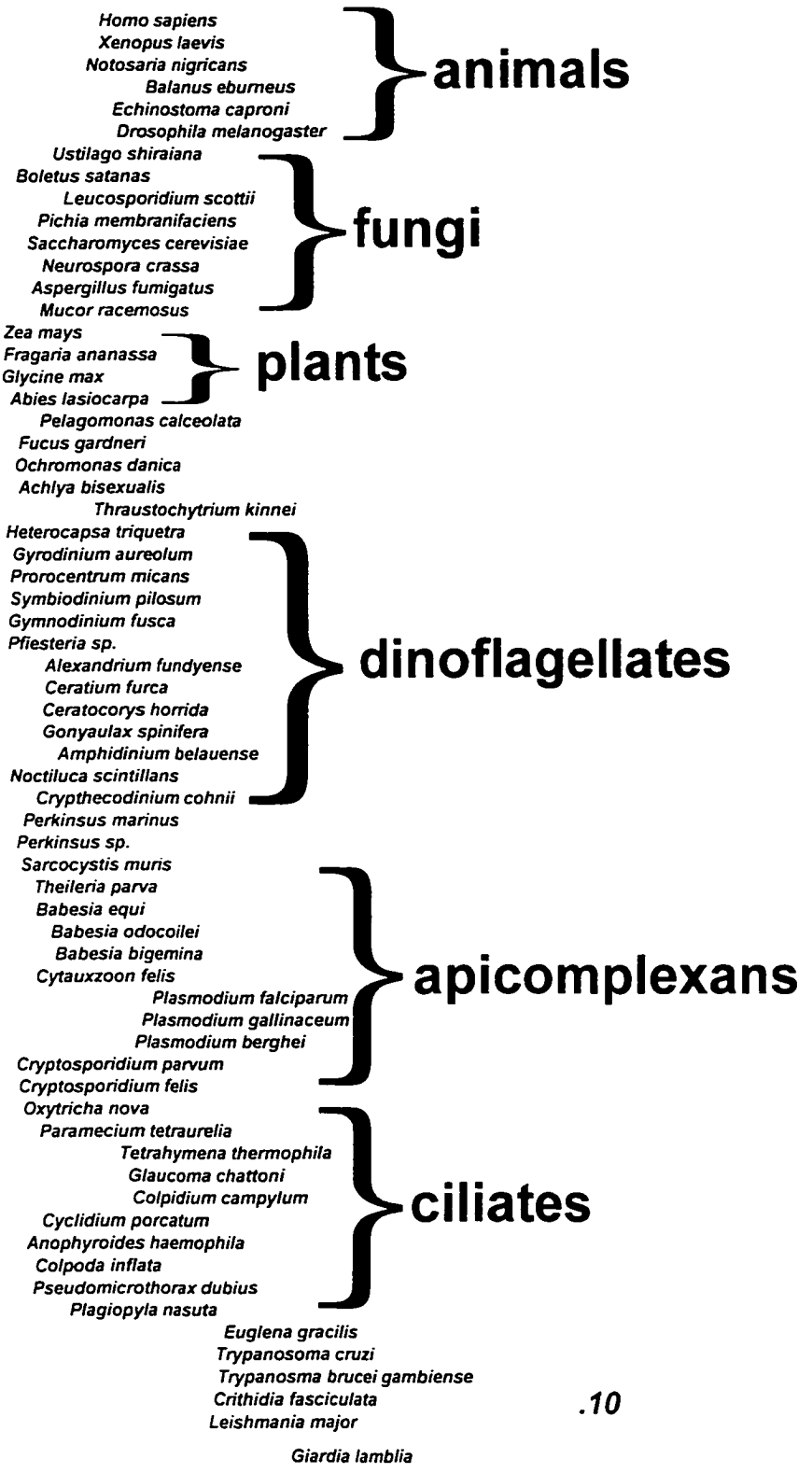

Homo sapiens
Xenopus laevis
Notosaria nigricans
Balanus eburneus
Echinostoma caproni
Drosophila melanogaster
animals
Ustilago shiraiana
Boletus satanas
Leucosporidium scottii
Pichia membranifaciens
Saccharomyces cerevisiae
Neurospora crassa
Aspergillus fumigatus
Mucor racemosus
fungi
Zea mays
Fragaria ananassa
Glycine max
Abies lasiocarpa
plants
Pelagomonas calceolata
Fucus gardneri
Ochromonas danica
Achlya bisexualis
Thraustochytrium kinnei
Heterocapsa triquetra
Gyrodinium aureolum
Prorocentrum micans
Symbiodinium pilosum
Gymnodinium fusca
Pfiesteria sp.
Alexandrium fundyense
Ceratium furca
Ceratocorys horrida
Gonyaulax spinifera
Amphidinium belauense
Noctiluca scintillans
Crypthecodinium cohnii
dinoflagellates
Perkinsus marinus
Perkinsus sp.
Sarcocystis muris
Theileria parva
Babesia equi
Babesia odocoilei
Babesia bigemina
Cytauxzoon felis
Plasmodium falciparum
Plasmodium gallinaceum
Plasmodium berghei
Cryptosporidium parvum
Cryptosporidium felis
apicomplexans
Oxytricha nova
Paramecium tetraurelia
Tetrahymena thermophila
Glaucoma chattoni
Colpidium campylum
Cyclidium porcatum
Anophyroides haemophila
Colpoda inflata
Pseudomicrothorax dubius
Plagiopyla nasuta
ciliates
Euglena gracilis
Trypanosoma cruzi
Trypanosma brucei gambiense
Crithidia fasciculata
Leishmania major
.10
Giardia lamblia

Figure 1.2

(1976) to place this taxon within the apicomplexans even though members of this genus share morphological similarities with both dinoflagellates and apicomplexans (Perkins, 1996). Phylogenetic reconstruction using SSU rRNA (Flores *et al.*, 1996; Kotob *et al.*, 1999) and actin (Reece *et al.*, 1997) genes show that *Perkinsus* has a closer affinity with the dinoflagellates than with apicomplexans but occupies a position that is intermediate between the two. Moreover, its phylogenetic placement suggests that study of this organism may reveal important clues concerning the nature of the common ancestor that gave rise to the dinoflagellates and apicomplexans. In particular, it may provide key information regarding how plastid and mitochondrial genomes evolved within the alveolates.

2. Phylogenetic relationships within the dinoflagellates.

Despite the high degree of interest in the origin of dinoflagellates, few investigators have examined phylogenetic relationships within the dinoflagellates. The first phylogenetic study, utilizing large subunit (LSU) rRNA divergent domains D1-D8, resolved several intra-dinoflagellate relationships. Although the taxon sampling was relatively small (13 species), it established that *Noctiluca* and *Oxyrrhis* are early-branching dinoflagellates (Lenaers *et al.*, 1991). *C. cohnii* also appears to be an early-evolving dinoflagellate species and usually branches as a member of the Gonyaulacales (McNally *et al.*, 1994; Saunders *et al.*, 1997). Other phylogenetic studies have focused on resolving species relationships within a single genus such as *Symbiodinium* or relationships among various symbiotic genera (Gast & Caron, 1996; McNally *et al.*, 1994; Rowan & Powers, 1992).

D. Overview of mitochondrial genome structure and gene content

a. The origin of mitochondria. The serial endosymbiotic theory (SET) proposes that mitochondria (and chloroplasts) arose when a bacterium was engulfed by a host cell. This resulted in a symbiotic relationship whereby each partner became dependent on the other (reviewed in Gray *et al.*, 1999; Lang *et al.*, 1999a). Phylogenetic reconstructions suggest that mitochondria are derived from a member of the α-proteobacteria (eubacteria) (Falah & Gupta, 1994; Gray & Spencer, 1996) whereas cyanobacteria (also eubacteria) gave rise to chloroplasts (Delwiche *et al.*, 1995; Gray & Spencer, 1996; Nelissen *et al.*, 1995).

What did the pre-mitochondrial host cell look like? One theory states that it was simply an amitochondriate eukaryote (i.e., possessed a large size, cytoskeleton and endomembrane system). Alternatively (or in addition), it has been proposed that eukaryotes originated from the fusion of an archaebacterium and proteobacterium (for review see Gray *et al.*, 1999; Lang *et al.*, 1999a). The first hypothesis was bolstered by the discovery of a group of amitochondrial protists that were thought to have diverged from other eukaryotes before the acquisition of mitochondria (Cavalier-Smith, 1983; Hirt *et al.*, 1999). These protists, collectively termed the Archezoa (Cavalier-Smith, 1983), are composed of Microsporidia, Diplomonada, Metamonada and Parabasalia. Recent findings, however, have cast doubt on the validity of Archezoa as the pre-mitochondrial progenitor that gave rise to higher eukaryotes. First, several nucleus-encoded heat shock protein genes have been found in Microsporidia (Germot *et al.*, 1997; Hirt *et al.*, 1997), Metamonada and Parabasalia (Bui *et al.*, 1996; Germot *et al.*, 1996; Roger *et al.*, 1996)

and Diplomonada (Roger *et al.*, 1998). These proteins typically function in mitochondria and this suggests that members of the Archezoa originally possessed mitochondria but lost them secondarily. Second, the evolutionary placement of several Archezoa as early diverging eukaryotes has been questioned. Typically, 18S rRNA (Sogin, 1991; Sogin, 1997), EF-1α and EF2 (Hashimoto *et al.*, 1998; Kamaishi *et al.*, 1996) trees place the three amitochondriate clades (Microsporidia, Metamonada and Parabasalia) at the base of the eukaryotes. More recent phylogenetic studies involving RNA polymerase II (Hirt *et al.*, 1999), tubulin (Keeling & Doolittle, 1996) and Hsp70 (Germot *et al.*, 1997; Hirt *et al.*, 1997) suggest that Microsporidia belong to the fungal group. Other studies also show that the basal placement of Metamonada and Parabasalia is poorly supported and thus is suspect (Hirt *et al.*, 1999). In fact all members of the Archezoa appear to experience accelerated rates of evolution as indicated by the gene sequences exhibiting long branch lengths. This also raises the possibility of a long-branch-attraction artifact (Felsenstein, 1978; Philippe & Forterre, 1999).

Recently the archaebacterium/proteobacterium fusion theory has received greater support. Phylogenetic analyses show that eukaryotic nuclear genomes are composed of operational genes that are largely of eubacterial origin whereas informational processing genes appear to be derived from archaebacteria (Rivera *et al.*, 1998). The fusion theory does not specifically address whether the endosymbiotic event that gave rise to mitochondria occurred at the same time or after the fusion event (most investigators assume it occurred later). However, if all modern-day eukaryotes evolved from a mitochondria-containing ancestor, then the time between the formation of the first

eukaryotic cell and the endosymbiotic event that gave rise to mitochondria may be short. Moreover, it is conceivable that both events occurred at the same time (Lang *et al.*, 1999a).

b. Size and structure of mtDNA. With more than 130 mitochondrial genomes completely sequenced it has become obvious that there is great diversity in mtDNA size, structure and gene content (reviewed in Gray *et al.*, 1998; Lang *et al.*, 1999a). The smallest mtDNAs (~6 kbp in size) belong to members of the apicomplexans (Fig. 1.3) (Feagin, 1994; Wilson & Williamson, 1997). At the other extreme are the land plants, whose genomes are up to 2400 kbp in size (muskmelon), although the largest completely sequenced land plant mtDNAs, those of *Arabidopsis thaliana* (small mustard) (Unseld *et al.*, 1997) and *Beta vulgaris* (sugar beet) (Kubo *et al.*, 2000), are only 366,924 and 368,799 bp, respectively.

Mitochondrial DNA structure is also quite variable. For example, vertebrate mtDNA comprises a 16-kbp circle that contains 37 genes (reviewed in Boore, 1999; Wolstenholme, 1992). In contrast, the mtDNA of another metazoan, *Dicymena*, consists of a series of 2- to 3-kbp circles, with one gene per circle (Watanabe *et al.*, 1999).

The most common mtDNA structure is a single circle between 16 and 70 kbp in size; however, linear mtDNAs also exist (e.g., in *Tetrahymena pyriformis*, *Ochromonas danica* and *Jakoba libera*) (Gray *et al.*, 1998; Lang *et al.*, 1999a). Interestingly, organisms that contain linear mtDNAs sometimes occur as a sister group to taxa that possess circular mtDNAs (e.g., within the genus *Chlamydomonas*) (Nedelcu, 1998; Nedelcu & Lee, 1998). This suggests that the change from a circular structure to a linear

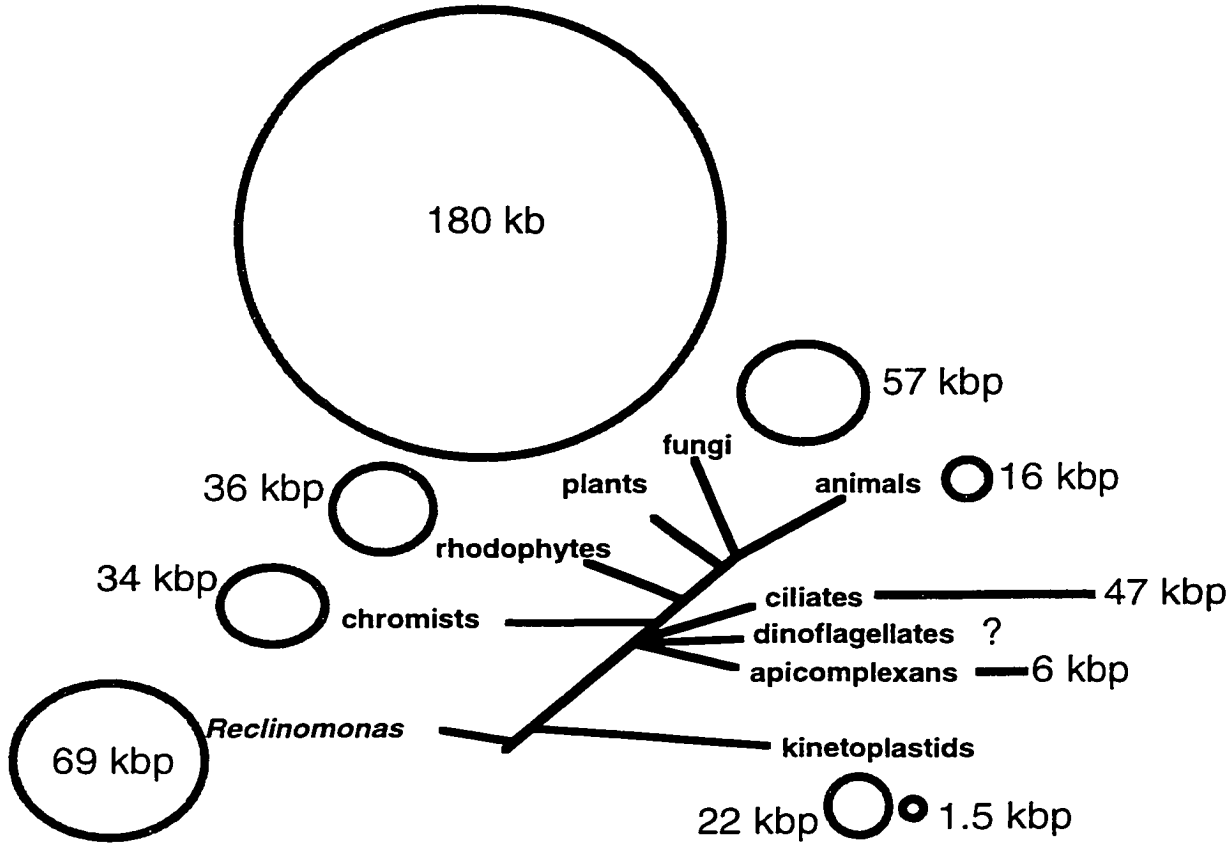


Fig. 1.3. Diagram showing representative examples of mtDNA shapes and sizes in the major eukaryotic lineages. Mitochondrial DNA molecules are shaded grey. Approximate sizes of individual mtDNAs are noted. Question mark (?) indicates that shape and size of mtDNA in dinoflagellates is unknown. Note: the fungal genome represented belongs to *Allomyces macrogynus* (accession number NC001715).

one may be relatively easy to accomplish, which limits the usefulness of this character as an evolutionary marker.

c. Gene content in mtDNA. Most mtDNAs encode the same basic set of genes, which can be categorized as falling into two groups. The first group consists of information processing genes, such as mitochondrial rRNA genes (*rnl*, *rns*) and tRNA genes (*trn*). The second set comprises protein-coding genes that are components of the inner membrane respiratory chain. In particular, genes for complex I (*nad*), complex III (*cob*), complex IV (*cox*) and complex V (*atp*) are typically encoded by mtDNAs (Gray & Spencer, 1996).

Mitochondrial genomes can be roughly classified into two general categories, ancestral or derived (Gray *et al.*, 1999). The best example of an ancestral mitochondrial genome is that of *Reclinomonas americana*. The mtDNA of this protist encodes the largest complement (97) of genes described to date in mitochondrial genomes (Lang *et al.*, 1997). The high number of mtDNA-encoded genes, as well as the distinct eubacterial nature of several of them, supports the view that the *R. americana* mitochondrial genome is ancestral (minimally diverged) (Gray *et al.*, 1998; Lang *et al.*, 1997; Lang *et al.*, 1999b). Land plants also have a large contingent of genes, with *Marchantia polymorpha* (liverwort) exhibiting the greatest number (71). Like *R. americana* mtDNA, land plant mitochondrial genomes can be classified as ancestral because they have genes not present in the mtDNA of metazoans. These 'missing' genes include complex II genes (*sdh*), ribosomal protein genes (*rps* and *rpl*) and additional complex I (*nad*) and complex V (*atp*) genes. Other identifiable features of ancestral mitochondrial genomes include a

standard genetic code, eubacteria-like gene clusters and eubacteria-like rRNA genes (*rns*, *rnl*, *rrn5*) (Gray *et al.*, 1999).

Derived mtDNAs are present in groups such as metazoans, some fungi, *Chlamydomonas* species and apicomplexans (Gray *et al.*, 1999). Some of the features that typify this group include gene loss, nonstandard genetic codes and accelerated rates of protein and rRNA gene evolution. In fact, rRNA divergences can be quite extreme, leading to gene truncations or even severe fragmentation (Gray *et al.*, 1999).

Not all mtDNAs fit neatly into ancestral or derived types. For example the mtDNA of the green alga *Scenedesmus obliquus* exhibits characteristics of both groups (Nedelcu *et al.*, 2000). Another interesting feature is that the derived-type mitochondrial genomes appear to have arisen independently several times.

1. Apicomplexan organellar genomes

a. Apicomplexan mtDNA. One of the first protist mtDNAs to be completely sequenced was that of the rodent malaria parasite, *Plasmodium yoelii* (Vaidya *et al.*, 1989). Since that time, the mtDNA sequences of six additional apicomplexans have been determined, including those of *Plasmodium falciparum* (Feagin *et al.*, 1992), *Plasmodium vivax* (McIntosh *et al.*, 1998; Sharma *et al.*, 1998), *Plasmodium berghei* (GenBank accession number AF014115), *Plasmodium chabaudi* (GenBank accession number AF014116), *Plasmodium reichenowi* (GenBank accession number AJ251941) and *Theileria parva* (Kairo *et al.*, 1994). The apicomplexan mitochondrial genome is the smallest known (approximately 6 kbp in length), containing only three protein-coding genes (*cox1*, *cox3* and *cob*) (Feagin, 1994; Wilson & Williamson, 1997) and highly

fragmented rRNA genes.

i. Apicomplexan mtDNA structure--In *Plasmodium* the mitochondrial genome consists of tandem arrays of head-to-tail repeats (Fig. 1.4) (Feagin, 1994; Feagin, 2000; Wilson & Williamson, 1997). In most cases these mtDNAs are linear, yet a small percentage has been shown to exist as 6-kbp circles (Preiser *et al.*, 1996). The average number of mtDNA molecules per cell ranges from 150 in *P. yoelii* (Vaidya & Arasu, 1987) to 15-20 in *P. gallinaceum* (Joseph *et al.*, 1989) and *P. falciparum* (Preiser *et al.*, 1996).

The mtDNA of *T. parva* is composed of discrete linear elements (6.6 kbp), the ends of which are flanked by 189-bp terminal inverted repeats (Fig. 1.4). Each telomeric end also contains smaller palindromic and direct repetitive sequences that are contained within the 189-bp repeats (Hall *et al.*, 1990; Kairo *et al.*, 1994; Shukla & Nene, 1998).

Another apicomplexan for which sequence data exist is *Toxoplasma gondii* (Ossorio *et al.*, 1991). So far only a complete *cob* gene has been reported (McFadden *et al.*, 2000). Unfortunately, attempts to isolate mtDNA have been hampered by the presence of multiple, partial *cox1* and *cob* genes dispersed throughout the nuclear genome. In all cases, these partial genes are flanked by a 91-bp repeat. The importance of these repeats is unknown, but their presence may provide a clue as to mtDNA organization and mitochondrial gene transfer to the nucleus (Ossorio *et al.*, 1991). Finally, a 2.6-kbp portion of *Babesia bovis* mtDNA has been sequenced (GenBank accession number AF053002).

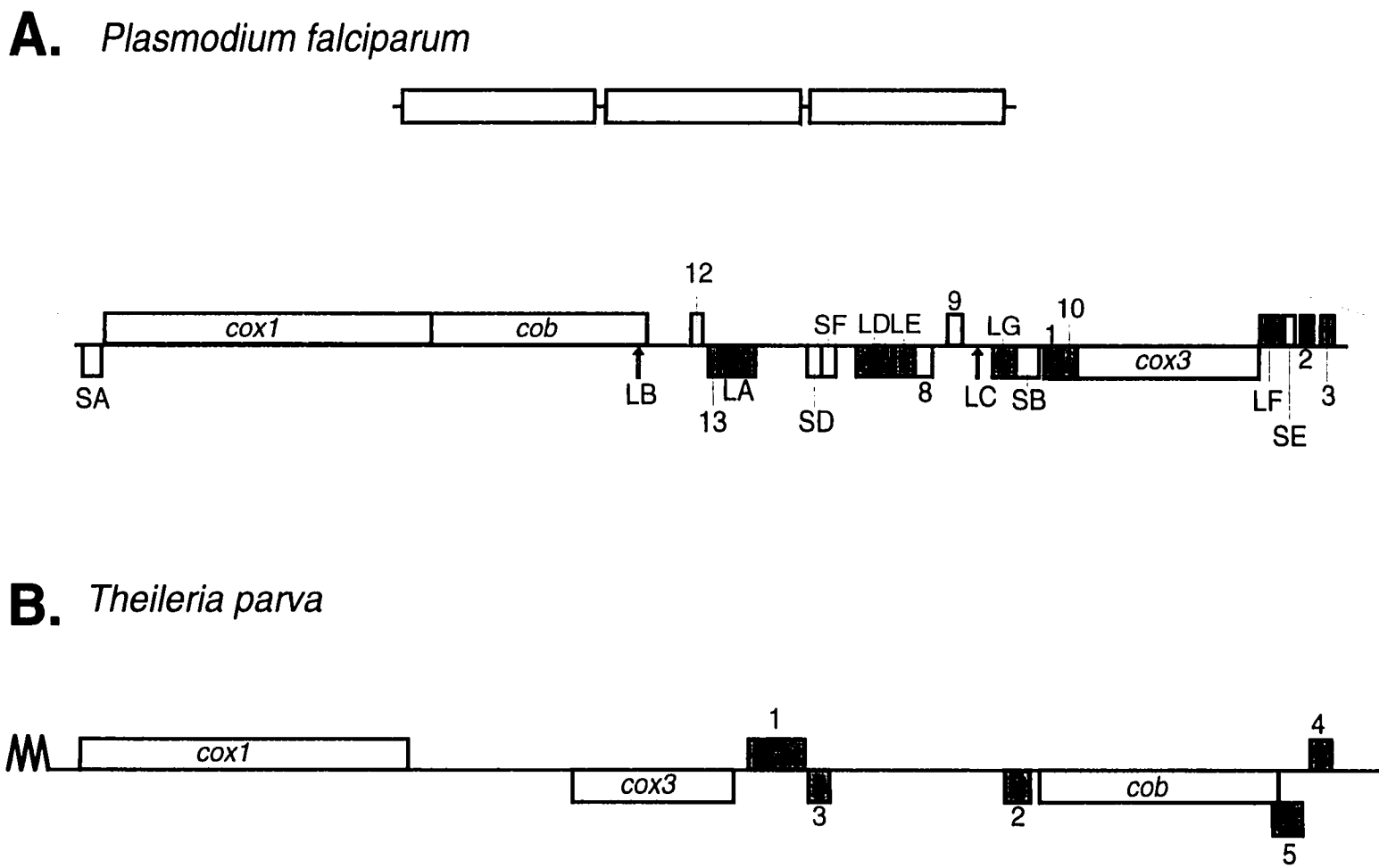


Fig. 1.4. Gene maps of (**A**) *P. falciparum* and (**B**) *T. parva* mtDNAs. Large open rectangles denote protein-coding genes, as labeled. Smaller shaded rectangles indicate LSU rRNA gene fragments. Smaller open rectangles mark SSU rRNA gene fragments. The letter 'L' denotes LSU and the letter 'S' represents SSU. Numbers indicate names assigned to rRNA genes. In *T. parva* mtDNA, terminal inverted repeats are indicated by a 'zig-zag' line. The three open rectangles above the *P. falciparum* map indicate that the mtDNA occurs as a head-to-tail tandem array. The proposed location of LSUB and LSUC are marked by shaded arrows. Modified from Wilson and Williamson (1997) and J.E. Feagin, personal communication.

ii. Mitochondrial gene order comparisons among apicomplexans--Within the genus *Plasmodium*, mtDNA divergence rates are low, as illustrated by an identical gene order in five *Plasmodium* species. Similarly, a high degree of sequence identity (90%) is maintained (McIntosh *et al.*, 1998). In contrast, as shown in Fig. 1.4, the genes in *Plasmodium* and *Theileria* mtDNAs have been extensively rearranged. In *Plasmodium*, the *cob* gene ORF is found downstream of *cox1* and is separated by a short spacer region; however, the 3' end of the *cox1* mRNA abuts the 5' end of the *cob* mRNA. In *T. parva*, *cob* occurs upstream of *cox1* and is separated by several pieces of fragmented rRNA genes. In *T. parva*, *cob* and *cox1* are encoded on opposite DNA strands, whereas in *Plasmodium* both genes are expressed from the same strand. There is also extensive rearrangement of the rRNA gene fragments.

iii. Codon usage--For the most part, the three protein-coding genes in *Plasmodium* are translated using the universal genetic code. The only deviation is the inferred use of ATA or ATT (isoleucine) as an initiation codon for *cox1* and *cox3* in *Plasmodium*. In addition, a putative start codon (methionine or isoleucine) has not been identified for the *cox1* gene in *P. yoelii* (Feagin, 1992; Feagin, 1994; McIntosh *et al.*, 1998). Nearly all *Plasmodium* genes use TAA as a stop codon, the only exception being *P. yoelii*, which also uses TAG as a mitochondrial termination codon. The TGG codon does not appear in *Plasmodium* mtDNA-encoded genes (Feagin, 1992; Feagin, 1994; McIntosh *et al.*, 1998).

Like most *Plasmodium* species, *T. parva* uses a slight modification of the genetic code. For example it appears to use AGT (serine) as an initiation codon for *cox1* (Kairo *et*

al., 1994).

iv. rRNA genes--The rRNA genes of *Plasmodium* and *Theileria* are fragmented into small pieces that are scattered throughout the genome (Feagin *et al.*, 1992; Feagin *et al.*, 1997; Gillespie *et al.*, 1999; Rehkopf *et al.*, 2000). So far 15 rRNA gene pieces have been identified in *Plasmodium*. These fragments are between 40 and 200 nt long and form stable transcripts (Feagin *et al.*, 1997; Gillespie *et al.*, 1999). Eight of these RNAs comprise part of the SSU rRNA whereas the remaining seven pieces form part of the LSU rRNA. Interestingly, RNAs for a number of functionally important rRNA regions have yet to be identified, including the GTPase center of the LSU rRNA (Feagin, 2000). However, a number of very small rRNA fragments have been indentifed in *P. falciparum* mitochondria that may correspond to the missing GTPase center (J. E. Feagin, personal communication). Besides the 15 rRNA transcripts, five other small RNAs have been isolated (Feagin *et al.*, 1997). These RNAs do not display homology to any of the missing rRNA regions; however, it is possible that they represent highly divergent and thus unrecognizable portions of rRNA that have yet to be accounted for. In *T. parva*, only five RNAs have been reported that are homologous to LSU rRNA and produce stable RNAs (Kairo *et al.*, 1994; Nene *et al.*, 1998). Ten rRNA gene pieces have also been identified as potential SSU fragments; however, only four of these have been confirmed by northern analysis (Nene *et al.*, 1998).

v. Transcription--The mtDNA of *Plasmodium* is gene-rich with few intergenic spacer regions. Genes are encoded on opposite strands and cluster together forming transcription units (Suplick *et al.*, 1990; Vaidya & Arasu, 1987). Currently very

little is known about transcription initiation or termination. In *Plasmodium*, full-length (6 kb) and gene-sized transcripts have been detected for protein-coding and fragmented rRNA genes (Aldritt *et al.*, 1989; Feagin *et al.*, 1997; Ji *et al.*, 1996). This has led researchers to speculate that *Plasmodium* mtDNA is polycistronically transcribed. Following this, individual, gene-sized rRNA and protein-coding transcripts are processed out of the larger transcripts. These findings are consistent with the view that each DNA strand contains a single promoter that gives rise to polycistronic transcripts. However, it is also possible that multiple promoters occur on each DNA strand, generating gene-specific transcripts (Ji *et al.*, 1996) as well as genome-length transcripts. In all *Plasmodium* species a 12 to 15 bp T-rich sequence motif has been identified near the transcription initiation site (McIntosh *et al.*, 1998; Rehkopf *et al.*, 2000). The location of the motif suggests that this region may be a transcription promoter element, a ribosome recognition element or a mRNA processing site (Feagin, 1994; Suplick *et al.*, 1990; Rehkopf *et al.*, 2000).

Information regarding cleavage of large polycistronic transcripts into gene-sized transcripts is largely lacking. Currently, a number of RNA structural features have been identified that may be involved in this process. One such feature is a hairpin structure that could form within the C-terminus of the *cox1* mRNA, 13 nt upstream of the putative start site for *cob*. It has been proposed that this could represent an RNA processing site that would allow cleavage into separate *cox1* and *cob* transcripts (Suplick *et al.*, 1990). Hairpins within the folded structure of rRNA may also provide additional sites for interaction with processing factors.

Within *Plasmodium* mitochondria there is an interesting pattern of RNA polyadenylation. RT-PCR data show that both mRNA (Rehkopf *et al.*, 2000) and many rRNA fragments (Gillespie *et al.*, 1999) are post-transcriptionally oligoadenylated. For rRNA, the number of residues added ranges from 0 to 21, and appears to be roughly the same for each rRNA fragment. For example, SSUA has on average 15 residues added whereas SSUD has only a few. The reason for oligoadenylation of rRNAs is unknown, but it has been suggested that this may provide a buffer against 3' exonuclease attack. As with rRNA, the number of A residues added post-transcriptionally to mRNAs varies in a gene-specific manner (Rehkopf *et al.*, 2000).

In *T. parva* the mtDNA is transcribed into low- and high-molecular-weight RNAs. Both DNA strands appear to be polycistronically transcribed (Kairo *et al.*, 1994). Promoter and transcription data have yet to be published for *T. parva*, but DNA sequence analysis shows that an LSU fragment overlaps the 5' end of the *cob* reading frame.

vi. DNA replication and recombination--Mitochondrial DNA replication has been shown to be fairly complex in *P. falciparum* and most similar to replication of bacteriophage T4 DNA (Mosig, 1987). For *P. falciparum* it has been proposed that the terminal ends of the mtDNA have 3' overhangs (Preiser *et al.*, 1996). These single-stranded overhangs invade homologous regions of other mtDNA molecules via recombination. This creates a replication fork that can be extended by using the invading strand as a 'primer' and the invaded strand as the template. A single mtDNA molecule can be attacked by more than one invading strand, thereby creating a net-like array of replicating DNA.

A small percentage of *P. falciparum* mtDNA occurs as circles. Invasion of circles by linear duplexes followed by replication could create a rolling circle-like system of DNA replication. This would also lead to the creation of a concatamer in which mtDNA molecules are joined together to form head-to-tail tandem arrays (Preiser *et al.*, 1996).

This model of DNA replication suggests that a discrete origin of replication is not required. Instead, replication would initiate at random points throughout the mtDNA. Another feature of this model is the creation of 3' overhangs as a result of the failure of DNA polymerase to replicate the 3' ends of DNA duplexes (Preiser *et al.*, 1996).

b. Apicomplexan plastid DNA. In addition to mitochondria, apicomplexans have a remnant plastid (reviewed in Blanchard & Hicks, 1999; McFadden *et al.*, 1997; Wilson & Williamson, 1997). The first apicomplexan ptDNA, a 35-kbp circular molecule, was characterized from *P. falciparum*. Since that time the complete plastid genome sequence has been determined for a second apicomplexan, *Toxoplasma gondii* (http://www.sas.upenn.edu/~jkissing/toxomap.html) (Denny *et al.*, 1998; Wilson, 1998). In addition, partial ptDNA sequence and gene order data are also available for *Eimeria tenella* and *Theileria annulata* (Denny *et al.*, 1998; Wilson, 1998).

Plasmodium ptDNA encodes ~60 genes, the majority (32) of which are tRNA genes. In addition, the 35-kbp circle contains duplicate SSU and two LSU rRNA genes, located within a large inverted repeat (Gardner *et al.*, 1993). The ptDNA also encodes three subunits of a eubacteria-like DNA-dependent RNA polymerase (encoded by the *rpo*B, *rpo*C1 and *rpo*C2 genes) (McFadden *et al.*, 1997; Wilson & Williamson, 1997). A 470-amino acid open reading frame (ORF) (ORF470) has also been identified in

Plasmodium ptDNA. This ORF is homologous to *ycf24*, a gene encoding a protein of unknown function that occurs in some bacteria as well as rhodophyte and chromophyte plastid genomes (reviewed in Wilson & Williamson, 1997). Finally, other plastid genes encode 17 ribosomal proteins (*rps* and *rpl*), a protease (*clpC*), an elongation factor (*tufA*) and a number of unidentified ORFs (Blanchard & Hicks, 1999; Wilson *et al.*, 1996).

2. Ciliates

The linear mitochondrial genomes of two ciliates, *Paramecium aurelia* (Pritchard *et al.*, 1990a; Pritchard *et al.*, 1990b) and *Tetrahymena pyriformis* (Burger *et al.*, 2000), have been completely sequenced. *T. pyriformis* mtDNA is slightly larger (at 47,172 bp) with an A+T content of 78.7% (Burger *et al.*, 2000), whereas *P. aurelia* mtDNA is only 40,469 bp in size, with an A+T content of 59% (Cummings, 1992; Pritchard *et al.*, 1990b). Genome sizes for the two ciliate mtDNAs exclude terminal telomeric repeats.

a. Structure of mtDNA in ciliates. In *T. pyriformis* mtDNA, the telomeric ends comprise tandem arrays of 31-bp identical direct repeats. The average number of repeats per mtDNA end is 31 in *T. pyriformis* (Morin & Cech, 1988a), with DNA molecules terminating within a repeat unit. Morin and Cech (1986, 1988a, 1988c) investigated the DNA sequence of telomeric ends in six *Tetrahymena* species representing 13 different strains. They found that repeat unit length varied from 31 to 54 bp, with four species exhibiting a different repeat sequence at the left end compared with the right end of the mtDNA. Intra-telomere analysis showed that, for the most part, there was little sequence variation among repeats and that most mtDNA elements contained a similar number of repeats per end (Morin & Cech, 1988a).

The majority of the mtDNA in *P. aurelia* occurs as single linear molecules (65%), with 10% of the population existing as linear dimers and a small percentage (3-6%) forming lariats (reviewed in Cummings, 1992). During DNA replication, initiation starts at one end of the linear molecule. The initiation end encompasses a high A+T region that is made up of direct tandem repeats. This repeat region is present in all *Paramecium* sp. examined to date but its sequence varies among different species. In particular, there is variation in repeat unit length and number of repeats. For example, in *Paramecium* species 4 the repeat unit length is 34 bp, with the number of repeats ranging from 4 to 16 copies. In other species the repeat unit length is between 3 and 72 bp. As found in the case of *Tetrahymena* mtDNA intra-species repeat sequences are highly conserved.

DNA replication terminates at the opposite end of the mtDNA element from where it was initiated. Attempts to recover the termination end of *Paramecium* mtDNA have been unsuccessful. Until this uncharacterized portion of the genome (~300 to 400 nt) is cloned and sequenced, the mechanism of termination of mtDNA replication in *Paramecium* will remain unknown (Cummings, 1992; Pritchard *et al.*, 1983).

b. Gene content in ciliate mtDNAs

i. Protein-coding genes--*T. pyriformis* mtDNA contains 21 protein-coding genes that have identifiable homologs in other mtDNAs. It also possesses bipartite small and large subunit rRNA genes as well as genes for seven tRNAs (Fig. 1.5). Twenty-two additional ORFs of length >60 amino acid residues have been identified; however, the function of these ORFs is unknown (Burger *et al.*, 2000). The gene contents of the *T. pyriformis* and *P. aurelia* mitochondrial genomes are very similar; in fact, of the 21

Fig. 1.5. Gene maps of *T. pyriformis* and *P aurelia* mtDNAs. Filled rectangles denote genes and ORFs. Grey rectangles indicate locations of terminal repeats. Non-homologous ORFs (*orf*) of the same size are set apart by the labels '-1' and '-2'. The *rnl* and *rns* gene fragments are noted (α and β). The *rnl* genes that are encoded next to the right (R) and left (L) telomeres in *T. pyriformis* are indicated. N- and C-terminal fragments of *nad1* are represented by the designation 'N-' and 'C-'. Modified from Burger *et al.* (2000).

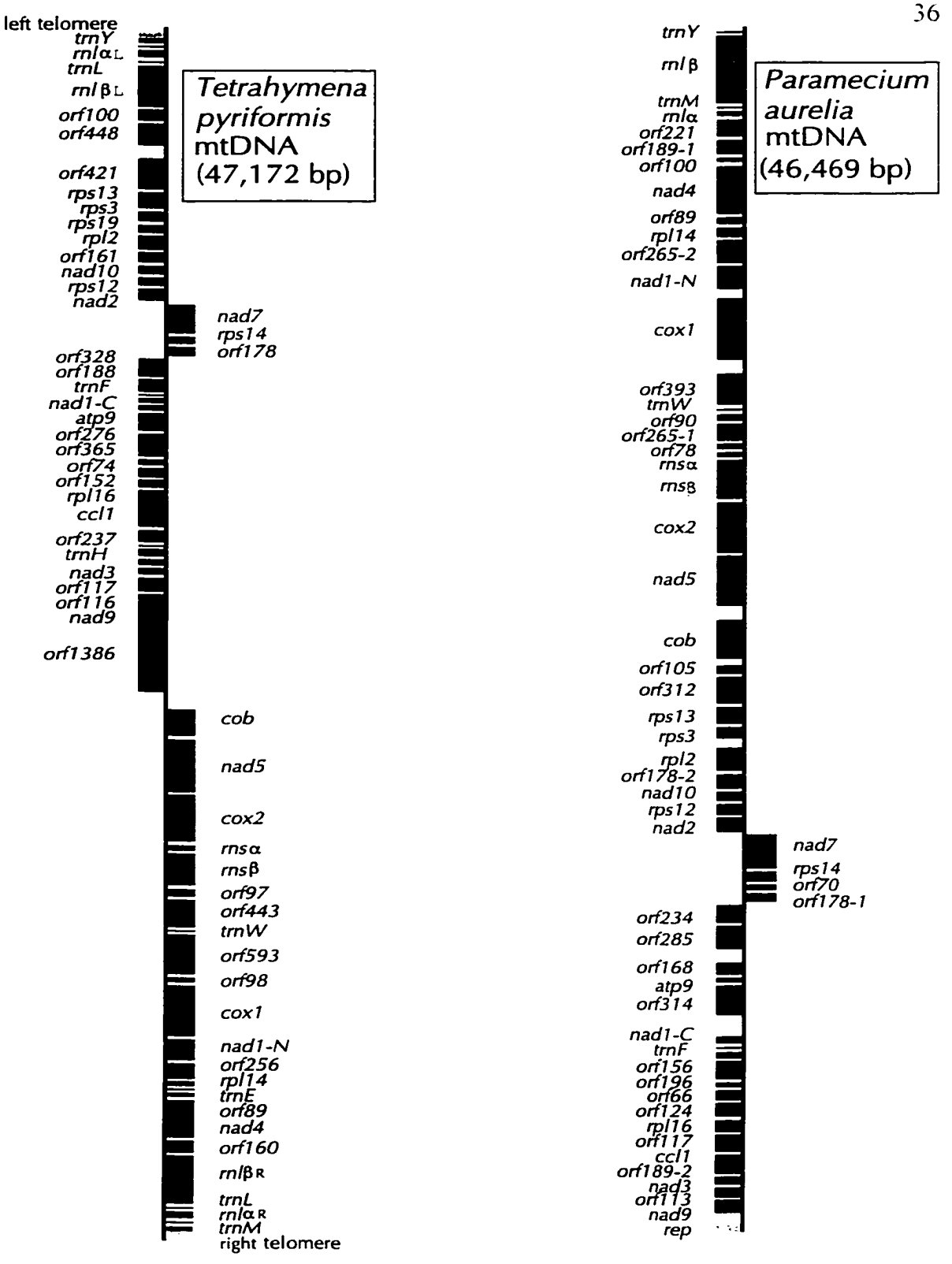

left telomere
trnY
rnlαL
trnL
rnlβL
orf100
orf448
orf421
rps13
rps3
rps19
rpl2
orf161
nad10
rps12
nad2
nad7
rps14
orf178
orf328
orf188
trnF
nad1-C
atp9
orf276
orf365
orf74
orf152
rpl16
ccl1
orf237
trnH
nad3
orf117
orf116
nad9
orf1386
cob
nad5
cox2
rnsα
rnsβ
orf97
orf443
trnW
orf593
orf98
cox1
nad1-N
orf256
rpl14
trnE
orf89
nad4
orf160
rnlβR
trnL
rnlαR
trnM
right telomere
Tetrahymena pyriformis mtDNA (47,172 bp)
trnY
rnlβ
trnM
rnlα
orf221
orf189-1
orf100
nad4
orf89
rpl14
orf265-2
nad1-N
cox1
orf393
trnW
orf90
orf265-1
orf78
rnsα
rnsβ
cox2
nad5
cob
orf105
orf312
rps13
rps3
rpl2
orf178-2
nad10
rps12
nad2
nad7
rps14
orf70
orf178-1
orf234
orf285
orf168
atp9
orf314
nad1-C
trnF
orf156
orf196
orf66
orf124
rpl16
orf117
ccl1
orf189-2
nad3
orf113
nad9
rep
Paramecium aurelia mtDNA (46,469 bp)

Figure 1.5

known protein-coding genes found in *T. pyriformis* mtDNA, all of them except *rps19* are also contained in the *P. aurelia* mitochondrial genome. The mtDNA of both ciliates also contains *yejR*, which codes for a component of the cytochrome *c* heme lyase. A gene for this protein has been described for only a few protist and plant mtDNAs (Burger *et al.*, 2000).

Comparisons of mtDNA organization and gene content between *T. pyriformis* and *P. aurelia* show that a number of distinctive changes must have occurred in the ancestor that gave rise to these two ciliates. First, both mitochondrial genomes lack *nad4L*, *nad6*, *cox3* and *atp6* genes, which are typically found in mtDNA. Second, they both possess a split *nad1* gene, whose two halves have been rearranged and are transcribed from different strands. Transcriptional analysis has shown that in *T. pyriformis* mitochondria, the two independently transcribed mRNAs are not trans-spliced, as occurs in plant mitochondria; instead each transcript appears to be translated independently (Edqvist *et al.*, 2000). Third, of the 22 unassigned ORFs found in *T. pyriformis* mtDNA, 13 have counterparts in *P. aurelia* mtDNA, making these ORFs putative ciliate-specific genes. Other similarities in mtDNA-encoded proteins between the two ciliates include large ciliate-specific insertions in COX1 (100 aa) and COX2 (200 aa); C-terminal extensions to RPS13 (*P. aurelia*, 100 aa; *T. pyriformis*, 135 aa); and N- and C-terminal deletions in NAD2 and NAD5, respectively (Burger *et al.*, 2000).

ii. rRNA genes--Like the *nad1* gene, the ciliate mitochondrial LSU (*rnl*) and SSU (*rns*) rRNA genes are also discontinuous. In *T. pyriformis*, the 5' end of the *rns* gene (*rns_a*, encoding the SSUα rRNA) is ~200 bp long and is separated from the 3'

portion of the gene (*rns_b*, encoding SSUβ) by a 54-bp A+T-rich spacer (Schnare *et al.*, 1986). The SSUβ rRNA is ~1400 nt long. *P. aurelia* also has a split *rns* gene, with the break in the corresponding SSU rRNA occurring at a position similar to that of its counterpart in *T. pyriformis*, near the 5' end (Cummings, 1992).

The mitochondrial *rnl* gene is also discontinuous in both *T. pyriformis* and *P. aurelia* (Heinonen *et al.*, 1990; Heinonen *et al.*, 1987; Seilhamer *et al.*, 1984). In *T. pyriformis*, the *rnl_a* portion encodes LSUα, which constitutes the 280-nt 5'-terminal portion of the LSU rRNA. The *rnl_b* region encodes the remainder of the LSU rRNA, LSUβ, which is 2315 nt in length. The *rnl_a* and *rnl_b* coding regions are rearranged relative to one another, with the *rnl_a* (the 5' portion) located downstream of *rnl_b* (the 3' portion) in the direction of transcription. Separating the two *rnl* pieces is a $tRNA^{leu}$ gene (Heinonen *et al.*, 1990; Heinonen *et al.*, 1987). In *P. aurelia*, the breakpoint in the mitochondrial *rnl* gene occurs at a position corresponding to that in *T. pyriformis*. This suggests that the *rnl* discontinuity was present in a common ancestor that gave rise these two lineage of ciliates; however, there has been no rearrangement of *rnl_a* relative to *rnl_b* in *P. aurelia*. Nevertheless, as in *T. pyriformis* mtDNA, the two halves of the *rnl* gene are separated by a tRNA gene (but $tRNA^{met}$ rather than $tRNA^{leu}$; Cummings, 1992; Seilhamer *et al.*, 1984).

Interestingly, *T. pyriformis* has two versions of the split *rnl* gene (both transcribed), one at each end of the linear mtDNA molecule and positioned in an inverted orientation relative to one another. The two *rnl* genes are not identical, differing at five positions, three of which localize to single-stranded regions of the LSU rRNA secondary

structure. The two differences that occur within paired regions do not disrupt RNA pairing in that substitution of A for G results in an **A**-U pair being replaced by a **G**-U pair.

The *rnl* repeat found in *T. pyriformis* mtDNA does not occur in all members of this genus. Three *Tetrahymena* species have been identified that possess the complete *rnl* gene duplication, whereas others are missing the LSUβ and $tRNA^{leu}$ duplication. In *T. pigmentosa*, a species that is missing the LSUβ duplication, the LSUα repeat is heterogeneous at four positions (Morin & Cech, 1988b), with sites of *rnl* heterogeneity evidently differing among *Tetrahymena* species (i.e., four LSUα variable sites in *T. pigmentosa* as compared to one in *T. pyriformis*). It has also been shown that heterogeneous sites in the *rnl* are not fixed, but can change over time. In *T. pyriformis*, five heterogeneous sites (one α and four β) were initially identified in the mtDNA (Heinonen *et al.*, 1990; Heinonen *et al.*, 1987). Preparations of mtDNA obtained eight years later revealed that only two *rnl* sites remained heterogeneous. One site was consistent with that found in the earlier isolation (LSUβ position 803) but a new heterogeneous site was identified at position 554 of the LSUβ fragment. Surprisingly, three of the previously heterogeneous sites (one LSUα and three LSUβ sites) had become homogeneous (Burger *et al.*, 2000). This observation suggests that variable sites within the *rnl* repeat may arise over time and that these sites can then become fixed, resulting in sequence heterogeneity; however, reversion to homogeneity is also possible.

II. Materials and Methods

A. Materials.

Restriction endonucleases were purchased from Life Technologies (Gibco BRL), New England Biolabs (NEB), Fisher Scientific and Pharmacia. Avian myeloblastosis virus (AMV) reverse transcriptase, *fmol*™ cycle sequencing system, pGEM®-T vector, proteinase K, PolyATract® mRNA Isolation System and Wizard™ *Plus* Miniprep DNA Purification System were obtained from Promega. MicroSpin™ G-24 and MicroSpin™ S-200 columns, deoxy- and dideoxynucleoside triphosphates, pUC18 *Eco*RI/BAP-prepared vector, T4 polynucleotide kinase, T4 DNA ligase and T4 DNA polymerase were purchased from Pharmacia. DNase I, exonuclease III, mung bean nuclease and the Klenow fragment of *Escherichia coli* DNA polymerase I were from Gibco BRL. Poly(A) polymerase was purchased from Amersham Life Science. *Taq* DNA polymerase was acquired from ID Labs, Life Technologies and Sigma Chemical Co. pBluescript II KS vector was from Stratagene and pT7Blue(R) vector was obtained from Novagen. Oligonucleotide primers were purchased from Dalton Chemicals, ID Labs and Life Technologies. [γ-^{32}P]ATP and [α-^{32}P]dATP were products of New England Nuclear (NEN)/Mandel.

B. Methods.

1. Culture of *C. cohnii*.

C. cohnii strain WHd was kindly provided by Dr. Carl Beam (Brooklyn College, New York, NY) and grown axenically in MLH liquid medium (Tuttle & Loeblich, 1975). Stock cultures were maintained by transferring 1.0 mL of inoculant to 5 mL of fresh

medium each month. Harvest cultures were prepared by adding 6.0 mL of inoculum (5 d old) to 1 L of MLH medium in a Fernbach flask. One-litre harvest cultures were grown in the dark on a New Brunswick Scientific rotary incubator at 27°C with mild shaking (setting of 2). Growth of *C. cohnii* was monitored by optical density (OD) readings at 600 nm. Cells were harvested during mid-log phase of the growth cycle. This typically occurred after 3-5 d of incubation (OD_{600} approximately 0.5-0.6). Alternatively, 1-L cultures were used to inoculate 15 L of MLH medium. Sixteen-litre cultures were grown in the dark at room temperature with light stirring (stir plate setting of 3) and filtered aeration. Again, cells were harvested after approximately 5 d when OD_{600} readings reached 0.5-0.7.

2. Nucleic acid purification procedures.

Mid-log phase cells were harvested by centrifugation (850x*g*, 10 m) and washed three times with 500 mL of resuspension buffer (25 mM Tris-HCl, 10 mM EDTA (pH 8.0)). Following resuspension, nucleic acids were purified using one of several different methods. All steps were carried out at 4°C except where otherwise stated.

a. Preparation of total nucleic acids by phenol extraction. Washed *C. cohnii* cells were resuspended in a mixture containing 10 mL of resuspension buffer per gram wet weight of cells. Total nucleic acids were prepared by lysing 100 mL of this slurry in a French pressure cell (Aminco) (2000 lb/in^2). The resulting lysate was combined with 0.1 vol. of 10% sarkosyl and 0.1 vol. of 1 M NaOAc. This solution was then extracted with 3 vol. of a mixture of phenol/cresol/8-hydroxyquinoline (500:70:0.5) (Parish & Kirby, 1966). Alternatively, resuspended cells were added to the detergent/phenol-cresol

solution directly and mixed on a table top shaker (high setting) at 4°C for 3 d.

Following the initial phenol extraction, total nucleic acids (in the aqueous phase) were re-extracted until the organic-aqueous boundary was clear (4-12 times). This was followed by two rounds of ethanol precipitation, an additional phenol-cresol extraction step and two further ethanol precipitations.

Ethanol precipitations were performed by adding 0.1 vol. of 3 M NaOAc and 3 vol. ice cold 95% ethanol to the nucleic acid solution. Samples were incubated at 4°C for 10 m. Alternatively, samples were placed at -20°C overnight (~16 h). Nucleic acids were collected by centrifugation (19,000x*g* at 4°C for 10 m) and the supernatant decanted. Nucleic acid-containing pellets were washed two times with 70% ethanol, dried under vacuum and resuspended in TE-8 (10 mM Tris-HCl, 0.1 mM Na_2EDTA (pH 8.0)).

b. Preparation of total DNA by guanidine/phenol extraction. Approximately 4 g (wet weight) of washed *C. colmii* cells were resuspended in 40 mL of guanidine extraction buffer (8 M guanidine-HCl, 100 mM Tris, 5 mM Na_2EDTA (pH 8.0)) and incubated at 45°C for 30 m. Following this step, cells were disrupted by shaking with glass beads (Lang *et al.*, 1977) and the resulting lysate was diluted 9-fold using 100 mM Tris-HCl (pH 8.0), 5 mM Na_2EDTA (pH 8.0). Eight millilitres of Triton X-100, 6.5 mg of proteinase-K and RNase A (see below) were added to the diluted lysate, after which the mixture was incubated at 60°C for 1 h. Unsolubilized material was removed by centrifugation (19,000x*g* for 10 m at 4°C) and the recovered supernatant was loaded onto a 500/G Genomic tip column (Qiagen). DNA was bound to the column, washed and eluted following the manufacturer's directions. Briefly, the column was equilibrated with

10 mL of QBT buffer (750 mM NaCl, 50 mM MOPS (pH 7.0), 15% isopropanol, 0.15% Triton X-100). The cell lysate supernatant was applied to the column and washed two times with 15 mL of QC buffer (1.0 mM NaCl, 50 mM MOPS (pH 7.0), 15% isopropanol). DNA was eluted in 15 mL of 50°C QF buffer (1.25 mM NaCl, 50 mM Tris-HCl (pH 8.5), 15% isopropanol). Following elution, 100 μg of proteinase-K was added to the sample, the mixture was dialysed against TE-8 for 12 h and concentrated in the dialysis bag by incubating at 37°C overnight. The concentrated DNA solution was recovered from the dialysis bag and stored at 4°C. Alternatively, DNA was precipitated with ethanol and resuspended in TE-8.

c. Preparation of mitochondrial nucleic acids. Nucleic acids were isolated from mitochondria that were purified by differential centrifugation of a whole-cell lysate as previously described (Spencer *et al.*, 1992). After harvesting and washing, 4 g (wet weight) of *C. colnii* cells were resuspended in 40 mL of buffer A (0.44 M sucrose, 50 mM Tris-HCl, 3 mM Na_2EDTA (pH 8.0), 1 mM β-mercaptoethanol, 0.1% BSA) and disrupted by shaking with glass beads (Lang *et al.*, 1977). Nuclei and unbroken cells were removed from the cell lysate by centrifugation (2,000x*g*, 10 m at 4°C; low-speed spin, LSS). Following this, the LSS supernatant was centrifuged at 19,000x*g* for 20 m at 4°C (high-speed spin, HSS). The pelleted material contained mitochondria. Crude mitochondrial pellets were resuspended in 50 mL of buffer A using a hand-held tissue homogenizer. The LSS and HSS steps were repeated. Following the second HSS, mitochondria were resuspended in 10 mL of buffer A and loaded onto a two-step discontinuous gradient comprising a 1.55 M sucrose cushion (1.55 M sucrose, 50 mM

Tris-HCl, 3 mM Na_2EDTA (pH 8.0), 1 mM β-mercaptoethanol, 0.1% BSA) overlaid with a 1.15 M sucrose solution (1.15 M sucrose, 50 mM Tris-HCl, 3 mM Na_2EDTA (pH 8.0), 1 mM β-mercaptoethanol, 0.1% BSA). Gradients were centrifuged for 1 h at 22,500 rpm in an SW25.1 swinging-bucket rotor (Beckman Instruments Inc.). Following sucrose gradient centrifugation, purified mitochondria were found in the pellet and slowly resuspended in 3 vol. of buffer C (50 mM Tris-HCl, 20 mM Na_2EDTA (pH 8.0)). Purified mitochondria were pelleted by centrifuging (19,000x*g* for 20 m) and resuspended in 9 mL of buffer D (10 mM Tris-HCl (pH 8.5), 50 mM KCl, 10 mM $MgCl_2$) using a tissue homogenizer. Mitochondrial membranes were disrupted by the addition of 2 mL of a 20% Triton X-100 solution followed by vigorous vortexing for 3 m. Disrupted membranes were removed by centrifugation and nucleic acids present in the supernatant were purified by phenol extraction and ethanol precipitation. Because the amounts of nucleic acids obtained using this procedure were low, RNA and DNA were often resolved from one another prior to nuclease treatment (see below).

d. DNA preparation.

i. RNase A treatment--Prior to treatment, RNase A was prepared by boiling for 5 m to inactivate contaminating DNases. Total and mitochondrial-enriched nucleic acids were treated with 1 µL of RNase A (10 µg/µL) per 100 µg of total nucleic acid. Hydrolysis was performed in TE-8 buffer at 37°C. Following incubation, DNA samples were extracted with phenol, precipitated with ethanol and resuspended in TE-8.

ii. Precipitation with polyethyleneglycol (PEG)--After the removal of salt-insoluble RNAs (see below), nucleic acids were incubated in the presence of 10%

PEG and 1.0 M NaCl overnight at 4°C, and then centrifuged at 12,000x*g* for 20 m. The supernatant was removed and stored at -20°C. The DNA-containing pellet was washed two times with 70% ethanol and resuspended in TE-8. The sample was then treated with RNase A followed by two rounds of phenol extraction. Finally, the DNA was resuspended in TE-8 and stored at 4°C.

iii. Isolation of DNA by CsCl density gradient centrifugation--To separate mtDNA from nDNA, total DNA was subjected to CsCl density gradient centrifugation. Between 2-3 mg of total DNA was dissolved in 10 mL of TE-8 buffer and added to 11 g of CsCl. Next, 2 mg of Hoechst dye (100 µL of a 20 mg/mL solution) was added to the DNA solution and mixed by inversion several times. Samples were placed in a quick-seal tube and centrifuged (40,000 rpm, 44 h at 25°C) in a Ti50 fixed angle rotor using a Beckman ultracentrifuge. DNA was visualized using a UV lamp and regions containing bands were collected using an 18-gauge needle connected to a 5 mL syringe. The Hoechst dye was removed from the DNA by repeated washing in water-saturated *n*-butanol until all of the dye was removed. Following this, DNA was precipitated in 3 vol. of absolute ethanol at room temperature, washed two times with 70% ethanol and resuspended in TE-8.

Alternatively, following centrifugation, CsCl gradients were fractionated into 200 µL aliquots using a Beckman gradient collection setup. Starting from the bottom of the gradient, 200 µL aliquots were individually collected into 1.5 mL microcentrifuge tubes. Individual samples were washed with *n*-butanol, precipitated with ethanol, washed with 70% ethanol and resuspended in TE-8 as previously described.

e. RNA preparation. RNA was prepared using two different protocols. For samples with large amounts of total nucleic acid, DNA was hydrolyzed using DNase I. For samples containing low amounts of nucleic acids (i.e., following subcellular fractionation), RNAs were typically isolated using salt fractionation before DNase I treatment.

i. DNase I treatment--Total and mitochondrial RNA fractions were prepared by incubating nucleic acids with RNase-free DNase I (Pharmacia) at 37°C for 30 m in a 50 μL reaction mixture containing 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$ and DNase I (1 unit per μg DNA). Samples were then extracted with phenol, precipitated with ethanol and the recovered RNA was stored as a 50% ethanol solution at -20°C.

ii. Salt fractionation--Total nucleic acid mixtures (>1 μg/μL) were incubated in the presence of 1.0 M NaCl at 4°C overnight (16 h) and then centrifuged (12,000x*g* for 20 m). The first supernatant was removed and kept at 4°C. The pellet, which mostly contained salt-insoluble RNA (iRNA), was washed two times with 70% ethanol and resuspended in dH_2O (distilled H_2O). The stored supernatant was subjected to PEG precipitation to remove DNA (see above). The resulting supernatant from the PEG precipitation procedure was then mixed with 100% ethanol, incubated at 4°C overnight and centrifuged at 12,000x*g* for 20 m. The final pellet, which contained the salt-soluble RNA fraction (sRNA), was washed with 70% ethanol and resuspended in dH_2O. Both the iRNA and sRNA fractions were subjected to DNase I treatment followed by two rounds of phenol extraction. The final product was stored at -70°C in a 50% ethanol solution.

3. PCR amplification.

a. Standard PCR. PCR amplifications were carried out in 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. Alternatively, 10 ng of mtDNA was used. PCR reaction mixtures were kept on ice until the thermocycler sample block reached 94°C, after which the cycling program was briefly interrupted until the PCR tubes were placed in the block (hot start). This method was used to reduce the chance of aberrant annealing and primer extension. Amplifications were carried out using a 2 m denaturation period (94°C) followed by 30 cycles of denaturation for 30 s at 94°C, 30 s annealing at 50°C and 30 s extension at 74°C. Following amplification, PCR products were stored at 4°C.

Initially, a 656-nt internal region of the *cox1* gene was amplified using forward (cox240) and reverse (cox448) primers (Table 2.1). These oligonucleotide primers, whose design is based on conserved regions of COX1 sequence (LFWFFGHPEV and MPRRIPDYPD, respectively), were a gift from Dr. B.F. Lang (Université de Montréal, Montréal, QC).

PCR products were visualized using agarose gel electrophoresis. Five to ten microlitres of PCR reaction were added to non-denaturing agarose gel loading buffer (6X stock: 15% Ficoll 400, 100 mM EDTA (pH 8.0), 0.1% xylene cyanol, 0.1% bromophenol blue) and electrophoresed at 1 V/cm in a 1.0% agarose gel containing 1X TAE (40 mM Tris-OAc, 1 mM Na_2EDTA (pH 8.0)) for 16 h. Alternatively, PCR products were

Table 2.1. The sequence of gene-specific as well as gene-flanking oligonucleotides.

cox1	
cox240	5'-TTATTTTGRTTTTTTGGTCATCCTGARGT-3'
cox442	5'-GCATTTCTTAGGATTTAATGTTATG-3'
cox448	5'-TCTGGGTAGTCTGGTATTCKTCKTGGCA-3'
P37	5'-ATTAATAGCTGGACATGTAG-3'
P6	5'-TGTGGAGCTATAAACCATAAATC-3'
cox3	
cox3 265	5'-TATTGGCATTTTCTTGAA-3'
cox3 270	5'-AAGAGCTATAAAAGACTAG-3'
cob	
cob296	5'-CTTCTAATGAATTATCTG-3'
P25	5'-AAGGATTTGGTTTCTTGATG-3'
P49	5'-TGGTCGTATATTAACATTAC-3'
P50	5'-CTGCCAGAGAATTATTGGTTAAC-3'
P51	5'-CTATCTAAATCCTATAAACAATG-3'
LSUG	
LSUG1	5'-TTAACCCAGCTCACGGATCG-3'
LSUG3	5'-GGTTAGAAACTGTCGCTG-3'
LSUG4	5'-AGAAGATTCCATTGGAAG-3'
LSUE	
LSUE2	5'-TTCATGCAGGACGGARMTTACCC-3'
LSUE3	5'-GGTAGCAAAATTCCTTGTCGGG-3'
LSUE4	5'-TAACGGTCCTAAGGTAG-3'

resolved in 1.0 % agarose gels containing 1X TBE (45 mM Tris, 45 mM boric acid, 1 mM EDTA (pH 8.0)). TBE gels were electrophoresed at 15 V/cm (1 h).

Resolved gels were stained in a bath containing ethidium bromide (100 μg/mL) for 30 m and washed in dH_2O for 2 m. PCR products were visualized using UV light (at 360 nm). Estimates of product sizes were determined by comparison with λ DNA *Hin*dIII size standards. Finally gels were photographed under UV light using an MP4 camera (Kodak).

b. Radiolabeling during PCR. Alternatively, PCR amplification reactions were carried out using ^{32}P-labeled dATP. PCR reactions were performed using the same protocols as outlined above with the following exceptions. First, dTTP, dCTP and dGTP concentrations were reduced to 0.02 mM. Second, dATP was replaced by 50 μCi [α-^{32}P]dATP. Third, PCR amplifications were performed in 20 μL reaction mixtures. Finally, PCR products were resolved in an 8% polyacrylamide gel, eluted from the gel matrix and used in hybridization experiments (see below).

4. Restriction enzyme hydrolysis.

Total DNA or fractions enriched in mtDNA were generally hydrolyzed with either *Eco*RI, *Eco*RV, *Sst*I or *Xba*I following manufacturer's directions. Following hydrolysis, agarose gel loading buffer was added to the DNA samples followed by incubation at 65°C for 10 m. Restriction products were electrophoresed in a 1.0% agarose gel containing 1X TAE.

5. Purification of DNA following gel electrophoresis.

Restriction fragments and/or PCR products were resolved in a 0.7% agarose gel.

Product bands were visualized under UV light and excised using a fresh razor blade. DNA was eluted from the agarose gel matrix by binding to silica powder. Gel slices were dissolved in 500 µL of 6.0 M NaI at 42°C for 10 m by gentle mixing. Once a gel slice had dissolved, 5 µL of a silica-milk slurry (50:50 mix of silica and celite particles suspended in dH_2O) was added to the mixture and incubated on ice for 10 m with occasional mixing. Following this, the silica-milk/DNA aggregate was collected by centrifugation (5 s at 14,000x*g*) and washed three times using 250 µL of 4°C NEET wash solution (50 mM NaCl, 1 mM Na_2EDTA, 50% ethanol, 10 mM Tris-HCl (pH 7.6)). DNA was eluted from the silica-milk by the addition of 20 µL of TE-8 buffer followed by incubation at 65°C for 5 m. The mixture was centrifuged at 14,000x*g* for 2 m, and the DNA-containing supernatant was collected with care taken not to disrupt the silica-milk pellet. The DNA elution step was repeated a second time and the two supernatants were pooled. Samples were stored at 4°C.

6. Cloning of *C. cohnii* mtDNA.

a. Preparation of competent cells. Chemically competent cells were prepared following the protocol of Hanahan (1983) as outlined in Sambrook *et al.* (1989). All steps in this procedure were performed using aseptic technique. A freezer stock of *E. coli* DH5α (kindly provided by D. F. Spencer) was used to inoculate a 5 mL culture tube containing SOB medium (2% Bacto-tryptone, 0.5% yeast extract, 0.05% NaCl, 2.5 mM KCl (pH 7.0)), which was then sterilized by autoclaving for 30 m. Sterile 2 M $MgCl_2$ was added to give a final concentration of 100 mM and the inoculum was grown with moderate agitation using a New Brunswick Scientific Rollodrum at 27°C (setting of 5).

The liquid culture was streaked onto an LB-agar plate (Luria-Bertani: 1% bactotryptone, 0.5% yeast extract, 0.5% NaCl, 1.5% agar (pH 7.0), sterilized by autoclaving for 30 min), which was incubated overnight (~16 h) at 37°C. The following day, a single *E. coli* colony from the plate was inoculated into 500 mL of SOB medium containing 20 mM $MgSO_4$. The culture was grown at 37°C with vigorous shaking for ~4-5 h, until an OD_{600} of 0.5 was reached. Next, the cells were aliquoted into sterile 50 mL polypropylene tubes (Corning), incubated on ice for 10 m and centrifuged (3,500x*g*, 4°C for 10 m). The supernatant was removed and the cells were suspended in 4°C FSB (Frozen Storage Buffer: 10 mM KOAc (pH 7.5), 45 mM $MnCl_2$, 10 mM $CaCl_2$, 100 mM KCl, 3 mM hexamminecobalt chloride, 10% glycerol, adjusted to pH 6.4 with HCl and sterilized by passage through a 0.2 μ filter attached to a 50 mL syringe), incubated on ice for 10 m and collected by centrifugation. The cells from each 50 mL tube were resuspended in 4 mL of cold FSB. Once the cells were resuspended, 140 μL of DMSO was slowly added to the suspension. This was accompanied by gentle mixing, followed by incubation on ice for 15 m. Next, an additional 140 μL DMSO was added to the cell suspension, the mixture was then distributed into 200 μL aliquots, snap frozen in a -70°C methanol bath and stored at -70°C.

b. Ligation of DNA into vector.

i. Ligation of PCR products into T-tailed vector--After gel purification, PCR products were ligated into pT7Blue-T vector (Novagen) or pGEM®-T vector (Promega). DNA ligations were carried out in 10 μL reaction mixtures containing DNA ligase reaction buffer (50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 2 mM ATP, 1 mM

DTT, 5% polyethylene glycol-8000), ~100 ng of PCR product, 50 ng of T-tailed vector (pT7Blue or pGEM) and 4 units T4 DNA ligase (Gibco BRL). Reaction mixtures were incubated at 14°C for at least 4 h and usually overnight. The T4 DNA ligase was inactivated at 65°C for 5 m, then the mix was cooled on ice for 5 m and stored at -20°C.

ii. Ligation of restriction enzyme products into vector--Two plasmid vectors were used in the cloning of restriction enzyme fragments. These included pBluescript II KS+ (pBKS, Stratagene) and pUC18 (Pharmacia). Briefly, pBKS plasmid DNA was amplified in *E. coli* and extracted (see below). Next, 2 µg of vector was hydrolyzed with a restriction enzyme that would produce compatible ends for ligation of the foreign DNA into the vector. Hydrolyzed plasmid DNA was treated with 0.1 unit BAP in 50 mM Tris-HCl (pH 8.0) for 30 m at 60°C. This was followed by standard gel isolation, phenol extraction and DNA precipitation protocols. Finally, treated vector DNA was resuspended in TE-8 and stored at -20°C. For the most part, pBKS vector was used in plasmid ligation. Alternatively, commercially prepared vector pUC18 plasmid DNA (hydrolyzed with *Eco*RI and dephosphorylated) was used.

Ligation reactions were performed using 50 ng of vector and ~100-300 ng of insert DNA. Reactions were carried out using standard ligation conditions (see above).

c. Transformation. Aliquots of competent cells were removed from -70°C storage just prior to transformation and allowed to thaw at room temperature. One microlitre of ligation mixture was gently added to 50 µL of competent cells and placed on ice for 30 m. Following this, the DNA/competent cell mixture was heat shocked for 30 s in a 42°C water bath, placed on ice for 2 m and added to 200 µL of SOC liquid medium

(SOC is prepared in the same way as SOB except that it also contains 20 mM glucose). Next, cells were incubated for 30 m at 37°C. Cells were then collected by centrifugation (14,000x*g* for 1 m), resuspended in 60 µL of fresh SOC, spread onto LB-agar plates containing ampicillin at 100 µg/mL and X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) at 80 µg/mL and incubated at 37°C overnight.

d. Alpha complementation and antibiotic resistance. All vectors used in this study encode the amino-terminal portion of the *E. coli* β-galactosidase gene (*lacZ*). Nestled within this coding region is a polycloning site that does not disrupt the reading frame. *E. coli* DH-5α cells are missing the amino-terminal portion of the protein but express the carboxy-terminal portion of *lacZ* (Sambrook *et al.*, 1989). Therefore, in *E. coli* host cells that contain a plasmid, the two halves of the β-galactosidase protein are present and can come together to produce a functional protein (α-complementation). Colonies that harbour a plasmid with an uninterrupted *lacZ* reading frame turn blue in the presence of chromogenic substances such as X-gal (Horwitz *et al.*, 1964). However, if foreign DNA is inserted into the polycloning site the β-galactosidase gene is typically disrupted. In such cases a functional amino-terminal β-galactosidase protein cannot be formed and α-complementation does not occur. Colonies produced in this manner will not turn blue in the presence of X-Gal. Therefore, insert-containing constructs can be identified based on blue/white selection.

Another selectable marker contained within the vector is the ampicillin resistance gene (*amp*^{r}). *E. coli* cells that are cultured in the presence of ampicillin will not reproduce unless they harbour a plasmid. Therefore, only transformed *E. coli* cells will grow.

e. Alkaline plasmid miniprep. Positive colonies were inoculated into 15 mL culture tubes containing 3 mL of LB-Amp (100 μg/mL ampicillin) liquid medium and grown overnight at 37°C in a rollodrum. Minipreps were performed using the Wizard™ *Plus* Miniprep DNA Purification System. Cells were pelleted in a swinging-bucket centrifuge (Hettich) at 3,000x*g* for 5 m. The supernatant was decanted and the cells were resuspended in 200 μL cell resuspension solution (50 mM Tris-HCl (pH 7.5), 10 mM EDTA, 100 μg/mL RNase A). Next, 200 μL of cell lysis solution (200 mM NaOH, 1% SDS) was added to tubes and the *E. coli* cells were disrupted by gentle vortexing. Following lysis, the cell suspension was neutralized by the addition of 200 μL neutralization solution (1.32 M KOAc (pH 4.8)). The mixture was then centrifuged (10 m at 3,000x*g*), the supernatant removed and placed in a 1.5 mL microcentrifuge tube. One millilitre of DNA purification resin was added to the supernatant and the mixture inverted several times. The suspension was loaded into the barrel of a 3 mL syringe that had a Wizard minicolumn attached to its Luer-Lok. The solution was passed through the minicolumn and plasmid DNA, which was bound to the column, was washed using 2 mL of column wash solution (80 mM KOAc, 8.3 mM Tris-HCl (pH 7.5), 40 μM EDTA, 55% ethanol). The minicolumn was then inserted into a 1.5 mL micro-centrifuge tube and centrifuged (14,000x*g* for 30 s). Next, the column was inserted into a new micro-centrifuge tube and the DNA was eluted by the addition of 50 μL of 65°C TE-8. Dissolved DNA was extracted from the column by centrifugation (14,000x*g* for 2 m), after which recombinant plasmids were hydrolyzed using appropriate restriction enzymes and visualized on TAE agarose gels. Plasmid DNA was stored at -20°C or 4°C.

f. Sub-cloning.

i. Nested deletion series--The insert contained within clone pc1#15 is 5.1 kbp in length. This construct was generated by inserting an *Eco*RI restriction fragment into pBKS vector. To manually sequence this insert, a series of overlapping deletion clones was generated by digestion with exonuclease III/mung bean nuclease (Henikoff, 1984). Briefly, 10 μg of construct DNA was hydrolyzed with 50 units of *Hin*dIII and *Kpn*I in React 4 buffer (Gibco BRL; 20 mM Tris-HCl (pH 7.4), 5 mM $MgCl_2$, 50 mM KCl) at 37°C for 4 h. *Hin*dIII produces a 3' overhang that can be further hydrolyzed by exonuclease III (exoIII), whereas *Kpn*I produces a 5' overhang that is exoIII resistant. The *Kpn*I site is located within the multiple cloning site and thereby protects the vector from exoIII attack. Hydrolysis products were isolated after gel electrophoresis, purified by phenol extraction and resuspended in 10 μL of TE-8. Approximately 6 μg of the hydrolyzed construct were added to a 160 μL reaction mixture containing 500 mM Tris-HCl (pH 8.0), 100 mM $MgCl_2$ and 100 units exoIII (Gibco BRL) and incubated at 37°C. Every 30 s (for a total of 20 m) a 4 μL aliquot of reaction mixture was removed and added to a separate microcentrifuge tube containing 1 μL of 10X mung bean nuclease buffer (500 mM NaCl, 100 mM NaOAc (pH 5.0), 1 mM $Zn(OAc)_2$, 1 mM L-cysteine, 50% glycerol (v/v)), 1 μL mung bean nuclease (20 units) and 4 μL of dH_2O on ice. Once all exoIII-treated samples had been collected, the samples were placed at 37°C for 10 m so that mung bean nuclease could hydrolyze the 3' single-stranded ends. The nuclease was inactivated by the addition of 1 μL of 500 mM Tris followed by heating to 70°C for 5 m and cooling on ice. The degree of DNA hydrolysis was evaluated by gel electrophoresis.

The remainder of each sample (5 μL, ~100 ng DNA) was added to a ligation reaction mixture. This allowed the free end of the insert to be rejoined to the vector. Constructs were transformed into *E. coli* and plasmid DNA isolated.

The DNA sequence of nested deletion clones prepared for pc1#15 and pc1#32 (~ 5.6 kbp in size) was determined. Gaps in the sequence were filled in by primer walking or sub-cloning using restriction enzyme-generated products.

ii. Restriction fragment sub-cloning--Restriction maps were determined for all recombinant clones. Large inserts (>1.0 kbp) were often sub-cloned by restriction enzyme hydrolysis, followed by fragment separation using gel electrophoresis, ligation into an appropriate vector and transformation.

7. DNA sequencing.

Clones were sequenced on both strands using the *fmol* DNA sequencing system (Promega). Oligonucleotides used in DNA sequencing were 5' end-labeled with ^{32}P. This was performed by incubating 10 pmol of primer with 500 μCi [γ-^{32}P]ATP and 7.5 units T4 polynucleotide kinase in a 10 μL reaction mixture containing 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine for 30 m at 37°C. T4 polynucleotide kinase was heat inactivated by placing the mixture at 90°C for 2 m, then stored at -20°C for up to 2 weeks. Vector-based oligonucleotides used in cycle sequencing included KS (5'-CTCGAGGTCGACGGTATCG-3'), SK (5'-GCTCTAGAACTAGTGGATC-3'), T3 (5'-ATTAACCCTCACTAAAG-3') and T7 (5'-AATACGACTCACTATAG-3') for pBKS; M13 forward and M13 reverse primers for pUC18; T7 and M13 universal primers for pT7 blue; and T7 or M13 forward primers for pGEM. Additional sequencing primers

were based on insert DNA sequence.

Cycle sequencing reactions were set up by adding 2 µL of one of the four nucleotide mixes (d/ddGTP, d/ddATP, d/ddTTP, d/ddCTP) to separate 0.2 mL thin-wall PCR tubes. Additional reagents were added to each 0.2 mL tube giving a final concentration of 0.4 pmol of labeled oligonucleotide, 1 unit sequencing-grade *Taq* DNA polymerase, 0.2 µg template DNA, 50 mM Tris-HCl (pH 9.0) and 2 mM $MgCl_2$. Samples were amplified in a thermocycler using previously outlined cycling parameters. Following amplification, 4 µL of stop solution (10 mM NaOH, 95% formamide, 0.05 % bromophenol blue and 0.05% xylene cyanol) was added to each mixture. Samples were stored at -20°C for up to 2 weeks, but were usually used the following day.

Radiolabeled DNA fragments were resolved in a 6% polyacrylamide denaturing gel (19 polyacrylamide: 1 *bis*acrylamide, 8.3 M urea, 90 mM TBE) attached to a 100S gel stand (Gibco BRL). Two microlitres of each sequencing reaction were loaded onto the gel and electrophoresis was carried out at constant power (35 W) for 1.5, 5 or 8 h. Gels were fixed in gel-fixing solution (10% methanol, 10% HOAc) for 30 m, blotted onto Whatman 3MM filter paper and dried under vacuum at 80°C for 1 h in a Hoefer gel drier. DNA ladders were visualized by autoradiography using XK-1 X-ray film (Kodak).

Alternatively, the DNA sequence of recombinant clones was determined using automated sequencing as performed at the Institute for Marine Biosciences, National Research Council, Halifax, Nova Scotia. Vector-based primer sequencing (M13 universal or M13 reverse) was performed using a LiCor 4200L automated sequencer. DNA sequence was also determined using insert-specific primers. In these cases sequencing

was performed with an ABI 373A automated sequencer

Automated sequencer traces were proof-read using the Sequencher ver. 3.1 (Gene Codes) program. Manual sequence data were also compared to automated sequence data using Sequencher.

8. Southern hybridization.

a. Transfer of DNA to nylon membranes. DNA samples were hydrolyzed with the appropriate restriction enzyme and the products resolved in a 1.0% agarose gel containing 1X TAE for 16 h. Resolved fragments were transferred to a Biotrans^{TM} (+) (ICN Biomedicals) nylon membrane using conventional alkaline transfer protocols (Ausubel *et al.*, 1987; Sambrook *et al.*, 1989). Briefly, DNA was depurinated by soaking gels in 0.25 M HCl for 5 m followed by rinsing with dH_20. Gels were positioned onto a kraft paper capillary stack that was pre-soaked in 0.4 mM NaOH. The gel was overlaid with nylon membrane followed by a stack of dry kraft paper. DNA was allowed to transfer to the membrane by capillary action for 4 h. The membrane was then neutralized in 2 X SSPE, baked at 65°C for several hours and stored at room temperature.

b. Colony lifts and screening. Transformed *E. coli* cells were spread onto large (15 cm) LB-agar plates containing ampicillin and X-Gal. The plates were incubated at 37°C overnight and stored at 4°C for at least 3 d prior to transfer. Positively charged nylon circles (Magna Lift; Fisher Scientific) were placed onto cold agar plates, the circles and agar plates were marked with an 18-gauge needle, and the membranes were removed, leaving a replica of the agar plate on the nylon disk. Plates were stored at 4°C. Membranes were air dried and placed (colony side up) onto a stack of Whatman 3MM

filter paper that had been pre-soaked in a solution of 0.5 M NaOH (Sambrook *et al.*, 1989). After 5 m, membranes were transferred to a second filter paper stack containing 0.5 M Tris-HCl (pH 7.4) and allowed to stand for 5 m. Next, the membranes were placed onto a final filter paper stack that contained 0.5 M Tris-HCl (pH 7.4)/1.5 M NaCl. After 5 m the nylon circles were air dried, sandwiched between dry kraft blotting paper and baked at 65°C overnight.

Following Southern hybridization and autoradiography (see below), positive constructs were identified by comparing hybridization signals to colonies located on colony plates (primary colony screening). Primary colonies were isolated and inoculated onto fresh LB-amp plates (secondary). Secondary colonies were subjected to an additional round of colony screening and colonies that gave positive signals in both cases were sequenced.

c. Preparation of randomly-labeled probe. All PCR and restriction fragment-based probes were radiolabeled using random hexamer extension (Ausubel *et al.*, 1987; Feinberg & Vogelstein, 1983). Template DNA (50 ng) was combined with 5 µg of random hexamer oligonucleotide (NNNNNN), denatured by boiling for 2 m and snap-cooled on ice. Randomly primed DNA was added to the labeling mix (50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT, 40 µg gelatin, 6 units Klenow fragment DNA Polymerase I (BRL), 50 mM dGTP, 50 mM dTTP, 50 mM dCTP and 50 µCi [α-^{32}P]dATP) and incubated at 22°C for at least 4 h. The labeling reaction was stopped by boiling for 2 m and snap-cooled on ice.

Probes were freed of unincorporated nucleotides by chromatography. Sephacryl

S-200 HR columns (Pharmacia) were prepared by centrifugation (400x*g* for 1 m). The volume of the probe-containing mixture was brought up to 50 µL by the addition of water and the mixture loaded onto the column, which was then centrifuged for 2 m at 400x*g*. The elute was boiled for 2 m, cooled on ice and added to the pre-hybridization solution.

Pre-hybridization and hybridization were carried out at 42°C in a rotary hybridization oven (Hybaid) using a rotation setting of 7. Nylon membranes were wetted with pre-hybridization solution (5X SSPE = 180 mM NaCl, 10 mM NaH_2PO_4, 1 mM EDTA (pH 7.7); 50% formamide; and 1X BLOTTO = 5% skim milk powder, 10% SDS) (Reed & Mann, 1985), sandwiched between two layers of nylon mesh, and placed into a 100 mL hybridization bottle, carefully avoiding the formation of air bubbles. Membranes were incubated in 30 mL of pre-hybridization solution for 1 h, the pre-hybridization solution was decanted, and 20 mL of fresh pre-hybridization solution were added. Labeled probe was mixed with 1 mL of pre-hybridization solution and added to the hybridization bottle. Annealing was allowed to continue overnight (~16 h). Following hybridization, membranes were rinsed in 2X SSPE and 0.1% SDS, then washed in a solution containing 0.1% SDS and 0.5X SSPE followed by a second wash in 0.1X SSPE and 0.1% SDS. Both washes were done at 25°C for 15 m. Finally, membranes were washed in 0.1X SSPE and 1.0% SDS for 30 m at 50°C before being subjected to autoradiography.

9. Northern hybridization analysis.

a. Northern analysis of mRNAs. Northern hybridization filters were prepared by resuspending 1-10 µg of RNA in 15 µL of gel loading buffer (50% deionized formamide,

1X MOPS buffer, 50 mM NaOAc (pH 5.2), 10 mM EDTA, 7% formaldehyde and 10 μg of ethidium bromide), heating to 55°C for 15 m and cooling on ice. Samples were electrophoresed for 3 h at 80 V in a 1.2% agarose gel containing 3% formaldehyde and 1X MOPS buffer. 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 randomly labeled DNA probes and Southern hybridization protocols as previously outlined. After autoradiography, sizes of RNA transcripts were estimated by comparison with RNA ladder size standards (Gibco BRL).

b. Northern analysis of rRNA. Five to 10 micrograms of DNase I-treated RNA were lyophilized, resuspended in NMF/urea loading buffer (60% v/v *N*-methylformamide, 5 M urea, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0), 0.1 % bromophenol blue and 0.1% xylene cyanol), heated to 65°C for 5 m and cooled on ice. RNA was loaded onto an 8% denaturing polyacrylamide gel (20 X 20 X 1.5 cm) and electrophoresed for 2.5 h at 400 V. Gels were stained in a 100 μg/mL ethidium bromide solution for 30 m and washed in dH_2O for 2 m. RNA species were visualized and photographed using UV light (at 360 nm). Size estimates were based on comparison with a 0.23- to 9.5-kb RNA ladder (Gibco BRL).

RNA was transferred to nylon membrane using capillary action as described for DNA, with a number of modifications. First, gels were laid on kraft paper stacks that had been soaked in 5 mM NaOH. Second, transfers were performed overnight (~16 h).

Finally, membranes were neutralized in 20X SSPE, baked at 65°C for 4 h and stored at room temperature.

Hybridization conditions for oligonucleotide probes were the same as those used for randomly-labeled probes, with the following exceptions. First, pre-hybridization and hybridizations were performed at 33°C instead of 42°C. Second, the final wash (30 m at 50°C in 0.1X SSPE and 1.0% SDS) was done at room temperature using 0.2X SSPE and 1.0% SDS. Alternatively, this step was omitted.

10. 3'-RACE analysis of *cox1* and rRNA transcripts.

The 3' ends of several RNAs (*cox1*, LSUG, LSUE) were mapped by the addition of a oligo(A) tail followed by cDNA synthesis, PCR amplification and DNA sequencing (Lingner & Keller, 1993; Zaug *et al.*, 1996), as outlined in Fig. 2.1. Initially, short (1-3 nt) CMP tails were added to the 3' ends of mtRNA species. This addition was performed in a 10 µL reaction mixture containing 500 ng mtRNA, 20 mM Tris-HCl (pH 7.0), 50 mM KCl, 0.7 mM $MnCl_2$, 0.2 mM EDTA, 1 µg of acetylated BSA, 10% glycerol, 500 units poly(A) polymerase (Amersham Life Science) and 0.5 mM CTP. The mixture was incubated at 30°C for 30 m. Oligo(A) tails (10-30 nt) were then added to the 3' ends of these modified RNAs by addition of 0.5 mM ATP to the reaction mixture and further incubation at 30°C for 10 m. The reaction was stopped by phenol extraction. The RNA was resuspended in dH_2O and stored in 50% ethanol at -20°C.

RNAs modified by CMP-AMP addition or untreated were used in cDNA synthesis. First, 500 ng of RNA were combined with a primer that were complementary to the newly synthesized oligo(A) region. The P-16 primer (3'-NVTTTTTTTTTTTTTT

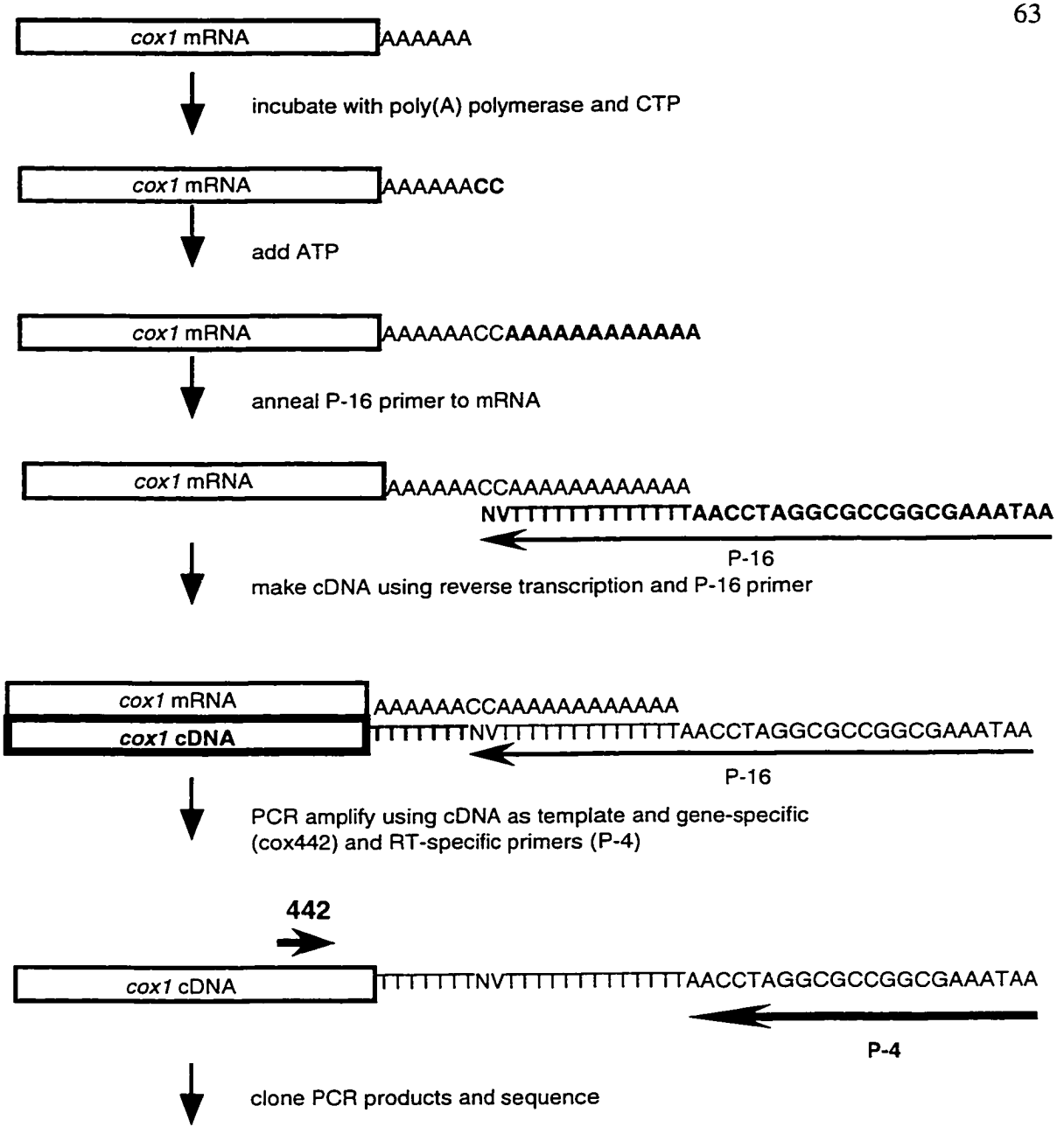


Fig. 2.1. Schematic representation of 3′-RACE using *cox1* mRNA as an example. Rectangles depict *cox1* mRNA or cDNA. Vertical arrows show progression from RNA to PCR products. Labels adjacent to vertical arrows show treatment of nucleic acid. Bolded characters (nucleotides, rectangles or arrows) denote changes in nucleic acid following treatment. The 3′-RACE methodology was the same for LSU rRNA analysis, except that *rnl*-specific primers were used. The letters 'N' (G, A, T or C) and 'V' (G, A or C) denote nucleotides.

TTAACCTAGGCGCCGGCGAAATAA-5') ensures that the primer will only anneal at the start of an oligo(A) tract (Borson *et al.*, 1992). The P-16 primer was annealed to the RNA by heating to 70°C for 3 m and incubating at 37°C for 20 m. Next, cDNA synthesis was performed in a 25 µL reaction mixture containing RNA and primer, 50 mM Tris-HCl (pH 8.3 at 25°C), 10 mM NaCl, 50 mM KCl, 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM spermidine, 10 units AMV reverse transcriptase (Promega), 26 units RNAguard ribonuclease inhibitor (Pharmacia) and 1 mM of each of the four dNTPs. Once the mixture reached 50°C, 2.5 µL of 40 mM sodium pyrophosphate was added and the mixture incubated for 1 h. Following cDNA synthesis, AMV reverse transcriptase was denatured at 95°C for 3 m.

Excess dNTPs were removed from the cDNA solution by spin chromotography using a G-25 spin column (Pharmacia) following the same protocol as outlined earlier for the S-200 column. PCR was performed using 50 ng of cDNA and standard protocols. Amplification primers included a forward gene-specific oligonucleotide and a reverse primer. The reverse primer was complementary to the 3' half of the P-16 oligo (P-4; 3'-AACCTAGGCGCCGGCGAAATAA-5') and was specific for cDNA products that were synthesized using the P16 oligo. Following amplification PCR products were cloned into pT7 vector and sequenced. Only one clone was characterized from each PCR reaction. Therefore, multiple (20) PCR reactions were produced for every primer combination (i.e., forward and reverse). This was done to reduce the chance of cloning aberrant PCR products from a single PCR reaction.

11. Isolation of polyadenylated RNAs.

Polyadenylated mRNAs were isolated from total RNA using the PolyATtract ® mRNA Isolation System (Promega). Briefly, 100 μg of RNA were resuspended in 500 μL of dH_2O and heated to 65°C for 10 m. The RNA was combined with 3 μL of biotinylated-oligo(dT) and 13 μL of 20X SSC. The mixture was then cooled at room temperature for 10 m. Streptavidin-paramagnetic beads were prepared by washing the beads in three 300 μL vol. 0.5 X SSC. The beads were resuspended in 100 μL of 0.5 X SSC, added to the RNA mixture and the suspension was incubated at room temperature (25°C) for 10 m. The beads and bound RNA were then captured magnetically, following which the supernatant was carefully decanted. Unbound RNA was removed by washing the beads with 0.1 X SCC followed by bead capture and supernatant removal. The washing step was repeated four times. Finally, RNA was eluted in 250 μL dH_2O. The RNA from all elution steps was precipitated with ethanol, resuspended in formaldehyde loading buffer, electrophoresed in a formaldehyde/agarose gel and blotted onto nylon membrane (see above).

12. Reverse transcriptase sequencing of rRNA genes.

Prior to sequencing, primers were ^{32}P-end-labeled and gel purified. Briefly, 100 pmol of end-labeled primer were electrophoresed in a 10% polyacrylamide gel (19:1 acrylamide: *bis*acrylamide, 8.3 M urea, 1 X TBE) for 3 h at 1500 V or until the bromophenol blue reached the bottom of the gel. The gel was exposed to X-ray film for 2 m, following which the primer band was excised from the gel using a razor blade. The radiolabeled oligonucleotide was eluted in a solution of 0.5 mM NH_4OAc and 1 mM

EDTA that was mixed with an equal volume of phenol/cresol/8-hydroxyquinoline. Purified primers were eluted from the gel using a table top shaker (high setting) at 4°C overnight. Primers were precipitated with ethanol and resuspended in dH_2O.

Purified primer was combined with 1 or 10 μg of RNA, in a 12 μL annealing mixture containing 250 mM KCl and 10 mM Tris-HCl (pH 8.3). RNA was denatured at 80°C for 3 m. Following this, annealing was performed at 55°C for 45 m. Next, 2 μL of the annealing mixture was transferred to each of four microcentrifuge tubes containing the AMV reverse transcriptase reaction mixture (25 μL reaction mixture containing RNA and primer, 50 mM Tris-HCl (pH 8.3 at 25°C), 10 mM NaCl, 50 mM KCl, 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM spermidine, 10 U AMV reverse transcriptase (Promega), 26 U RNAguard ribonuclease inhibitor (Pharmacia), 420 μM dGTP, 210 μM dATP, 210 μM dTTP μM 210 dCTP and either 160 μM ddGTP, 160 μM dATP, 320 μM dTTP or 160 μM dCTP). A fifth reaction was usually included that did not contain dideoxy nucleotides. Reverse transcription was carried out at 50°C for 45 m and stopped by the addition of 3 μL of stop solution (10 mM NaOH, 95% formamide, 0.05% bromophenol blue and 0.05% xylene cyanol) and the products stored at -20°C for up to 2 weeks. Samples were heated to 80°C for 3 m, snap-cooled on ice and resolved in a 6% polyacrylamide gel.

13. Analysis of potential secondary structure.

Ribosomal RNA secondary structure was inferred by comparing primary sequence alignments between *C. cohnii* and other mitochondrial and bacterial rRNA genes. Similarities were then mapped onto known secondary structures, with this information

then being used to model the secondary structure of *C. colnii* mitochondrial rRNAs.

Secondary structure associated with protein-coding transcripts (*cox1*) was inferred based on stability of potential folded structures. In particular, longer stem regions were chosen over shorter ones, pairing preferences were ranked (G-C > A-U > G·U) and bulges >3 nt were favored over 1- and 2-nt loops. RNA secondary structure diagrams were generated using the XRNA program (B. Weiser and H. Noller, University of California, Santa Cruz)

14. Data and phylogenetic analyses.

Sequence data were assembled using the GDE software package (Smith *et al.*, 1994) on a Sun SPARCstation 4. Alternatively, sequence data were assembled and scanned for open reading frames using the Sequencher ver. 3.1 program. Additional protein-coding, rRNA and tRNA gene searches were carried out by querying NCBI databases using BLAST (Altschul *et al.*, 1997). Analysis of repeat sequences was performed using the MicroGenie (SciSoft) software package.

Amino acid alignments were generated using CLUSTAL W (Thompson *et al.*, 1994), and were manually optimized and examined using SeqVu 1.0.1 (J. Gardner, Garvan Institute of Medical Research, Sydney, Australia). Phylogenetic trees were inferred using neighbor-joining (NEIGHBOR; Felsenstein, 1993), parsimony (PROTPARS; Felsenstein, 1993) and maximum likelihood (ProtML; Adachi & Hasegawa, 1996a) methods. Distance matrices were calculated using a Dayhoff PAM amino acid substitution model (Dayhoff *et al.*, 1979) for PROTDIST and a mtREV (Adachi & Hasegawa, 1996b) amino acid substitution model for ProtML. Distance and

parsimony bootstrap resampling analyses (100 data sets) were performed (SEQBOOT: Felsenstein, 1993) to assess branch support (Felsenstein, 1985). Maximum likelihood support values (RELL) were estimated using the resampling of log-likelihood method (Adachi & Hasegawa, 1996a).

III. Results

C. cohnii cultures were harvested during mid-log phase of growth. This typically occurred three to five days after inoculation (OD_{600} 0.5-0.7). Light microscope examinations showed that these cultures are healthy with few dead cells. At this stage of growth, cells are rapidly dividing and appear to contain many small mitochondria with tubular cristae (Fig. 3.1A and 3.1B). Notably, bacterial symbionts were not seen using microscopic investigation.

Kubai and Ris (1969) identified a chloroplast-like organelle in *C. cohnii*. However, all other studies (including the present one, using electron microscopy) have failed to detect plastids in this dinoflagellate. The reason for this discrepancy is unknown. One possibility is that only certain strains of *C. cohnii* have plastids. Another is that plastids are rare or only occur under certain environmental conditions.

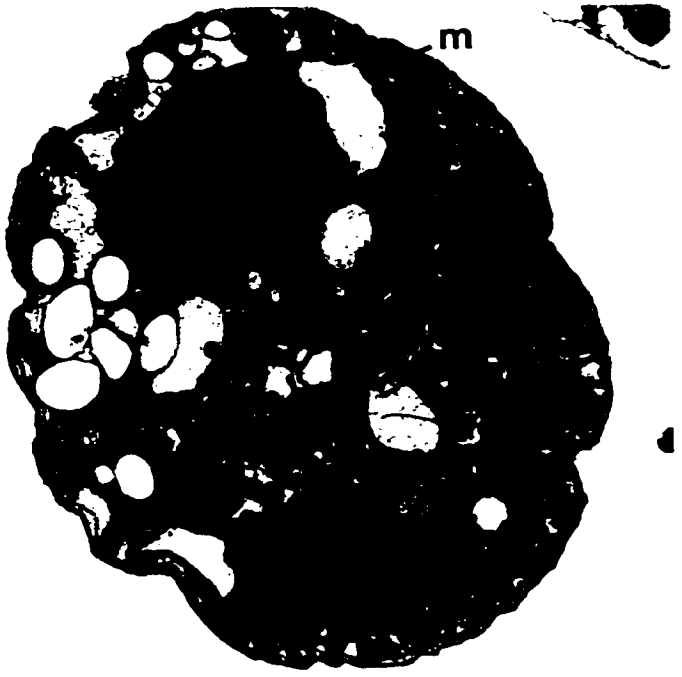
A. Isolation of nucleic acids from *C. cohnii*.

Approximately 2.5 mg of total nucleic acids were routinely isolated from 1-L cultures of *C. cohnii* using a phenol extraction procedure. A large portion of this material is composed of nDNA (~1 mg); however, rRNA and tRNAs are also present (Fig. 3.2, lane 1). Total DNA isolated in this manner displays little evidence of degradation (Fig. 3.2, lanes 1 and 2). However, hydrolysis with *Eco*RI, *Eco*RV, *Sst*I and *Xba*I restriction enzymes converts the tight band of nuclear DNA to a lower-molecular-weight, heterogeneous smear (Fig. 3.2, lanes 3 to 6).

Variable yields of 1 to 70 μg of mtDNA and 1 to 30 μg of mtRNA were prepared

Fig. 3.1. Transmission electron micrographs of *C. cohnii*.(**A**) Micrograph of a whole cell (mag. 6200 X) isolated from a healthy culture. (**B**) Increased magnification (30.000 X) of a *C. cohnii* cell showing mitochondria with tubular cristae. Representative mitochondria are marked by the letter 'm'.

A.



B.

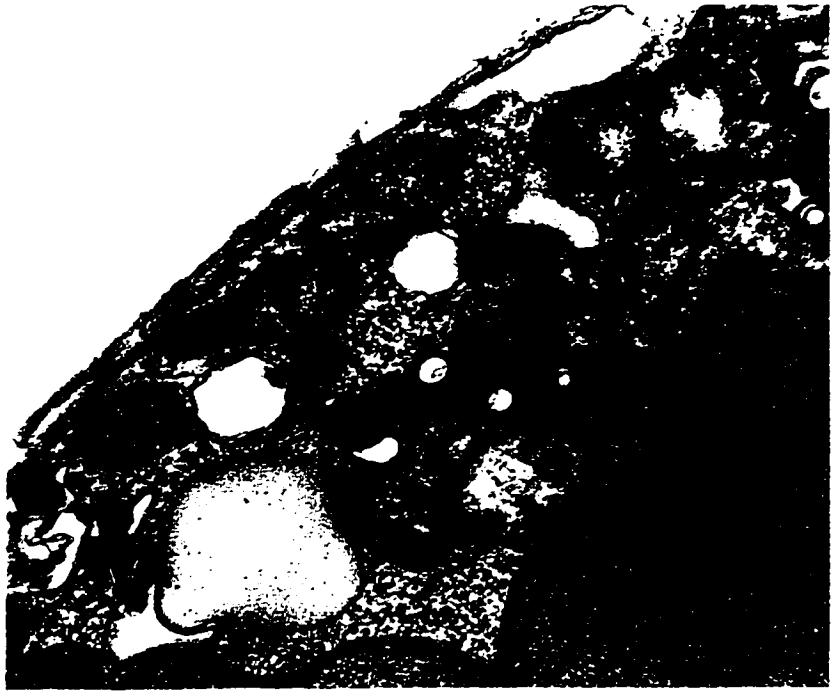


Figure 3.1

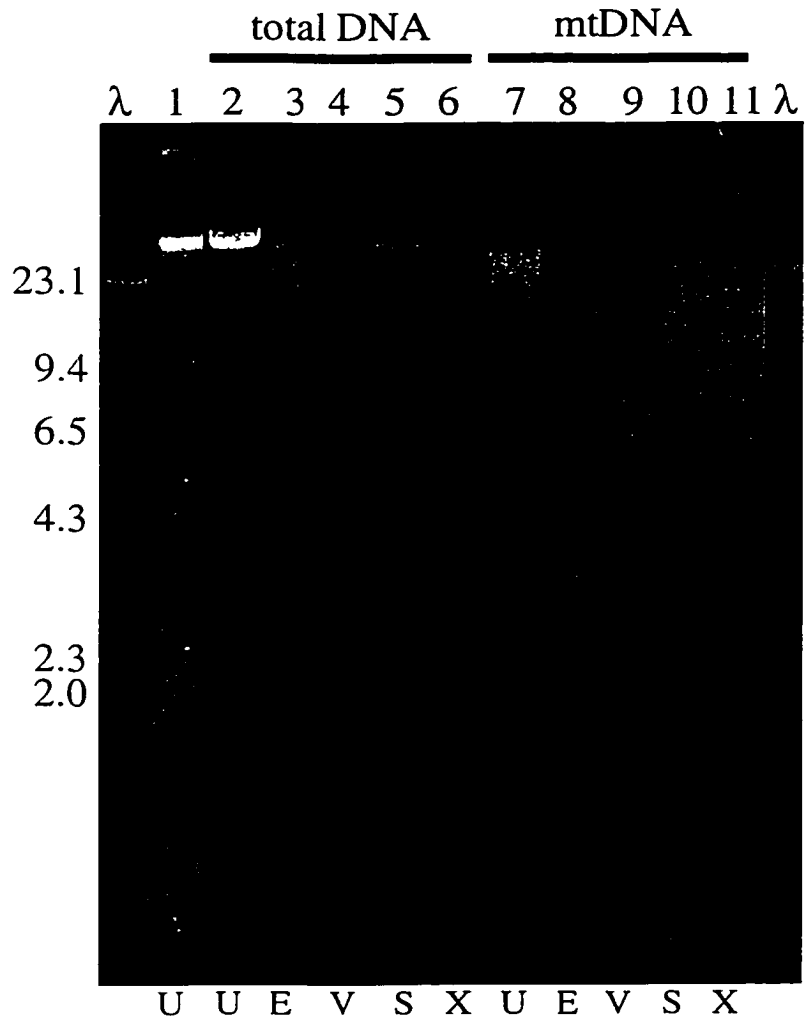


Fig. 3.2. Photograph of a 1% agarose gel of *C. cohnii* nucleic acids. Lane 1 contains total nucleic acids. Total (lanes 2 to 6) and mitochondrial (lanes 7 to 11) DNA fractions are indicated (solid bars). DNA samples were hydrolyzed with: E, *Eco*RI (lanes 3 and 8); V, *Eco*RV (lanes 4 and 9); S, *Sst*I (lanes 5 and 10) or X, *Xba*I (lanes 6 and 11). Undigested control lanes are labeled U (lanes 1, 2 and 7). Lanes marked 'λ' contain the size marker (λ DNA hydrolyzed with *Hin*dIII). The sizes of marker fragments are indicated to the left of the figure.

from 16-L cultures of *C. cohnii* using subcellular fractionation protocols. Mitochondria isolated using this technique are prone to rupture or are heavily contaminated with membranes from other cellular components (Fig. 3.3). However, intact nuclei were not seen in this fraction. Despite this observation, it is possible that nDNA may be associated with membranes present within the mitochondrial fraction.

The mtDNA-enriched fraction produced a smear (Fig. 3.2, lane 7) in the undigested control lane as well as in restriction-enzyme-treated fractions (Fig. 3.2, lanes 8 to 11). This suggests that mtDNA is degraded during isolation or that the mitochondrial genome of *C. cohnii* is composed of a population of small-sized genetic elements. Three of the restriction enzymes (*Eco*RI, *Eco*RV and *Xba*I) produced faint bands with the mtDNA fractions (Fig. 3.2, lanes 8, 9 and 11). Similar-size bands can also be seen in the total DNA samples (Fig. 3.2, lanes 3, 4 and 6); however, the bands are less intense in the latter sample.

The location of mtDNA and nDNA bands in CsCl density gradients is different in *C. cohnii* compared to other organisms. In *C. cohnii*, only one distinct UV-visible band can be seen (Fig. 3.4). In many other eukaryotes, mtDNA is found in a second discrete band located just above the major one. In *C. cohnii*, however, a faint diffuse zone was detected above the main band. Based on this observation, it is presumed that *C. cohnii* mtDNA does not resolve into a single tight band, but is scattered throughout the middle portion of the gradient. Additional data that support this conclusion is presented in section III B.

Fig. 3.3. Electron micrograph of a *C. cohnii* mitochondrial fraction isolated using sub-cellular fractionation. The sample was recovered from the bottom of the sucrose gradient (pellet) following centrifugation (mag. 23,000 X).

Figure 3.3

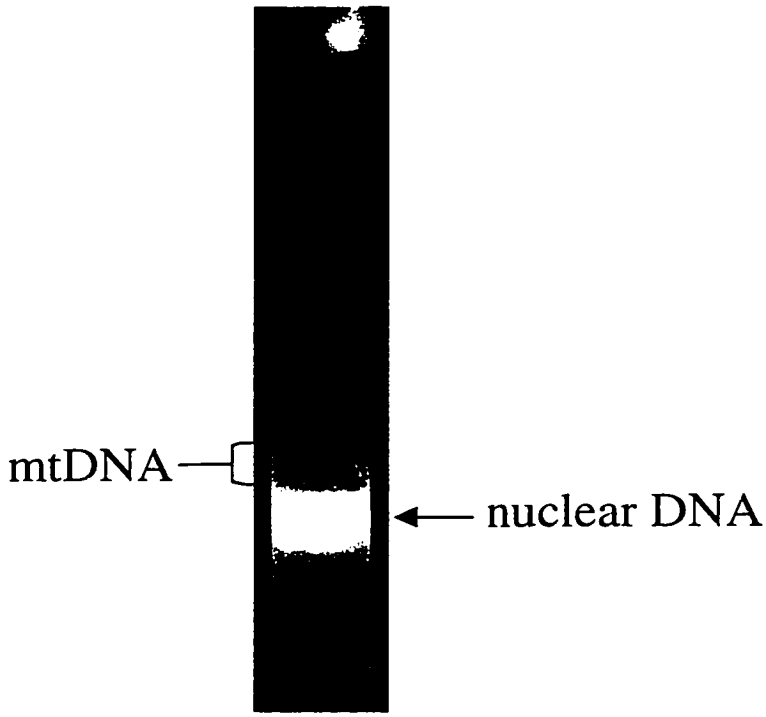


Fig. 3.4. Photograph showing results of CsCl density gradient centrifugation using total DNA isolated from *C. cohnii*. Presumed locations of nDNA and mtDNA are noted.

B. Organization, sequence and expression of *cox1* in *C. cohnii*.

1. Southern hybridization analysis: determination of *cox1* gene organization.

An internal portion (656 nt) of *C. cohnii cox1* was amplified using cox240 (forward) and cox448 PCR primers (Table 2.1). These primers were designed based on two highly conserved regions of COX1 sequence, subunit 1 of the cytochrome oxidase protein complex. Sequence analysis revealed that this product was homologous to *cox1*, thereby allowing it to be used as a probe in Southern analysis. In Southern hybridizations, the cox240-448 probe produced a smear (from <2 kbp to 23 kbp) in the undigested control and *Xba*I lanes (Fig. 3.5, lanes 1 and 5, respectively). However, the smear was more uniformly distributed over the *Xba*I lane. In the *Eco*RI digest, four fragments of approximate size 3.5, 4.1, 5.1 and 5.7 kbp were observed (Fig. 3.5, lane 2), superimposed on a faint trailing smear starting at the 5.7-kbp band. The stoichiometries of the four *Eco*RI bands were unequal. Each of the two smaller bands displayed roughly twice the intensity of each of the two larger bands. With the same probe, *Eco*RV produced a single 2.6-kbp band overlaid on a faint trailing smear starting at approximately 10 kbp and continuing to < 2 kbp in size (Fig. 3.5, lane 3). *Sst*I generated two hybridizing fragments, 1.5 and 2.0 kbp in size (Fig. 3.5, lane 4), with the smaller band displaying twice the hybridization intensity of the larger band. Because the cox240-448 region does not contain *Eco*RI or *Sst*I sites, the multiple banding pattern seen in these lanes is not due to the probe hybridizing across internal restriction sites. Thus, the presence of multiple *cox1*-hybridizing bands in the *Eco*RI and *Sst*I digests strongly suggests that the *cox1* gene in *C. cohnii* is present in different contexts that are stoichiometrically unequal.

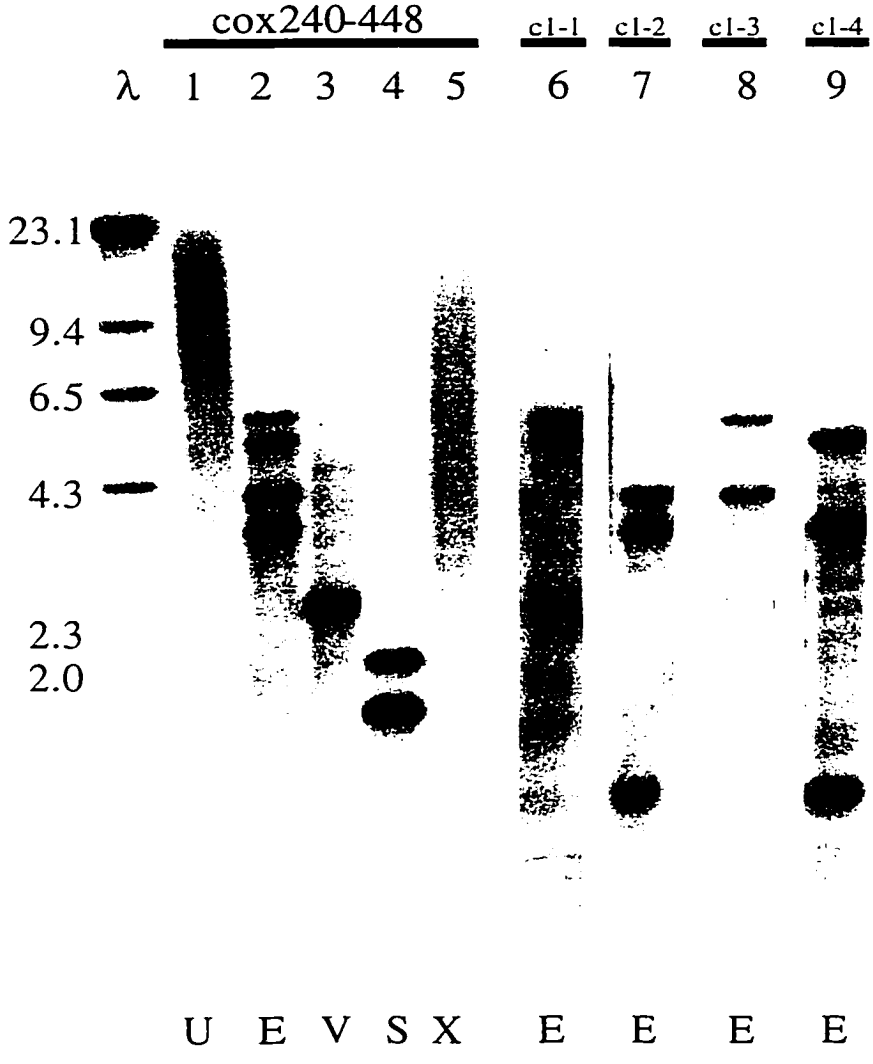


Fig. 3.5. Autoradiograms showing results of Southern hybridization analysis of *C. cohnii* mtDNA using a *cox1*-specific probe. Samples were hydrolyzed with: E, *Eco*RI (lanes 2, 6, 7, 8 and 9); V, *Eco*RV (lane 3); S, *Sst*I (lane 4); or X, *Xba*I (lane 5). Undigested control lane is labeled U (lane 1). Blots were hybridized with radiolabeled probes as indicated (cox240-448, lanes 1-5; c1-1, lane 6; c1-2, lane 7; c1-3, lane 8; and c1-4, lane 9). The positions of size marker fragments (kbp; λ DNA hydrolyzed with *Hin*dIII) are indicated to the left of the figure.

With total DNA resolved on CsCl density gradients, Southern analysis revealed that the *cox1* gene was distributed over a broad region of the gradient. Fine-scale fractionation of a CsCl gradient (200 µL aliquots X 50) showed that the major DNA band elutes in fractions 16-23 (Fig. 3.6A, lanes containing visible amounts of DNA). Based on Southern analysis, the majority of *cox1*-containing DNA was located just above the main DNA band (fractions 25-28). However, *cox1*-containing DNA also appeared to be spread over the middle portion of the gradient (Fig. 3.6B, lanes 20 to 30). In fact, a fairly intense *cox1* hybridization signal could be seen in the visible DNA fractions (Fig 3.6B, lanes 22 and 23).

The cox240-448 probe produced a much weaker Southern hybridization signal with total DNA compared to DNA isolated from mitochondria (Fig. 3.7, lanes 1 to 3). Interestingly, preparation of total DNA by guanidine/proteinase K extraction generated a stronger *cox1* hybridization signal compared with methods based on detergent/phenol-cresol extraction (Fig. 3.7). This difference may be due to more efficient removal of DNA-bound proteins using guanidine and proteinase K treatment.

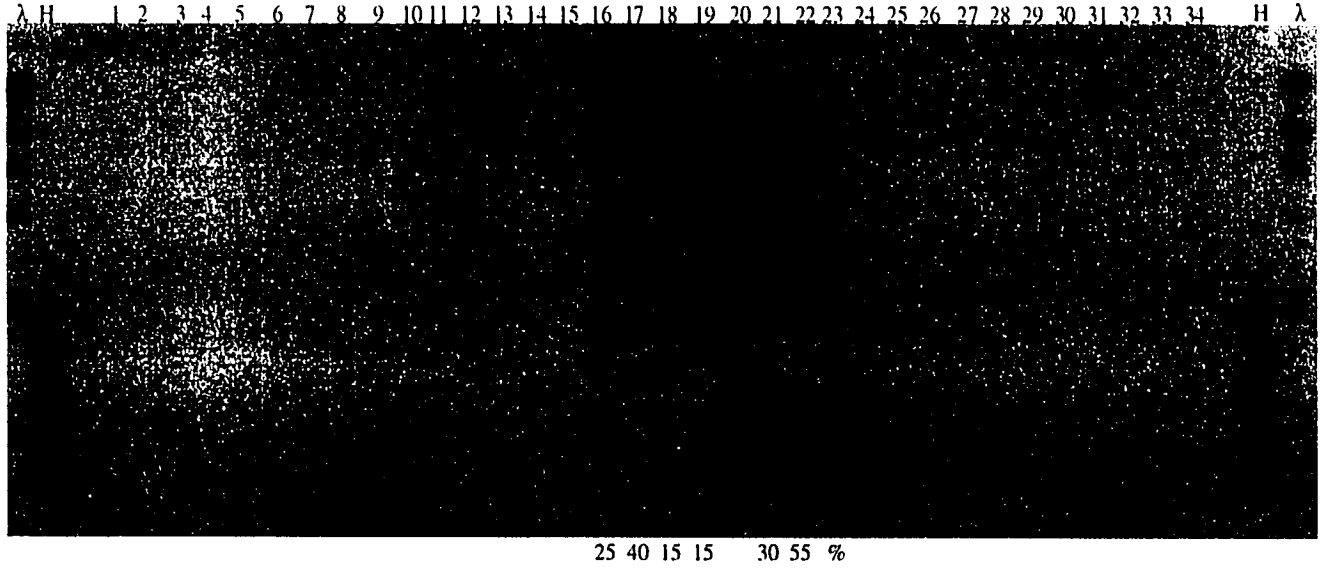
Throughout the course of this study (five years), different DNA preparations consistently produced the same cox240-448 banding pattern (Fig. 3.7, lanes 4 and 5). Therefore, despite exhibiting a complex hybridization pattern, the *cox1* arrangement in *C. cohnii* appears to be stable.

2. Characterization of the major *cox1*-containing elements.

A total of 17 *cox1*-containing clones were isolated by colony screening of a mtDNA *Eco*RI library using the cox240-448 probe. Mapping and sequence analysis

Fig. 3.6. Evaluation of CsCl-density gradient separation of mtDNA from nuclear DNA. (**A**) Gel profile showing fractions collected from a CsCl gradient. Up to 1 μg of *Eco*RI hydrolyzed-DNA was loaded per lane. Those samples that contained more than 1 μg of DNA are marked with an asterisk (*). For these fractions, a number is given (numbers to the left of the percent sign, %, and below the asterisks) that represents the percentage of the sample collected that was loaded on the gel. (**B**) Autoradiogram of *cox1* Southern hybridization using a membrane prepared from the gel shown in panel A. The location of the major UV-fluorescent samples in panel A is marked (wire frame). Blots were hybridized with radiolabeled cox240-448 PCR products. The first sample collected (starting at the bottom of the gradient) is found in lane 1. The subsequent lanes contain sequentially collected samples. A total of 50 fractions were collected but only the first 34 are shown. Fractions 35-50 did not exhibit UV-fluorescence or hybridization to the cox240-448 probe. Lanes marked with 'λ' or the letter 'H' contain size markers (λ DNA hydrolyzed with *Hin*dIII or 100-bp ladder, respectively).

A.


λ H 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 H λ
25 40 15 15 30 55 %

B.

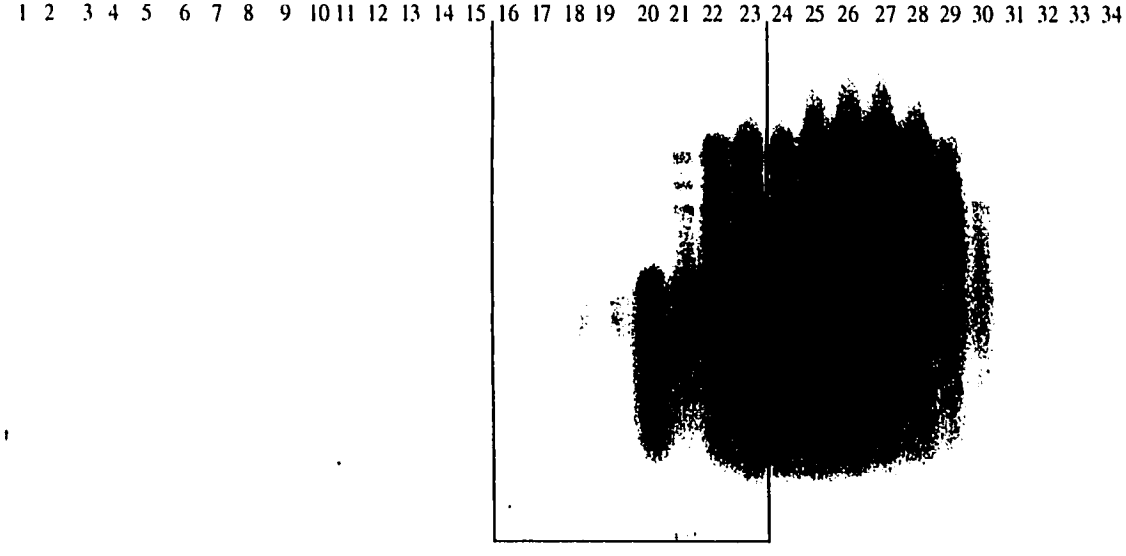

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34

Figure. 3.6

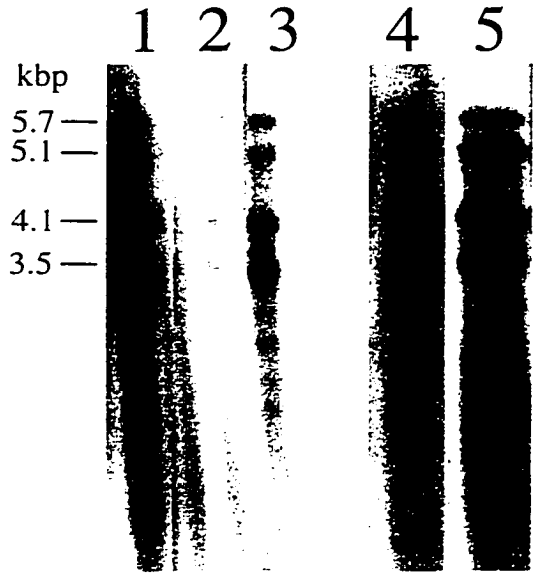


Fig. 3.7. Autoradiograms showing results of Southern hybridization analysis of different *C. cohnii* DNA preparations using a *cox1*-specific probe. All samples contained 1 µg of DNA and were hydrolyzed with *Eco*RI. Fractions enriched in mtDNA (lanes 1, 4 and 5). Lane 4 contains mtDNA that was isolated ~4 years earlier than that in lane 5. Total DNA samples were prepared using detergent/phenol-cresol (lane 2) or guanidine/proteinase K (lane 3) extraction methods. Lanes 1 to 3 were subjected to the same hybridization conditions. Membranes were hybridized with radiolabeled cox240-448 probe. The sizes of the four *cox1* fragments are indicated to the left of the figure.

showed that each clone contained an insert between 0.2 and 6.0 kbp in size and that most of these inserts (12) appeared to be representative of the four *Eco*RI bands visualized in Southern analysis.

a. Characterization of sequence elements containing large repeats. Sequence analysis of clones representing the four *cox1*-hybridizing *Eco*RI fragments (pc1#2, pc1#15, pc1#8, pc1#3; Fig. 3.8) revealed that *C. cohnii* mtDNA contains two variant *cox1* ORFs. The *cox1* gene is composed of an identical N-terminal portion (1.3 kbp) followed by one of two different in-frame C-terminal regions. Besides the *cox1* gene, no introns or protein-coding, rRNA or tRNA genes were present. Instead, scattered throughout the regions flanking the *cox1* ORF are large stretches of non-coding DNA that contain numerous short (< 27-nt) inverted and direct repetitive sequence elements (see section 2.b below). All *cox1* clones contained a core 2814-nt central repeat (CR) unit. This central repeat houses most of the *cox1* gene (Fig. 3.8) and is flanked by two different upstream and two different downstream domains. These four *cox1*-flanking domains are arranged such that those upstream, flank-1 (F1, 1882 nt) and flank-2 (F2, 302 nt), each occurs in pairwise combination with either of the two downstream regions, flank-3 (F3, 963 nt) or flank-4 (F4, 382 nt). This allows 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. 3.8, the two largest fragments (5.7 and 5.1 kbp) contain the same 5' flanking sequence, as do the two smallest fragments (4.1 and 3.5 kbp); conversely, the 5.7- and 4.1-kbp fragments have the same 3' ends, as do the 5.1- and 3.5-kbp fragments.

To confirm that the four isolated *Eco*RI clones are indeed representative of the

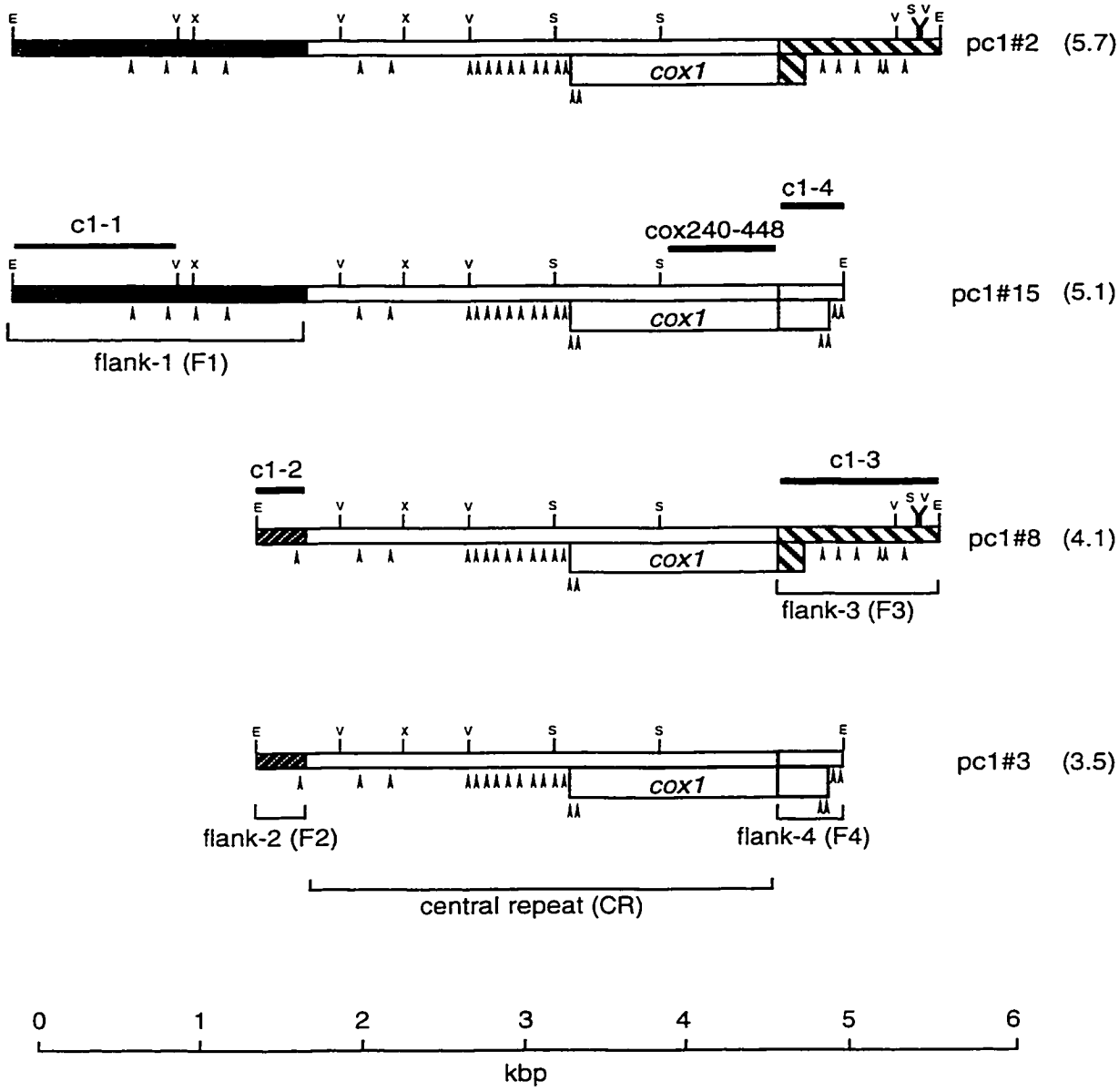


Fig. 3.8. Physical maps of the four major *cox1*-containing *Eco*RI fragments that were cloned and characterized from *C. cohnii* mtDNA. Sizes and names of constructs are listed to the right of the figure. Open rectangles indicate identical regions in all four clones (central repeat, CR). Shaded rectangles 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 inverted repeats. Solid black lines above each map indicate the positions of five probes (cox240-448, c1-1, c1-2, c1-3, c14) used in Southern hybridizations. Restriction sites are: E, *Eco*RI; V, *Eco*RV; S, *Sst*I; X, *Xba*I. A size scale bar is located along the bottom of the figure.

four major *Eco*RI bands visualized in Southern analysis, this analysis was repeated using probes (c1-1, c1-2, c1-3 and c1-4) specific for each of the flanking regions. Two probes were prepared using the pc1#15 clone. Hydrolysis of pc1#15 with *Eco*RI and *Eco*RV produces a 997-nt fragment (labeled c1-1 in Fig. 3.8) that covers the 5' end of this insert, whereas hydrolysis of pc1#15 with *Eco*RI and *Dra*I gave a 236-nt fragment (c1-4, Fig. 3.8) that encompasses the terminal portion (3' end). Two other probes were prepared by PCR amplification using pc1#8 as template. The first, c1-2, is 371 nt in length and was amplified from clone pc1#8 (Fig. 3.8) using the vector-based T7 and internal P37 primers (Table 2.1). Similarly, a 1029-nt PCR product (c1-3, Fig. 3.8) was amplified from clone pc1#8 using the two primers T3 and P6 (Table 2.1).

The results showed that 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, probe c1-1 detected two bands, 5.7 and 5.1 kbp in size (Fig. 3.5, lane 6); probe c1-2 revealed two bands of 4.1 and 3.5 kbp as well as a third faint band (~0.8 kb) (Fig. 3.5, lane 7); probe c1-3 hybridized to two bands, 5.7 and 4.1 kbp in size (Fig. 3.5, lane 8); and probe c1-4 identified two major bands, 5.1 and 3.5 kbp in size, plus an additional minor band approximately 0.6 kbp in size (Fig. 3.5, lane 9). 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. 3.9A, the 5' ends of the CR region align perfectly until the divergence

Fig. 3.9. Nucleotide sequence showing the location of divergence points and inverted repeats within *cox1*-containing *Eco*RI fragments cloned from *C. cohnii* mtDNA. Nomenclature of flanking regions and clone designation are as described in Fig. 3.8. Sequence co-ordinates are relative to the inferred *cox1* initiation codon. (**A**) Sequence alignment showing divergence points (marked by I) between the 3' end of flank-1 and flank-2 leading into the 5' end of the central repeat (enclosed by a rectangle). Dots indicate positions of nucleotide identity in the bottom sequence relative to the top one. (**B**) Alignment showing divergence points (marked by II) between the 3' end of the central repeat (enclosed by a rectangle) leading into the 5' ends of flank-3 and flank-4. (**C**) Nucleotide sequence showing the location of inverted repeats in the central repeat region close to the *cox1* initiation codon (which is shown in bold and underlined). Uppercase letters denote sequence contained within the *cox1* ORF. Closely-spaced horizontal arrows pointing in opposite directions located above (pc1#15) and below (pc1#8) the sequence mark the regions of inverted repeats that can potentially base pair to form the stem region of the repeat secondary structure (see Table 3.1 and Fig. 3.10). Names assigned to each inverted repeat follow the nomenclature of Table 3.1. The three forward-pointing arrows (labelled D.68.1, D68.2, D68.3) denote the DNA sequence of direct repeats. Stop codons are underlined and highlighted by three asterisks (***).

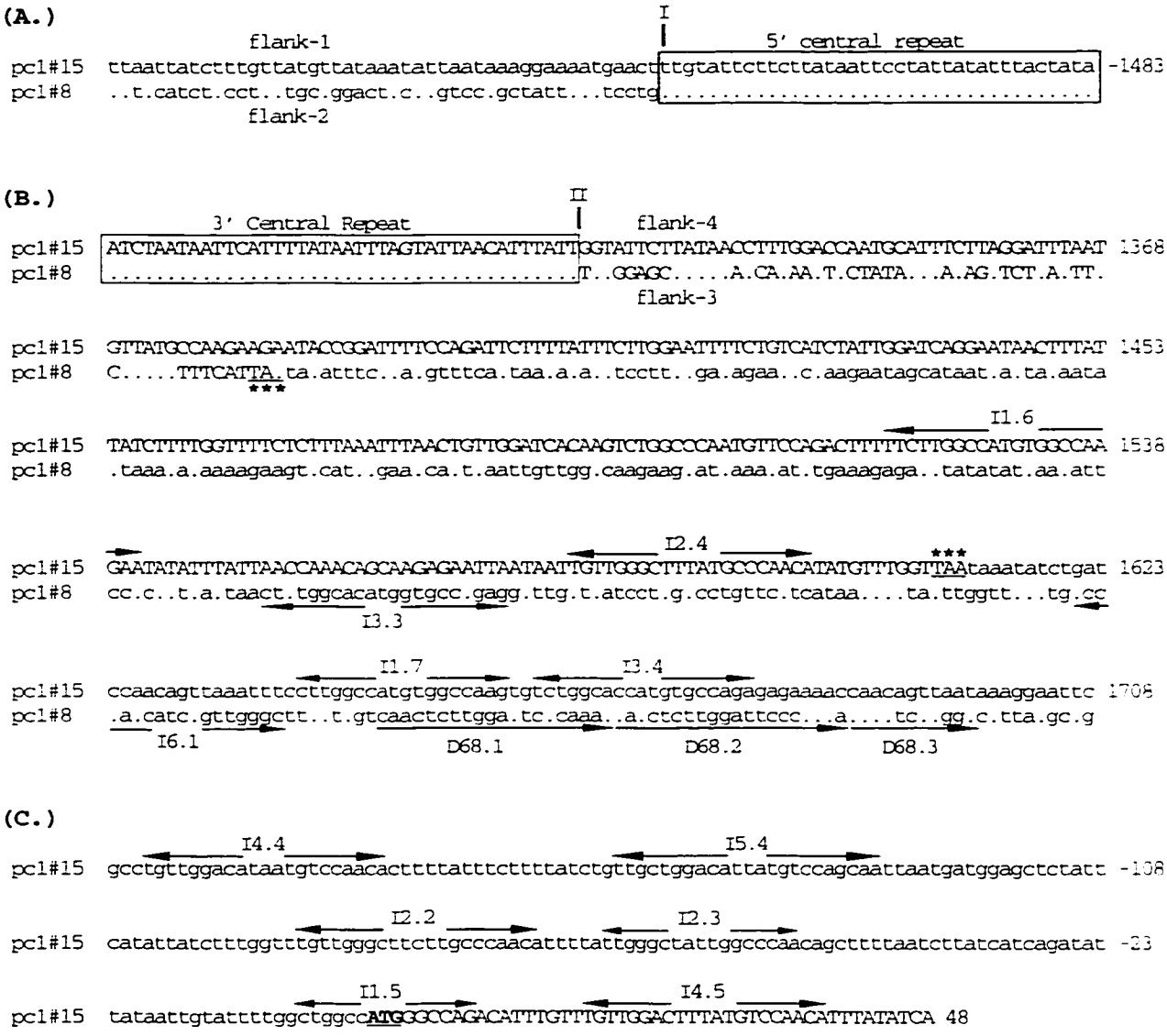

(A.)
I
flank-1
5' central repeat
pc1#15 ttaattatctttgttatgttataaatattaataaaggaaaatgaacttgtattcttcttataattcctattatatttactata -1483
pc1#8 ..t.catct.cct..tgc.ggact.c..gtcc.gctatt...tcctg.....
flank-2
(B.)
II
3' Central Repeat
flank-4
pc1#15 ATCTAATAATTCATTTTATAATTTAGTATTAACATTTATTGGTATTCTTATAACCTTTGGACCAATGCATTTCTTAGGATTTAAT 1368
pc1#8 .....T..GGAGC.....A.CA.AA.T.CTATA...A.AG.TCT.A.TT.
flank-3
pc1#15 GTTATGCCAAGAAGAATACCGGATTTTCCAGATTCTTTTATTTCTTGGAATTTTCTGTCATCTATTGGATCAGGAATAACTTTAT 1453
pc1#8 C.....TTTCATTA.ta.atttc..a.gtttca.taa.a.a..tcctt..ga.agaa..c.aagaatagcataat.a.ta.aata
I1.6
pc1#15 TATCTTTTGGTTTTCTCTTTAAATTTAACTGTTGGATCACAAGTCTGGCCCAATGTTCCAGACTTTTTCTTGGCCATGTGGCCAA 1538
pc1#8 .taaa.a.aaaagaagt.cat..gaa.ca.t.aattgttgg.caagaag.at.aaa.at.tgaaagaga..tatatat.aa.att
I2.4
pc1#15 GAATATATTTATTAACCAAACAGCAAGAGAATTAATAATTGTTGGGCTTTATGCCCAACATATGTTTGGTTAAtaaatatctgat 1623
pc1#8 cc.c..t.a.taact.tggcacatggtgcc.gagg.ttg.t.atcct.g.cctgttc.tcataa....ta.ttggtt...tg.cc
I3.3
I1.7
I3.4
pc1#15 ccaacagttaaatttccttggccatgtggccaagtgtctggcaccatgtgccagagagaaaaccaacagttaataaaggaattc 1708
pc1#8 .a.catc.gttgggctt..t.gtcaactcttgga.tc.caaa..a.ctcttggattccc...a....tc..gg.c.tta.gc.g
I6.1
D68.1
D68.2
D68.3
(C.)
I4.4
I5.4
pc1#15 gcctgttggacataatgtccaacacttttatttcttttatctgttgctggacattatgtccagcaattaatgatggagctctatt -108
I2.2
I2.3
pc1#15 catattatctttggtttgttgggcttcttgcccaacattttattgggctattggcccaacagcttttaatcttatcatcagatat -23
I1.5
I4.5
pc1#15 tataattgtattttggctggccATGGGCCAGACATTTGTTTGTTGGACTTTATGTCCAACATTTATATCA 48

Figure 3.9

point (labeled 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 (Fig. 3.9 B); however, in this case, the split into flanking domains 3 and 4 generates two *cox1* reading frames that differ in the C-terminal region of the encoded amino acid sequence. This feature, coupled with the presence of numerous small repetitive sequence elements located upstream and downstream of the two divergence points (see section 2.b below), suggests that the complex *cox1* organization observed in *C. cohnii* is likely a consequence of repeat-mediated recombination events. Initially, illegitimate recombination within small repeats could have promoted large-scale genomic rearrangements, leading to the creation of duplicate *cox1*-containing elements (a two-membered repeat as described by Coulthart *et al.* (1990) for rye mtDNA). These *cox1* repeats would then be capable of undergoing further rearrangement by recombination within the CR, generating the four different but related *cox1*-containing elements described here.

b. Characterization of sequence elements containing small repeats. Analysis of repetitive sequence elements, scattered throughout the four major *Eco*RI *cox1*-containing clones, revealed that inverted repeats (IRs) exhibit a non-random pattern that can be characterized on the basis of shared DNA sequence similarity. In total, 29 IRs were identified (Table 3.1), ranging from 15 to 22 nt in length. Fourteen IRs were found within the CR region, with the remainder occurring in the four flanking regions (F1-F4).

Nucleotide sequence comparison revealed that the stem region of most inverted repeats is GC-rich, unlike the majority of *C. cohnii* mtDNA, which is AT-rich. As illustrated by Table 3.1, inverted repeats can be assembled into six different groups based

Table 3.1. Nucleotide sequence comparison of all inverted repeats identified in the four major *cox1*-containing *Eco*RI fragments.

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

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 based on similarity in nucleotide sequence. A consensus sequence is shown in bold at the top of each group. Exact nucleotide matches between individual repeats and the consensus sequences are shown as dots. Dashes denote gaps in the sequence. Names assigned to individual repeats are indicated along the left of the figure. The coordinates and location of inverted repeats are based on the nomenclature of Fig. 3.8. Underlined consensus nucleotides show exact sequence matches among groups 1-5.

on shared DNA sequence similarity. Within each group, the central stem portion of individual repeats is highly conserved whereas the stem termini and unpaired loops are more variable (Table 3.1, Fig. 3.10). Inter-group sequence alignments indicate that groups 1-5 possess a common stem motif (-TGG-//-CCA-, underlined in Table 3.1) that is not shared by group 6. In addition, group 6 elements only occur within flank-3 (F3).

c. Expression of *cox1*. As illustrated in Figs. 3.8 and 3.9B, *C. cohnii* mtDNA contains two variant *cox1* ORFs. The longer ORF (536 amino acid residues) is found in pc1#15 and pc1#3, where 285 nt of the 1608-nt *cox1* reading frame extend into flank-4 (Fig. 3.9B). In contrast, the shorter *cox1* ORF, which occurs in pc1#2 and pc1#8, extends into flank-3 by only 57 nt, creating a reading frame of length 1437 nt (479 amino acids) (Fig. 3.9B). 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 (Fig. 3.11); moreover, the shorter COX1 sequence appears to lack the last two membrane-spanning domains that are present in all other COX1 sequences, including the longer *C. cohnii* COX1, characterized to date (Trumpower & Gennis, 1994).

Upstream of divergence point II, pc1#2 and pc1#8 maintain 100% sequence identity with clones that possess the longer version of the *cox1* gene. 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-kbp *Sst*I band (Fig. 3.5, lane 4) is more intense than the longer *Sst*I fragment that encodes what appears to be the authentic *cox1*. Similarly, the 4.1-kbp *Eco*RI fragment, which contains the truncated *cox1* version, hybridizes more intensely than the other small (3.2-

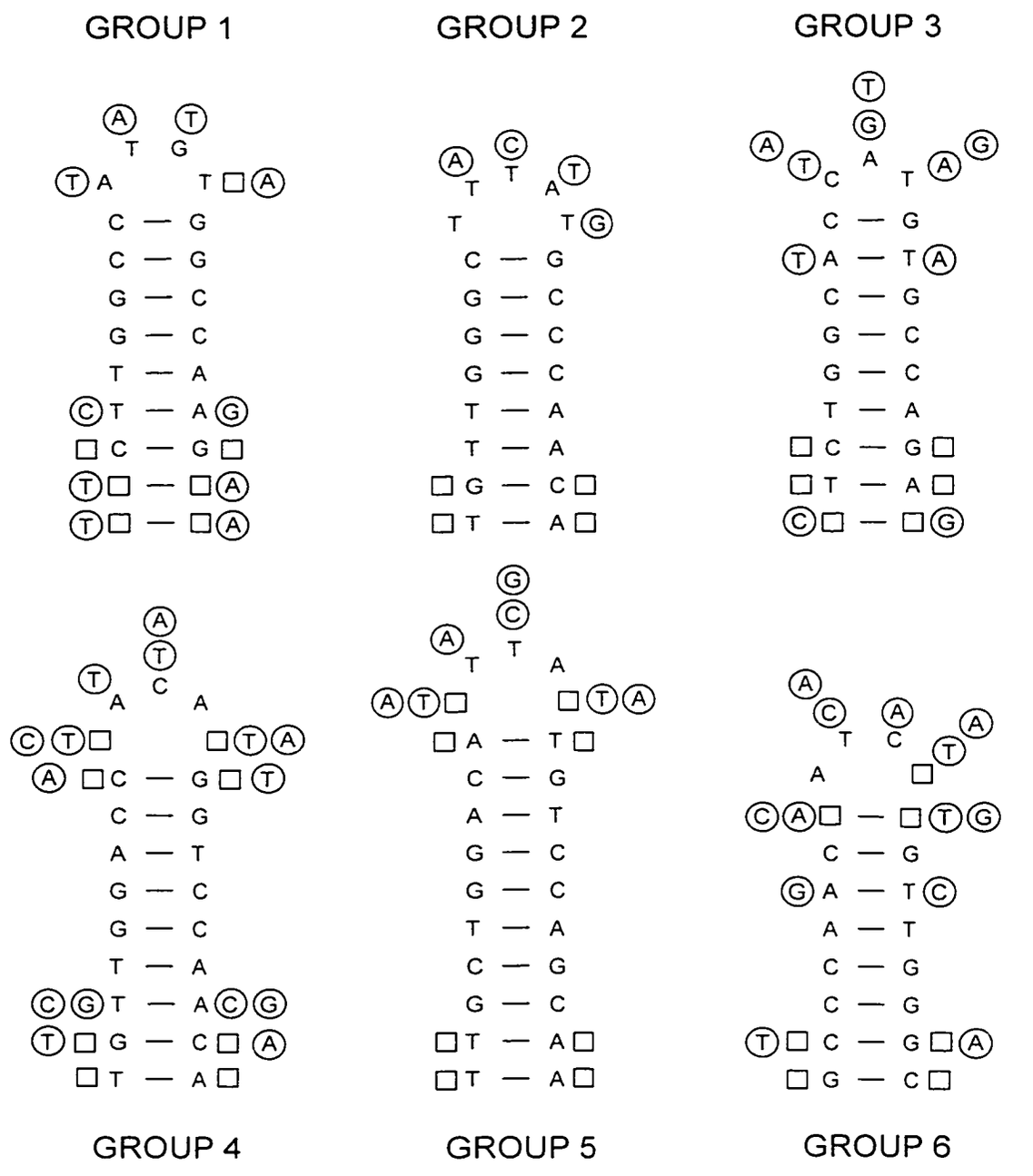


Fig. 3.10. Diagram depicting secondary structure models for inverted repeats listed in Table 3.1. Each structure is drawn using the group consensus sequence (see Table 3.1). Individual repeat nucleotide positions that differ from the consensus sequence are shown around the outside of the consensus structure as circled letters. Squares indicate secondary structure positions for which nucleotides are missing in either the consensus or individual structures. RNA secondary structures were drawn using the XRNA program (B. Weiser and H. Noller, University of California, Santa Cruz).

Fig. 3.11. 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 vertical line (|) 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_coh, *Crypthecodinium cohnii* (AF012554); P_fal, *Plasmodium falciparum* (M76611); T_pyr, *Tetrahymena pyriformis* (X06133); T_bru, *Trypanosoma brucei* (X01094); B_vul, *Beta vulgaris* (X57693); N_cra, *Neurospora crassa* (X14669); and B_tau, *Bos taurus* (V00654). Numbers within parentheses refer to EMBL accession numbers. The C-terminal portion of the alignment contains two version of *C. cohnii* COX1. 'C_coh full' refers to the presumed authentic (complete) version of COX1. 'C_coh trunc' denotes the shorter version. Up to residue position 397, the two versions are identical. Amino acid differences between the two versions of COX1 are highlighted by underlining.

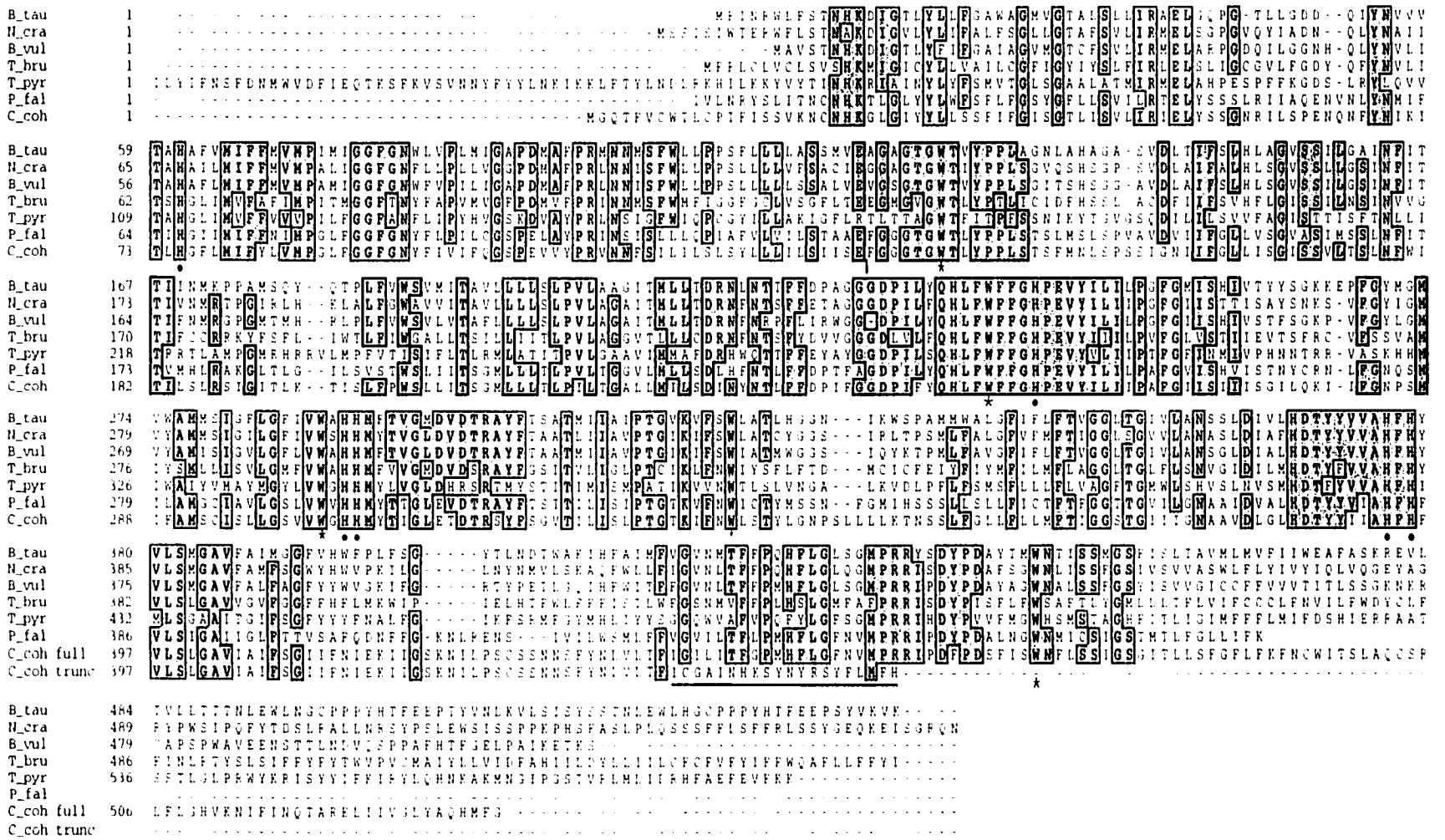


Figure 3.11

kbp) *Eco*RI band.

To determine whether the truncated *cox1* gene is functional in *C. cohnii*, evidence was sought for the presence of corresponding stable transcripts. A probe specific for the truncated version of *cox1* (flank-3) did not hybridize to northern blots (Fig. 3.12, lane 5), despite the fact that it generated a strong hybridization signal in Southern analysis performed at the same time (Fig. 3.5, lane 9). This suggests that stable transcripts are not produced from this ORF. In contrast, when northern filters were re-probed using c1-4, which is specific for the C-terminal portion of the longer *cox1* gene, the probe hybridized strongly to a 1.7-kb RNA (Fig. 3.12, lane 4). A 1.7-kb transcript is consistent with the size predicted for the longer *cox1* ORF, and is the same size as a band detected using an internal *cox1* probe (Fig. 3.12 , lane 1). Notably, a *cox1* hybridization signal was not detected in the oligo(dT)-bound RNA fraction (Fig. 3.12, lane 3).

A further difference between flank-3 and flank-4 is in the density of inverted repeats located within the two regions (Fig. 3.9B). Approximately 118 nt downstream of divergence point II and within *cox1* of flank-4 are several stretches of DNA that have the potential to fold into an elaborate and highly stable stem-loop structure (Fig. 3.13). In fact, as shown in Fig 3.13, the stop codon for the complete *cox1* gene is nestled within a helical region consisting of 18 consecutive base pairs. In contrast, only a few inverted repeats are located within the 5' end of flank-3 (Fig. 3.9B), and these occur well downstream (178 nt) of the termination codon. From these results it is inferred that flank-4 sequence, and potentially its folded RNA structure, acts to stabilize *cox1* transcripts.

Mapping of the C-terminal end of *cox1* using 3'-RACE supports the secondary

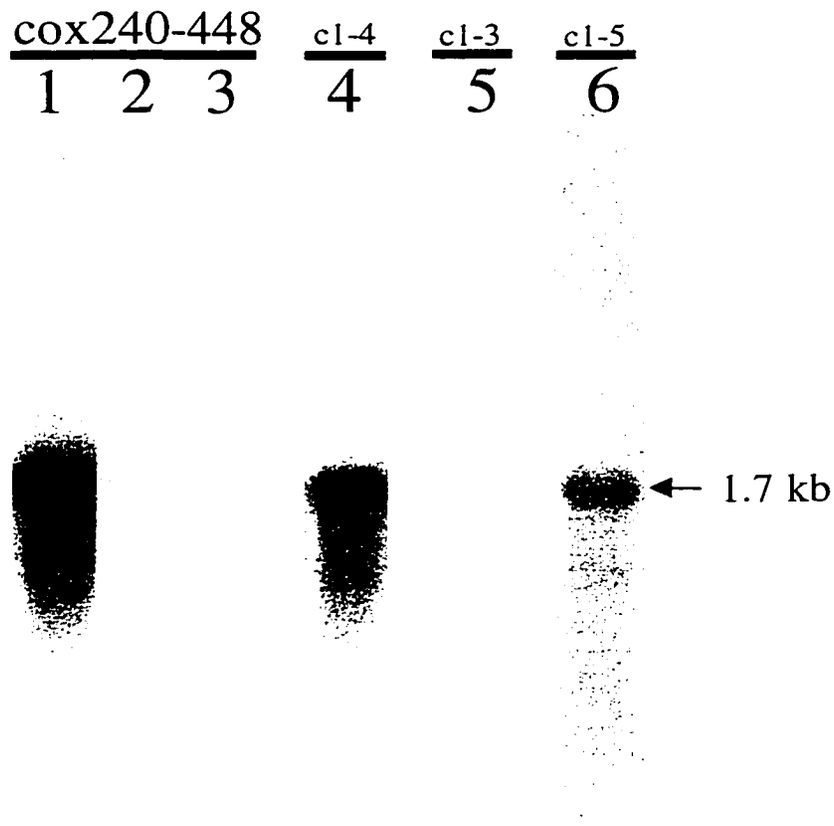


Fig. 3.12. Autoradiograms showing the results of northern hybridization analysis of *cox1* mRNA isolated from *C. cohnii*. Lanes 1, 4, 5 and 6 contain RNA that was prepared from mitochondria isolated by subcellular fractionation. Total (lane 2) and PolyATract®-purified RNAs (lane 3) were also included in the analysis. Each lane contains 1 µg of RNA. Blots were hybridized with radiolabeled cox240-448, c1-3, cl-4 (Fig. 3.8) and c1-5 (Fig. 3.13) probes as indicated.

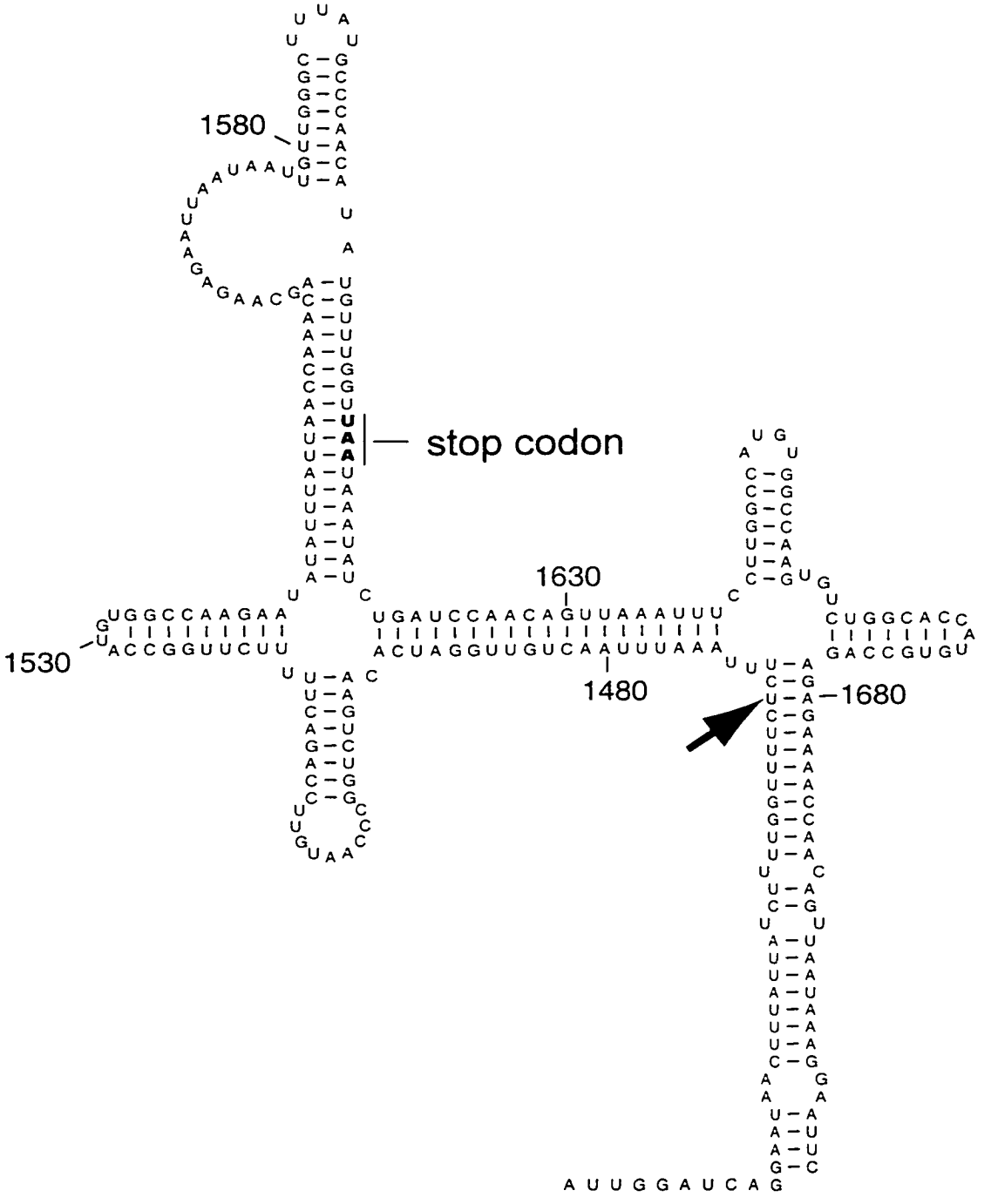


Fig. 3.13. Predicted secondary structure of the 3' end and downstream flanking sequence of the RNA transcribed from the longer (presumably authentic) *cox1* gene. Sequence numbering is relative to the inferred *cox1* initiation codon. The UAA termination codon is indicated (positions 1609-1611). Arrow denotes predicted 3' end of *cox1* mRNA based on 3'-RACE PCR (Table 3.2). The c1-5 probe (Fig. 3.12) starts at the first C nucleotide following the arrow and terminates at the 3' end of the predicted structure. RNA secondary structure was drawn using the XRNA program.

structure modeled in Fig. 3.13. RT-PCR produced a single band (~200 nt) in the poly(A) polymerase modified and untreated RNA lanes (Fig. 3.14, lanes 2 and 4). In contrast, a 300-nt product was predicted, based on the size of the 'complete' *cox1* gene. DNA sequence analysis of 3'-RACE clones derived from RT-PCR indicated that the 3' end of *cox1* occurred within the 5' end of the folded structure (Table 3.2), imbedded within the variable 3'-terminal extension of *cox1*. Because this mapped-end localizes to a variable region of *cox1*, it could be interpreted as the end of the transcript; however it lacks a stop codon and northern analysis suggests that the *cox1* mRNA extends beyond this site (Fig. 3.12, lane 6). A more likely explanation for this finding is that the elaborate folded RNA structure impedes RT from reading through the authentic *cox1* transcript. The 3'-RACE experiments suggest that the 3' end of *cox1* degradation products can be reverse transcribed because the absence of a highly folded RNA enhances the ability of RT to remain bound to RNA. Interestingly, these putative *cox1* degradation products appear to be polyadenylated *in vivo*, as suggested by the ability of oligo(dT) to prime cDNA synthesis and by the number of 3'-terminal A residues (between 6 and 27 residues, Table 3.2). In addition, all of the clones have the same oligo(A) addition site, suggesting that this site may be a specific target for mRNA degradation.

Although the majority of 3'-RACE clones contained *cox1*, several variants were detected (Table 3.1), four of which exhibited homology to 3' end of *cox3*. One other (Table 3.1, clone #14) contained unique sequence. The origin of these clones is unknown; however, all five variants appear to originate from polyadenylated transcripts.

Because the possibility of RT- or PCR-based artifacts has not been ruled out, the

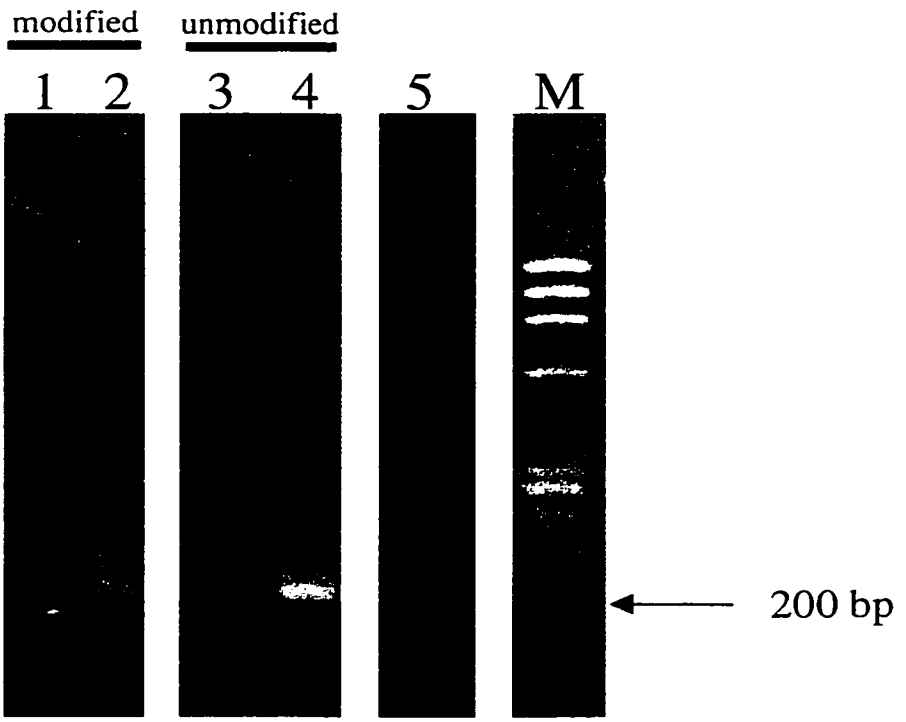


Fig. 3.14. Agarose gel displaying PCR amplification products generated from 3'-RACE (outlined in Fig. 2.1). Amplification was performed using cox442 and P4 primers. The word 'modified' denotes those RNA samples whose 3' ends were lengthened by the addition of CMP and AMP residues prior to cDNA synthesis. Samples marked 'unmodified' were not modified by CMP-AMP addition. Lanes 1 and 3 are controls (no DNA). Reaction mixtures in lanes 2 and 4 contained cDNA as well as primers. For lanes 2 and 4, first-strand synthesis was performed using the lock-dock primer (P16). The sample in lane 5 used mtDNA as template for PCR. M denotes the 100-bp DNA ladder that served as size standard. The 200-bp band is highlighted with an arrow.

Table 3.2. DNA sequence of *C. cohnii* 3'-RACE clones generated using a *cox1* gene-specific primer (cox442).

1. 3'-RACE clones prepared using unmodified* RNA			
cox1	DNA	*CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCTTTAAATTTAACTGTTGGATCA*	
	1	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCT***(A)***$_n$	
	2	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCT***(A)***$_n$	
	3	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCT***(A)***$_n$	
	4	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCT***(A)***$_n$	
	5	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCT***(A)***$_n$	
2. 3'-RACE clones prepared using modified* RNA			
cox1	DNA	*CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCTTTAAATTTAACTGTTGGATCA*	
	1	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCTAAAAAAAAAAAAAAAAAAAAAAAAAAA***CC(A)***$_n$	27
	2	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCTAAAAAAAAAAAAAAAAA***CACCC(A)***$_n$	17
	3	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCTAAAAAAAAAAAAAAAAA***CC(A)***$_n$	17
	4	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCTAAAAAAAAAAAAAAAAA***CACC(A)***$_n$	16
	5	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCTAAAAAAAAAAAA***CC(A)***$_n$	12
	6	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCTAAAAAAAAAA***CC(A)***$_n$	10
	7	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCTAAAAAAAAA***CAC(A)***$_n$	9
	8	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCTAAAAAAAA***CCC(A)***$_n$	8
	9	CTATTGGATC AGGAATAACT TTATTATCTT TTGGTTTTCTAAAAAA***CACC(A)***$_n$	6
cox3	pc3#17	ATCTTATGGT TATTTATCTT TCTAGTCTTT TATAGCTGTTAAAAAAAAAAAAAAAAA***C(A)***$_n$	17
	11	ATCTTATGGT TATTTATCTT TCTAGTCTTT TATAGCTGTTAAAAAAAAAA***CC(A)***$_n$	10
	12	ATCTTATGGT TATTTATCTT TCTAGTCTTT TATAGCTGTTAAAAAAAA***CC(A)***$_n$	7
	13	ATCTTATGGT TATTTATCTT TCTAGTCTTT TATAGCTGTTAAAAAA***C(A)***$_n$	5
unknown	14	TCTTTTCATT ATTTATTTCT ATTATTTTAA TGACCACTATAAAAAAAAAAAAAAAAAA***CC(A)***$_n$	18

The sequence given in italics represents *cox1* mtDNA sequence. Modified RNA refers to RNA that was subjected to CMP-AMP addition prior to reverse transcription. Unmodified RNA was not treated with poly(A) polymerase before cDNA synthesis. Clones were generated as outlined in Fig. 2.5. Numbers in front of sequences indicate the name given to individual 3'-RACE clones. The oligo(A) tail is underlined and tail size is indicated to the right of the sequence. The size and location of the oligo(A) tail was determined by CMP-AMP addition (noted in bold and italics) and is also noted in brackets. Location of oligo(dT) primer binding sites are noted ('A_n').

3'-RACE results have to be interpreted with caution. Despite this potential short-coming, however, the 3'-RACE data mark the length and location of pre-existing oligo(A) tails.

While the density of IRs is high within the 5' end of flank-4, the only region that exhibits a higher concentration of inverted repeats is the segment of mtDNA that encompasses the 5' end of *cox1*. As shown in Fig. 3.9C, two repeats occur within the *cox1* ORF itself, whereas four others are located within the first 200 nt upstream of *cox1*. In fact, 41% (12) of all identified inverted repeats occur within the first 550 nt of *cox1*, a region that constitutes only 8% of the total DNA comprising the *cox1*-containing elements. The placement of these inverted repeats is intriguing and may be a reflection of their importance in regulating gene expression near the 5' end of the gene, similar to what is observed at the 3' end of *cox1*.

The presence of the truncated *cox1* gene raises questions about its origin. One possibility is that it arose when flank-3, which was present elsewhere in the genome, replaced flank-4 through a spurious recombination event (e.g., between short repeated sequences). In such an event, 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. 3.5 (lane 9), probe c1-4, which targets only the C-terminal region of the longer *cox1* gene, hybridized to three bands, of which the two largest (5.1 and 3.5 kbp in size) correspond to the *Eco*RI inserts contained in pc1#15 and pc1#3; however, a third unidentified 0.8-kbp band is also visible. This 0.8-kbp fragment may represent the portion of *cox1* that is missing from the truncated version of the gene. While this result is intriguing, numerous attempts to clone this particular *Eco*RI

fragment have been unsuccessful.

3. The arrangement of minor *cox1*-containing elements in *C. cohnii* mtDNA.

In addition to the four major *Eco*RI fragments, six other *Eco*RI clones were isolated that had insert sizes that did not correspond to pc1#2, pc1#15, pc1#8 or pc1#3. Analysis of these six clones revealed that two (pc1#7 and pc1#2.2) were terminally truncated versions of the 5.7- and 5.1-/3.5-kbp fragments, respectively. Two other inserts (pc1#11 and pc1#26.1) possessed internally deleted regions (Fig. 3.15). In particular, the region that is missing in pC1#11 is 20 nt in length. This corresponds to the second (D68.2) of three tandemly duplicated direct repeats (Fig. 3.9B). Finally, two constructs (pc1#32 and pc1#11) contained unique 5' sequence flanking the *cox1* gene. Comparisons of the points of divergence in the latter two recombinant clones and in flank-1 and flank-2 revealed no identity among them. In fact, the start of the unique region occurred within the CR region (Fig. 3.15). These unique *cox1*-flanking regions also contained repetitive sequence elements similar to those seen in all other *cox1*-flanking regions. Again, these minor elements appear to have arisen by recombination events. However, their low copy number suggests that, individually, they comprise a minor component of the mitochondrial genome. These results suggest that while 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 fact, one of the previously published versions of *C. cohnii cox1* (Inagaki *et al.*, 1997) does not resemble any of the constructs described here outside of conserved protein-coding domains, implying that this copy of *cox1*, which was cloned via PCR, may

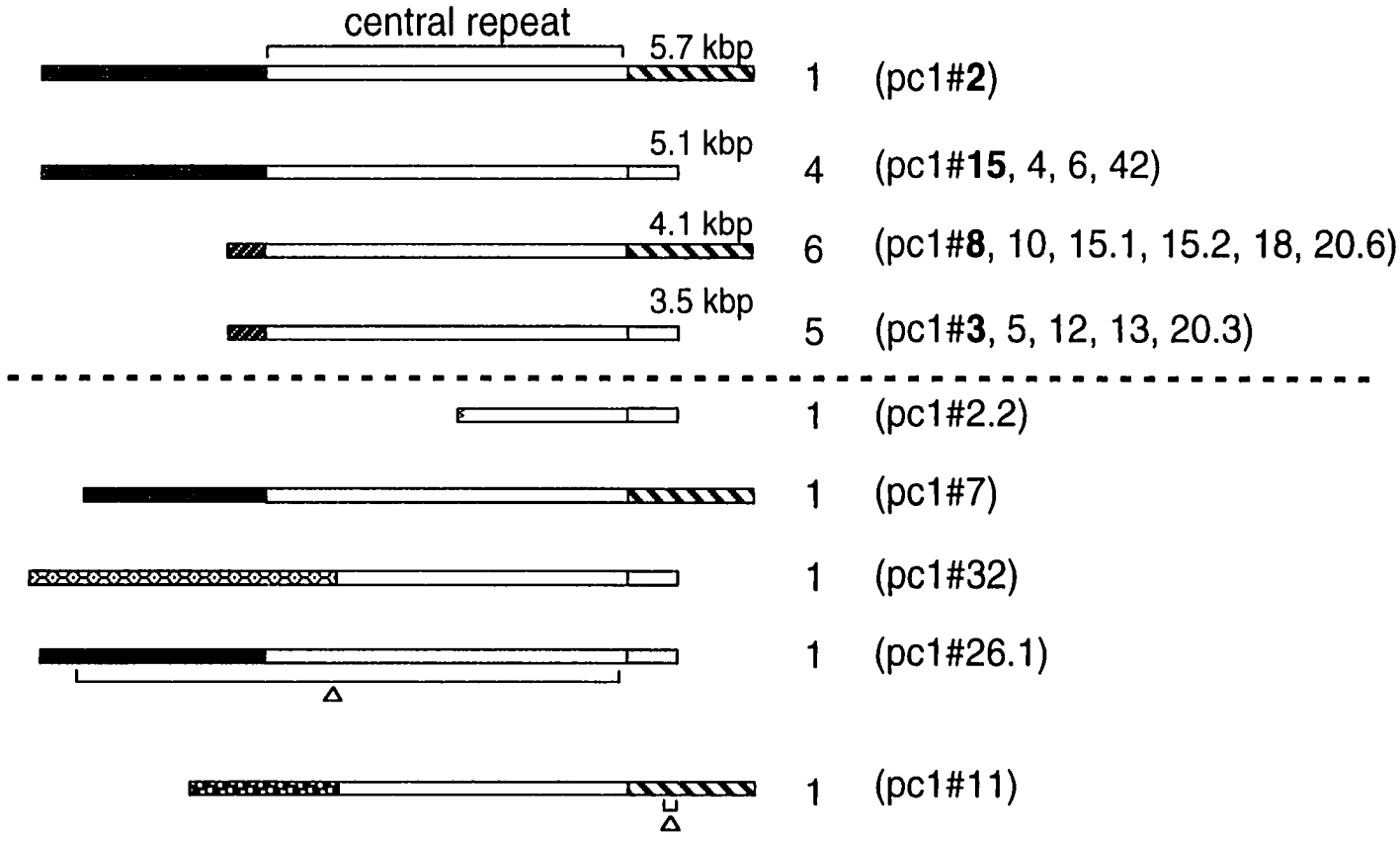


Fig. 3.15. Schematic representation of all *cox1 Eco*RI fragments that were cloned and characterized from *C. cohnii* mtDNA in this study. To the right of each map are the number and name of all clones that contain a particular insert. Rectangles that have the same shading denote flanking regions that are shared by those constructs. The central repeat is delineated by the open rectangles. Internally deleted regions are indicated by a triangle. The dotted line separates the four major *Eco*RI fragments characterized here (above) from the minor variants identified by cloning (below).

represent a minor component of *C. cohnii* mtDNA.

4. The *cox1* ORF.

The deduced amino acid sequence of the stably expressed *C. cohnii cox1* gene predicts a protein of 536 residues, beginning with a standard initiation codon, ATG (Met), and terminating in a TAA stop codon. Fig. 3.11 shows an alignment of COX1 sequences from *C. cohnii*, *P. falciparum*, *T. pyriformis*, *Trypanosoma brucei*, *Beta vulgaris*, *Neurospora crassa* and *Bos taurus*. The highest amino acid identity for the *C. cohnii* protein is with *P. falciparum* COX1 (59%) whereas the lowest value (32%) is with *T. pyriformis* COX1 (Table 3.3). Notably, COX1 from the more distantly related organisms *T. brucei* (kinetoplastid), *B. vulgaris* (plant), *N. crassa* (fungus) and *B. taurus* (animal) show intermediate levels of identity (36%, 45%, 44% and 40%, respectively) with *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 (Fig 3.11, dots) that bind heme *a*, Cu_B and heme a_3 (reviewed in Trumpower & Gennis, 1994). 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.

Table 3.3. Pairwise comparisons of inferred COX1 amino acid sequences.

	C_coh	P_fal	T_pyr	T_bru	B_vul	N_cra	B_tau
C_coh	100	59	32	36	45	44	40
P_fal	59	100	36	40	50	49	46
T_pyr	32	3	100	34	38	38	39
T_bru	36	40	34	100	47	45	46
B_vul	45	50	38	47	100	74	72
N_cra	44	49	38	45	74	100	68
B_tau	40	46	39	46	72	68	100

Values in the table represent percent identity. Names and protein accession numbers are referenced in legend of Fig. 3.11.

C. Organization of the *cox3* gene in *C. cohnii* mtDNA.

1. 3'-RACE PCR clones of *cox3*.

As previously mentioned, while mapping the 3' end of *cox1* (using 3'-RACE methods), five clones were isolated that did not exhibit *cox1* homology. Four of these clones contained a small region that was homologous to *cox3* (Table 3.2). While the insert is 167 nt long, only the last 57 nt (19 amino acid residues) are homologous to *cox3* (Fig. 3.16). One interesting feature of these constructs is that the DNA sequence does not appear to encode a stop codon: the last DNA-encoded nucleotide is a T residue. However, a UAA stop codon is created in the corresponding transcript by the addition of a poly(A) tail (Fig. 3.17A). Although such a situation has not been reported previously in protists, creation of stop codons by post-transcriptional oligoadenylation has been reported for a number of animal mitochondrial RNAs (Anderson *et al.*, 1981; Ojala *et al.*, 1981).

The pc3#17 clone was used to prepare a *cox3* Southern hybridization probe. Hydrolysis of pc3#17 with *Bam*HI produces a 300-nt fragment that covers the 3' end of *cox3* (labeled c3-1 in Fig. 3.17B). Colony screening using the c3-1 probe identified three colonies (pc3#2, pc3#1, and pc3#5).

Only a portion of the c3-1 probe was homologous to *cox3*. Therefore a second probe (labeled cox3 265-270 in Fig. 3.17B), which only contained sequence that was clearly homologous to *cox3*, was generated. A 58-nt radiolabeled PCR product was prepared using clone pc3#2 as template with cox3 265 and cox3 270 as primers (Table 2.1) in the presence of [α-^{32}P]dATP (Fig. 3.17A). This fragment was purified by polyacrylamide gel electrophoresis and used in hybridization analysis.

Fig. 3.16. C-terminal alignment of COX3 amino acid sequences. Boxed segments indicate regions that show >60% amino acid identity among all compared taxa. Taxa included in the alignment are: C_coh, *Crypthecodinium cohnii*; P_fal, *Plasmodium falciparum* (G1197815); T_par, *Theileria parva* (AAA73631) T_bru, *Trypanosoma brucei* (A28782); R_ame, *Reclinomonas americanus* (AAA11871); B_vul, *Beta vulgaris* (BAA99481); N_ cra, *Neurospora crassa* (P00422); and B_taurus, *Bos taurus* (AAB59274). Numbers within parentheses refer to EMBL accession numbers. Two *C. cohnii* samples are shown. C_coh RT denotes amino acid sequence inferred from 3'-RACE clones (Fig. 14, labeled pc3#17, 11, 12 and 13). C_coh DNA marks COX3 sequence inferred from mtDNA clones pc3# 1, 2 and 5). The underlined portions highlight sequence differences between mtDNA and clones generated via 3'-RACE.

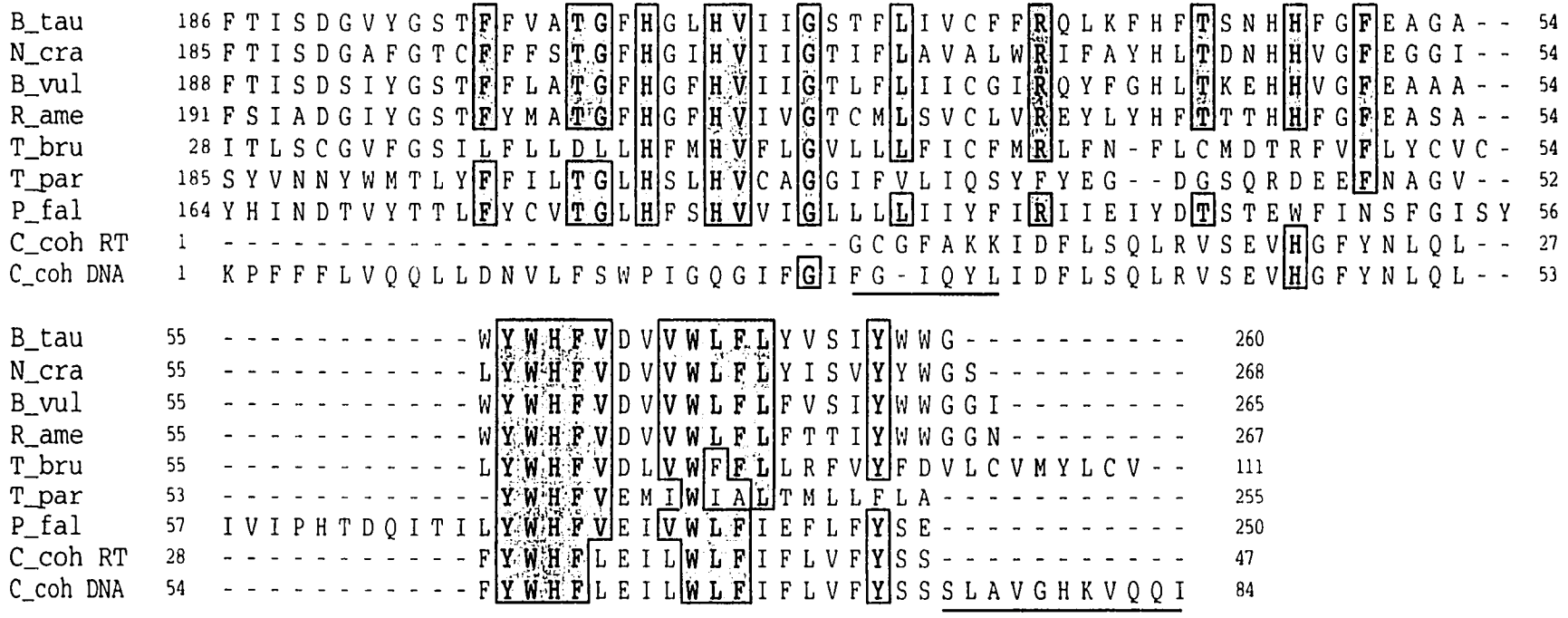

B_tau 186 F T I S D G V Y G S T F F V A T G F H G L H V I I G S T F L I V C F F R Q L K F H F T S N H H F G F E A G A - - 54
N_cra 185 F T I S D G A F G T C F F F S T G F H G I H V I I G T I F L A V A L W R I F A Y H L T D N H H V G F E G G I - - 54
B_vul 188 F T I S D S I Y G S T F F L A T G F H G F H V I I G T L F L I I C G I R Q Y F G H L T K E H H V G F E A A A - - 54
R_ame 191 F S I A D G I Y G S T F Y M A T G F H G F H V I V G T C M L S V C L V R E Y L Y H F T T T H H F G F E A S A - - 54
T_bru 28 I T L S C G V F G S I L F L L D L L H F M H V F L G V L L L F I C F M R L F N - F L C M D T R F V F L Y C V C - 54
T_par 185 S Y V N N Y W M T L Y F F I L T G L H S L H V C A G G I F V L I Q S Y F Y E G - - D G S Q R D E E F N A G V - - 52
P_fal 164 Y H I N D T V Y T T L F Y C V T G L H F S H V V I G L L L L I I Y F I R I I E I Y D T S T E W F I N S F G I S Y 56
C_coh RT 1 - G C G F A K K I D F L S Q L R V S E V H G F Y N L Q L - - 27
C_coh DNA 1 K P F F F L V Q Q L L D N V L F S W P I G Q G I F G I F G - I Q Y L I D F L S Q L R V S E V H G F Y N L Q L - - 53
B_tau 55 - - - - - - - - - - - - W Y W H F V D V V W L F L Y V S I Y W W G - - - - - - - - - - 260
N_cra 55 - - - - - - - - - - - - L Y W H F V D V V W L F L Y I S V Y Y W G S - - - - - - - - - 268
B_vul 55 - - - - - - - - - - - - W Y W H F V D V V W L F L F V S I Y W W G G I - - - - - - - - 265
R_ame 55 - - - - - - - - - - - - W Y W H F V D V V W L F L F T T I Y W W G G N - - - - - - - - 267
T_bru 55 - - - - - - - - - - - - L Y W H F V D L V W F F L L R F V Y F D V L C V M Y L C V - - 111
T_par 53 - - - - - - - - - - - - Y W H F V E M I W I A L T M L L F L A - - - - - - - - - - - 255
P_fal 57 I V I P H T D Q I T I L Y W H F V E I V W L F I E F L F Y S E - - - - - - - - - - - 250
C_coh RT 28 - - - - - - - - - - - - F Y W H F L E I L W L F I F L V F Y S S - - - - - - - - - - - 47
C_coh DNA 54 - - - - - - - - - - - - F Y W H F L E I L W L F I F L V F Y S S S L A V G H K V Q Q I 84

Figure 3.16

Fig. 3.17. Location of *cox3* probes used in hydridization analysis of *C. cohnii* mitochondrial nucleic acids. (**A**) Nucleotide sequence alignment showing location of divergence points and inverted repeats within 3'-RACE (pc3#17) and mtDNA (pc3#2) *cox3* clones. Clone designations are as described in Fig. 3.16. Dots and open rectangle indicate long stretches of nucleotide identity in the bottom sequence relative to the top one. Uppercase letters denote sequence contained within the *cox3* ORF. Closely-spaced horizontal arrows pointing in opposite directions mark the regions of inverted repeats that can potentially base pair to form the stem region of the repeat secondary structure. Long arrows denote locations of primer binding sites. The cox 442 primer only refers to sequence contained with the pc3#17 clone. Stop codons are underlined and highlighted by three asterisks (***). The last nucleotide shared by both set of clones is shown in reverse contrast. Poly(A) tail sequence is underlined and in italics. C and A residues that were added by poly(A) polymerase are offset in bold and italics. (**B**) Schematic diagram showing a portion of the pc3#2 clone. Solid black lines indicate location of cox3 265-270 and c3-1 probes. Grey line illustrates that a region of the c3-1 probe is not homologous to pc3#2. Thick zig-zag lines mark the end of *cox3* homology. Thin zig-zag lines indicate that only a portion of the pc3#2 clone is represented.

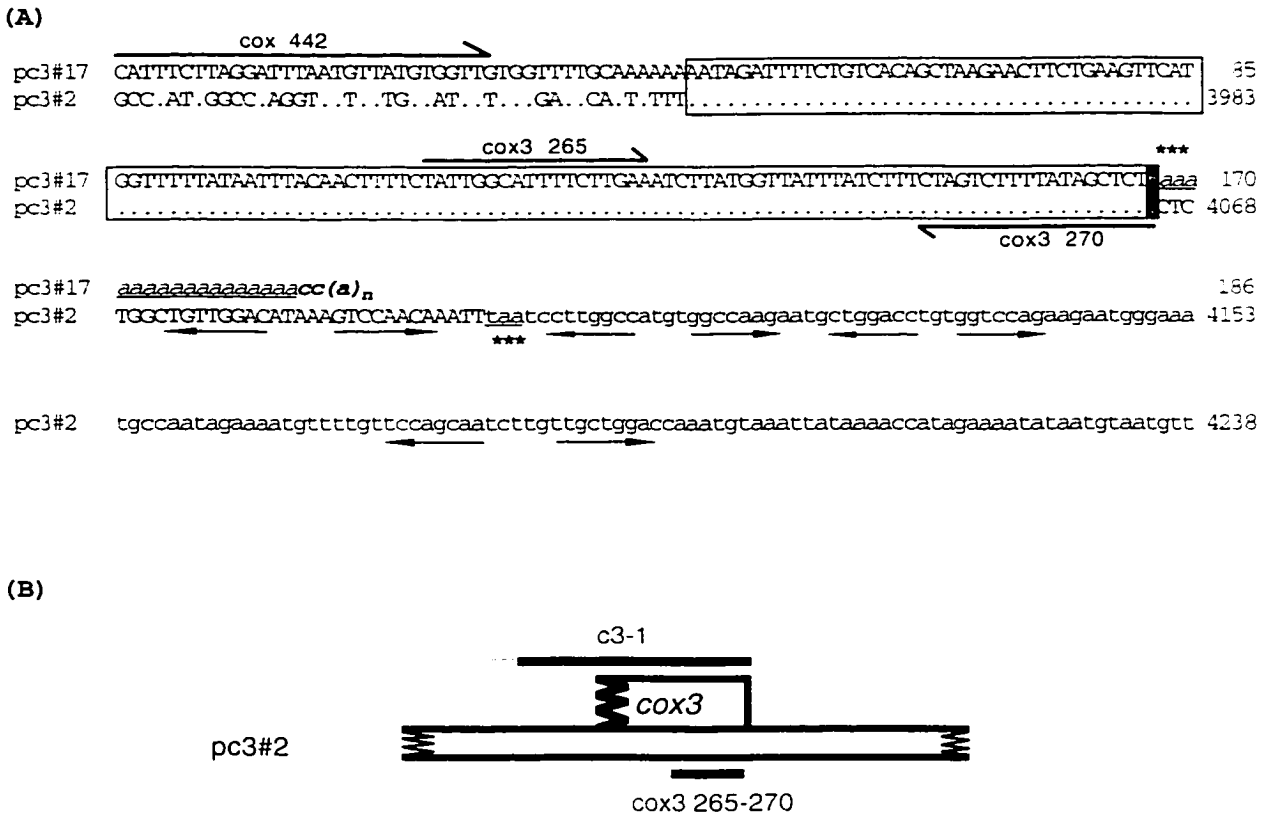

(A)
cox 442
pc3#17 CATTTCTTAGGATTTAATGTTATGTGGTTGTGGTTTTGCAAAAAAAATAGATTTTCTGTCACAGCTAAGAACTTCTGAAGTTCAT 85
pc3#2 GCC.AT.GGCC.AGGT..T..TG..AT..T...GA..CA.T.TTT..... 3983
cox3 265
***
pc3#17 GGTTTTTATAATTTACAACTTTTCTATTGGCATTTTCTTGAAATCTTATGGTTATTTATCTTTCTAGTCTTTTATAGCTCTaaaa 170
pc3#2 .....CTC 4068
cox3 270
pc3#17 aaaaaaaaaaaaaacc(a)n 186
pc3#2 TGGCTGTTGGACATAAAGTCCAACAAATTtaatccttggccatgtggccaagaatgctggacctgtggtccagaagaatgggaaa 4153
***
pc3#2 tgccaatagaaaatgttttgttccagcaatcttgttgctggaccaaatgtaaattataaaaccatagaaaatataatgtaatgtt 4238
(B)
c3-1
cox3
pc3#2
cox3 265-270

Figure 3.17

In Southern hybridizations, both probes gave the same results. In the uncut control lane, a smear between <2 kbp and 23 kbp was produced (Fig. 3.18, lane 1). Similar smears were also observed in the *Sst*I and *Xba*I lanes (Fig 3.18, lanes 4 and 5). In the *Eco*RI lane, four fragments of approximate size 4.4, 2.9, 2.8 and 2.5 kbp were observed (Fig. 3.18, lane 2), superimposed on a faint trailing smear starting at approximately 9 kbp. The stoichiometries of the four *Eco*RI bands were unequal. The three smaller bands were equally intense; however, the larger band (4.4 kbp) was fainter. The *Eco*RV lane produced a 4.0-kbp band overlaid on a faint trailing smear starting at ~10 kbp and continuing to < 2 kbp in size (Fig. 3.17, lane 3). In addition several very faint bands can be seen in this lane.

As with *cox1*, there are no *Eco*RI or *Eco*RV restriction sites within either of the *cox3* probes. Therefore, the multiple banding pattern observed by Southern hybridization suggests that this portion of *cox3* is present in different contexts that are stoichiometrically unequal in *C. cohnii* mtDNA.

Northern hybridization using the c3-1 and cox3 265-270 probes detected an RNA product (0.8 kb) in the mtRNA lane (Fig. 3.19, lanes 1 and 2). A similar-size band, displaying a less intense hybridization signal, was also found in the total RNA lane (Fig. 3.19, lane 3). Finally, no bands were seen in the lane that contained mRNA that was isolated using the PolyATract® mRNA Isolation System (Fig. 3.19, lane 4). Northern results using *cox3* probes were similar to those obtained using *cox1* probes, with the one notable difference that the *cox3* band was not as discrete. In particular, the c3-1 probe detects a smear of RNA ranging from 0.8 to 0.24 kb in size. This may suggest that the

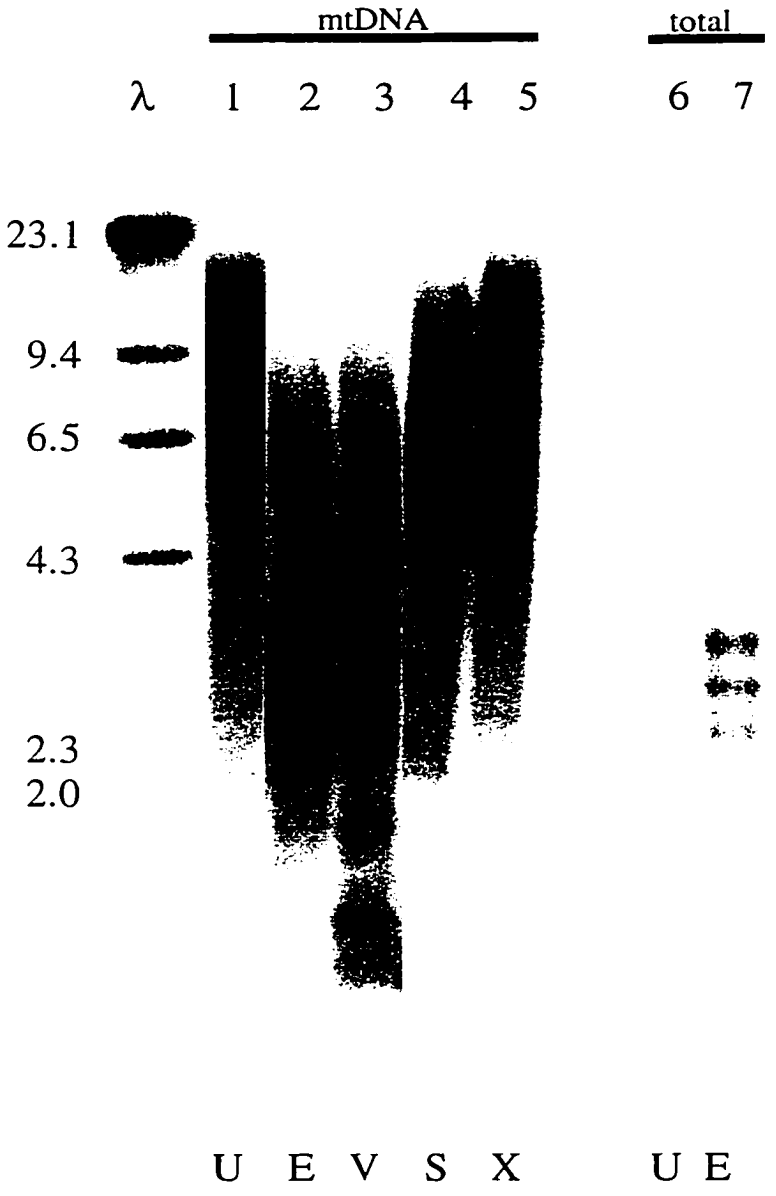


Fig. 3.18. Autoradiograms showing results of Southern blot analysis of *C. cohnii* DNA using a *cox3*-specific probe. DNA was hydrolyzed with: *Eco*RI, E (lanes 2 and 7); *Eco*RV, V (lane 3); *Sst*I, S (lane 4); or *Xba*I, X (lane 5). Undigested controls are marked with the letter 'U' (lanes 1 and 6). Blots were hybridized with radiolabeled cox3 265-270 PCR products. Hydrolysis was performed using mtDNA (lanes 1 to 5) or total DNA (lanes 6 and 7) as noted. Autoradiograms of total DNA samples were exposed for ~5 times as long as those of mtDNA samples. The positions of size marker fragments (kbp; λ DNA hydrolyzed with *Hin*dIII) are indicated to the left of the figure.

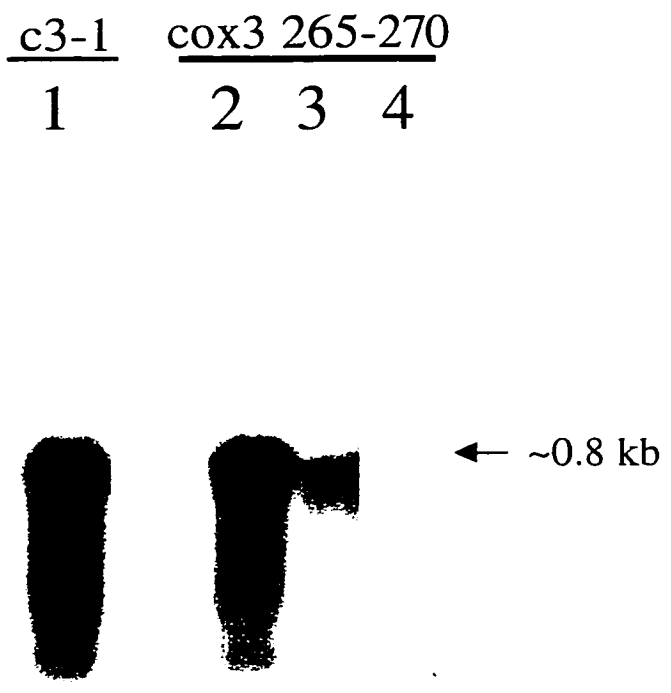


Fig. 3.19. Autoradiograms showing the results of northern hybridization analysis of mRNA isolated from *C. cohnii* using a *cox3*-specific probe. Lanes 1 and 2 contain RNA that was isolated from mitochondria. Lane 3 contains total RNA. Lane 4 contains RNA that was purified using the PolyATract® purification system. Blots were hybridized with radiolabeled cox3 265-270 PCR product or with c3-1 restriction fragment (Fig 3.17).

cox3 mRNA is particularly prone to degradation. Alternatively, *cox3* transcript processing may be less tightly regulated than *cox1* mRNAs, thereby allowing partially degraded *cox3* RNAs to accumulate *in vivo*.

A 0.8-kb *cox3* transcript is consistent with the size of *cox3* mRNAs predicted for other organisms. For example, the *cox3* gene is 750 nt long in *P. falciparum*. Northern analysis suggests that stable *C. cohnii cox3* transcripts are produced. In addition, these mRNAs contain the C-terminal portion of *cox3* found in the 3'-RACE PCR clones.

2. Characterization of *cox3*-containing clones by colony library screening.

DNA sequence analysis revealed that all three constructs (labeled pc3#2, pc3#1 and pc3#5 in Fig. 3.20) contained the same *cox3* gene fragment (C-terminal portion) as well as a 54-nt *cox1* gene fragment (N-terminal portion). With the exception of pc3#2, no other genes were found within the pc3 clones. Besides the *cox3* and *cox1* gene fragments, pc3#2 also contained a full-length *cob* gene and an *rnl* gene fragment.

The three *cox3*-containing clones (pc3#2, pc3#1 and pc3#5) have the same *cox3* fragment. This version of *cox3*, however, was slightly different than that found in the 3'-RACE PCR clones (see Figs. 3.16 and 3.17). The first in frame termination codon in the *cox3-Eco*RI clones is farther downstream (11 amino acids) than that predicted for the 3'-RACE PCR clones. This extension is also longer than that predicted for other *cox3* genes (Fig. 3.16). Notably, there are several inverted repeats surrounding the C-terminus of the *cox3* gene fragment (Fig. 3.17), similar to that seen with the stably expressed version of *cox1*.

Mapping and sequence analysis of the three '*cox3*'-containing *Eco*RI clones

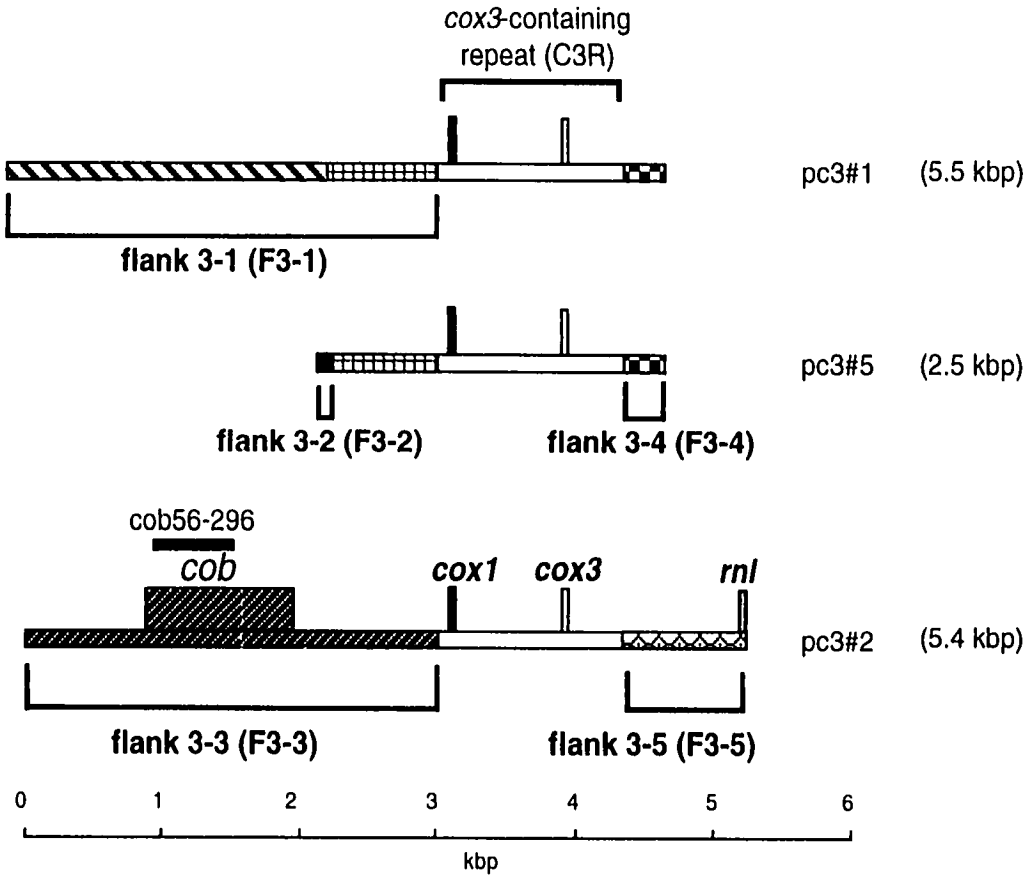


Fig. 3.20. Schematic drawings of the three '*cox3*'-containing *Eco*RI fragments that were cloned and characterized from *C. cohnii* mtDNA. Sizes and names of constructs are listed to the right of the figure. Open rectangles indicate identical regions in all three clones (central repeat cox3, C3R). Shaded rectangles denote flanking regions that contain DNA sequence different from at least one other construct (flank 3-1 to flank 3-5; F3-1 to F3-5). Large rectangles delineate the positions of the various encoded genes. The *cox1*, *cox3* and *rnl* sequences correspond to short partial gene fragments. The location of the cob56-296 probe is indicated by the solid bar. The *cob* ORF appears to be complete. A size scale is located along the bottom of the figure.

revealed a similar mtDNA arrangement as that found with *cox1*. Specifically, all constructs contained a common central repeat unit (C3R) that is shared by the three inserts (labeled C3R in Fig. 3.20). Upstream of the C3R repeat, three different flanking domains were identified (labeled F3-1, F3-2 and F3-3, Fig. 3.20). Downstream of C3R two different flanking domains were observed (F3-4 and F3-5). The C3R region is 1348 nt in length, extending 661 nt upstream of the *cox3* partial ORF and 433 nt downstream.

As with *cox1*, the boundary between the C3R central repeat and flanking domains is abrupt and maps to a specific site. These findings further suggest that *C. cohnii* mtDNA has undergone numerous rearrangements. This has resulted in the creation of a complicated array of large repeats. These repeats are not only peculiar to large gene-containing regions (i.e., *cox1*) but smaller ones as well (i.e., small *cox3* fragments).

The origin of these C-terminal *cox3* fragments is unknown. One possibility is that recombination has resulted in the creation of different versions of the *cox3* gene. Presumably, at least one version of *cox3* remains intact. Evidence in support of this conclusion is the detection of *cox3* transcripts of the predicted size. Truncated versions of *cox3* may also be created by homologous recombination, similar to that seen with *cox1*. The *cox3* data also suggest that mtDNA rearrangements are directed toward the 3' ends of genes. In fact, the C-terminal portion of *cox3* identified in this study may be analogous to the 'missing' end of *cox1* detected in Southern analysis. Finally, the *cox3* fragment spans a highly conserved COX3 domain; therefore, it is presumed that versions of COX3 that are missing this protein domain will be non-functional.

D. Organization, sequence and expression of the *cob* gene.

1. Organization based on Southern analysis of *cob*.

Based on the DNA sequence of pc3#2, a *cob*-specific probe was prepared (labelled cob56-296 in Fig. 3.20) and used in Southern analysis. Initially, a 2170-nt PCR product was amplified from clone pc3#2 using the vector-based SK and internal cob296 primers (Table 2.1). Hydrolysis of the PCR product with *Eco*RV produced a 740-nt fragment (cob56-296) that covers the 5' end of *cob*. In Southern hybridizations, the cob56-296 probe produced a smear in the undigested control, *Sst*I and *Xba*I lanes (Fig. 3.21, lanes 1, 4 and 5). In the *Eco*RI digest, one intense band (~3.7 kb) was observed (Fig. 3.21, lane 2). In addition, at least three other faint bands (~4.6, 3.2 and 2.4 kbp in size) were visible. These faint bands, however, were nearly obscured by the presence of a trailing smear between 5.0 and 2.0 kbp in size. Samples hydrolyzed with *Eco*RV produced five faint bands (4.5, 3.8, 3.5, 2.6 and 2.3 kb) overlaid on a faint trailing smear starting at ~10 kbp and continuing to < 2.0 kbp in size (Fig. 3.21, lane 3). In this lane, the three larger bands displayed approximately twice the hybridization intensity of the smaller bands. As seen in Fig. 3.21 (lanes 1, 2, 6 and 7), hybridization performed at the same time showed that the *cob*-specific probe hybridized to the mtDNA-enriched lanes more intensely than to total DNA lanes. As with *cox1* and *cox3* fragments, the *cob* gene is present in different contexts that are stoichiometrically unequal.

2. Characterization of *cob*-containing clones by colony library screening.

Southern hybridization analysis suggested that the *cob* gene arrangement in *C. cohnii* mitochondria is simpler than that seen for *cox1*. Only one predominant *cob*-

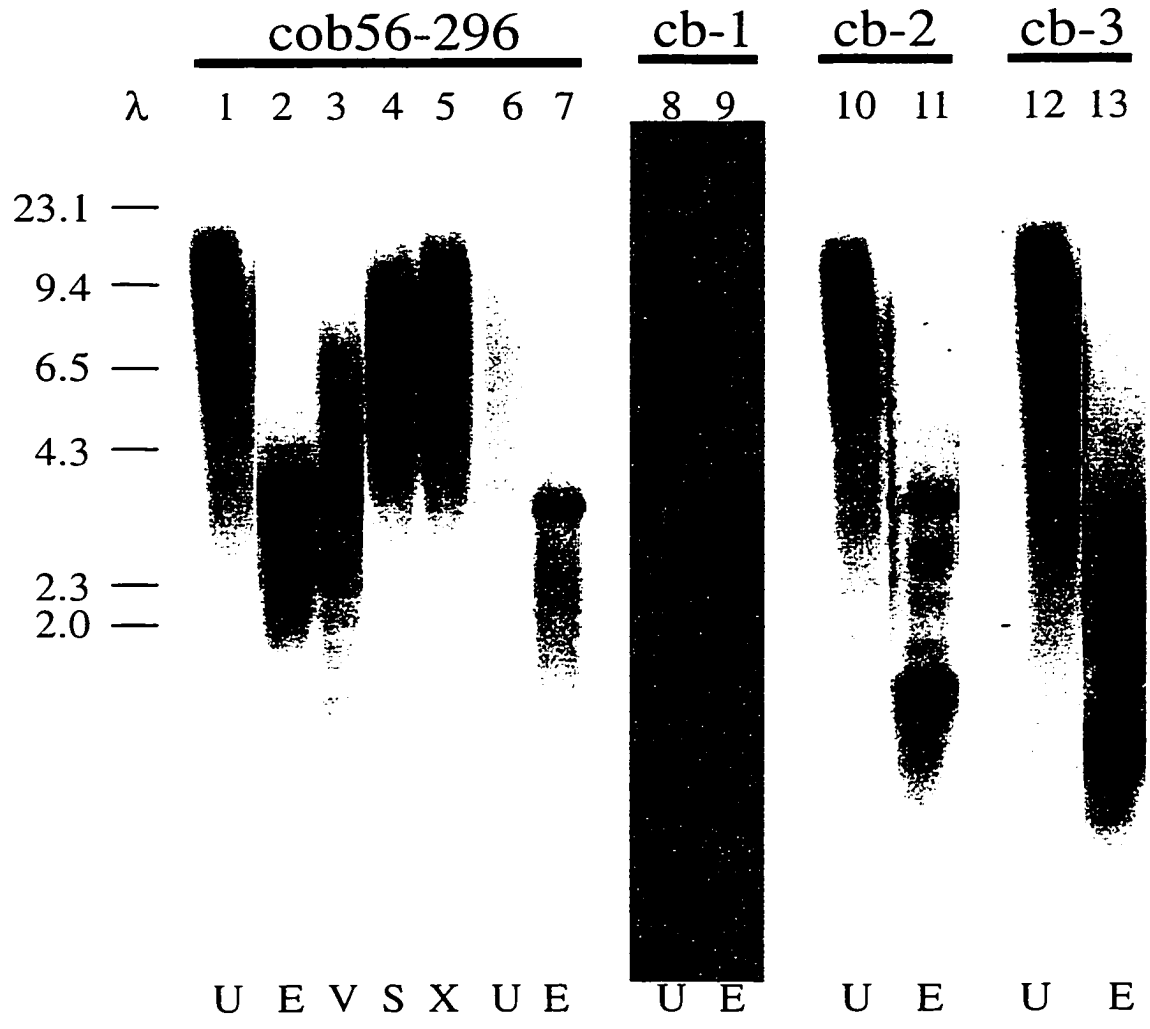


Fig. 3.21. Autoradiographic results of Southern blot analysis of *C. cohnii* DNA using *cob* probes. DNA was hydrolyzed with: E, *Eco*RI (lanes 2, 7, 9, 11 and 13); V, *Eco*RV (lane 3); S, *Sst*I (lane 4); or X, *Xba*I (lane 5). Undigested controls are present in lanes 1, 6, 8, 10 and 12 (labeled 'U'). Blots were hybridized with radiolabeled cob56-296, cb-1, cb-2 or cb-3 PCR products (Fig 3.22) as indicated. All lanes contained mtDNA with the exception of lanes 6 and 7, which contained total DNA. Total DNA samples exhibited 10% the hybridization signal intensity of mtDNA samples. The approximate positions of size marker fragments (kbp; λ DNA hydrolyzed with *Hin*dIII) are indicated to the left of the figure.

hybridizing *Eco*RI band (3.7 kbp) was seen in Southern analysis. However, several less intense bands were also detected in the *Eco*RI lane. This suggests that there is one major as well as a number of minor *cob*-containing *Eco*RI fragments. To confirm this observation, several *cob*-containing clones were isolated from an *Eco*RI clone library of mtDNA using cob56-296 as a probe. In total, 12 *cob*-containing constructs were characterized using restriction mapping and partial sequencing techniques. As shown in Fig. 3.22, insert sizes ranged from 5.4 to 2.4 kbp, with eight of the 12 clones having a 3.7-kbp insert. This result is consistent with the size of the major *Eco*RI fragment (3.7 kbp) seen in Southern analysis (Fig. 3.21, lane 2).

DNA sequence analysis of *cob* clones revealed that all clones share a central *cob*-containing repeat (labeled CBR in Fig. 3.22) whose sequence is identical among all clones. Upstream of the CBR are three different flanking domains (labeled Fb-1, Fb-2 and Fb-3 in Fig. 3.22), whereas five different flanking domains (labeled Fb-4, Fb-5, Fb-6, Fb-7 and Fb-8 in Fig. 3.22) are located downstream of the CBR. Half of the clones (pcb#1, 3, 6, 7, 10, 11) contain the same insert. Likewise, a large majority of the clones (10 of 12) have the same upstream flanking domain (Fb-2), the only exceptions being pc3#2 and pcb#8.

The *cob* gene comprises the majority of the *cob*-containing repeat (CBR). In all 12 cases the 5' end of *cob* is the same; however, there is some 3' end variability. The five downstream flanking domains, if translated, would result in the generation of three different COB proteins.

These differences occur within the C-terminal extension of COB, just downstream

Fig. 3.22. Schematic diagram showing *cob*-containing *Eco*RI fragments that were cloned and characterized from *C. cohnii* mtDNA. Sizes of inserts (in parentheses) and names of constructs are listed to the right of the figure. Open rectangles indicate identical regions in all clones (*cob*-containing central repeat, CBR). Large arrow and bolded clone names indicate the predominant *cob*-containing *Eco*RI fragment isolated (i.e. pcb#1). Similarly shaded rectangles denote regions with identical DNA sequence. Uniquely shaded boxes indicate DNA regions that are not shared with any other *cob*-containing clone. Characterization of flanking domains is based on the pcb#1 clone (Fb-1 and Fb-4). DNA regions that are different from those found in pcb#1 are assigned a different flank number (Fb-2 to Fb-3 and Fb5 to Fb-8). The open rectangles labeled '*cob*' and adjoining shaded rectangles above each map delineate the length of the *cob* ORF. The locations of three gene fragments (*cox1*, *cox3* and *rnl*) are also denoted by rectangles. Solid black lines above each map indicate the locations of four probes (cob51-256, cb1, cb2, cb3) used in Southern and northern hybridizations. A size scale is located along the bottom of the figure.

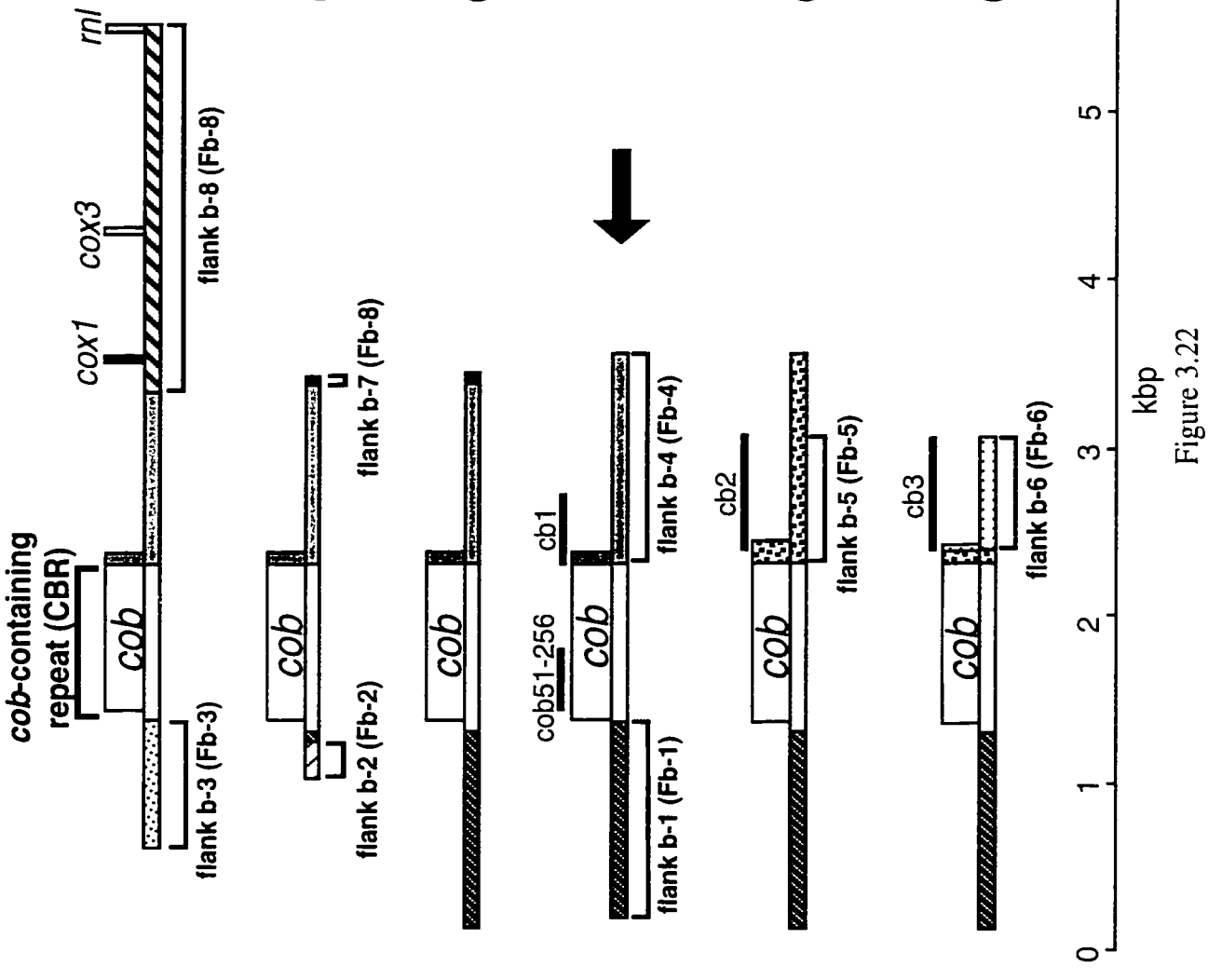

cob-containing repeat (CBR)
cox1
cox3
rnl
cob
flank b-3 (Fb-3)
flank b-8 (Fb-8)
(5.4)
pc3#2
flank b-2 (Fb-2)
flank b-7 (Fb-8)
(2.5)
pcb#8
(3.6)
pcb#5
cb1
cob51-256
flank b-1 (Fb-1)
flank b-4 (Fb-4)
(3.7)
pcb#1, 3, 6, 7, 10, 11
cb2
flank b-5 (Fb-5)
(3.7)
pcb#4, 9
cb3
flank b-6 (Fb-6)
(3.2)
pcb#2
0
1
2
3
4
5
6
kbp

Figure 3.22

of the last conserved COB domain. Notably, the version of COB that was present in the majority of the clones (pcb#1, 3, 5, 6, 7, 8, 10, 11) does not contain a sequence motif that other versions of *C. colnii* COB share with their *P. falciparum* homolog (Fig. 3.23)[1]. Instead, the COB sequence encoded by pcb#4, pcb#9 and pcb#2 is most similar to that found in other eukaryotes.

3. Variants of the *cob* gene and their expression.

Of the three *cob* variants, the longest, cob-2 (labeled cb2 in Fig. 3.23 and represented by pcb#4 in Fig. 3.24A), is 1200 nt in length and is present in clones pcb#4 and pcb#9. In this case, the last 204 nt of the ORF extend into flank-5 (Fb-5, Fig. 3.24). In contrast, the shortest *cob* gene, cob-1 (labeled cb1 in Fig. 3.23), is only 1020 nt long. The final 24 nt of this ORF extend into flank-4 (Fb-4, Fig. 3.24). The final *cob* variant, cob-3 (labeled cb3 in Fig. 3.22), is 1092 nt long and is contained within clone pcb#2. This version shares a portion of flank-5 (75 nt) with cb2; however, the end of the ORF extends into flank-6 (Fb-6, Fig. 3.24).

Unlike cob-2 and cob-3, the C-terminal stretch of amino acids encoded by cob-1 is shorter than the C-terminal region of other COB proteins (Fig. 3.23). Yet despite these C-terminal difference, the three COB variants appear to encode all of the expected, functionally significant protein-coding domains.

To confirm that the version of *cob* seen in the majority of *cob* clones (cob-1) is indeed representative of the single major *Eco*RI band visualized by Southern hybridization, the analysis was repeated using three different probes (cb-1, cb-2 and cb-3)

---

[1] The same 3-amino acid stretch is also found in *Toxoplasma gondii* COB (g2564669).

Fig. 3.23. Alignment of COB amino acid sequences. Shaded boxed segments indicate regions that show >60% amino acid identity among all compared taxa. Taxa included in the alignment are: C_coh, *Crypthecodinium cohnii*; P_fal, *Plasmodium falciparum* (M76611); T_pyr, *Tetrahymena pyriformis* (AAD41943); *Trypanosoma brucei* (X00017); B_vul, *Beta vulgaris* (BAA99352); N_cra, *Neurospora crassa* (AAA31961); and B_taurus, *Bos taurus* (P00157). Numbers within parentheses refer to EMBL accession numbers. The three *C. cohnii* COB variants identified in this study are labeled cb1 (pcb#7), cb2 (pcb#4) and cb3 (pcb#2). Sequence differences among *C. cohnii* COB ORFs are highlighted (underlined). Sequence simply labeled C_coh is identical among all *C. cohnii* isolates. The three potential methionine start sites are numbered and marked by circles. Two *C. cohnii* tryptophan codons that are encoded by TGA are denoted with small open rectangles and asterisks. The large open rectangle denotes a small region of amino acid sequence that is identical among *P. falciparum* and two of the three *C. cohnii* COB variants (cb2 and cb3).

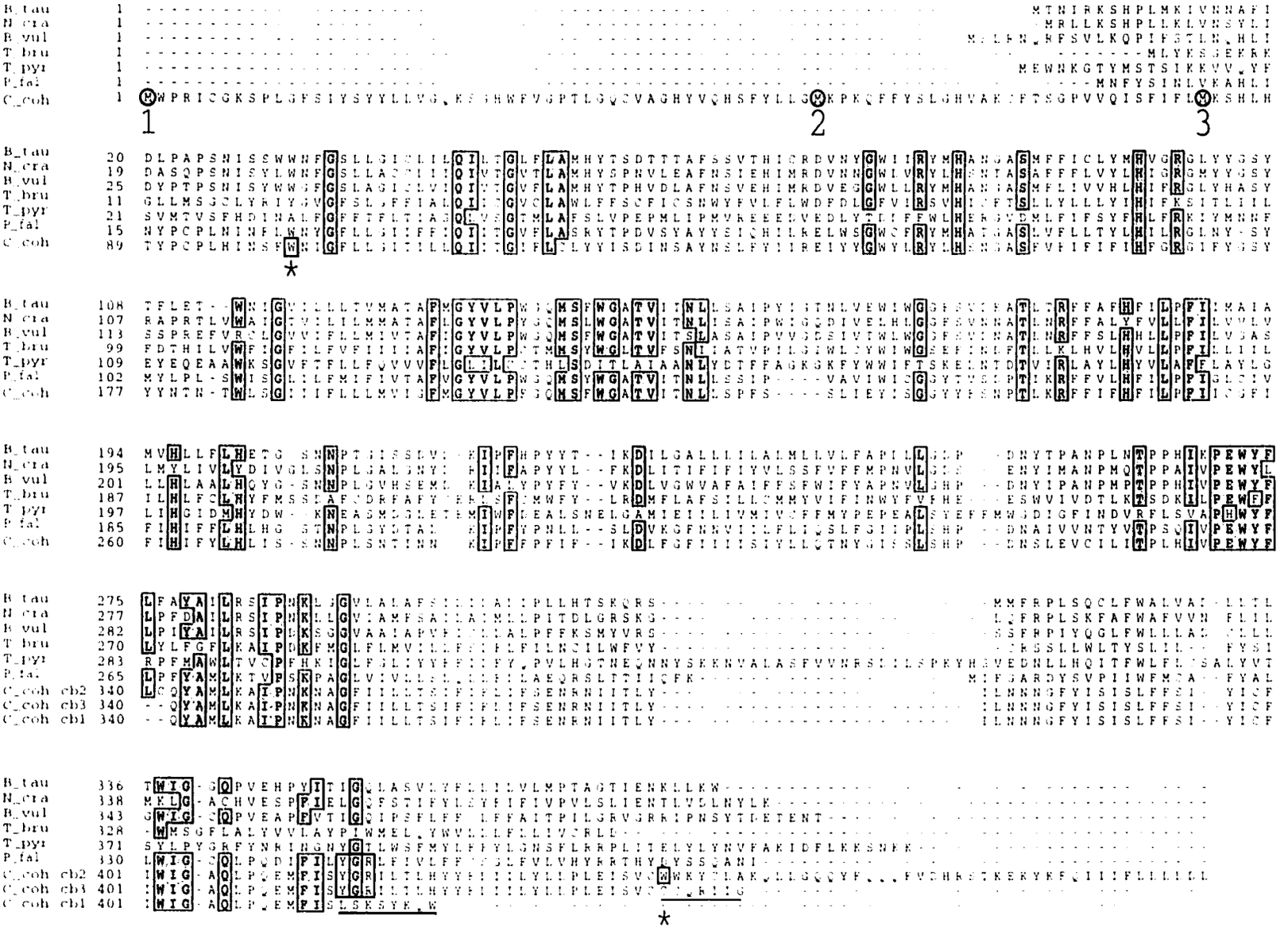

B_tau
N_cra
B_vul
T_bru
T_pyr
P_fal
C_coh
C_coh cb2
C_coh cb3
C_coh cb1
1
2
3
*
*

Figure 3.23

Fig. 3.24. Nucleotide sequence showing the locations of divergence points and inverted repeats within *cob*-containing *Eco*RI fragments cloned from *C. cohnii* mtDNA. (**A**) Alignments show the 3' end of the central repeat (enclosed by a large rectangles) leading into the 5' ends of flank b-4, flank b-5 and flank b-6. Identical sequence shared by pcb#4 and pcb#2 is enclosed by a smaller rectangle. Nomenclature of flanking regions and clones is as described in Fig. 3.22. The divergence points between flanks b-4 and b-5 (marked I) as well as flanks b-5 and b-6 (marked by II) are indicated. (**B**) Nucleotide sequence showing the locations of inverted repeats in the central repeat region close to the predicted *cob* initiation codon (shown in bold, underlined and numbered three). A second possible intiation codon is marked with the number two as referenced in Fig. 3.23. Sequence co-ordinates are relative to the inferred *cob* initiation codon. Uppercase letters denote sequence contained within the *cob* ORF. Stop codons are underlined and highlighted by three asterisks (***). Closely-spaced horizontal arrows pointing in opposite directions located below the sequence mark the regions of inverted repeats that can potentially base pair to form the stem region of the repeat secondary structure.

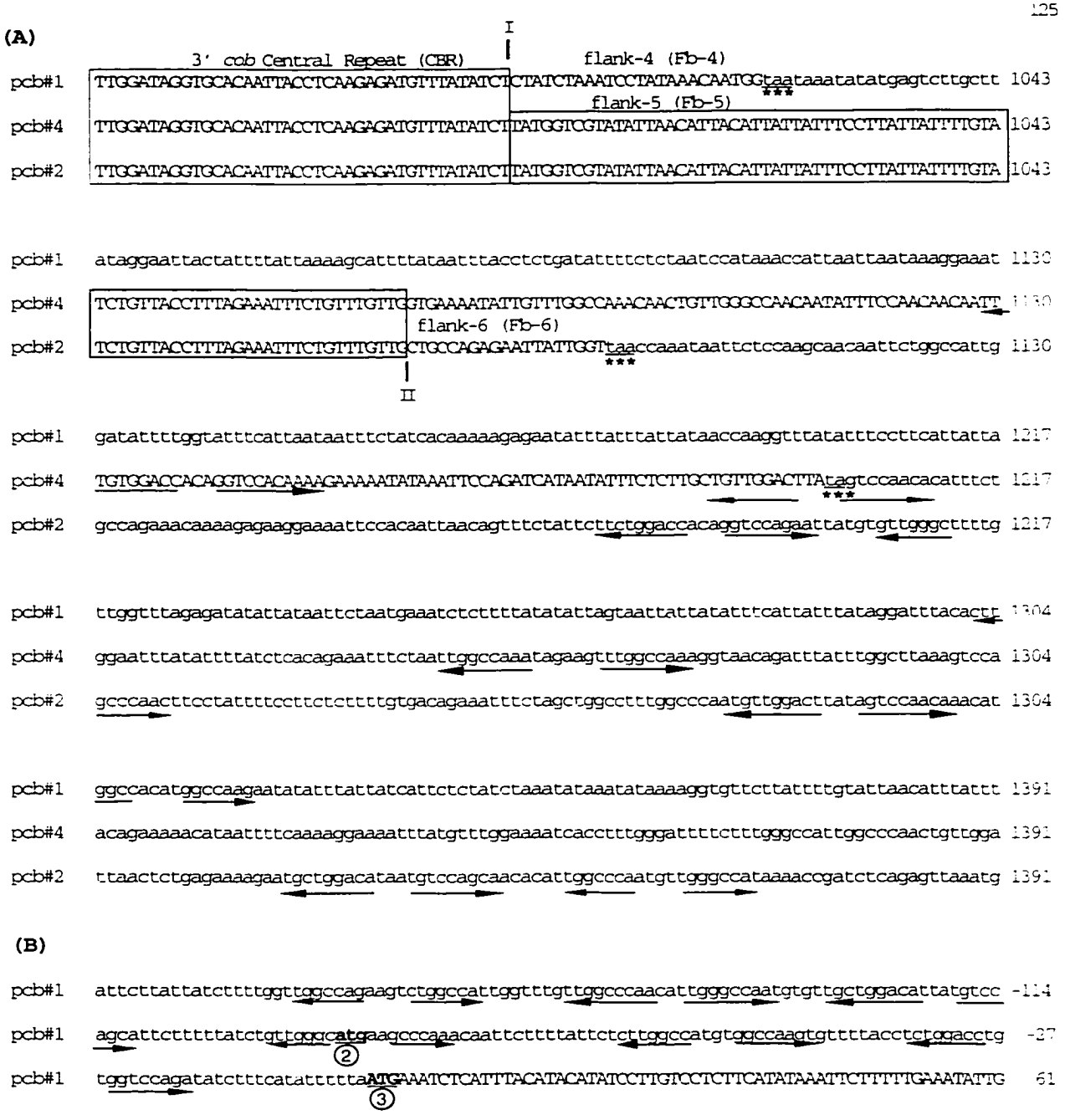

(A)
I
3' cob Central Repeat (CBR)
flank-4 (Fb-4)
flank-5 (Fb-5)
flank-6 (Fb-6)
II
pcb#1
pcb#4
pcb#2
(B)
1043
1130
1217
1304
1391
-114
-27
61

Figure 3.24

specific for the *cob* gene. The first probe, cb-1 (Fig. 3.22), is 432 nt in length and was amplified from clone pcb#7 using P51 and P25 primers (Table 2.1). The second probe (cb-2, Fig. 3.22), amplified from clone pcb#4 using the vector-based M13 forward and internal P49 primers (Table 2.1), is 1010 nt in length. Similarly, a 716-nt PCR product (cb-3, Fig. 3.22) was generated from clone pcb#2 using P50 (Table 2.1) and M13 reverse vector-based primers.

Southern hybridization using the cob-1-specific probe (cb-1, Fig 3.21) gave the same *Eco*RI banding pattern as was seen with cob56-296 (Fig. 3.21, lanes 2), in contrast to cob-2- and cob-3-specific probes. For example, cb-3 detected 10 bands (ranging between 3.7 kb and 0.5 kbp in size) in the *Eco*RI lane (Fig. 3.21, lane 13). The cb2 probe identified three bands, the least intense of which had a size that is consistent with the majority of the *cob*-containing clones (3.7 kbp). In addition, two intense bands were visible, ~1.0 and 0.8 kbp in size (Fig. 3.21, lane 11). Therefore, the major *cob*-containing *Eco*RI band encodes the cob-1 version of *cob*.

To determine whether all *cob* genes are functional in *C. cohnii*, northern hybridization experiments were used to test for stable transcripts. Surprisingly, probes specific for cob-1 (cb-1), cob-2 (cb-2) and cob-3 (cb-3) did not hybridize to northern blots (Fig. 3.25, lanes 4, 5 and 6). However, all three probes produced hybridization signals in Southern analysis performed at the same time (Fig. 3.21, lanes 9, 11, 13). This may suggest that stable transcripts are not produced from these ORFs. In contrast, reprobing of northern filters using an internal *cob* probe (cob56-296) produced a very weak 1.3-kb hybridization signal (Fig. 3.25, lanes 1 and 2). These results suggest that the *cob* gene is

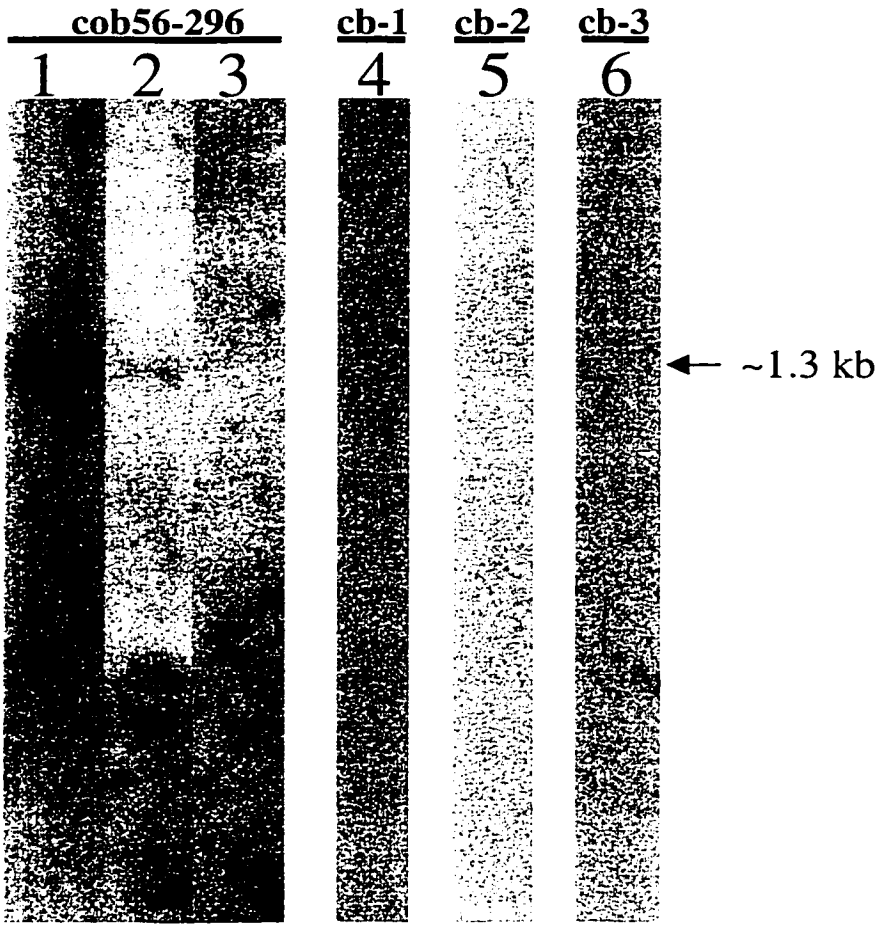


Fig. 3.25. Autoradiograms showing the results of northern hybridization analysis of *C. cohnii* mRNA using *cob* probes. Lanes 1, 4, 5 and 6 contain RNA that was isolated from mitochondria prepared by subcellular fractionation. Lane 2 contains total RNA, whereas lane 3 contains mRNA that was purified using the PolyATract® purification system. Blots were hybridized with radiolabeled cob56-296, cb-1, cb-2 and cb-3 PCR products (as marked in Fig. 3.22). The approximate size of the *cob* transcript is noted to the right of the figure.

expressed, producing a transcript of the predicted size. However, none of the versions of *cob* identified in this study appears to produce a stable transcript. Alternatively, the 3' mRNA processing site may be very close to the end of the *cob* ORF, providing the probes very little complementary sequence with which to hybridize (only 24 and 31 nt, in the case of the cb-1 and cb-3 probes).

Examination of the 3' ends of all three versions of *cob* reveals regions of sequence that can potentially fold into hairpin structures (Fig. 3.24A). However, an elaborate secondary structure, as seen with stable *cox1* transcripts, cannot be modeled. This observation supports the conclusion that the three versions of *cob* investigated in this study are not stably expressed. However, these findings have to be interpreted with caution. Because the internal *cob*-specific probe produced only a weak signal, variants of *cob* may support the production of only low-abundance transcripts (due to a high rate of turnover) that are not easily detectable.

Examination of DNA sequence surrounding the 5' end of *cob* revealed a situation similar to that found for *cox1*. In particular, numerous inverted repeats are present just upstream and possibly within the N-terminal portion of the ORF (Fig. 3.24B). Examination of repetitive sequences within the *cob*-containing clones revealed a lower number of repeats compared to *cox1*-containing clones. However, more than 50% of the inverted repeats identified occur near the 5' end of *cob*. Moreover, one of the putative *cob* start codons (labeled 2, Fig. 3.24B) occurs within the loop portion of a hairpin structure, identical to that found for *cox1*. These findings further support the hypothesis that these hairpins may be important in mRNA transcription and/or processing ie., these putative

mRNA processing sites, which occur near the initiation and termination codons, may play a role in directing the production of gene-sized transcripts.

4. *cob* ORF.

Analysis of the *cob* ORF suggests that there are three potential ATG (methionine) initiation codons (Fig. 3.23), all of which are in frame. However, examination of COB amino acid alignments suggests that the third initiation codon (nucleotide position 1415 based on pcb#7) is likely the actual start codon. This would produce a COB N-terminal extension that is consistent in length with that found in other organisms (Fig. 3.23).

Characterization of *cob*-containing clones suggests that the deduced COB protein is between 340 and 400 amino acid residues long. Two versions of *cob*, cob-1 and cob-3, terminate using the TAA stop codon; however, cob-2 uses TAG. All three ORFs are continuous and encode a protein of expected length. Similar to COX1, the highest amino acid identity exhibited by *C. cohnii* COB is with its *P. falciparum* homolog (48%, Table 3.4). Likewise, the lowest identity value (21%) is with *T. pyriformis* COB, with the more distantly related organisms showing intermediate levels of identity to *C. cohnii* COB.

For the most part, the universal genetic code is used by *C. cohnii* to translate COB. The ORF contains three tryptophan codons (W) (five in cob-2), that are translated using TGG. Two tryptophan residues, however, are specified by TGA, one of the standard stop codons. The first occurs in a region that is common to all three versions of *cob* (amino acid position 18, Fig. 3.23). The other is specific to cob-2 and occurs at residue position 357 (Fig. 3.23). Based on this information, *C. cohnii* does not appear to adhere to the standard genetic code, even though *cox1* uses TGG exclusively to encode

Table 3.4. Pairwise comparisons of inferred COB amino acid sequences.

	C_coh	P_fal	T_pyr	T_bru	B_vul	N_cra	B_tau
C_coh	100	48	21	27	37	34	43
P_fal	48	100	20	27	43	37	44
T_pyr	21	20	100	19	21	19	19
T_bru	27	27	19	100	24	22	26
B_vul	37	43	21	24	100	52	56
N_cra	34	37	19	22	52	100	51
B_tau	43	44	19	26	56	51	100

Values in the table represent percent identity. Names and protein accession numbers are as referenced in Fig. 3.23.

tryptophan.

E. Sequence and expression analysis of *rnl* gene fragments.

1. LSUG rRNA.

a. Comparison of the primary sequence of LSUG rRNA. The pc3#2 clone contains a 99-nt gene fragment that is homologous to mitochondrial *rnl* genes (Fig. 3.26). In particular, this fragment shares ~80% sequence identity with a similar-size (104 nt) rRNA gene (LSUG) identified in *P. falciparum* mtDNA. This value is much higher than the 54% nucleotide identity shared between *C. cohnii* LSUG and the same region of the nucleus-encoded LSU rRNA gene from *Prorocentrum micans*. The LSUG rRNA gene codes for a portion of the peptidyltransferase center, a region that is highly conserved among a wide spectrum of organisms.

C. cohnii LSUG can be folded into a secondary structure that conforms to the basic model adopted for *E. coli* 23S rRNA (Fig. 3.27A and B). This secondary structure is also very similar to that predicted for the homologous *P. falciparum* sequence. In fact, the middle portion (positions 10-60) contains only four differences compared to its *P. falciparum* counterpart, two of which give rise to compensating base pair changes in a paired stem region (Fig. 3.27A).

Two oligonucleotides were ^{32}P-end-labeled and used as probes in Southern hybridization. The first, LSUG3 (Table 2.1), corresponds to a sequence located near the 3' end of the LSUG gene (Fig. 3.26, labeled LSUG3). The second primer, LSUG1 (Table 2.1), also corresponds to a sequence near the 3' end of the LSUG gene fragment, just

```
mt   C_cohG     1 ---------* TTGTAGAAGA TTCCATTGGA AGGAGAGTTC ACCCTTAGAA GCGATCCGTG
mt   P_falG     1 -----*GAAC ...A.C...G .......... .T........ ...G...... .....G....
mt   T_par      1 ---------- ---------- ------*..T .T........ T.ATA.TA.T AT.G.A....
mt   T_pyr   2043 CATCCTGATG G..G.C..TC .AT..AG..T CC.GCT.... G..GG.TA.. .T.G.A....
mt   R-ame   2385 CATCCTGGAG C..C.....G .....AG..T TT.GCT.... G..AA.TA.. .T.G.A....
mt   B-vul   2989 CATCCTGGGG ...A.....G .C...AG..T TC.GTT.... G..GA.TC.. .T.G.A....
n    P_mic   2922 TATCATTGAG AC.C.....T CC.A.AGT.T C...TT.... ....GCCA.T AG.GAA....
     E.coli  2517 CATCCTGGGG C..A..T..G .C...AG..T .T.GCT.... G..A..TA.. .T.G.A..C.
                                   LSUG4                              LSUG1

mt   C_cohG    51 AGCTGGGTTA ATAGCGTCCA GCGACAGTTT CTAACCAATG GTCCGAATT*_        99
mt   P_falG    55 .......... .G.A....TT .A.G...... G.TC..T..C TA...TT..A*       104
mt   T_par     34 ..T....... .G.A..C.GC AA.G.T...C G.TC..T..C TAGAT.T*---        80
mt   T_pyr   2104 .........T .A.A....GT .A........ GGTC..T..C TGTT.T.A         2154
mt   R_ame   2446 .........T .G.A....GT .A.......C GGTC..T..C TG...TTGGTG      2495
mt   B_vul   3050 ..T......T .G.A....GT .A.......C GGTT..T..C TA...TTGGTG      3100
n    P_mic   2983 .........T .G.C....GT .A.....G.. AGTTTT.CCC TA-.TG..GAA      2503
     E.coli  2578 .........T .G.A....GT .A.......C GGTC..T..C TG...TGGGCG      2627
                                        LSUG3
```

Fig. 3.26. Nucleotide sequence alignment of the *rnl* gene (LSUG portion). Dots (.) denote the same nucleotide as shown for E_col. Numbers to the right and left of the figure show nucleotide position based on the start of the gene. Inferred gaps are indicated by dashes (-). Asterisks (*) denote the predicted ends (5' and 3') of the LSUG gene in C_cohG, P_falG and T_par4. Taxa included in the alignment are: E_col, *E. coli* (D12649); P_mic, *Prorocentrum micans*, (X16108);B_vul, *B. vulgaris* (9558480); R_ame, R. americana (AF007261); T_pyr, *T. pyriformis* (M58010); T-par4, *T. parva* (Z23263); P_falE, *P. falciparum* (M76611); and C_cohE, *C. cohnii*. Numbers within parentheses refer to EMBL accession numbers. The location of three primers (LSUG1, LSUG3 and LSUG4) is marked by arrows. Note: The 3′ ends of *C. cohnii* and *P. falciparum* LSUG rRNAs have been experimentally determined whereas rRNA end points for *T. parva* are inferred.

Fig. 3.27. Diagrams showing the predicted secondary structure of mitochondrial and bacterial LSU rRNAs. (**A**) The inferred secondary structure of LSUG and LSUE rRNAs from *C. cohnii*. RNA sequence was deduced from DNA and cDNA sequences. The 5' and 3' end of each rRNA fragment is noted. The letter 'A', enclosed in a circle denotes poly(A) addition to rRNAs based on 3'-RACE PCR analysis. Letters enclosed in rectangles denote a subset of nucleotide differences with the corresponding *P. falciparum* rRNA species. (**B**) Schematic diagram depicting the potential secondary structure of *E. coli* 23S rRNA. Four consecutive dots (....) indicate that a portion of the structure has been omitted. Wireframe highlights rRNA regions that have been identified in *C. cohnii* mitochondria. Numbers denote rRNA nucleotide position starting at the 5' end of the gene.

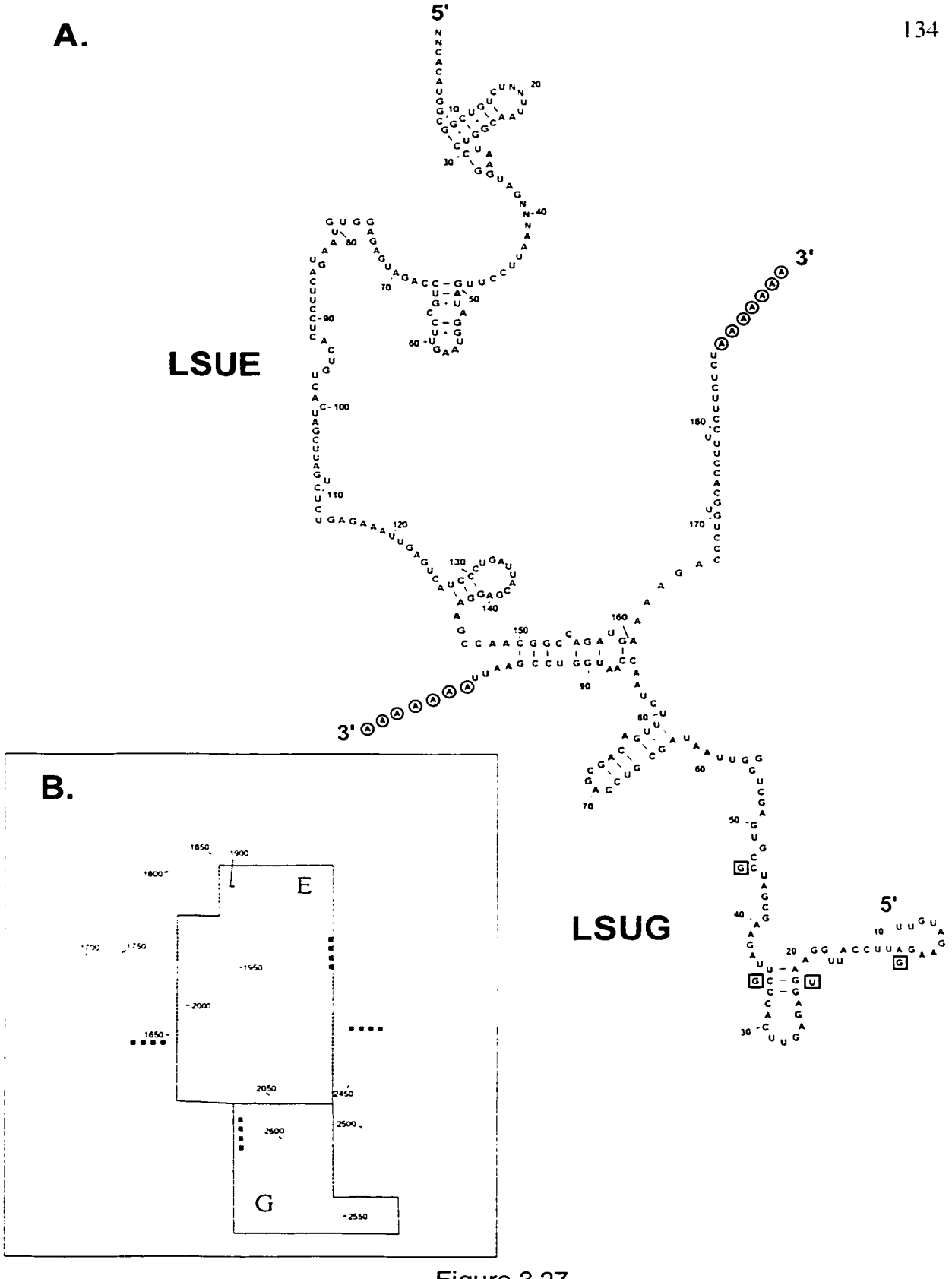

A.
5'
LSUE
3'
3'
LSUG
5'
B.
E
G
1850
1900
1800
1750
1950
2000
1650
2050
2450
2500
2600
2550

Figure 3.27

upstream of the LSUG3-specific sequence (Fig. 3.26, labeled LSUG1).

To increase hybridization signal strength, a fourth probe (82 nt) was prepared using radiolabeled PCR products. The LSUG3-4 probe was generated using pc3#2 as amplification template and LSUG3 and LSUG4 primers (Table 2.1). Following amplification, PCR products were purified in an 8% polyacrylamide gel.

In Southern hybridization experiments the LSUG3-4 probe produced a smear in the unhydrolyzed control lanes (Fig. 3.28, lane 1). In the *Eco*RI digest, seven fragments (~3.5, 2.6, 2.5, 1.9, 1.8 1.0 and 0.9 kbp) were observed (Fig. 3.28, lane 2). The stoichiometries of the seven *Eco*RI bands are unequal, with the three smaller bands exhibiting a greater intensity than the others. Total DNA hydrolyzed with *Eco*RI exhibited a much weaker hybridization signal compared to the mtDNA-enriched fraction (Fig. 3.28, lane 3). As with *cob* and *cox1*, the presence of multiple *rnl*-hybridizing bands in the *Eco*RI digest suggests that the LSUG gene in *C. cohnii* is present in different contexts that are stoichiometrically unequal.

b. Expression of LSUG rRNA. RNA extracted from mitochondria is of high quality with little contaminating cytosolic SSU or LSU rRNAs (Fig. 3.29, lanes A to C). Unfortunately, detection of small mitochondrial rRNAs by ethidium bromide staining is complicated by the presence of contaminating small nuclear RNAs (snRNAs) (Fig. 3.29, lane A).

Northern analysis using LSUG-specific probes (LSUG1, LSUG3 and LSUG3-4) revealed that this gene produces a single RNA species ~108 nt in size (Fig. 3.29). A band (~108 nt) was also seen in the total RNA lane, but the hybridization signal from this

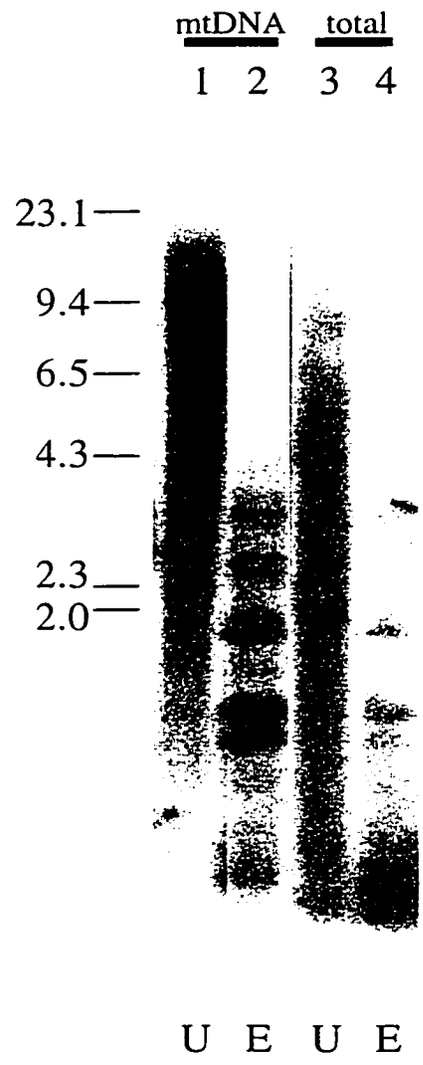


Fig. 3.28. Autoradiograms showing results of Southern hybridization analysis of *C. cohnii* DNA using an LSUG-specific probe. DNA was hydrolyzed with E, *Eco*RI (lanes 2 and 4). Unhydrolyzed control lanes are labeled with the letter 'U' (lanes 1 and 3). Blots were hybridized with radiolabeled LSUG3-4 PCR product. Hydrolysis was performed using mtDNA (lanes 1 and 2) or total DNA (lanes 3 and 4) as indicated. Total DNA lanes were exposed for the same length of time as mtDNA samples. The positions of size marker fragments (kbp; λ DNA hydrolyzed with *Hin*dIII) are indicated to the left of the figure.

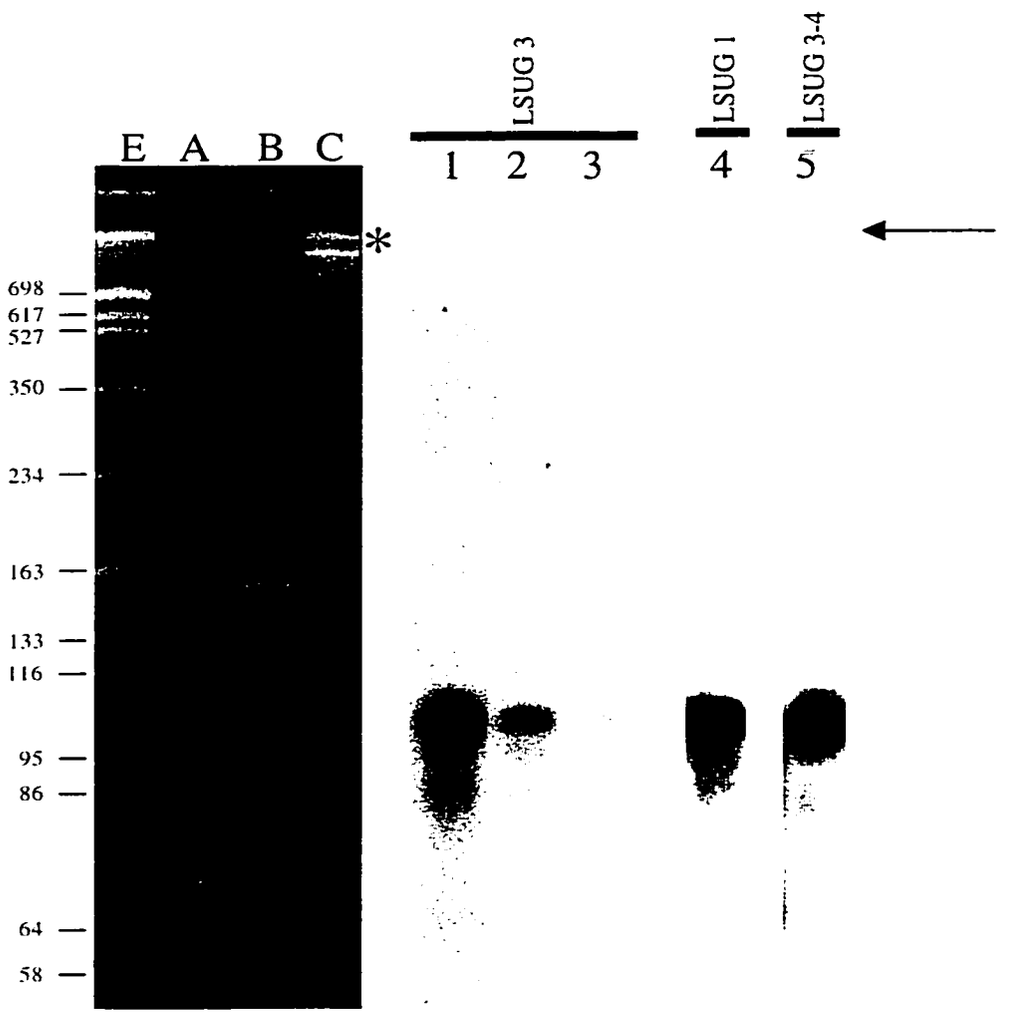


Fig. 3.29. Ethidium bromide-stained polyacrylamide gel (lanes E to C) and autoradiogram (lanes 1 to 5) showing results of northern hybridization analysis of RNA isolated from *C. cohnii*. RNA (lanes A, B, 1, 2, 4 and 5) was isolated from mitochondria prepared by subcellular fractionation. Lanes A, 1, 4 and 5 contain RNA present in the supernatant fraction following disruption of mitochondria. RNA present in the pellet fraction is shown in lanes B and 2. Total cellular RNA is present in lanes C and 3. Blots were hybridized with radiolabeled oligonucleotides (LSUG3 or LSUG1) or radiolabelled PCR product (LSUG3-4) as indicated. Arrow denotes expected location of a band had LSGU probes hybridized to cytoplasmic LSU rRNA (marked with *). Sizes of polyacrylamide gel and autoradiograms are proportionate. Lane E contains *Euglena* cytoplasmic RNA used as a size marker. *Euglena* RNA fragment sizes (Smallman *et al.*, 1996) are indicated to the left of the figure.

fraction was much weaker (Fig. 3.29, lane 3). A 108-nt rRNA fragment is only slightly larger than the size of the LSUG gene found within clone pc3#2 (Fig. 3.19). A similar-sized transcript has been detected in *P. falciparum* (Feagin *et al.*, 1997) as well as *T. parva* (~79 nt; Kairo *et al.*, 1994) for LSUG.

As shown in Fig. 3.29, cytosolic LSU rRNA can be seen in the ethidium-stained gel. None of the northern probes, however, hybridized to this region of the filter. This demonstrates that the LSUG probes are specific for mitochondrial rRNAs.

A much stronger hybridization signal is seen with the mitochondrial supernatant fraction as compared to the mitochondrial pellet fraction. Hence, LSUG rRNA is not membrane-bound but is present within the inner mitochondrial matrix.

The 3' end of LSUG was mapped by sequencing 3'-RACE PCR clones. As shown in Fig. 3.27, the end of the RNA maps to a location that is just outside of the peptidyltransferase domain. The same 3' end was identified in all ten 3'-RACE clones (Table 3.5). The 3' end of LSUG in *C. colnii* is only 1 nt shorter than that determined for LSUG in *P. falciparum* (Gillespie *et al.*, 1999). 3'-RACE also suggests that non-encoded A residues (between 5 and 15) are added post-trancriptionally onto the end of LSUG rRNA (Table 3.5). This result has also been observed for *P. falciparum*, where an 11- to 12-nt oligo(A) tail appears to be post-transcriptionally added onto the 3' end of LSUG rRNA (Gillespie *et al.*, 1999).

The size of the LSUG rRNA is slightly larger based on northern analysis than that predicted from the mtDNA-encoded rDNA sequence. This size discrepancy may be due to oligo(A) addition to the LSUG RNAs. Alternatively, a slightly longer version of

Table 3.5. DNA sequence of *C. cohnii* 3'-RACE clones generated using a primer specific for LSUG4.

1. 3'-RACE clones prepared using unmodified* RNA

LSUG	DNA		
		ATAGCGTCCA GCGACAGTTT CTAACCAATG GTCCGAATTC	
	1	ATAGCGTCCA GCGACAGTTT CTAACCAATG GTCCGAATT***(A)***$_n$	

2. 3'-RACE clones prepared using modified* RNA

LSUG	DNA		
		ATAGCGTCCA GCGACAGTTT CTAACCAATG GTCCGAATTC	
	1	ATAGCGTCCA GCGACAGTTT CTAACCAATG GTCCGAATTAAAAAAAAAAAAAAA***CC(A)***$_n$	15
	2	ATAGCGTCCA GCGACAGTTT CTAACCAATG GTCCGAATTAAAAAAAAAAA***CCC(A)***$_n$	11
	3	ATAGCGTCCA GCGACAGTTT CTAACCAATG GTCCGAATTAAAAAAAAAA***CC(A)***$_n$	10
	4	ATAGCGTCCA GCGACAGTTT CTAACCAATG GTCCGAATTAAAAAAAAAA***CC(A)***$_n$	10
	5	ATAGCGTCCA GCGACAGTTT CTAACCAATG GTCCGAATTAAAAAAAA***CC(A)***$_n$	8
	6	ATAGCGTCCA GCGACAGTTT CTAACCAATG GTCCGAATTAAAAAAA***CC(A)***$_n$	7
	7	ATAGCGTCCA GCGACAGTTT CTAACCAATG GTCCGAATTAAAAAAA***CC(A)***$_n$	7
	8	ATAGCGTCCA GCGACAGTTT CTAACCAATG GTCCGAATTAAAAAA***CCC(A)***$_n$	5
	9	ATAGCGTCCA GCGACAGTTT CTAACCAATG GTCCGAATTAAAAAA***CC(A)***$_n$	5

PCR amplification was performed using P4 and LSUG4 primers. Reverse transcription was carried out using the P16 primer. Letters and numbers are as referenced in Table 3.2.

LSUG rDNA may occur in the mtDNA of *C. cohnii*, giving rise to a 108-nt transcript.

In *C. cohnii*, the 5' end of LSUG was inferred from primary and secondary structure comparisons as well as from transcript size measurements. Sequence alignments suggest that the 5' end of LSUG in *C. cohnii* is only 3 nt shorter than that determined for *P. falciparum* (Fig. 3.26). In both cases, the rRNA breakpoints occur within the same stem region (Fig. 3.27A).

2. LSUE rRNA.

The LSUE2 oligonucleotide (Table 2.1) was designed based on a conserved region of mitochondrial LSU rRNA genes. Using this oligonucleotide as a probe, northern hybridization identified a single RNA that is ~200 nt long (Fig. 3.30A, lanes 1 and 2). This is similar to the size (~190 nt) determined for the LSUE transcript in *P. falciparum* (Gillespie *et al.*, 1999). In *C. cohnii*, bands were detected in the mtRNA (Fig. 3.30, lane 1) and total RNA (Fig. 3.30, lane 2) fractions, although the mtRNA band displayed ~4-fold greater intensity. The LSUE2 probe did not detect cytosolic rRNAs using normal exposure times (Fig. 3.30A); however, a very weak signal with cytosolic LSU rRNA could be visualized after prolonged periods of autoradiography (Fig. 3.30B).

The sequence of LSUE rRNA was established using a combination of methods. The 5' end of the LSUE rRNA was determined by cDNA sequencing using the LSUE2 primer. As seen in Fig. 3.31, three RT stops were detected. The first two correspond to rRNA loops (Fig. 3.27A); however, the final one occurs in a paired stem region. This may represent the authentic 5' end of the LSUE RNA. The fact that the sequencing ladder does not continue beyond this site supports this view. Longer electrophoresis times revealed

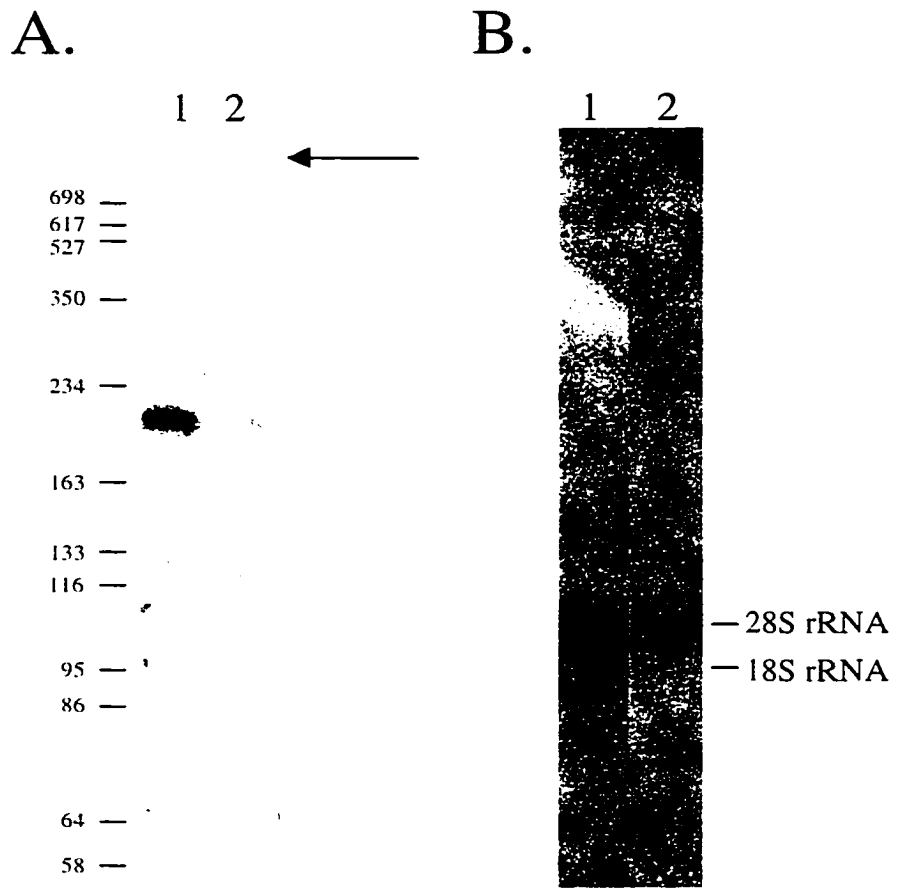


Fig. 3.30. Autoradiogram showing results of northern hybridization analysis of rRNA isolated from *C. cohnii* using an LSUE-specific probe. **A**. Northern blot prepared using 1 μg of mitochondrial (lane 1) or total (lane 2) RNA. RNA was resolved in an 8% polyacrylamide denaturing gel. The sizes of *Euglena* marker RNAs (Smallman *et al.*, 1996) is shown to the left of the panel. The position of *C. cohnii* cytosolic LSU rRNA is indicated with an arrow. **B**. Northern hybridization analysis results using 20 μg (lane 1) or 5 μg (lane 2) of total RNA. RNA was resolved in a 1% formaldehyde agarose gel. The positions of cytosolic rRNAs are indicated (28S rRNA and 18S rRNA). For panel B, autoradiography was performed for an extended period of time (~1 week). Both membranes were hybridized with ^{32}P-end-labeled LSUE2 primer.

Fig. 3.31. Sequence ladders generated by RT sequencing using the LSUE2 primer. (**A**). Complete sequence ladder produced using 1 μg of mtRNA. (**B**). Two sequencing ladders generated using 10 μg of total RNA (marked by 'total') or 1 μg of mtRNA (marked 'mtRNA'). The samples in panel B were subjected to a longer electrophoresis time: therefore only the top part of the gel in panel A (denoted by a bracket) is reproduced in panel B. The putative 5' end of LSUE rRNA is marked (large arrow). The letter E denotes primer extension lane. The other lanes are labeled based on the order of sequence terminators (ddGTP, ddATP, ddTTP, ddCTP). The RNA sequence is given to the left of each set of sequencing reactions reading from bottom to top. Differences between total and mitochondrial RNA sequence in panel B are indicated by an asterisks (*).

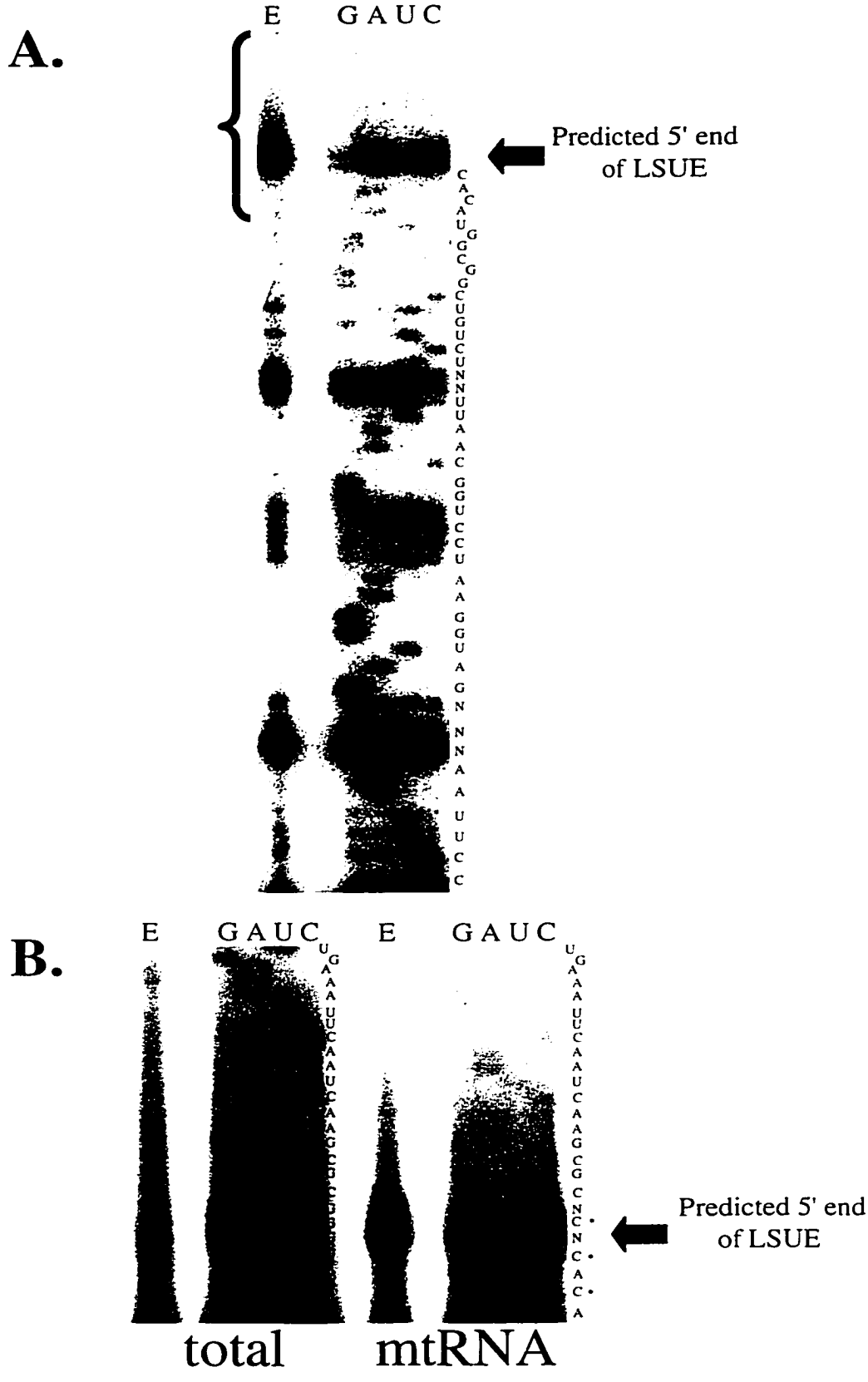

A.
E GAUC
Predicted 5' end of LSUE
B.
E GAUC E GAUC
Predicted 5' end of LSUE
total
mtRNA

Figure 3.31

that the LSUE2 primer could reverse transcribe beyond the strong RT stop, but the sequence ladder produced is much weaker and corresponds to the cytoplasmic LSU rRNA. Other supporting evidence can be found in *P. falciparum*, where the 5' end of LSUE is within 2 nt of that predicted for its *C. cohnii* counterpart (Fig. 3.32)

The sequence of the 3' end of LSUE rRNA was determined using 3'-RACE. Two primers were designed based on the sequence of the 5' end of LSUE (Fig. 3.32, labeled LSUE3, LSUE4). These gene-specific primers were then used to generate 3'-RACE PCR products that were cloned and sequenced. As shown in Fig. 3.33, one clone exhibited LSU sequence homology over the entire length of the insert. Pairwise comparisons (Fig. 3.32) revealed that the *C. cohnii* sequence exhibits the highest degree of identity (76%) with *P. falciparum* LSUE rRNA. In contrast, the lowest identity value (52%) was between *C. cohnii* LSUE rRNA and *P. micans* cytoplasmic 24S rRNA.

Six other 3'-RACE constructs were identified that exhibited LSUE rRNA homology; however, the homologous region was limited to the 5' end of the PCR product (Fig. 3.33). All 3'-RACE clones contained the same 5' portion of LSUE (18 or 35 nt). Two divergence points were found. The sequence contained in the majority of clones diverges from the authentic LSUE gene (Fig. 3.33, divergence point I). A second divergence point occurs much later, with one of the remaining six clones exhibiting a different sequence (Fig. 3.33, divergence point II). These clones do not exhibit homology to any known gene other than LSUE (Feagin *et al.*, 1997).

While it is possible that the LSUE clones may be the result of PCR artifact, the likelihood of this is reduced by the fact that all six constructs were cloned from six

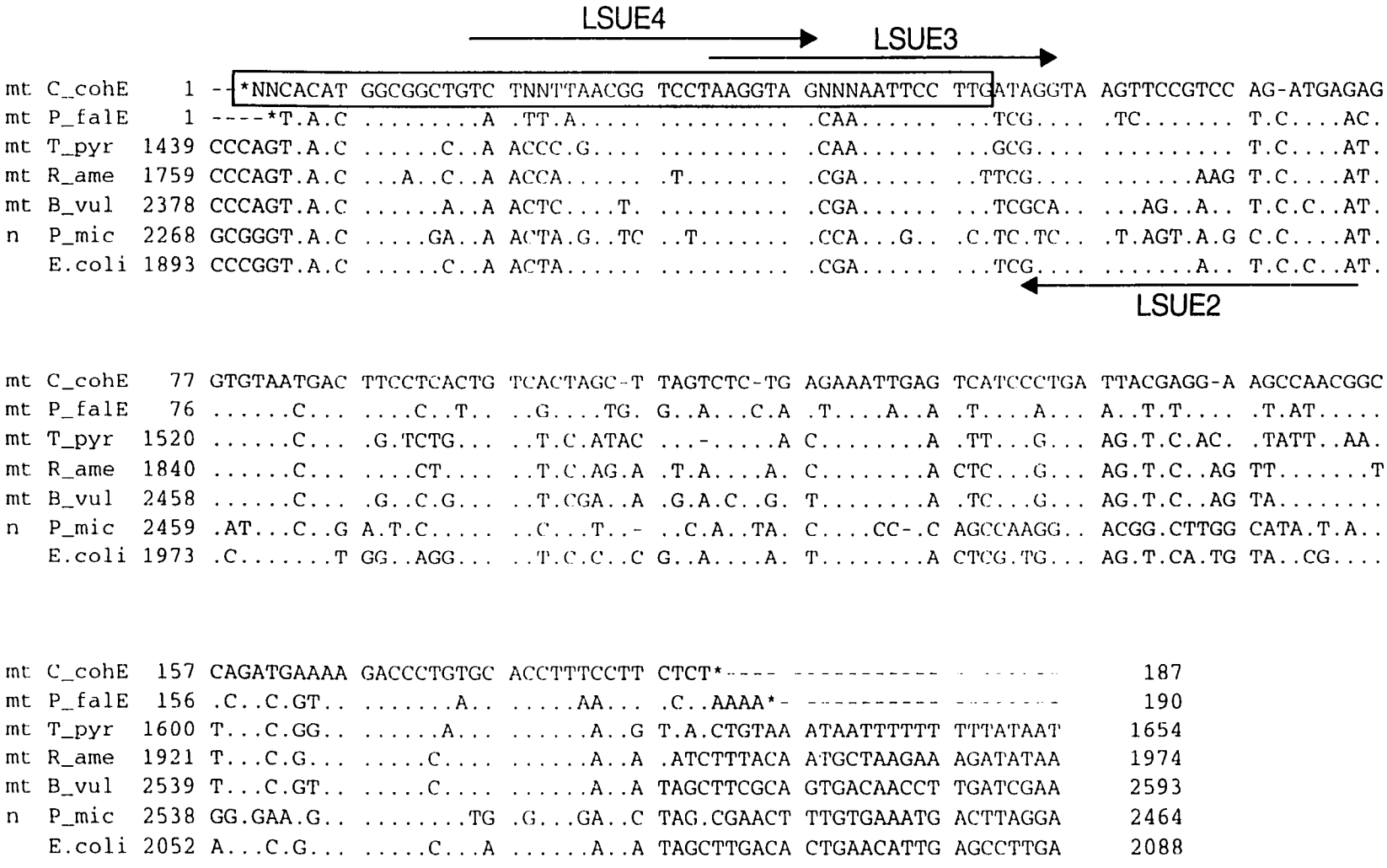


Fig. 3.32. Nucleotide alignment of *rnl* sequence (LSUE portion). Symbols, letters and numbers are as in Fig. 3.26. Arrows above the sequence indicate the position of 3′-RACE primers (LSUE3 and 4). Arrow below the top sequence panel marks the position of the LSUE2 primer. Sequence data obtained by RT sequencing using the LSUE2 primer are enclosed by an open rectangle. Note: due to the highly divergent nature of the *T. parva* LSU1 rRNA gene (term used for LSUG in *T. parva*), this sequence was not included in the alignment. Note: The 3′ ends of *C. cohnii* and *P. falciparum* LSUE rRNAs have been experimentally determined.

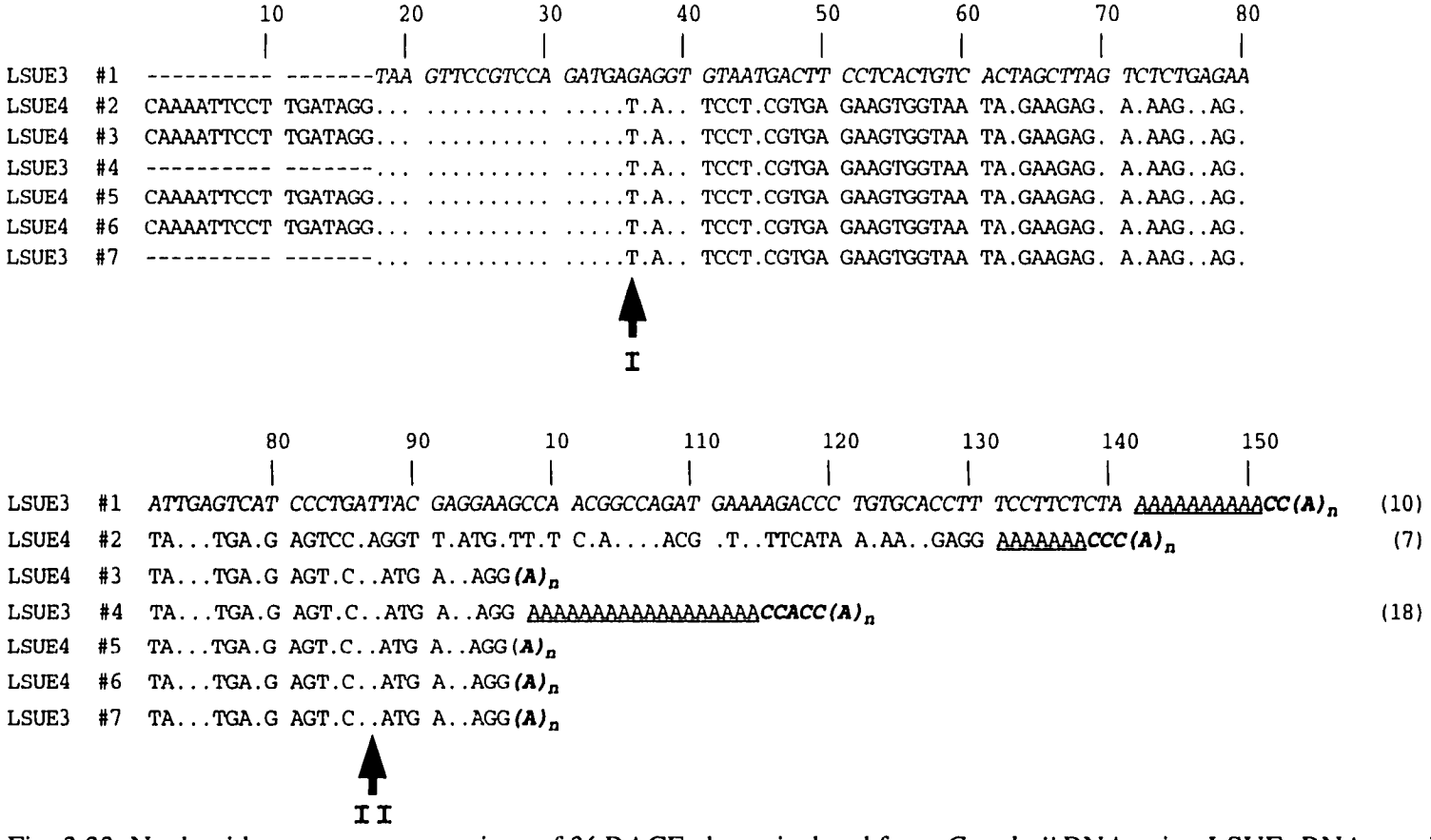


Fig. 3.33. Nucleotide sequence comparison of 3′-RACE clones isolated from *C. cohnii* RNA using LSUE rRNA-specific primers. Clone names are preceded by a number sign (#). The LSUE-specific primer used to amplify each clone is indicated to the far left of the figure (LSU3 or LSUE4). Sequence identical to clone #1 is indicated by a dot (.). Dashes represent the LSUE3 primer binding site. The 'authentic' rRNA gene (clone #1) is highlighted in italics. Nucleotide divergence points are marked with an arrow. Divergence point I separates #1 from all other clones. Divergence point II separates clone #2 from clones #3 to #7. Poly(A) tail length (underlined A residues) was determined by CMP-AMP addition (noted in bold and italics)and is also noted in brackets. Numbers above the sequence denote nucleotide position based on the 3′ end of the LSUE primer. Location of oligo(dT) primer binding site is indicated('$(A)_n$').

different PCR reactions. Thus if these clones are in fact artifactual, the same artifact arose independently six times.

Similar to regions flanking *cox1*, *cox3* and *cob*, the LSUE-flanking domains appear to be unique, and presumably they arose by way of homologous recombination. Based on 3'-RACE data, the LSUE gene appears to exist in a number of different forms: at least one version appears to be full length (i.e., the same size as seen in *P. falciparum*) but copies truncated at the 3' end also occur. These truncated clones are nevertheless expressed and potentially polyadenylated. Several 3'-RACE clones did not contain C residues 5' to the oligo(dT) primer binding site (Fig. 3.33, clones #3, #5 , #6 and #7). Therefore, it is not known whether these RNAs already possessed oligo(A) tails or whether their oligo(A) tails were generated by poly(A) polymerase.

The combined sequence data (5' and 3' end) suggest that the LSUE rRNA is 187 nt in length (Fig. 3.32). This size is only slightly shorter than that predicted using northern analysis data (~200 nt). As with LSUG, this size discrepancy may be attributable to polyadenylation. However, in the absence of gene sequence data it is difficult to predict which A residues are mtDNA-encoded and which ones may have been added to the RNA post-transcriptionally. Moreover, the possibility of RT, PCR or poly(A) polymerase artifacts has not been ruled out.

The predicted 3' end of *C. cohnii* LSUE rRNA is only slightly shorter (4 nt) than that determined for *P. falciparum* (Fig. 3.32). In addition, both rRNAs may undergo post-transcriptional polyadenylation.

The *C. cohnii* LSUE rRNA can be folded into a predicted secondary structure that

resembles the *E. coli* one for the same region (Fig. 3.27). As seen in Fig. 3.27, a highly conserved stem region (just outside of the peptidyltransferase center) is also maintained. This stem is formed by complementary base pairing of the LSUE and LSUG fragment. This demonstrates that even though the *rnl* gene is fragmented, conserved secondary structure is maintained by the pairing of separate RNAs. A similar interaction is also seen in *P. falciparum*.

These results suggest that the LSU rRNA gene in *C. cohnii* mitochondria is highly fragmented. In addition, these gene pieces are expressed to produce small but stable transcripts. LSUE and LSUG data suggest that the *rnl* gene fragmentation seen in *C. cohnii* is remarkably similar to that characterized in *P. falciparum* mitochondria.

F. Analysis of codon usage.

C. cohnii cox1 (76% A+T) and *cob* (77% A+T) genes exhibit a preference for A or T in the third (wobble) position of codons (87%) whereas first- (excluding Met) and second-position A+T biases are lower (70% and 65%, respectively) (Table 3.6). First-position preferences for A or T in Leu and Arg codons also contribute to the high A+T content of protein-coding genes. Of the 16 Arg residues present, 12 are encoded by AGA, two by AGG and two by CGT; however, CGC, CGA and CGG codons are not used at all. Similarly, of 133 Leu residues, 88 are specified by TTA whereas seven are encoded by TTG, with CTC, CTA and CTG together accounting for only 13 codons. Although codon usage in *cox1* and *cob* 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

Table 3.6. Codon usage in *C. cohnii cox1* and *cob* genes.

aa	codon	#	%	aa	codon	#	%	aa	codon	#	%	aa	codon	#	%
Phe	TTT	74	77	**Ser**	TCT	41	52	**Tyr**	TAT	51	100	**Cys**	TGT	14	93
	TTC	22	23		TCC	3	3.8		TAC	0	0		TGC	1	6.7
Leu	TTA	88	66		TCA	35	44	***	TAA	1	100	**Trp**	TGA	2	12
	TTG	7	5.3		TCG	0	0	***	TAG	1	100		TGG	15	88
Leu	CTT	25	19	**Pro**	CCT	15	41	**His**	CAT	23	96	**Arg**	CGT	2	13
	CTC	3	2.3		CCC	1	2.7		CAC	1	4.2		CGC	0	0
	CTA	5	3.8		CCA	20	54	**Gln**	CAA	18	86		CGA	0	0
	CTG	5	3.8		CCG	1	2.7		CAG	3	14		CGG	0	0
Ile	ATT	55	38	**Thr**	ACT	17	36	**Asn**	AAT	48	92	**Ser**	AGT	9	82
	ATC	10	6.9		ACC	2	4.3		AAC	4	7.7		AGC	2	18
	ATA	79	55		ACA	28	60	**Lys**	AAA	17	77	**Arg**	AGA	12	75
Met	ATG	19	100		ACG	0	0		AAG	5	23		AGG	2	13
Val	GTT	14	50	**Ala**	GCT	7	35	**Asp**	GAT	11	85	**Gly**	GGT	32	43
	GTC	2	7.1		GCC	4	20		GAC	2	15		GGC	4	5.3
	GTA	10	36		GCA	9	45	**Glu**	GAA	13	81		GGA	36	48
	GTG	2	7.1		GCG	0	0		GAG	3	19		GGG	3	4

Percentage (%) indicates the proportion of codons used for each amino acid. Asterisks (***) denote stop codons. Open reading frames used in this analysis are contained within the pc1#15 (*cox1*) and pcb#1 (*cob*) clones.

content as the gene-coding regions.

G. Phylogenetic reconstruction.

1. Analysis of COX1 trees.

Phylogenetic trees were constructed using several members of each major taxonomic group to accurately gauge COX1 global topology. A COX1 phylogenetic tree based on neighbor joining analysis (Fig. 3.34) shows that members of the kinetoplastids and alveolates branch together. For the most part all three phylogenetic programs (NEIGHBOR, PROTPARS and PROTDIST) produced the same topology (but see the case of *T. parva*, below). Support for this clade is strong (100%); however, the long branch lengths within the two groups suggest than this affiliation is more likely indicative of a 'long branch attraction' artifact (Felsenstein, 1993) than of a close evolutionary relationship. In fact, the phylogenetic distances between bacterial COXA (the closest COX1 homolog in eubacteria) 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 (97 to 100%) monophyletic group. However, overall support for the alveolate clade is low (Fig. 3.34). Surprisingly, the inclusion of *T. parva* within the apicomplexans is not supported by COX1 trees. In fact, parsimony places *T. parva* basal to the dinoflagellate/apicomplexan assemblage. Once again, *T. parva* appears on a long branch, making this result (which is at odds with biological and other molecular evidence) suspect.

Fig. 3.34. Phylogenetic analysis of COX1 amino acid sequences. The depicted tree was generated using neighbor joining. Numbers in rectangles indicate percent support for the inferred topology. Top number was produced using parsimony bootstrap resampling (100). Middle value represents neighbor joining bootstrap resampling. Bottom value indicates maximum likelihood support values. Two dashes (--) indicate <50% support. A dot (.) represents a node that was not present using that phylogenetic algorithm. Complete amino acid alignment includes only the COX1 core region (amino acid residues 59-520, based on the alignment in Fig. 3.11). The scale bar measures 10 substitutions per 100 amino acids. Taxa included in the analysis are those mentioned in Fig. 3.11 as well as *Rickettsia prowazekii* (CAA74167), *Nitrobacter winogradskyi* (X89566), *Paracoccus denitrificans* (X0529), *Leishmania tarentolae* (P14544), *Trypanoplasma borreli* (AAA73454), *Euglena gracilis* (CAA69263), *Theileria parva* (S41689), *Plasmodium yoelii* (A38891), *Paramecium aurelia* (X15917), *Prototheca wickerhamii* (X68721), *Zea mays* (X02660), *Glycine max* (M16884), *Pisum sativum* (X14409), *Oenothera berteriana* (X05465), *Marchantia polymorpha* (AAC09441), *Triticum aestivum* (CAA39651), *Sorghum bicolor* (P05502), *Oryza sativa* (P14578), *Kluyveromyces marxianus* (S17998), *Allomyces macrogynus* (AAC49234), *Hansenula wingei* (BAA06563), *Saccharomyces cerevisiae* (NP009305), *Aspergillus nidulans* (CAB38220), *Podospora anserina* (AAA32006), *Homo sapiens* (V00662), *Strongylocentrotus purpuratus* (X12631), *Paracentrotus lividus* (J04815) *Cyprinus carpio* (NP007084), *Xenopus laevis* (NP008136), *Rattus norvegicus* (SO4749), *Mus musculus* (ODMS1), *Balaenoptera physalus* (C58850), *Acanthamoeba castellanii* (AAD11820), *Malawimonas jakobiformis* (B. F. Lang, unpublished data), *Ochromonas danica* (B. F. Lang, unpublished data), *Chrysodidymus synuroideus* (AAF36937), *Porphyra purpurea* (AAD03116), and *Chondrus crispus* (NP062482). Numbers within parentheses refer to EMBL, GenBank or Swiss-Prot accession numbers.

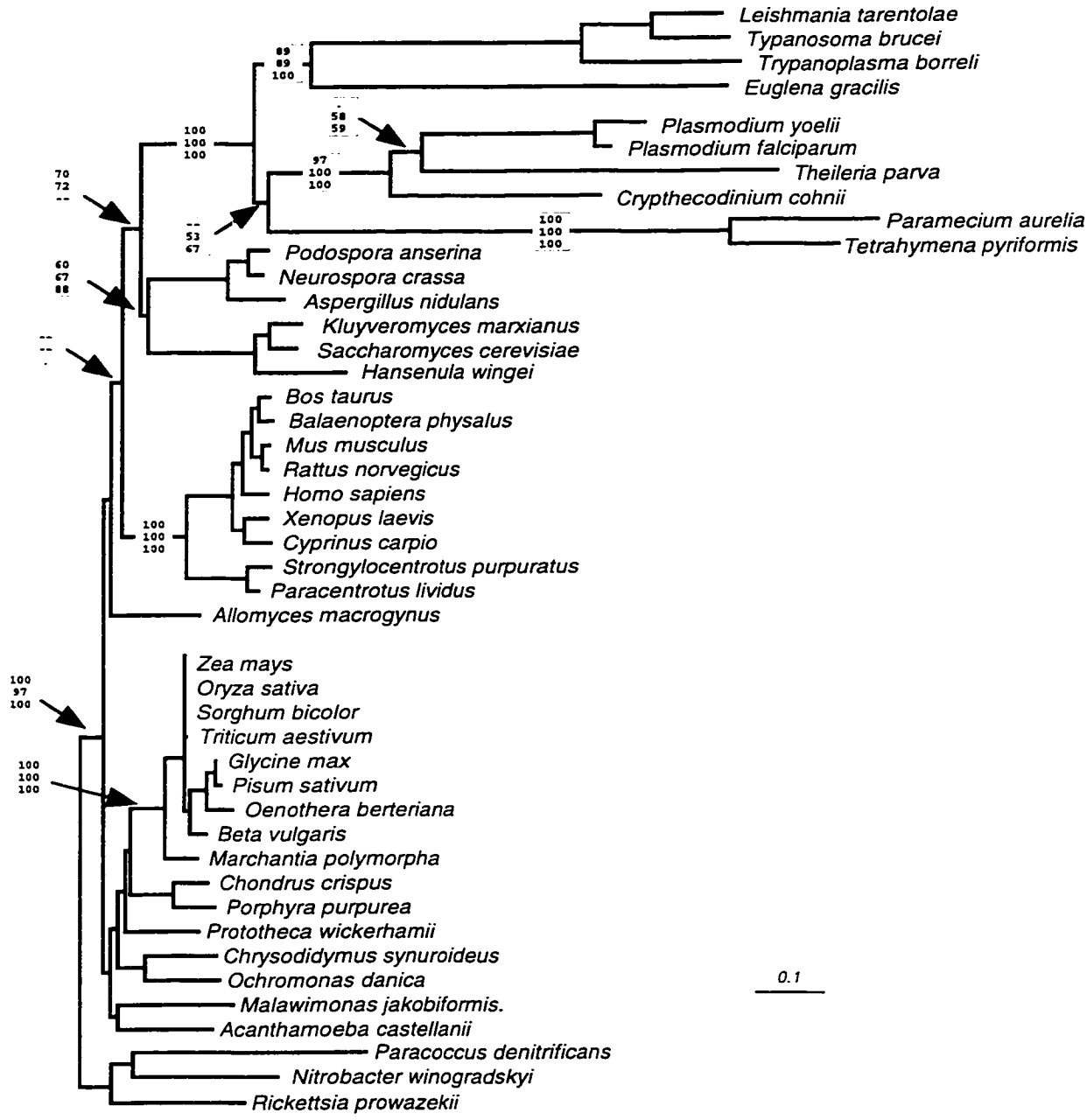

Leishmania tarentolae
Typanosoma brucei
Trypanoplasma borreli
Euglena gracilis
Plasmodium yoelii
Plasmodium falciparum
Theileria parva
Crypthecodinium cohnii
Paramecium aurelia
Tetrahymena pyriformis
Podospora anserina
Neurospora crassa
Aspergillus nidulans
Kluyveromyces marxianus
Saccharomyces cerevisiae
Hansenula wingei
Bos taurus
Balaenoptera physalus
Mus musculus
Rattus norvegicus
Homo sapiens
Xenopus laevis
Cyprinus carpio
Strongylocentrotus purpuratus
Paracentrotus lividus
Allomyces macrogynus
Zea mays
Oryza sativa
Sorghum bicolor
Triticum aestivum
Glycine max
Pisum sativum
Oenothera berteriana
Beta vulgaris
Marchantia polymorpha
Chondrus crispus
Porphyra purpurea
Prototheca wickerhamii
Chrysodidymus synuroideus
Ochromonas danica
Malawimonas jakobiformis.
Acanthamoeba castellanii
Paracoccus denitrificans
Nitrobacter winogradskyi
Rickettsia prowazekii
0.1

Figure 3.34

2. Analysis of COB trees.

Phylogenetic reconstructions using COB produced similar results to those obtained with COX1. Once again, all three phylogenetic analyses infer that members of the kinetoplastids and alveolates branch together as a monophyletic group, with 100% support (Fig. 3.35). COB trees, however, do not show strong support for an apicomplexan/dinoflagellate lineage. This result is somewhat surprising because branch lengths within this group are shorter compared to COX1 reconstructions. Support for the apicomplexan clade also appears to be weak; however, maximum likelihood analysis indicates strong support for this group (97%).

In COB trees, the apicomplexan *T. parva* branches with moderate support as a sister group to the kinetoplastids (Fig. 3.35). This relationship is not supported by nuclear gene phylogenies and is presumed to be artifactual. In particular, branch length comparisons suggest that this association is most likely due to long branch attraction rather than phylogenetic affinity.

The ciliates were omitted from this analysis because neighbor joining and maximum likelihood trees suggested that the phylogenetic distance separating the ciliates from the next closest clade (kinetoplastids) was greater than all other distances combined. This indicates that mitochondrial protein-coding genes in ciliates have undergone accelerated evolutionary change.

3. Analysis of COX1 and COB combined data sets.

Phylogenetic reconstruction using a concatenated data set (COX1 and COB) resulted in a tree that closely resembles that of COX1. All three algorithms infer that

Fig. 3.35. Phylogenetic analysis of COB amino acid sequences. The depicted tree was generated using neighbor joining. Numbers are as indicated in Fig. 3.23. Complete amino acid alignment includes only the COB core region (amino acid residues 89-414, based on the alignment in Fig. 3.23). The scale bar indicates 10 substitutions per 100 amino acids. Taxa included in the analysis are those mentioned in Fig. 3.35 as well as *Rickettsia prowazekii* (O54070), *Rhodospirillum rubrum* (CBQFR), *Rhodopseudomonas viridis* (JQ0346), *Leishmania tarentolae* (H22848), *Theileria parva* (S41691), *Plasmodium vivax* (AAD05206), *Toxoplasma gondii* (g2564669), *Prototheca wickerhamii* (AAD12650), *Chlamydomonas reinhardtii* (CAA36421), *Zea mays* (224151), *Pisum sativum* (CAA10613) *Oenothera berteriana* (P09843), *Marchantia polymorpha* (AAC09441), *Oryza sativa* (CAA37747), *Allomyces macrogynus* (AAC49221), *Hansenula wingei* (BAA06572), *Saccharomyces cerevisiae* (P00163), *Aspergillus nidulans* (Q33798), *Rhizopus stolonifer* (F.B. Lang, unpublished data), *Spizellomyces punctatus* (B. F. Lang, unpublished data), *Strongylocentrotus purpuratus* (NP006977), *Xenopus laevis* (AAA65520), *Balaenoptera physalus* (P24950), *Lumbricus terrestris* (NP008243), *Katharina tunicata* (NP008180), *Drosophila melanogaster* (NP008288), *Acanthamoeba castellanii* (Q37378), *Reclinomonas americana* (AAD11921), and *Chondrus crispus* (NP062486). Numbers within parentheses refer to EMBL, GenBank or Swiss-Prot accession numbers.

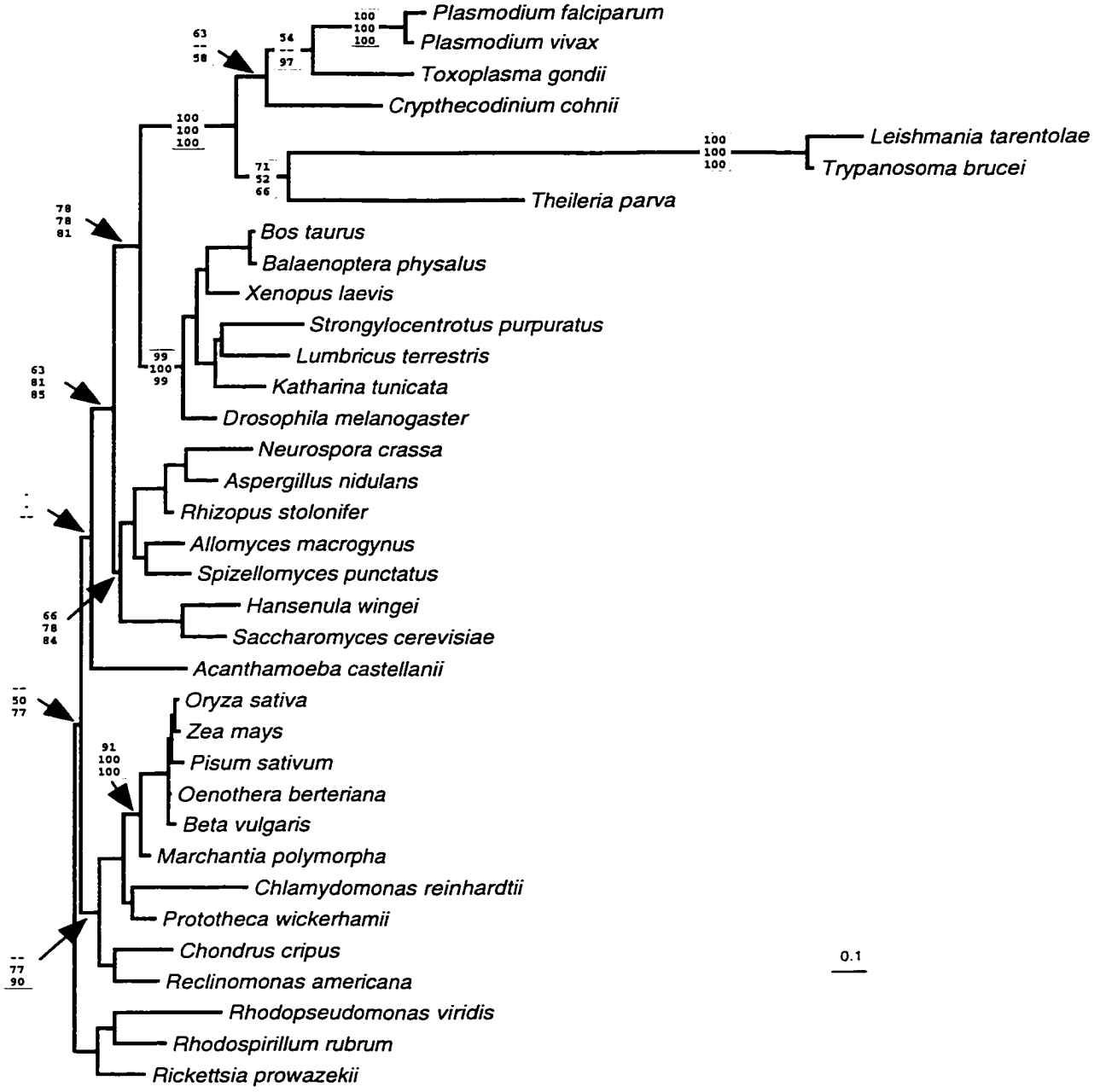

Plasmodium falciparum
Plasmodium vivax
Toxoplasma gondii
Cryptheco dinium cohnii
Leishmania tarentolae
Trypanosoma brucei
Theileria parva
Bos taurus
Balaenoptera physalus
Xenopus laevis
Strongylocentrotus purpuratus
Lumbricus terrestris
Katharina tunicata
Drosophila melanogaster
Neurospora crassa
Aspergillus nidulans
Rhizopus stolonifer
Allomyces macrogynus
Spizellomyces punctatus
Hansenula wingei
Saccharomyces cerevisiae
Acanthamoeba castellanii
Oryza sativa
Zea mays
Pisum sativum
Oenothera berteriana
Beta vulgaris
Marchantia polymorpha
Chlamydomonas reinhardtii
Prototheca wickerhamii
Chondrus cripus
Reclinomonas americana
Rhodopseudomonas viridis
Rhodospirillum rubrum
Rickettsia prowazekii
0.1

Figure 3.35

members of the kinetoplastids and alveolates branch together. Support for this clade is strong (Fig. 3.36) although long branch lengths tend to challenge the validity of these inferences. In the combined data set the dinoflagellate/apicomplexan clade is strongly supported. However, once again, the placement of *T. parva* is unresolved. Based on neighbor and parsimony analyses, *T. parva* forms a sister group to the dinoflagellate/apicomplexan group. Another relationship that is poorly resolved is the placement of the ciliates.

Overall, analysis of evolutionary relationships within the alveolates using COX1 and COB phylogenetic reconstruction is informative but the results must be interpreted with considerable caution. A dinoflagellate/apicomplexan clade appears to be supported, although in some cases, *C. colnii* branches within the apicomplexan clade. However, using mitochondrially encoded proteins, resolution of the branching position of the ciliate clade is not as clear. One reason for this is that mitochondrial protein-coding sequences of members of this group appear to have experienced an accelerated rate of evolution. Evidence to support this inference can be seen in pairwise comparisons of COX1 and COB amino acid sequences, where the ciliates show the lowest level of amino acid identity with their counterparts in all other eukaryotic groups.

Fig. 3.36. Phylogenetic analysis of combined COX1/COB amino acid sequence. The depicted tree was generated using maximum likelihood distance. Numbers and names are as referenced in Figs. 3.34 and 3.35.

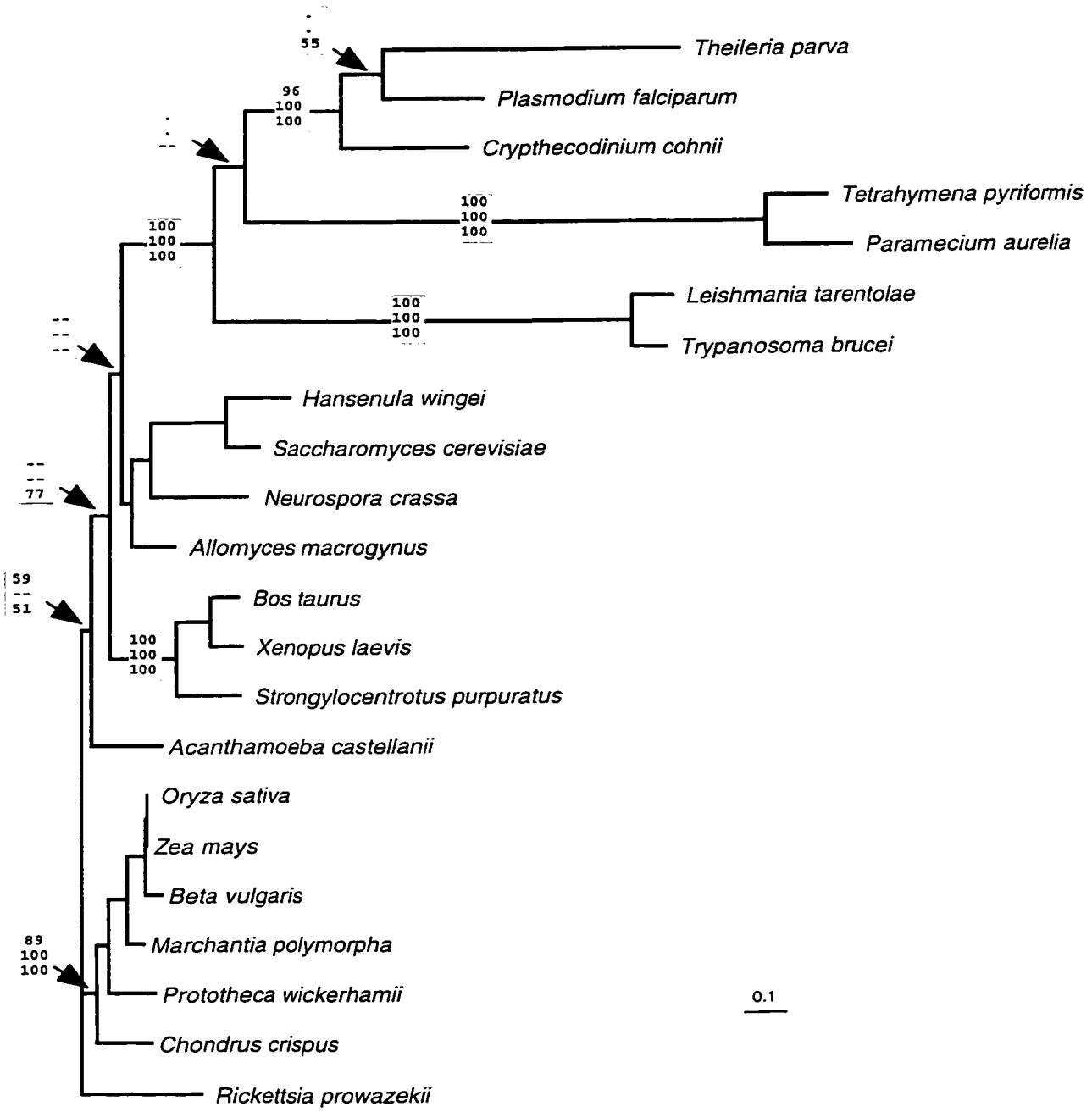

Theileria parva
Plasmodium falciparum
Crypthecodinium cohnii
Tetrahymena pyriformis
Paramecium aurelia
Leishmania tarentolae
Trypanosoma brucei
Hansenula wingei
Saccharomyces cerevisiae
Neurospora crassa
Allomyces macrogynus
Bos taurus
Xenopus laevis
Strongylocentrotus purpuratus
Acanthamoeba castellanii
Oryza sativa
Zea mays
Beta vulgaris
Marchantia polymorpha
Prototheca wickerhamii
Chondrus crispus
Rickettsia prowazekii
0.1

Figure 3.36

IV. Discussion.

The goal of this project was to provide information regarding the structure, organization, expression and evolutionary relationships of mitochondrial genes in the non-photosynthetic dinoflagellate *C. cohnii*. In doing so, I expected to gain further knowledge about mitochondrial genome evolution within the alveolates, a group that is composed of dinoflagellates, apicomplexans and ciliates. Within this assemblage, the dinoflagellates and apicomplexans share a common ancestry, whereas the dinoflagellate-apicomplexan clade is specifically allied with the ciliates (Gajadhar *et al.*, 1991; Sadler *et al.*, 1992; Van de Peer *et al.*, 1996).

Within the alveolates, mtDNA has been completely sequenced and characterized only within the ciliates and apicomplexans. The results reveal very different and even bizarre complements of mitochondrial genes. For instance, apicomplexan mitochondrial genomes (~6 kbp in size) contain only 5 genes (*cox1*, *cox3*, *cob*, *rnl* and *rns*). In contrast, the 40- to 50-kbp mtDNA of ciliates encodes more than 40 ORFs, half of which are of unknown function. Thus, the specific question that I addressed in my thesis research was: Is the mitochondrial genome of *C. cohnii* more similar to that of the apicomplexans or the ciliates? I anticipated that the answer to this question would provide valuable information regarding mitochondrial genome evolution within an early-branching protistan group. Moreover, it was of interest to learn whether the mitochondrial genome present within the alveolate common ancestor already possessed a derived-type mitochondrial genome or whether this character arose independently within the ciliate and apicomplexan lineages.

On balance, in terms of organization and expression, the mitochondrial genome of

C. cohnii resembles that of apicomplexan mtDNA to a greater degree than that of the ciliates. In particular, the split *rnl* gene found in *C. cohnii* mtDNA bears a striking resemblance to the heavily fragmented mitochondrial *rnl* gene present in the apicomplexans. Phylogenetic analysis and protein-coding gene (COX1 and COB) comparisons also support a close evolutionary relationship between the dinoflagellates and apicomplexans. Yet *C. cohnii* exhibits several unique features that clearly set it apart at the mitochondrial level from its closest relatives. For example, the mitochondrial genome of *C. cohnii* is not composed of a single, gene-rich, genome-sized molecule as is found in both the ciliates and apicomplexans. Instead, homologous recombination appears to have generated a structurally heterogeneous mitochondrial genome. Rearrangements have been so extensive that different versions of the *cox1* and *cob* genes have been generated. This kind of mtDNA organization is novel for protists; however, large-scale mtDNA rearrangements of this type have been extensively documented in plants (Hanson & Folkerts, 1992; Wolstenholme & Fauron, 1995).

A. Isolation of nucleic acids from *C. cohnii* mitochondria.

One of the most challenging aspects to characterizing the mitochondrial genome of *C. cohnii* was obtaining sufficient quantities of mitochondrial nucleic acids. Despite the fact that large culture volumes (16 L) were used, relatively small amounts of mtDNA and mtRNA were typically obtained using various nucleic acid extraction protocols.

Mitochondrial DNA did not resolve into a discrete band in CsCl density gradients, but was instead spread throughout the middle portion of the gradient. While the reason

for this is not known, a number of possible explanations may be suggested. First, the *C. cohnii* mitochondrial genome may be composed of multiple, separate DNA elements. In this case, slight differences in base composition would produce a broad mtDNA band in CsCl density fractionations. This explanation seems the most plausible based on the heterogeneous nature of *cox1*, *cob* and partial-*cox3* gene-containing elements identified in *C. cohnii* mtDNA. Second, *C. cohnii* mtDNA duplexes may contain variable stretches of single-stranded DNA, resulting in a range of mtDNA buoyant densities. In this regard, it is notable that the mtDNA of *P. falciparum* also forms a broad band in CsCl gradients (Preiser *et al.*, 1996). This broadening is thought to be due to variable stretches of single-stranded DNA that are attached to mtDNA duplexes.

Quantities of mitochondrial nucleic acids isolated using subcellular fractionation were also low. Again, the reasons for this are unknown, but a number of possibilities can be entertained. One possibility is that *C. cohnii* cells contain few mitochondria or mitochondria with small amounts of mtDNA. However, electron micrographs (see Fig. 3.1) suggest that *C. cohnii* cells have many mitochondria. Another explanation is that the techniques necessary to break the hard outer covering of *C. cohnii* cells may inadvertently disrupt mitochondrial membranes. Interestingly, *C. cohnii* mitochondria did not behave like 'typical' mitochondria during subcellular fractionation. Using standard isolation procedures, mitochondria normally collect at the 1.55 M sucrose shelf following two-step discontinuous gradient centrifugation (Spencer *et al.*, 1992). In the present study, however, *C. cohnii* mitochondria passed through the sucrose cushion and collected at the bottom of the centrifugation tube. This may indicate that mitochondria in *C. cohnii* are

more dense compared to mitochondria of other organisms. Another consideration is that *C. cohnii* mitochondria may form a contiguous reticulum (net-like structure), thereby making them more susceptible to disruption. In this regard, there is some evidence to suggest that *Plasmodium* mitochondria form a network of connecting tubes as opposed to the typical 'cigar-shaped' structure found in animal mitochondria (Slomianny & Prensier, 1986).

B. Evolutionary affiliation of the mitochondrial protein-coding genes of *C. cohnii*.

1. Evaluation using phylogenetic tree reconstruction.

Amino acid sequence comparisons (COB and COX1) demonstrate that *C. cohnii* mitochondrial protein-coding genes share highest amino acid identity with those of *P. falciparum* and lowest with their counterparts in *T. pyriformis*. The relatively low degree of identity between dinoflagellate and ciliate COX1 and COB sequences is inconsistent with SSU rRNA phylogenies (Cavalier-Smith, 1993; Gajadhar *et al.*, 1991; Schlegel, 1991; Van de Peer *et al.*, 1996) but may be explained by the highly derived nature of many ciliate mitochondrial genes (Cummings, 1992; Pritchard *et al.*, 1986; Pritchard *et al.*, 1990b).

The COX1 and COX1/COB concatenated phylogenetic trees (in agreement with amino acid identity values) show support for a dinoflagellate and apicomplexan clade, but display poor resolution within the alveolate clade. This low degree of support in part may reflect an artifactual grouping of unrelated taxa (e.g., alveolates and kinetoplastids) as a result of long branch attraction (Felsenstein, 1978), a consequence of an accelerated rate

of sequence divergence. Agreement of SSU rRNA trees (Cavalier-Smith, 1993; Gajadhar *et al.*, 1991; Schlegel, 1991; Van de Peer *et al.*, 1996) and morphological data (Sadler *et al.*, 1992; Van de Peer *et al.*, 1996) with COX1 and COX1/COB concatenated trees implies that members of the alveolates share a common ancestry; however, their *cox1* and *cob* genes are obviously evolving at accelerated and unequal rates. In both trees (COX1 and COX1/COB concatenated), the branch separating the ciliates from the other alveolates is quite long. Similar results have been shown in trees generated using concatenated protein (COB/COX1/COX2/COX3; Burger *et al.*, 2000) as well as mitochondrial rRNA (Gray & Spencer, 1996) sequences.

Besides the ciliates, many members of the alveolates exhibit unequal and accelerated rates of evolution as implied by branch lengths (e.g., see *T. parva*, Fig. 3.35). Moreover, the inclusion of alveolates in the phylogenetic analysis may affect overall tree topology and support values. For example, if members of the alveolates and kinetoplastids are excluded, COB trees become highly resolved and consistent with 18S rRNA trees (David F. Spencer, personal communication). Because alveolate mitochondrial genes are highly divergent, phylogenetic reconstruction and assessment of relationships is difficult. In such cases, it is important to identify other mitochondrial characters, such as gene structure, organization, arrangement and mode of expression, that can provide additional information regarding the evolution of dinoflagellate, apicomplexan and ciliate mtDNAs.

Interestingly, the plastid genes of dinoflagellates and alveolates also appear to be highly divergent. A recent phylogenetic reconstruction (Zhang *et al.*, 2000) using 23S

rRNA and *psbA* gene sequences showed that peridinin-containing dinoflagellates and apicomplexans form a monophyletic group. However, the strength of this conclusion is tempered by the fact that this affiliation may again reflect a long branch attraction artifact. Intriguingly, nuclear gene phylogenies do not exhibit the same problem.

2. Codon usage.

Examination of nucleotide sequence in the vicinity of the predicted N-terminus of COB identified three potential ATG (Met) initiation codons. Similarly, a single ATG codon was found near the predicted 5' end of COX1. The initiation codon data for *cox1* are in contrast to data from apicomplexans (reviewed in Feagin, 1994) and ciliates (reviewed in Cummings, 1992), which suggest that alternative *cox1* initiation codons are used. Alternative initiation codons have been proposed for *P. falciparum*, *P. gallinaceum* (ATT) (Feagin *et al.*, 1992; reviewed in Feagin, 1994), *T. parva* (AGT) (Kairo *et al.*, 1994) and *T. pyriformis* (ATA) (Ziaie & Suyama, 1987), but the situations in *P. yoelii* (Suplick *et al.*, 1990), *P. aurelia* (Cummings, 1992) and *P. vivax* (McIntosh *et al.*, 1998) are unresolved.

Why *C. cohnii* uses conventional start codons when other members of the alveolates do not is unclear. This finding may simply be fortuitous, considering that only two genes in *C. cohnii* mtDNA have been examined. However, apicomplexans and ciliates also possess mitochondrial genes that have ATG start codons. As with *C. cohnii*, all apicomplexans (Feagin *et al.*, 1992; Kairo *et al.*, 1994; Suplick *et al.*, 1990; Vaidya *et al.*, 1993) use an ATG initiation codon for *cob*. Likewise, universal initiation codons have been determined for a number of ciliate mitochondrial genes (i.e., *atp9*; Cummings, 1992;

Edqvist *et al.*, 2000; Pritchard *et al.*, 1990b). This suggests that 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 (Karlovsky & Fartmann, 1992; Osawa *et al.*, 1992). *C. cohnii* COX1 and COB exhibit a bias for A or T in codon third positions (Table 3.7); 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 3.7 also shows that the vast majority of the 15 Trp residues are encoded using the universal TGG codon. In fact, the entirety of COX1 and all but two COB Trp codons (Fig 3.23) appear to be encoded using TGG. In many mitochondrial systems, TGA (a universal stop codon) also codes for Trp (reviewed in Osawa *et al.*, 1992). As in *C. cohnii cox1*, *Plasmodium* mitochondrial protein genes do not appear to use TGA (Feagin, 1994; Feagin *et al.*, 1992). In contrast, *T. pyriformis* mtDNA uses TGA almost exclusively (Burger *et al.*, 2000; Ziaie & Suyama, 1987) whereas the *P. aurelia* mitochondrial genome uses both TGA and TGG equally (Cummings, 1992; Pritchard *et al.*, 1990b).

It has been suggested (Jukes & Osawa, 1990; Osawa *et al.*, 1992) that virtually all non-plant mitochondria (those of the oomycete *Phytophthora infestans* being an exception (Karlovsky & Fartmann, 1992)) 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 do use TGA to specify Trp. Examples are the chytrid *Allomyces* contrasted with other fungi (Paquin & Lang, 1996) and *Euglena* contrasted with kinetoplastids (Yasuhira & Simpson, 1997) (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 (Lang *et al.*, 1987; Paquin & Lang, 1996).

C. Fragmentation of the mitochondrial *rnl* gene in *C. cohnii*.

One of the most intriguing findings of this study is the close similarity in the mitochondrial LSU rRNA gene fragmentation patterns exhibited by *C. cohnii* and members of the apicomplexans. The LSUG rRNA gene in *C. cohnii* is 99 nt long and encodes a portion of the peptidyltransferase domain. Examination of mtDNA sequence 5' to LSUG rRNA did not identify any rRNA-homologous regions. This suggests that the *rnl* gene in *C. cohnii* is discontinuous and rearranged. A similar finding has been documented in the apicomplexans. In *Plasmodium* and *T. parva*, the *rnl* and *rns* genes are discontinuous and scrambled, being scattered throughout the entire genome (reviewed in Feagin, 1994; Feagin, 2000; Wilson & Williamson, 1997). One major difference is that genes appear to be tightly clustered in the apicomplexans, with rRNA gene pieces being interspersed with one another and conventional protein-coding genes.

Northern analysis and cDNA sequence comparisons show that the LSUG and LSUE rRNA genes in *C. cohnii* are expressed as small stable RNAs (~108 and 200 nt, respectively). Larger mitochondrial rRNA products were not identified. This supports the view that the *rnl* gene in *C. cohnii* is highly fragmented, a finding that is also consistent with results obtained from apicomplexans. In *Plasmodium* (Feagin *et al.*, 1997; Suplick *et al.*, 1990) and *T. parva* (Nene *et al.*, 1998), the LSU rRNA is expressed as a set of small RNAs (78-190 nt).

One question that immediately comes to mind is: Do the LSU rRNA fragments form part of a functional mitochondrial ribosome in *C. cohnii*? For this issue to be tested, much more data will be required. In particular, all missing LSU rRNA fragments will have to be found. When this has been accomplished, biochemical investigations into ribosome assembly and function can be initiated. The data available, while limited at this point, do not suggest that the LSUE and LSUG rRNAs are derived from pseudogenes in *C. cohnii*. First, LSUE and LSUG rRNAs are expressed as stable transcripts. Second, primary sequence alignments show a high degree of sequence identity with functional LSU rRNAs. Third, LSUE and LSUG have the potential to associate by base pairing, producing a secondary structure that conforms to a large portion of the highly conserved, and functionally indispensable, peptidyltransferase domain. These facts do not automatically lead to the conclusion that the fragmented rRNAs are functional in *C. cohnii*. For example, secondary structure predictions cannot distinguish between recently emerged pseudogenes and functional rRNAs (Boer & Gray, 1988). However, both *C. cohnii* and *Plasmodium* exhibit a high degree of primary and secondary structure

conservation. This observation coupled with the fact that the dinoflagellate and apicomplexan lineages diverged from one another a long time ago suggests that their rRNA genes must be under some sort of evolutionary pressure to maintain conserved nucleotides and secondary structure. If one or both lineages possessed rRNA pseudogenes, then it is presumed that random mutation would have reduced the degree of similarity shared between the mitochondrial *rnl* genes of *C. cohnii* and *Plasmodium*.

As in dinoflagellates and apicomplexans, the mitochondrial *rnl* gene in *Chlamydomonas*, a chlorophyte algae, is highly fragmented. Nevertheless, *Chlamydomonas* rRNA fragments are capable of forming particles that are thought to be functional ribosomes (Denovan-Wright & Lee, 1995). This evidence has been used to support the notion that the fragmented rRNAs of apicomplexans are also capable of being incorporated into functional ribosomes (Feagin, 2000). In *Plasmodium*, the various rRNA fragments are able to interact, forming most of the highly conserved domains that are thought to be required for a functional ribosome. As of this date, however, an RNA species that corresponds to the GTPase domain has not been found (Feagin *et al.*, 1997).

One of the most surprising findings of this study is that the LSUG and LSUE rRNA 'modules' in *C. cohnii* and *Plasmodium* are so similar. In fact, in this respect, *P. falciparum* is more like *C. cohnii* than like *T. parva*. For instance, LSUE transcript sizes are similar between *C. cohnii* (~200 nt) and *P. falciparum* (~190 nt; Feagin *et al.*, 1997). In contrast, although rRNA transcript sizes have not been published for *T. parva*, mtDNA alignments suggest that the LSUE rRNA is considerably larger (301 nt) in this apicomplexan (Kairo *et al.*, 1994). This longer LSUE rRNA is manifested by a predicted

5' breakpoint that occurs ~100 bp upstream of that inferred for the corresponding *C. cohnii* and *P. falciparum* rRNAs. It should be noted, however, that the LSUE gene in *T. parva* is homologous to two mitochondrial rRNA genes (LSUD and LSUE) that are present in *P. falciparum* mtDNA. Notably, the LSUD gene occurs just upstream of LSUE rRNA gene and directly abuts it in *P. falciparum* mtDNA.

Other LSUE rRNA similarities can be seen in the high degree of sequence conservation between the *C. cohnii* and *Plasmodium* homologs (~80%). On the other hand, the corresponding *T. parva* sequence is divergent enough that a core LSUE domain (*T. parva*, LSU1 positions 100-180) is difficult to align with the corresponding *C. cohnii* and *Plasmodium* sequences. In addition, the predicted secondary structure of this region is atypical, even for a mitochondrial rRNA (Kairo *et al.*, 1994). Interestingly, the same mitochondrial rRNA species of another apicomplexan, *Babesia bovis* (GenBank accession AF053002), resembles *T. parva* in predicted secondary structure (Murray N. Schnare, personal communication).

The LSUG fragment is more conserved among *C. cohnii*, *Plasmodium* and *T. parva*, in that RNA sizes are comparable (approx. 108, 100 and 84 nt, respectively; Feagin *et al.*, 1997; Kairo *et al.*, 1994). Likewise, predicted rRNA ends are close to one another. This may mean that only the LSUE gene is divergent in *T. parva*; however, sequence alignments reveal greater conservation between *C. cohnii* and *P. falciparum* LSUG homologs than between either of these and the *T. parva* LSUG RNA. Primary sequence conservation is further mirrored by a greater likeness in the secondary structures of *C. cohnii* and *P. falciparum* LSUG RNAs.

It is not clear why *C. cohnii* and *Plasmodium* should exhibit a higher degree of LSUE and LSUG similarity than *Plasmodium* does with *T. parva*. The most probable explanation is that *T. parva* mitochondrial *rnl* genes are evolving at a faster rate compared to those of *C. cohnii* and *P. falciparum*. A similar result is also seen with protein-coding genes (i.e., COB phylogenetic reconstruction). Another possibility is that the apicomplexans are paraphyletic, with the dinoflagellates and *Plasmodium* being more closely related than either is to *T. parva*. This scenario, however, is highly unlikely because it is not supported by nuclear phylogenies or mitochondrial genome organization and arrangement data.

The mitochondrial rRNA data strongly unite the dinoflagellate and apicomplexan clades. They also suggest that the common ancestor to the dinoflagellate/apicomplexan lineage possessed a fragmented mitochondrial *rnl* gene. Interestingly, the ciliates also possess a discontinuous and slightly rearranged *rnl* gene (a tRNA gene separates the two rRNA modules). However, fragmentation is quite limited, occurring at a single site near the 5' end of the mitochondrial LSU rRNA in *T. pyriformis* (nucleotide position 280; Heinonen *et al.*, 1990; Heinonen *et al.*, 1987) and *P. aurelia* (nucleotide position 283; Seilhamer *et al.*, 1984). This evidence suggests that the common ancestor to the alveolates may have possessed discontinuous rRNA genes. Alternatively, there may be another process occurring in alveolate mitochondria that increases the likelihood of rRNA gene fragmentation. Supporting evidence for this statement is the fact that *rnl* gene breakpoints are not the same among the ciliates, dinoflagellates and apicomplexans. This is highly suggestive that *rnl* fragmentation has arisen independently in the alveolates. A

similar argument has been made previously regarding mitochondrial rRNA gene fragmentation in a group of related chlorophyte algae (Turmel *et al.*, 1999). Members of the Chlorophyta (*Pedinomonas minor* (Turmel *et al.*, 1999); *C. reinhardtii* (Boer *et al.*, 1985; Boer & Gray, 1988); *Chlamydomonas eugametos* (Denovan-Wright & Lee, 1995); *Scenedesmus obliquus* (Nedelcu *et al.*, 2000); and *Chlorogonium elongatum* (Kroymann & Zetsche, 1998)) represent the only other mitochondrial lineage to exhibit fragmented *rnl* genes. Despite sharing a common ancestor, the location and extent of *rnl* fragmentation is different for each member. These observations have been interpreted as support for the idea that *rnl* discontinuity arose independently in each taxa. On the other hand, these diverse patterns of *rnl* discontinuity are unlike that seen in dinoflagellates and *Plasmodium*, where there is striking similarity between *rnl* breakpoints in the rRNA characterized to date. Nevertheless, the results reported here should be extrapolated with caution because only two LSU rRNA modules have been identified in *C. cohnii* so far.

One interesting finding of this study is the possibility that rRNAs in *C. cohnii* are post-transcriptionally polyadenylated. Although still preliminary, 3'-RACE data suggest that LSUG rRNAs contain oligo(A) tails that are not mtDNA-encoded. At present, the possibility of cDNA or PCR-generated artifacts has not been ruled out. However, amplification of polyadenylated RNAs in the absence of *in vitro* CMP-AMP suggests that 3' terminal A residues were already present *in vivo*. In addition, the predicted size of LSUG rRNAs based on gel mobility appears to be slightly larger than that determined using mtDNA sequence data. One possibility is that only a small percentage of rRNAs is polyadenylated. Supporting evidence is the fact that oligo(A) tail length differences were

noted using 3'-RACE, although LSUG size heterogeneity was not obvious from northern analysis. In *P. falciparum*, it has been demonstrated that many of the mitochondrial rRNA fragments contain oligo(A) tails that are not DNA-encoded (Gillespie *et al.*, 1999). It is thought that other members of the apicomplexans also exhibit this characteristic. For example, mitochondrial rRNAs were isolated from an oligo(dT)-bound RNA fraction in both *P. yoelii* (Vaidya & Arasu, 1987) and in *T. parva* (Nene *et al.*, 1998). The ciliates, however, do not share this feature: in *T. pyriformis* judging from RNA sequence data, the 3' ends of all four mitochondrial rRNA pieces are identical to the mtDNA sequence (Schnare *et al.*, 1986; Heinonen *et al.*, 1987).

The purpose of A addition to the end of rRNAs is unknown. As discussed by Gillespie *et al.* (1999), there are a number of reasons why this may happen. First, it may simply be the result of non-specific addition of A residues by mitochondrial polyadenylation proteins. Second, it may be important in ribosome assembly or in maintenance of tertiary structure. Finally, it has been suggested that residues added to the ends of rRNAs may protect it to some degree against degradation of the RNA by exonuclease activity. This explanation appears to be less likely in *C. cohnii* because it is hypothesized that *cox1* mRNAs that are in the process of being degraded are oligoadenylated. Based on this information, non-specific oligo(A) addition seems the more plausible possibility in *C. cohnii*.

D. Arrangement and organization of mtDNA in *C. cohnii*.

It is unlikely that the mitochondrial genome of *C. cohnii* is composed of a single,

gene-rich element that encodes one version of each gene. In fact, the direct opposite seems to be the case. *C. cohnii* contains a mitochondrial genome that is composed of multiple, gene-poor elements that appear to encode several possible variants of each gene. As a result, mtDNA organization in *C. cohnii* is very different from that observed in apicomplexans and ciliates.

In *C. cohnii*, *cox1* is part of a central repeat unit (containing most of the *cox1* sequence) flanked by one of two upstream (flank-1 or flank-2) and one of two downstream (flank-3 or flank-4) regions. While most of *cox1* is located within the central repeat, the C-terminal portion of the ORF extends into flanking regions 3 or 4, thereby creating two distinct *cox1* ORFs. The organization of the *cob* gene is seemingly much simpler, with this gene found predominantly in a single prominent *Eco*RI fragment of 3.7 kbp. However, the detection of minor *cob*-containing *Eco*RI fragments indicates that *cob* gene organization is still quite complex. Based on all *cob* gene arrangements identified, the *cob* gene is part of a central repeat unit that is flanked by different 5' (flank b-1, flank b-2 or flank b-3) and 3' (flank b-4, flank b-5, flank b-6, flank b-7 or flank b-8) ends. Once again, as in the case of *cox1*, almost the entire *cob* gene is located within the central repeat; however, its C-terminus stretches into flanking domains 4, 5 or 6. This results in the creation of 3 discrete *cob* ORFs; one is present within the major *Eco*RI fragment (cob-1), with the others (cob-2 and cob-3) occurring within minor *cob*-containing constituents. Finally, a small portion of the *cox3* gene was also identified in *C. cohnii* mtDNA. As with *cox1*, the *cox3* fragment appeared to form part of a central repeat 'cassette' that was flanked by different 5' and 3' domains.

This arrangement of mitochondrial genes is quite different from that found in the apicomplexans and ciliates, where *cox1*, *cob* and *cox3* constitute unique ORFs that are flanked by the same sequence in all mtDNA copies. Another difference is the nature of the DNA sequence flanking these genes. In *C. cohnii*, *cox1* and *cob* are flanked upstream and downstream by long stretches of non-coding DNA. In contrast, in both apicomplexan and ciliate mtDNAs, *cox1* and *cob* are separated from flanking genes by relatively short intergenic spacer regions. In fact, in the apicomplexans *P. falciparum* (Feagin *et al.*, 1992; Vaidya *et al.*, 1993), *P. yoelii* (Vaidya *et al.*, 1989), *P. gallinaceum* (Aldritt *et al.*, 1989) *P. vivax* (McIntosh *et al.*, 1998; Sharma *et al.*, 1998), *P. reichenowi* (GenBank accession number AJ251941) and *T. parva* (Kairo *et al.*, 1994), the complete linear mitochondrial genomes are only slightly larger (6-7 kbp) than the two largest *cox1* elements described here in *C. cohnii*. This pronounced variation in non-coding sequence suggests that mtDNA size constraints differ considerably in the two groups.

Despite differences in *cox1* and *cob* organization, the mitochondrial genomes of *C. cohnii* and the apicomplexans share a number of superficial similarities that appear to be centered around repetitive sequence elements. An example of this can be found in *T. parva* mtDNA (Kairo *et al.*, 1994), where *cox1* is located immediately downstream of one of two large terminal inverted repeat regions that are composed of smaller (12-bp) direct and inverted repeats. Along somewhat different lines, the 6-kbp mtDNA core is itself repeated in *P. falciparum*, but in a head-to-tail manner, forming tandem arrays of oligomeric units (reviewed in Feagin, 1994; Wilson & Williamson, 1997). Although the mtDNA repeat arrangements in *C. cohnii* and *P. falciparum* are different in detail,

Southern hybridization analyses using *cox1*- and *cob*-specific probes revealed 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-kbp to <6 kbp. Endonucleases having a single restriction site in the *Plasmodium* 6 kbp element produce a band of this size (Preiser *et al.*, 1996), superimposed on a background smear starting at 6 kbp. Similarly, a *cob*-specific probe produces a single 3.7-kbp band superimposed on a trailing smear in *C. cohnii*. However, a *cox1*-specific probe produces four bands superimposed on a trailing smear, unlike the single-band pattern seen in *P. falciparum*. In all three cases, the trailing smear is thought to be largely the result of terminal heterogeneity. In *C. cohnii*, I infer the presence of mtDNA molecules in which different sequences flank *cox1* and *cob*, resulting in different *Eco*RI sites and fragment sizes. The terminal heterogeneity present in *P. falciparum* has been attributed to DNA degradation or non-specific initiation of mtDNA replication that coincides with homologous recombination. In this case, it is thought that incompletely replicated mtDNAs that have overhanging 3' ends can invade homologous regions of other DNA duplexes (Fig. 4.1A: Y-branched, three-way DNA junction), thus creating a replication fork for DNA replication (Preiser *et al.*, 1996). There is no evidence for such a mechanism in *C. cohnii*; however, if such a recombination-mediated replication system were present in dinoflagellates, it might explain the apparent heterogeneity of *cox1* and *cob* element ends, as well as the occurrence of unique genes and elements containing partial genes.

The arrangement of *cox1*-containing *Eco*RI fragments as a two-membered family

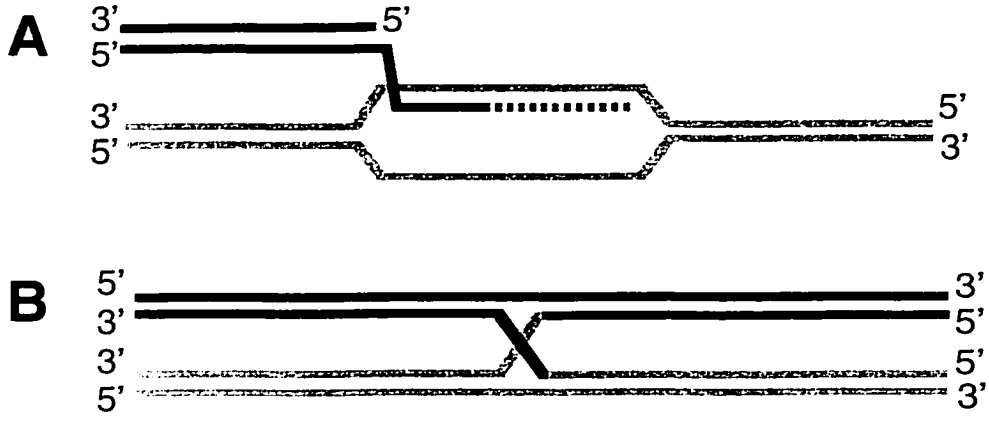


Fig. 4.1. Models showing two types of recombination-generated DNA junctions. (A) Y-branched, three-way DNA junction is created when one segment of DNA (black) that has a 3' overhang invades an homologous region present in a second DNA duplex (grey), thereby creating a replication fork that allows the 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 (both strands black or grey) starting at the cross-over point.

of repeats is consistent with homologous recombination via Holliday junction formation (Fig. 4.1B: X-branched, four-way DNA junction). It is assumed that the four *cox1*-hybridizing bands that I observe in *C. cohnii* originated as a single *cox1* element. Duplication of the *cox1* central repeat elsewhere in the genome would have created two different *cox1*-containing repeat elements, each having unique flanking domains. Homologous recombination within the central repeat of two *cox1* regions would then generate four different contexts, with the central repeat flanked by one of two upstream and one of two downstream domains. The occurrence of *cox1* in multiple distinct contexts, as documented here, strongly suggests that *C. cohnii* mtDNA does undergo repeat-mediated homologous recombination.

Another striking finding of this study is that *cox1*, *cob* and '*cox3*' gene organization in *C. cohnii* exhibits a number of striking similarities to gene arrangements seen in angiosperm (flowering plant) mtDNAs. First, the presence of four gene-specific restriction fragments (*cox1*), each the result of pairwise combinations of a central repeat unit flanked by two sets of upstream and downstream flanking regions, is quite common in plant mtDNAs (Hanson & Folkerts, 1992; Wolstenholme & Fauron, 1995). Second, comparisons among repeat flanking domains within *C. cohnii* show that very little sequence similarity is shared within these regions, and that the transition between the repeat unit and the adjacent flanking domains is abrupt. This has also been documented in plants, where discrete divergence points are the result of large-scale genomic rearrangements that shuffle adjacent segments of DNA to other parts of the genome (Hanson & Folkerts, 1992; Wolstenholme & Fauron, 1995). Third, within *C. cohnii*

mtDNA, many small direct and inverted repetitive sequences were detected, < 27 nt in length, scattered throughout central repeat and flanking regions. Similar small repeats have also been described in the mitochondrial genome of numerous plants (reviewed in André *et al.*, 1992). Fourth, in *C. cohnii* I detected a low-abundance sub-population of *cox1*- and *cob*-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 irreversible recombination between small repeats, have been described in angiosperm mtDNAs (Small *et al.*, 1987). Finally, short portions of disrupted protein-coding genes (*cox3* and *cox1*) were found to occur in *C. cohnii* mtDNA. Similar cases of short regions of protein-coding genes, scattered throughout plant mtDNAs (Bonen, 1993), have been reported. Because the mitochondrial genomes of plants and dinoflagellates are only distantly related, these similar recombining repeat arrangements presumably arose independently in *C. cohnii* and the flowering plant lineages.

The putative identification of *cox3* in *C. cohnii* is highly suggestive that the dinoflagellates have at least as many protein-coding genes as identified in the apicomplexans (*cox1*, *cob* and *cox3*). This finding further establishes that the dinoflagellates are more closely related to the apicomplexans than are the ciliates, because the ciliate mitochondrial genome does not encode a *cox3* gene (although it remains possible that one of the ciliate-specific unidentified ORFs in *T. pyriformis* and *P. aurelia* mtDNA may represent a highly divergent *cox3* gene). This suggests that the loss of *cox3* occurred within an ancestor that gave rise to *T. pyriformis* and *P. aurelia* and

possibly the ciliates as a group.

The origin of the isolated C-terminal portion of *cox3* in *C. cohnii* mtDNA is unknown. One possibility is that this element represents a portion of the *cox3* ORF that was separated from the rest of the gene by homologous recombination. In fact, it may be analogous to the situation seen with *cox1*, the difference being that most of the truncated *cox1* ORF was identified, whereas the C-terminal portion was not. Based on this observation I predict that the *cox3* gene will be found to exhibit a gene organization similar to that of *cox1*. In particular, I expect that the mtDNA of *C. cohnii* contains full-length as well as C-terminally truncated versions of the *cox3* ORF.

C. cohnii is not the only member of the alveolates that contains mtDNA-encoded protein-coding gene fragments. Notably, the *nad1* gene (*nad1_a* and *nad1_b*) is discontinuous and scrambled in *T. pyriformis* and *P. aurelia* mtDNAs (Burger *et al.*, 2000).

E. Expression of protein-coding genes in *C. cohnii*.

In *C. cohnii*, single discrete *cox1*, *cox3* and *cob* transcripts were detected, suggesting that the corresponding genes have their own promoter and transcription initiation sites. If post-transcriptional processing occurs, it must happen rapidly because larger *cox1*, *cox3* and *cob* transcripts are not detected in the steady-state RNA population. In contrast, the 6-kbp element in *P. falciparum* mitochondrion is polycistronically transcribed, with individual mature RNAs being generated by processing (Ji *et al.*, 1996; Rehkopf *et al.*, 2000). The situation in *T. pyriformis* mitochondria is not as well

characterized, but appears to resemble that in *P. falciparum* mitochondria. In *T. pyriformis*, it is thought that transcription of mtDNA is initiated from a small number of promoters, to produce polycistronic transcripts. So far attempts to delineate transcription initiation sites in ciliate mtDNA have been unsuccessful. This failure has been interpreted to mean that transcript processing may occur very rapidly. However, small polycistronic precursor and mature transcripts have been detected, suggesting that multiple genes can be transcribed using a single transcription initiation site (Edqvist *et al.*, 2000).

C. cohnii mtDNA contains two major but slightly different versions of the *cox1* coding element, only the longer of which (present in pc1#15 and pc1#3) specifies all 12 expected membrane-spanning domains and is stably expressed. The fact that the truncated version of *cox1* (present in pc1#2 and pc1#8) does not produce a stable *cox1* transcript is somewhat surprising, as the same mtDNA sequence is present for more than 1.5 kbp upstream of the translation initiation site in each *cox1* variant and extends more than 1.3 kbp into the coding region. Therefore, we would expect that both genes should have the same transcription initiation and promoter sites and would be transcribed. My results therefore suggest that *cox1* transcript abundance is regulated by sequences occurring at or downstream of the C-terminus. Similar to the situation described for *cox1*, *C. cohnii* mtDNA also contains three slightly different versions of the *cob* ORF. However, only one version (cob-1 as found in pcb#1) appears to be contained within a major *cob*-containing *Eco*RI fragment, the rest occurring in minor *cob*-containing *Eco*RI fragments. Surprisingly, neither *cob* variant produced a detectable transcript, despite the fact that they each possess a complete set of conserved protein-coding domains. This finding is

even more surprising considering that a discrete *cob* mRNA was detected using an internal *cob*-specific probe. One explanation is that the 'authentic' *cob* gene has not been characterized. Alternatively, *cob* transcripts specific for these *cob* variants may be present in low quantity or for technical reasons are simply undetectable using the current protocols. Support for the former interpretation comes from the fact that the internal *cob* probe generated a northern band of relatively low intensity. An even lower-intensity *cob* hybridization signal is seen in the total RNA fraction. These results, in conjunction with the fact that the predicted COB protein contains eight membrane-spanning domains (Delgi-Esposti *et al.*, 1993), makes it highly unlikely that the 'authentic' *cob* gene is nucleus-encoded.

In *C. cohnii*, inverted repeats are present in the region surrounding the C-terminus of the longer *cox1* gene but are absent in the truncated version. These repeats can be folded into an elaborate secondary structure. Accordingly, I hypothesize 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 RNA sequence that confers stability. Inverted repeats also surround the 3' end of all three *cob* ORFs. Unlike *cox1*, however, an elaborate secondary structure cannot be formed for any of the *cob* genes. The absence of a highly folded RNA structure may explain why none of the three versions of *cob* appear to be stably expressed. The presence of small inverted repeats contrasted with the lack of mRNA stability may indicate that a particular required conformation cannot be formed in the *cob* transcript. A 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, inverted repeats present in 3'-untranslated regions fold into stable stem-loop structures, thereby providing signals for mRNA processing factors as well as enhancing upstream sequence stability. 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* and *cob* transcription is under the control of a sequence element that is located within the 3' end of these ORFs. So far, however, no such mechanism has been reported in mitochondria.

The observation that inverted repeats also surround the N-terminal coding region of *cox1* and *cob* in *C. cohnii* is intriguing and implies that these repeats may be important in regulating gene expression at the 5' as well as the 3' ends of *C. cohnii* protein-coding genes. As in *C. cohnii*, the *P. yoelii cox1* gene also contains an inverted repeat 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. Suplick *et al.* (1990) also suggested that the *cox1* and *cob* genes may be transcribed from independent promoters. This is more consistent with what is found in *C. cohnii* mtDNA, where *cox1* and *cob* appear to be independently transcribed.

Despite the fact that a complete *cox3* gene has yet to be identified in *C. cohnii*, northern hybridization suggests that *cox3* is expressed as a transcript of the predicted size. Attempts to determine the complete sequence of *cox3* have so far been unsuccessful. While the reason for this is not known, a number of possibilities can be entertained. First, the C-terminal portion of *cox3* identified in this study may not be required for a functional

COX3 protein. However, this seems unlikely because of the highly conserved nature of this region and the fact that it is detected in full-length *cox3* mRNA transcripts. Second, the N- and C-termini of COX3 may be translated independently, giving rise to separate proteins. A similar explanation has been proposed for the split and rearranged *nad1* gene in *T. pyriformis* (Edqvist *et al.*, 2000). This explanation also seems unlikely, because once again full-length *cox3* mRNAs were detected using northern analysis. Third, a mature transcript may be formed by *trans*-splicing of two independent *cox3* pre-mRNAs. A similar mechanism has been described in wheat (Chapdelaine & Bonen, 1991), *Petunia* (Conklin *et al.*, 1991) and *Oenothera* (Wissinger *et al.*, 1991) mitochondria. In these cases, separate *nad1* 'modules' are flanked by incomplete group II introns. Full-length transcripts are produced when independent pre-mRNAs are brought together by association in *trans* of group II intron 'halves' followed by intron removal. It is unlikely, however, that *trans*-splicing occurs in the case of *C. cohnii cox3* transcripts because sequence characteristic of group II introns were not found flanking *cox3*. In addition, few group II introns have been identified in the protist mitochondrial lineage (Gray *et al.*, 1998). The most plausible explanation is that both full-length and fragmented versions of *cox3* are quite abundant in *C. cohnii* mtDNA, but that the full-length version was not among the small number of *cox3* clones that were isolated and characterized in the current study.

F. Future directions.

The mitochondrial genome of *C. cohnii* appears to contain few genes, which are

interspersed by long stretches of repeat-containing, intergenic spacer DNA. Therefore, the general 'shot-gun' approach to DNA sequence determination is not a particularly good strategy to use with dinoflagellate mtDNAs. In particular, the presence of numerous small repeats as well as gene pieces would make contiguous DNA assembly of randomly sequenced mtDNAs monumentally difficult, if not impossible. In this situation, a gene-targeted approach, as used in this study, is the only way to generate significant amounts of data with which to identify genes as well as make inferences about mtDNA organization and gene arrangement.

So far only two LSU rRNAs (LSUE and LSUG) have been identified in *C. cohnii* mtDNA. Each of these rRNAs is expected to interact by way of complementary base pairing to other rRNA fragments. In *Plasmodium*, LSUE pairs with LSUD whereas LSUG interacts with LSUF and LSUE (Feagin *et al.*, 1992). Based on potential pairing with *C. cohnii* LSUE and LSUG rRNAs as well as comparison with *Plasmodium* sequences, the LSUD and LSUF rRNAs and their coding elements are obvious rRNAs/rRNA genes to attempt to characterize in *C. cohnii* mitochondria. In addition, conserved portions of the SSU rRNA gene (or genes) such as the 530 and 790 loops could be targeted for characterization.

DNA sequence determination for rRNA genes will also provide valuable insight into the evolution of mtDNA organization in *C. cohnii*. Southern analysis based on *Eco*RI hydrolysis reveals that LSUG occurs in a number of different contexts. In addition, LSUE rRNA 3'-RACE results may also indicate that the 3' flanking domain as well as the 3' end of the LSUE rRNA gene are variable. In particular, it will be interesting to learn if 3' end

truncation of rRNA genes occurs, as has been identified for protein-coding genes.

One final area that deserves additional investigation is characterization of the 5' ends of *cox1*, *cob* and *cox3* mRNAs using primer extension. This information should shed further light on mitochondrial transcription and mRNA stability in *C. cohnii* mitochondria. Moreover, it should provide important information regarding which *cob* genes are potentially functional, as well as help in the identification of the authentic *cox3* gene.

G. Conclusions.

This study provides the first insights into mitochondrial genome organization in a member of the dinoflagellates, a major protist lineage whose mtDNA has resisted isolation and characterization. On balance, the mitochondrial genome of *C. cohnii* is most like that of its closest relatives, the apicomplexans. First, the *rnl* gene fragmentation pattern, as deduced from LSUE and LSUG rRNAs, is surprisingly similar between *C. cohnii* and *Plasmodium*. Second, the complement of protein-coding genes (*cox1*, *cob* and *cox3*) as well as their evolutionary affiliation, as determined by phylogenetic analysis, shows that the dinoflagellates and apicomplexans are closely allied.

However, unlike what is seen in the other members of the alveolates, *C. cohnii* mtDNA comprises long stretches of intergenic spacer regions that are sparsely populated with complete and partial genes. This feature is also unlike that of all other characterized protist mitochondrial genomes, with the exception of *Euglena gracilis*, which exhibits a similar type of mtDNA organization (David F. Spencer, personal communication).

Typically, protist mtDNAs are gene-rich. In fact, many of the features identified in *C. cohnii* mtDNA have thus far only been found in plant mtDNAs. In particular, mitochondrial gene organization in *C. cohnii* exhibits a repeat arrangement most like that found in angiosperms. This similarity in gene organization may also explain similarities in protein-coding gene expression and mRNA stability, for example, the use of inverted repeats to maintain mRNA stability.

The fact that this type of complex organization has not been documented in any other protist mtDNA suggests that it arose independently within the dinoflagellates and land plants, with recombination between direct repeats resulting in a parallel overall arrangement in the two cases. Therefore, it is surprising to find that the apicomplexans, the closest relatives to the dinoflagellates, contain the smallest known mitochondrial genome. This disparity in mtDNA organization is one the largest ever recorded between two closely related groups and resembles the degree of disparity seen between the reduced-derived and ancestral type mtDNAs found within the chlorophyte lineage (Turmel *et al.*, 1999). These results suggest that after the dinoflagellates and apicomplexans diverged, one or both lineages underwent a rapid change in genome structure.

V. Appendix

'complete' *coxI* nucleotide sequence.

```
   1 ATGGGCCAGA CATTTGTTTG TTGGACTTTA TGTCCAACAT TTATATCATC AGTAAAAAAT TGTAATCATA   70
  71 AAGGCTTAGG AATCTATTAT TTATTATCAT CATTTATATT TGGTATATCT GGTACATTAA TTTCAGTCCT  140
 141 TATAAGAATA GAATTATATT CATCAGGAAA TAGGATTTTA TCTCCAGAAA ACCAGAACTT CTATAATATA  210
 211 AAAATAACAT TACATGGGTT TCTTATGATC TTCTATTTAG TAATGCCAGG ATTATTTGGT GGATTTGGTA  280
 281 ATTATTTTAT TGTTATATTT CAAGGATCTC CAGAAGTTGT ATATCCTAGA GTTAATAATT TTTCTATCTT  350
 351 AATTCTTTCT CTTTCATATC TTTTATTAAT ATTATCTATA ATATCAGAGT TTGGAGGTGG AACAGGATGG  420
 421 ACATTATATC CACCATTAAG TACTTCTTTT ATGAATTTAT CACCATCAAG TATAGGAAAT ATTATATTTG  490
 491 GATTATTAAT ATCTGGTATA TCATCAGTTC TTACATCTCT TAATTTTTGG ATAACTATTT TATCTCTTAG  560
 561 ATCTATTGGA ATAACATTAA AGACTATATC ATTATTCCCT TGGTCATTAT TAATTACATC TGGAATGCTA  630
 631 TTATTAACAT TACCAATATT AACAGGAGCT CTATTAATGA TATTATCTGA TATAAATTAT AATACACTAT  700
 701 TCTTTGATCC AATATTTGGA GGAGATCCTA TATTCTATCA ACATTTATTT TGGTTTTTTG GTCATCCAGA  770
 771 AGTTTATATC TTAATTATTC CAGCATTTGG TATAATTAGC ATAATAATTT CTGGTATTTT ACAAAAAATA  840
 841 ATCTTTGGGA ATCCATCAAT GATCTTTGCT ATGAGCTGTA TTTCTCTTCT TGGAAGTGTT GTTTGGGGTC  910
 911 ATCATATGTA TACTATAGGA TTAGAAACTG ATACAAGATC TTATTTTTCT GGAGTTACTA TATTAATATC  980
 981 CTTACCAACA GGTACAAAAA TATTTAATTG GTTAAGTACA TATCTTGGTA ATCCTTCATT ATTATTACTT 1050
1051 AAAACTAATT CATCATTATT TGGACTTCTC TTTTTATTAA TGTTTACAAT TGGTGGTTCA ACAGGTATAA 1120
1121 TTATTGGTAA TGCAGCAGTA GATTTAGGAT TACATGATAC ATATTATATT ATAGCTCATT TTCATTTTGT 1190
1191 ATTATCATTA GGTGCTGTAA TAGCTATATT CTCTGGTATT ATATTTAATA TTGAAAAGAT TATTGGTTCT 1260
1261 AAGAATATAT TACCTTCATG CTCATCTAAT AATTCATTTT ATAATTTAGT ATTAACATTT ATTGGTATTC 1330
1331 TTATAACCTT TGGACCAATG CATTTCTTAG GATTTAATGT TATGCCAAGA AGAATACCGG ATTTTCCAGA 1400
1401 TTCTTTTATT TCTTGGAATT TTCTGTCATC TATTGGATCA GGAATAACTT TATTATCTTT TGGTTTTCTC 1470
1471 TTTAAATTTA ACTGTTGGAT CACAAGTCTG GCCCAATGTT CCAGACTTTT TCTTGGCCAT GTGGCCAAGA 1540
1541 ATATATTTAT TAACCAAACA GCAAGAGAAT TAATAATTGT TGGGCTTTAT GCCCAACATA TGTTTGGTTA 1610
1611 A
```

'short' *coxI* nucleotide sequence.

```
   1 ATGGGCCAGA CATTTGTTTG TTGGACTTTA TGTCCAACAT TTATATCATC AGTAAAAAAT TGTAATCATA   70
  71 AAGGCTTAGG AATCTATTAT TTATTATCAT CATTTATATT TGGTATATCT GGTACATTAA TTTCAGTCCT  140
 141 TATAAGAATA GAATTATATT CATCAGGAAA TAGGATTTTA TCTCCAGAAA ACCAGAACTT CTATAATATA  210
 211 AAAATAACAT TACATGGGTT TCTTATGATC TTCTATTTAG TAATGCCAGG ATTATTTGGT GGATTTGGTA  280
 281 ATTATTTTAT TGTTATATTT CAAGGATCTC CAGAAGTTGT ATATCCTAGA GTTAATAATT TTTCTATCTT  350
 351 AATTCTTTCT CTTTCATATC TTTTATTAAT ATTATCTATA ATATCAGAGT TTGGAGGTGG AACAGGATGG  420
 421 ACATTATATC CACCATTAAG TACTTCTTTT ATGAATTTAT CACCATCAAG TATAGGAAAT ATTATATTTG  490
 491 GATTATTAAT ATCTGGTATA TCATCAGTTC TTACATCTCT TAATTTTTGG ATAACTATTT TATCTCTTAG  560
 561 ATCTATTGGA ATAACATTAA AGACTATATC ATTATTCCCT TGGTCATTAT TAATTACATC TGGAATGCTA  630
 631 TTATTAACAT TACCAATATT AACAGGAGCT CTATTAATGA TATTATCTGA TATAAATTAT AATACACTAT  700
 701 TCTTTGATCC AATATTTGGA GGAGATCCTA TATTCTATCA ACATTTATTT TGGTTTTTTG GTCATCCAGA  770
 771 AGTTTATATC TTAATTATTC CAGCATTTGG TATAATTAGC ATAATAATTT CTGGTATTTT ACAAAAAATA  840
 841 ATCTTTGGGA ATCCATCAAT GATCTTTGCT ATGAGCTGTA TTTCTCTTCT TGGAAGTGTT GTTTGGGGTC  910
 911 ATCATATGTA TACTATAGGA TTAGAAACTG ATACAAGATC TTATTTTTCT GGAGTTACTA TATTAATATC  980
 981 CTTACCAACA GGTACAAAAA TATTTAATTG GTTAAGTACA TATCTTGGTA ATCCTTCATT ATTATTACTT 1050
```

```
1051 AAAACTAATT CATCATTATT TGGACTTCTC TTTTTATTAA TGTTTACAAT TGGTGGTTCA ACAGGTATAA  1120
1121 TTATTGGTAA TGCAGCAGTA GATTTAGGAT TACATGATAC ATATTATATT ATAGCTCATT TTCATTTTGT  1190
1191 ATTATCATTA GGTGCTGTAA TAGCTATATT CTCTGGTATT ATATTTAATA TTGAAAAGAT TATTGGTTCT  1260
1261 AAGAATATAT TACCTTCATG CTCATCTAA
```

cob-1 nucleotide sequence.

```
   1 ATGAAATCTC ATTTACATAC ATATCCTTGT CCTCTTCATA TAAATTCTTT TTGAAATATT GGTTTTCTTC    70
  71 TTGGAATTAC TATTTTATTA CAAATTATTA CTGGAATATT CTTATGTTTA TATTATATAT CAGATATTAA   140
 141 TTCAGCATAT AATAGTCTTT TTTATATTAT TAGAGAAATA TATTATGGAT GGTATTTACG TTATCTTCAT   210
 211 TCTAATGGAG CATCATTTGT ATTTATATTT ATATTTATAC ATTTTGGAAG AGGTATATTT TATGGTTCAT   280
 281 ATTATTATAA TACAAATACT TGGTTATCAG GAATTATTAT TTTCTTATTA TTAATGGTTA TAGGATTTAT   350
 351 GGGTTATGTC TTACCTTTTG GACAAATGAG TTTTTGGGGA GCAACTGTAA TTACTAATTT ATTATCTCCT   420
 421 TTTTCATCTT TAATAGAATA TATATCTGGA GGATATTATT TTTCTAATCC AACATTGAAG AGATTCTTTA   490
 491 TATTTCATTT TATATTACCA TTTATAATTT GTGGTTTTAT ATTTATACAT ATATTTTATC TACATTTAAT   560
 561 ATCTTCTAAT AATCCTTTAA GTAATACAAT AAATAATAAA ATACCATTTT TCCCTTTCAT ATTTATAAAA   630
 631 GATTTATTTG GTTTCATTAT AATTATCTCT ATTTATCTTC TTCAAACTAA TTATGGTATC TCTTCTTTAT   700
 701 CACATCCAGA TAATTCATTA GAAGTTTGTA TATTAATTAC TCCTTTACAT ATAGTACCTG AATGGTATTT   770
 771 CCTATGTCAA TATGCTATGT TAAAAGCTAT TCCCAATAAA AATGCAGGAT TCATAATTTT ATTAACTTCT   840
 841 ATATTTATTT TCTTAATATT TAGTGAGAAT AGAAATATAA TAACCTTGTA TATATTAAAT AATAATGGTT   910
 911 TTTATATATC TATTTCATTA TTCTTCTCTA TTTATATTTG TTTCATTTGG ATAGGTGCAC AATTACCTCA   980
 981 AGAGATGTTT ATATCTCTAT CTAAATCCTA TAAACAATGG TAA
```

cob-2 nucleotide sequence.

```
   1 ATGAAATCTC ATTTACATAC ATATCCTTGT CCTCTTCATA TAAATTCTTT TTGAAATATT GGTTTTCTTC    70
  71 TTGGAATTAC TATTTTATTA CAAATTATTA CTGGAATATT CTTATGTTTA TATTATATAT CAGATATTAA   140
 141 TTCAGCATAT AATAGTCTTT TTTATATTAT TAGAGAAATA TATTATGGAT GGTATTTACG TTATCTTCAT   210
 211 TCTAATGGAG CATCATTTGT ATTTATATTT ATATTTATAC ATTTTGGAAG AGGTATATTT TATGGTTCAT   280
 281 ATTATTATAA TACAAATACT TGGTTATCAG GAATTATTAT TTTCTTATTA TTAATGGTTA TAGGATTTAT   350
 351 GGGTTATGTC TTACCTTTTG GACAAATGAG TTTTTGGGGA GCAACTGTAA TTACTAATTT ATTATCTCCT   420
 421 TTTTCATCTT TAATAGAATA TATATCTGGA GGATATTATT TTTCTAATCC AACATTGAAG AGATTCTTTA   490
 491 TATTTCATTT TATATTACCA TTTATAATTT GTGGTTTTAT ATTTATACAT ATATTTTATC TACATTTAAT   560
 561 ATCTTCTAAT AATCCTTTAA GTAATACAAT AAATAATAAA ATACCATTTT TCCCTTTCAT ATTTATAAAA   630
 631 GATTTATTTG GTTTCATTAT AATTATCTCT ATTTATCTTC TTCAAACTAA TTATGGTATC TCTTCTTTAT   700
 701 CACATCCAGA TAATTCATTA GAAGTTTGTA TATTAATTAC TCCTTTACAT ATAGTACCTG AATGGTATTT   770
 771 CCTATGTCAA TATGCTATGT TAAAAGCTAT TCCCAATAAA AATGCAGGAT TCATAATTTT ATTAACTTCT   840
 841 ATATTTATTT TCTTAATATT TAGTGAGAAT AGAAATATAA TAACCTTGTA TATATTAAAT AATAATGGTT   910
 911 TTTATATATC TATTTCATTA TTCTTCTCTA TTTATATTTG TTTCATTTGG ATAGGTGCAC AATTACCTCA   980
 981 AGAGATGTTT ATATCTTATG GTCGTATATT AACATTACAT TATTATTTCC TTATTATTTT GTATCTGTTA  1050
1051 CCTTTAGAAA TTTCTGTTTG TTGGTGAAAA TATTGTTTGG CCAAACAACT GTTGGGCCAA CAATATTTCC  1120
1121 AACAACAATT TGTGGACCAC AGGTCCACAA AAGAAAAATA TAAATTCCAG ATCATAATAT TTCTCTTGCT  1190
1191 GTTGGACTTA TAG
```

cob-3 nucleotide sequence.

```
   1 ATGAAATCTC ATTTACATAC ATATCCTTGT CCTCTTCATA TAAATTCTTT TTGAAATATT GGTTTTCTTC    70
```

```
  71 TTGGAATTAC TATTTTATTA CAAATTATTA CTGGAATATT CTTATGTTTA TATTATATAT CAGATATTAA   140
 141 TTCAGCATAT AATAGTCTTT TTTATATTAT TAGAGAAATA TATTATGGAT GGTATTTACG TTATCTTCAT   210
 211 TCTAATGGAG CATCATTTGT ATTTATATTT ATATTTATAC ATTTTGGAAG AGGTATATTT TATGGTTCAT   280
 281 ATTATTATAA TACAAATACT TGGTTATCAG GAATTATTAT TTTCTTATTA TTAATGGTTA TAGGATTTAT   350
 351 GGGTTATGTC TTACCTTTTG GACAAATGAG TTTTTGGGGA GCAACTGTAA TTACTAATTT ATTATCTCCT   420
 421 TTTTCATCTT TAATAGAATA TATATCTGGA GGATATTATT TTTCTAATCC AACATTGAAG AGATTCTTTA   490
 491 TATTTCATTT TATATTACCA TTTATAATTT GTGGTTTTAT ATTTATACAT ATATTTTATC TACATTTAAT   560
 561 ATCTTCTAAT AATCCTTTAA GTAATACAAT AAATAATAAA ATACCATTTT TCCCTTTCAT ATTTATAAAA   630
 631 GATTTATTTG GTTTCATTAT AATTATCTCT ATTTATCTTC TTCAAACTAA TTATGGTATC TCTTCTTTAT   700
 701 CACATCCAGA TAATTCATTA GAAGTTTGTA TATTAATTAC TCCTTTACAT ATAGTACCTG AATGGTATTT   770
 771 CCTATGTCAA TATGCTATGT TAAAAGCTAT TCCCAATAAA AATGCAGGAT TCATAATTTT ATTAACTTCT   840
 841 ATATTTATTT TCTTAATATT TAGTGAGAAT AGAAATATAA TAACCTTGTA TATATTAAAT AATAATGGTT   910
 911 TTTATATATC TATTTCATTA TTCTTCTCTA TTTATATTTG TTTCATTTGG ATAGGTGCAC AATTACCTCA   980
 981 AGAGATGTTT ATATCTTATG GTCGTATATT AACATTACAT TATTATTTCC TTATTATTTT GTATCTGTTA  1050
1051 CCTTTAGAAA TTTCTGTTTG TTGCTGCCAG AGAATTATTG GTTAA
```

VI. References

Adachi, J. & Hasegawa, M. (1996a). MOLPHY version 2.3. (Programs for molecular phylogenetics based on maximum likelihood). Distributed by the authors, Institute of Statistical Mathematics, University of Tokyo, Japan.

Adachi, J. & Hasegawa, M. (1996b). Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J Mol Evol* **42**, 459-68.

Aldritt, S. M., Joseph, J. T. & Wirth, D. F. (1989). Sequence identification of cytochrome *b* in *Plasmodium gallinaceum*. *Mol Cell Biol* **9**, 3614-20.

Allen, J. R., Roberts, T. M., Loeblich, A. R. III. & Klotz, L. C. (1975). Characterization of the DNA from the dinoflagellate *Crypthecodinium cohnii* and implications for nuclear organization. *Cell* **6**, 161-9.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-402.

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. & Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457-65.

André, C., Levy, A. & Walbot, V. (1992). Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet* **8**, 128-32.

Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidman, J., Smith, J. & Struhl, K. (1987). *Current Protocols in Molecular Biology*, 1, John Wiley and Sons, New York.

Barbrook, A. C. & Howe, C. J. (2000). Minicircular plastid DNA in the dinoflagellate *Amphidinium operculatum*. *Mol Gen Genet* **263**, 152-8.

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**, 5057-68.

Bhaud, Y., Guillebault, D., Lennon, J., Defacque, H., Soyer-Gobillard, M. & Moreau, H. (2000). Morphology and behaviour of dinoflagellate chromosomes during the cell cycle and mitosis. *J Cell Sci* **113**, 1231-9.

Blanchard, J. L. & Hicks, J. S. (1999). The non-photosynthetic plastid in malarial parasites and other apicomplexans is derived from outside the green plastid lineage. *J Eukaryot Microbiol* **46**, 367-75.

Boer, P. H., Bonen, L., Lee, R. W. & Gray, M. W. (1985). Genes for respiratory chain proteins and ribosomal RNAs are present on a 16-kilobase-pair DNA species from *Chlamydomonas reinhardtii* mitochondria. *Proc Natl Acad Sci U S A* **82**, 3340-4.

Boer, P. H. & Gray, M. W. (1988). Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *Cell* **55**, 399-411.

Bonen, L. (1993). Trans-splicing of pre-mRNA in plants, animals, and protists. *FASEB J* **7**, 40-6.

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Res* **27**, 1767-80.

Borson, N. D., Salo, W. L. & Drewes, L. R. (1992). A lock-docking oligo(dT) primer for 5' and 3' RACE PCR. *PCR Methods Appl* **2**, 144-8.

Bui, E. T., Bradley, P. J. & Johnson, P. J. (1996). A common evolutionary origin for mitochondria and hydrogenosomes. *Proc Natl Acad Sci U S A* **93**, 9651-6.

Burger, G., Zhu, Y., Littlejohn, T. G., Greenwood, S. J., Schnare, M. N., Lang, B. F. & Gray, M. W. (2000). Complete sequence of the mitochondrial genome of *Tetrahymena pyriformis* and comparison with *Paramecium aurelia* mitochondrial DNA. *J Mol Biol* **297**, 365-80.

Cavalier-Smith, T. (1975). The origin of nuclei and of eukaryote cells. *Nature* **256**, 463-8.

Cavalier-Smith, T. (1983). A 6-kingdom classification and a unified phylogeny. In *Endocytobiology* (Schwemmler, W. & Schenk, H. E. A., eds.), pp. 1027-34. De Gruyter, Berlin.

Cavalier-Smith, T. (1993). Kingdom protozoa and its 18 phyla. *Microbiol Rev* **57**, 953-94.

Cavalier-Smith, T. (1999). Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Euk Micro* **46**, 347-66.

Chapdelaine, Y. & Bonen, L. (1991). The wheat mitochondrial gene for subunit I of the NADH dehydrogenase complex: a trans-splicing model for this gene-in-pieces. *Cell* **65**, 465-72.

Chesnick, J. M., Morden, C. W. & Schmieg, A. M. (1996). Identity of the endosymbiont of *Peridinium foliaceum* (Pyrrophyta): Analysis of the *rbcLS* operon. *J Phycol* **32**, 850-75.

Conklin, P. L., Wilson, R. K. & Hanson, M. R. (1991). Multiple *trans*-splicing events are required to produce a mature *nad1* transcript in a plant mitochondrion. *Genes Dev* **5**, 1407-15.

Coulthart, M. B., Huh, G. S. & Gray, M. W. (1990). Physical organization of the 18S and 5S ribosomal RNA genes in the mitochondrial genome of rye (*Secale cereale* L.). *Curr Genet* **17**, 339-46.

Cummings, D. J. (1992). Mitochondrial genomes of the ciliates. *Int Rev Cytol* **141**, 1-64.

Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1979). A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed.), pp. 345-52. National Biomedical Research Foundation, Washington, D.C.

Delgi-Esposti, M. D., De Vries, S., Crimi, M., Ghelli, A., Patarnello, T. & Meyer, A. (1993). Mitochondrial cytochrome *b*: evolution and structure of the protein. *Biochim Biophys Acta* **26**, 243-71.

Delwiche, C. F. (1999). Tracing the thread of plastid diversity through the tapestry of life. *The American Naturalist [Suppl.]* **154**, 164-77.

Delwiche, C. F., Kuhsel, M. & Palmer, J. D. (1995). Phylogenetic analysis of *tufA* sequences indicates a cyanobacterial origin of all plastids. *Mol Phylogenet Evol* **4**, 110-28.

Denny, P., Preiser, P., Williamson, D. & Wilson, I. (1998). Evidence for a single origin of the 35 Kb plastid DNA in Apicomplexans. *Protist* **149**, 51-9.

Denovan-Wright, E. M. & Lee, R. W. (1995). Evidence that the fragmented ribosomal RNAs of *Chlamydomonas* mitochondria are associated with ribosomes. *FEBS Lett* **370**, 222-6.

Dodge, J. D. (1965). Chromosome structure in the dinoflagellates and the problem of the mesokaryotic cell. *International Congress Series Excerpta Medica* **91**, 264-5.

Dodge, J. D. (1985). *Atlas of Dinoflagellates: A Scanning Electron Microscope Survey*, Farrand, London.

Dodge, J. D. (1989). Phylogenetic relationships of dinoflagellates and their plastids. In *The Chromophyte Algae: problems and perspectives.* (Green, J. C., Leadbeater, B. S. C. & Diver, W. L., eds.), pp. 207-27. Clarendon Press, Oxford.

Dombrowski, S., Brennicke, A. & Binder, S. (1997). 3'-Inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators. *EMBO J* **16**, 5069-76.

Durnford, D. G., Deane, J. A., Tan, S., McFadden, G. I., Gantt, E. & Green, B. R. (1999). A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution. *J Mol Evol* **48**, 59-68.

Edqvist, J., Burger, G. & Gray, M. W. (2000). Expression of mitochondrial protein-coding genes in *Tetrahymena pyriformis*. *J Mol Biol* **297**, 381-93.

Falah, M. & Gupta, R. S. (1994). Cloning of the *hsp70* (*dnaK*) genes from *Rhizobium meliloti* and *Pseudomonas cepacia*: phylogenetic analyses of mitochondrial origin based on a highly conserved protein sequence. *J Bacteriol* **176**, 7748-53.

Feagin, J. E. (1992). The 6-kb element of *Plasmodium falciparum* encodes mitochondrial cytochrome genes. *Mol Biochem Parasitol* **52**, 145-8.

Feagin, J. E. (1994). The extrachromosomal DNAs of apicomplexan parasites. *Annu Rev Microbiol* **48**, 81-104.

Feagin, J. E. (2000). Mitochondrial genome diversity in parasites. *Int J Parasitol* **30**, 371-90.

Feagin, J. E., Mericle, B. L., Werner, E. & Morris, M. (1997). Identification of additional rRNA fragments encoded by the *Plasmodium falciparum* 6 kb element. *Nucleic Acids Res* **25**, 438-46.

Feagin, J. E., Werner, E., Gardner, M. J., Williamson, D. H. & Wilson, R. J. (1992). Homologies between the contiguous and fragmented rRNAs of the two *Plasmodium falciparum* extrachromosomal DNAs are limited to core sequences. *Nucleic Acids Res* **20**, 879-87.

Feinberg, A. P. & Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* **132**, 6-13.

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401-10.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783-91.

Felsenstein, J. (1993). Phylip (Phylogenetic inference Package) version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle.

Flores, B. S., Siddall, M. E. & Burreson, E. M. (1996). Phylogeny of the Haplosporidia (Eukaryota: Alveolata) based on small subunit ribosomal RNA gene sequence. *J Parasitol* **82**, 616-23.

Gaines, G. & Elbrachter, M. (1987). Heterotrophic nutrition. In *The biology of Dinoflagellates* (Taylor, F. J. R., ed.), pp. 224-68. Blackwell Scientific Publication, Oxford.

Gajadhar, A. A., Marquardt, W. C., Hall, R., Gunderson, J., Ariztia-Carmona, E. V. & Sogin, M. L. (1991). Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Mol Biochem Parasitol* **45**, 147-54.

Gardner, M. J., Feagin, J. E., Moore, D. J., Rangachari, K., Williamson, D. H. & Wilson, R. J. (1993). Sequence and organization of large subunit rRNA genes from the extrachromosomal 35 kb circular DNA of the malaria parasite *Plasmodium falciparum*. *Nucleic Acids Res* **21**, 1067-71.

Gast, R. J. & Caron, D. A. (1996). Molecular phylogeny of symbiotic dinoflagellates from planktonic foraminifera and radiolaria. *Mol Biol Evol* **13**, 1192-7.

Germot, A., Philippe, H. & Le Guyader, H. (1996). Presence of a mitochondrial-type 70-kDa heat shock protein in *Trichomonas vaginalis* suggests a very early mitochondrial endosymbiosis in eukaryotes. *Proc Natl Acad Sci USA* **93**, 14614-7.

Germot, A., Philippe, H. & Le Guyader, H. (1997). Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. *Mol Biochem Parasitol* **87**, 159-68.

Gillespie, D. E., Salazar, N. A., Rehkopf, D. H. & Feagin, J. E. (1999). The fragmented mitochondrial ribosomal RNAs of *Plasmodium falciparum* have short A tails. *Nucleic Acids Res* **27**, 2416-22.

Gold, K. & Baren, C. F. (1966). Growth requirements of *Gyrodinium cohnii*. *J Protozool* **13**, 255-7.

Gordon, G. S. & Wright, A. (1998). DNA segregation: putting chromosomes in their place. *Curr Biol* **8**, R925-7.

Gray, M. W., Burger, G. & Lang, B. F. (1999). Mitochondrial evolution. *Science* **283**, 1476-81.

Gray, M. W., Lang, B. F., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M., Brossard, N., Delage, E., Littlejohn, T. G., 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**, 865-78.

Gray, M. W. & Spencer, D. F. (1996). Organellar evolution. In *Evolution of Microbial life* (Roberts, D. M., Sharp, P., Alderson, G. & Collins, M., eds.), pp. 109-26. Cambridge University Press, Cambridge, UK.

Haapala, O. K. & Soyer, M. O. (1974). Size of circular chromatids and amounts of haploid DNA in the dinoflagellates *Gyrodinium cohnii* and *Prorocentrum micans*. *Hereditas* **76**, 83-9.

Hall, R., Coggins, L., McKellar, S., Shiels, B. & Tait, A. (1990). Characterisation of an extrachromosomal DNA element from *Theileria annulata*. *Mol Biochem Parasitol* **38**, 253-60.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557-80.

Hansen, P. J. (1991). Quantitative importance and trophic role of heterotrophic dinoflagellates in a coastal pelagial food web. *Marine Ecology Progress Series* **73**, 253-61.

Hansen, P. J. & Calado, A. J. (1999). Phagotrophic mechanisms and prey selection in free-living dinoflagellates. *J Euk Micro* **46**, 382-9.

Hanson, M. R. & Folkerts, O. (1992). Structure and function of the higher plant mitochondrial genome. *Int Rev Cytol* **141**, 129-72.

Hashimoto, T., Sanchez, L. B., Shirakura, T., Muller, M. & Hasegawa, M. (1998). Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny. *Proc Natl Acad Sci USA* **95**, 6860-5.

Heinonen, T. Y., Schnare, M. N. & Gray, M. W. (1990). Sequence heterogeneity in the duplicate large subunit ribosomal RNA genes of *Tetrahymena pyriformis* mitochondrial DNA. *J Biol Chem* **265**, 22336-41.

Heinonen, T. Y., Schnare, M. N., Young, P. G. & Gray, M. W. (1987). Rearranged coding segments, separated by a transfer RNA gene, specify the two parts of a discontinuous large subunit ribosomal RNA in *Tetrahymena pyriformis* mitochondria. *J Biol Chem* **262**, 2879-87.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351-9.

Herzog, M. & Maroteaux, L. (1986). Dinoflagellate 17S rRNA sequence inferred from the gene sequence: evolutionary implications. *Proc Natl Acad Sci USA* **83**, 8644-8.

Hinnebusch, A. G., Klotz, L. C., Blanken, R. L. & Loeblich, A. R. D. (1981). An evaluation of the phylogenetic position of the dinoflagellate *Crypthecodinium cohnii* based on 5S rRNA characterization. *J Mol Evol* **17**, 334-7.

Hirt, R. P., Healy, B., Vossbrinck, C. R., Canning, E. U. & Embley, T. M. (1997). A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. *Curr Biol* **7**, 995-8.

Hirt, R. P., Logsdon, J. M., Jr., Healy, B., Dorey, M. W., Doolittle, W. F. & Embley, T. M. (1999). Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc Natl Acad Sci USA* **96**, 580-5.

Holm-Hansen, O. (1969). Algae: amounts of DNA and organic carbon in single cells. *Science* **163**, 87-8.

Horwitz, J. P., Chua, J., Curby, R. J., Tomson, A. J., DaRooge, M. A., Fisher, B. E., Mauricio, J. & Klundt, I. (1964). Substrates for cytochemical demonstration of enzyme activity. I. Some substituted 3-indolyl-β-D-glycopyranosides. *J Med Chem* **7**, 574.

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.

Jacobson, D. M. (1999). A brief history of dinoflagellate feeding research. *J Euk Micro* **46**, 376-81.

Jeong, H. J. (1999). The ecological roles of heterotrophic dinoflagellates in marine planktonic community. *J Euk Micro* **46**, 390-6.

Ji, Y. E., Mericle, B. L., Rehkopf, D. H., Anderson, J. D. & Feagin, J. E. (1996). The *Plasmodium falciparum* 6 kb element is polycistronically transcribed. *Mol Biochem Parasitol* **81**, 211-23.

Joseph, J. T., Aldritt, S. M., Unnasch, T., Puijalon, O. & Wirth, D. F. (1989). Characterization of a conserved extrachromosomal element isolated from the avian malarial parasite *Plasmodium gallinaceum*. *Mol Cell Biol* **9**, 3621-9.

Jukes, T. H. & Osawa, S. (1990). The genetic code in mitochondria and chloroplasts. *Experientia* **46**, 1117-26.

Kairo, A., Fairlamb, A. H., 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.

Kamaishi, T., Hashimoto, T., Nakamura, Y., Masuda, Y., Nakamura, F., Okamoto, K., Shimizu, M. & Hasegawa, M. (1996). Complete nucleotide sequences of the genes encoding translation elongation factors 1 alpha and 2 from a microsporidian parasite, *Glugea plecoglossi*: implications for the deepest branching of eukaryotes. *J Biochem (Tokyo)* **120**, 1095-103.

Karlovsky, P. & Fartmann B. (1992). Genetic code and phylogenetic origin of oomycetous mitochondria. *J Mol Evol* **34**, 254-8.

Keeling, P. J. & Doolittle, W. F. (1996). Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Mol Biol Evol* **13**, 1297-305.

Kotob, S. I., McLaughlin, S. M., van Berkum, P. & Faisal, M. (1999). Characterization of two *Perkinsus* spp. from the softshell clam, *Mya arenaria* using the small subunit ribosomal RNA gene. *J Eukaryot Microbiol* **46**, 439-44.

Kroymann, J. & Zetsche, K. (1998). The mitochondrial genome of *Chlorogonium elongatum* inferred from the complete sequence. *J Mol Evol* **47**, 431-40.

Kubai, D. F. & Ris, H. (1969). Division in the dinoflagellate *Gyrodinium cohnii* (Schiller). A new type of nuclear reproduction. *J Cell Biol* **40**, 508-28.

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**, 2571-6.

Lang, B., Burger, G., Doxiadis, I., Thomas, D. Y., Bandlow, W. & Kaudewitz, F. (1977). A simple method for the large-scale preparation of mitochondria from microorganisms. *Anal Biochem* **77**, 110-21.

Lang, B. F., Burger, G., O'Kelly, C. J., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M. & Gray, M. W. (1997). An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* **387**, 493-7.

Lang, B. F., Cedergren, R. & Gray, M. W. (1987). The mitochondrial genome of the fission yeast, *Schizosaccharomyces pombe*. Sequence of the large-subunit ribosomal RNA gene, comparison of potential secondary structure in fungal mitochondrial large-subunit rRNAs and evolutionary considerations. *Eur J Biochem* **169**, 527-37.

Lang, B. F., Gray, M. W. & Burger, G. (1999a). Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet* **33**, 351-97.

Lang, B. F., Seif, E., Gray, M. W., O'Kelly, C. J. & Burger, G. (1999b). A comparative genomics approach to the evolution of eukaryotes and their mitochondria. *J Eukaryot Microbiol* **46**, 320-6.

Leadbeater, B. & Dodge, J. D. (1967). Fine structure of the dinoflagellate transverse flagellum. *Nature* **213**, 421-2.

Lenaers, G., Scholin, C., Bhaud, Y., Saint-Hilaire, D. & Herzog, M. (1991). A molecular phylogeny of dinoflagellate protists (pyrrhophyta) inferred from the sequence of 24S rRNA divergent domains D1 and D8. *J Mol Evol* **32**, 53-63.

Lingner, J. & Keller, W. (1993). 3'-end labeling of RNA with recombinant yeast poly(A) polymerase. *Nucleic Acids Res* **21**, 2917-20.

Liu, M. H., Reddy, R., Henning, D., Spector, D. & Busch, H. (1984). Primary and secondary structure of dinoflagellate U5 small nuclear RNA. *Nucleic Acids Res* **12**, 1529-42.

McFadden, G. I, Tomavo, S., Berry, E. A. & Boothroyd, J. C. (2000). Characterization of cytochrome *b* from *Toxoplasma gondii* and Q(o) domain mutations as a mechanism of atovaquone-resistance. *Mol Biochem Parasitol* **108**, 1-12.

McFadden, G. I., Waller, R. F., Reith, M. E. & Lang-Unnasch, N. (1997). Plastids in apicomplexan parasites. *Pl. Syst. Evol. [Suppl.]* **11**, 261-87.

McIntosh, M. T., Srivastava, R. & Vaidya, A. B. (1998). Divergent evolutionary constraints on mitochondrial and nuclear genomes of malaria parasites. *Mol Biochem Parasitol* **95**, 69-80.

McNally, K. L., Govind, N. S., Thome, P. E. & Trench, R. K. (1994). Small-subunit ribosomal DNA sequence analyses and a reconstruction of the inferred phylogeny among symbiotic dinoflagellates (Pyrrophyta). *J Phycol* **30**, 316-29.

Morin, G. B. & Cech, T. R. (1986). The telomeres of the linear mitochondrial DNA of *Tetrahymena thermophila* consist of 53 bp tandem repeats. *Cell* **46**, 873-83.

Morin, G. B. & Cech, T. R. (1988a). Mitochondrial telomeres: surprising diversity of repeated telomeric DNA sequences among six species of *Tetrahymena*. *Cell* **52**, 367-74.

Morin, G. B. & Cech, T. R. (1988b). Phylogenetic relationships and altered genome structures among *Tetrahymena* mitochondrial DNAs. *Nucleic Acids Res* **16**, 327-46.

Morin, G. B. & Cech, T. R. (1988c). Telomeric repeats of *Tetrahymena malaccensis* mitochondrial DNA: a multimodal distribution that fluctuates erratically during growth. *Mol Cell Biol* **8**, 4450-8.

Morrill, L. C. & Loeblich, A. R. III. (1983). Ultrastructure of the dinoflagellate amphiesma. *Int Rev Cytol* **82**, 151-80.

Mosig, G. (1987). The essential role of recombination in phage T4 growth. *Annu Rev Genet* **21**, 347-71.

Nedelcu, A. M. (1998). Contrasting mitochondrial genome organizations and sequence affiliations among green algae: potential factors, mechanisms and evolutionary scenarios. *J Phycol* **34**, 16-28.

Nedelcu, A. M. & Lee, R. W. (1998). Short repetitive sequences in green algal mitochondrial genomes: potential roles in mitochondrial genome evolution. *Mol Biol Evol* **15**, 690-701.

Nedelcu, A. M., Lee, R. W., Lemieux, C., Gray, M. W. & Burger, G. (2000). The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res* **10**, 819-31.

Nelissen, B., Van de Peer, Y., Wilmotte, A. & De Wachter, R. (1995). An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences. *Mol Biol Evol* **12**, 1166-73.

Nene, V., Morzaria, S. & Bishop, R. (1998). Organisation and informational content of the *Theileria parva* genome. *Mol Biochem Parasitol* **95**, 1-8.

Norman, J. E. & Gray, M. W. (1997). The cytochrome oxidase subunit 1 gene (*cox1*) from the dinoflagellate, *Crypthecodinium cohnii*. *FEBS Lett* **413**, 333-8.

Ojala, D., Montoya, J. & Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**, 470-4.

Osawa, S., Jukes, T. H., Watanabe, K. & Muto, A. (1992). Recent evidence for evolution of the genetic code. *Microbiol Rev* **56**, 229-64.

Ossorio, P. N., Sibley, L. D. & Boothroyd, J. C. (1991). Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of *Toxoplasma gondii*. *J Mol Biol* **222**, 525-36.

Palmer, J. D. & Delwiche, C. F. (1998). The origin and evolution of plastids and their genomes. In *Molecular Systematics of Plants. II. DNA Sequencing*. (Soltis, P. S., Soltis, D. E. & Doyle, J. J., eds.), pp. 375-409. Kluwer Academic, Boston.

Paquin, B. & Lang, B. F. (1996). The mitochondrial DNA of *Allomyces macrogynus*: the complete genomic sequence from an ancestral fungus. *J Mol Biol* **255**, 688-701.

Parish, J. H. & Kirby, K. S. (1966). Reagents which reduce interactions between ribosomal RNA and rapidly labelled RNA from rat liver. *Biochim Biophys Acta* **129**, 554-62.

Perkins, F. O. (1976). Zoospores of the oyster pathogen, *Dermocystidium marinum*. I. Fine structure of the conoid and other sporozoan-like organelles. *J Parasitol* **62**, 959.

Perkins, F. O. (1996). The structure of *Perkins marinus* (Mackin, Owen and Collier, 1950) Levine, 1978 with comments on taxonomy and phylogeny of *Pekinsus* spp. *J Shellfish Res* **15**, 67-87.

Philippe, H. & Forterre, P. (1999). The rooting of the universal tree of life is not reliable. *J Mol Evol* **49**, 509-23.

Preiser, P. R., Wilson, R. J., Moore, P. W., McCready, S., Hajibagheri, M. A., Blight, K. J., Strath, M. & Williamson, D. H. (1996). Recombination associated with replication of malarial mitochondrial DNA. *EMBO J* **15**, 684-93.

Pritchard, A. E., Laping, J. L., Seilhamer, J. J. & Cummings, D. J. (1983). Inter-species sequence diversity in the replication initiation region of *Paramecium* mitochondrial DNA. *J Mol Biol* **164**, 1-15.

Pritchard, A. E., Sable, C. L., Venuti, S. E. & Cummings, D. J. (1990a). Analysis of NADH dehydrogenase proteins, ATPase subunit 9, cytochrome *b*, and ribosomal protein L14 encoded in the mitochondrial DNA of *Paramecium*. *Nucleic Acids Res* **18**, 163-71.

Pritchard, A. E., Seilhamer, J. J. & Cummings, D. J. (1986). *Paramecium* mitochondrial DNA sequences and RNA transcripts for cytochrome oxidase subunit I, URF1, and three ORFs adjacent to the replication origin. *Gene* **44**, 243-53.

Pritchard, A. E., Seilhamer, J. J., Mahalingam, R., Sable, C. L., Venuti, S. E. & Cummings, D. J. (1990b). Nucleotide sequence of the mitochondrial genome of *Paramecium*. *Nucleic Acids Res* **18**, 173-80.

Rae, P. M. (1973). 5-Hydroxymethyluracil in the DNA of a dinoflagellate. *Proc Natl Acad Sci USA* **70**, 1141-5.

Rae, P. M. (1976). Hydroxymethyluracil in eukaryote DNA: a natural feature of the pyrrophyta (dinoflagellates). *Science* **194**, 1062-4.

Rae, P. M. & Steele, R. E. (1978). Modified bases in the DNAs of unicellular eukaryotes: an examination of distributions and possible roles, with emphasis on hydroxymethyluracil in dinoflagellates. *Biosystems* **10**, 37-53.

Ragan, M. A. & Chapman, D. J. (1978). *A Biochemical Phylogeny of Protists*. Academic Press, New York.

Reddy, R., Spector, D., Henning, D., Liu, M. H. & Busch, H. (1983). Isolation and partial characterization of dinoflagellate U1-U6 small RNAs homologous to rat U small nuclear RNAs. *J Biol Chem* **258**, 13965-9.

Reece, K. S., Siddall, M. E., Burreson, E. M. & Graves, J. E. (1997). Phylogenetic analysis of *Perkinsus* based on actin gene sequences. *J Parasitol* **83**, 417-23.

Reed, K. C. & Mann, D. A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res* **13**, 7207-21.

Rehkopf, D. H., Gillespie, D. E., Harrell, M. I. & Feagin, J. E. (2000). Transcriptional mapping and RNA processing of the *Plasmodium falciparum* mitochondrial mRNAs. *Mol Biochem Parasitol* **105**, 91-103.

Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. (1998). Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci USA* **95**, 6239-44.

Rizzo, P. J. & Burghardt, R. C. (1982). Histone-like protein and chromatin structure in the wall-less dinoflagellate *Gymnodinium nelsoni*. *Biosystems* **15**, 27-34.

Rizzo, P. J. & Nooden, L. D. (1974a). Isolation and partial characterization of dinoflagellate chromatin. *Biochim Biophys Acta* **349**, 402-14.

Rizzo, P. J. & Nooden, L. D. (1974b). Partial characterization of dinoflagellate chromosomal proteins. *Biochim Biophys Acta* **349**, 415-27.

Roger, A. J., Clark, C. G. & Doolittle, W. F. (1996). A possible mitochondrial gene in the early-branching amitochondriate protist *Trichomonas vaginalis*. *Proc Natl Acad Sci USA* **93**, 14618-22.

Roger, A. J., Svärd, S. G., Tovar, J., Clark, C. G., Smith, M. W., Gillin, F. D. & Sogin, M. L. (1998). A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proc Natl Acad Sci USA* **95**, 229-34.

Rowan, R. & Powers, D. A. (1992). Ribosomal RNA sequences and the diversity of symbiotic dinoflagellates (zooxanthellae). *Proc Natl Acad Sci USA* **89**, 3639-43.

Sadler, L. A., McNally, K. L., Govind, N. S., Brunk, C. F. & Trench, R. K. (1992). The nucleotide sequence of the small subunit ribosomal RNA gene from *Symbiodinium pilosum*, a symbiotic dinoflagellate. *Curr Genet* **21**, 409-16.

Sambrook, J., Fritsch, E. & Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual*. 2nd edit. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Saunders, G. W., Hill, D. R., Sexton, J. P. & Anderson, R. A. (1997). Small-subunit ribosomal RNA sequence from selected dinoflagellates: testing classical evolutionary hypothesis with molecular systematic methods. *Pl. Sys. Evol. [Suppl.]* **11**, 237-59.

Schlegel, M. (1991). Protist evolution and phylogeny as discerned from small subunit ribosomal RNA sequence comparisons. *Eur J Protistol* **27**, 207-19.

Schnare, M. N., Heinonen, T. Y., Young, P. G. & Gray, M. W. (1986). A discontinuous small subunit ribosomal RNA in *Tetrahymena pyriformis* mitochondria. *J Biol Chem* **261**, 5187-93.

Schnepf, E. (1993). From prey via endosymbiont to plastids: comparative studies in the dinoflagellates. In *Origins of Plastids* (Lewin, R. A., ed.), pp. 53-76. Chapman and Hall, New York.

Schnepf, E. & Elbrachter, M. (1992). Nutritional strategies in dinoflagellates: A review with emphasis on cell biological aspects. *Eur J Protistol* **28**, 3-24.

Seilhamer, J. J., Gutell, R. R. & Cummings, D. J. (1984a). *Paramecium* mitochondrial genes. II. Large subunit rRNA gene sequence and microevolution. *J Biol Chem* **259**, 5173-81.

Sharma, I., Pasha, S. T. & Sharma, Y. D. (1998). Complete nucleotide sequence of the *Plasmodium vivax* 6 kb element. *Mol Biochem Parasitol* **97**, 259-63.

Shukla, G. C. & Nene, V. (1998). Telomeric features of *Theileria parva* mitochondrial DNA derived from cycle sequence data of total genomic DNA. *Mol Biochem Parasitol* **95**, 159-63.

Siddall, M. E., Stokes, N. A. & Burreson, E. M. (1995). Molecular phylogenetic evidence that the phylum Haplosporidia has an alveolate ancestry. *Mol Biol Evol* **12**, 573-81.

Sleigh, M. A. (1989). *Protozoa and Other Protists*. Edward Arnold, London.

Slomianny, C. & Prensier, G. (1986). Application of the serial sectioning and tridimensional reconstruction techniques to the morphological study of the *Plasmodium falciparum* mitochondrion. *J Parasitol* 72, 595-8.

Small, I. D., Isaac, P. G. & Leaver, C. J. (1987). Stoichiometric differences in DNA molecules containing the *atpA* gene suggest mechanisms for the generation of mitochondrial genome diversity in maize. *EMBO J* **6**, 865-9.

Small, I., Suffolk, R. & Leaver, C. J. (1989). Evolution of plant mitochondrial genomes via substoichiometric intermediates. *Cell* **58**, 69-76.

Smallman, D. S., Schnare, M. N. & Gray, M. W. (1996). RNA:RNA interactions in the large subunit ribosomal RNA of *Euglena gracilis*. *Biochim Biophys Acta* **1305**, 1-6.

Smith, S. W., Overbeek, R., Woese, C. R., Gilbert, W. & Gillevet, P. M. (1994). The genetic data environment an expandable GUI for multiple sequence analysis. *Comput Appl Biosci* **10**, 671-5.

Sogin, M. L. (1991). Early evolution and the origin of eukaryotes. *Curr Opin Genet Dev* **1**, 457-63.

Sogin, M. L. (1997). History assignment: when was the mitochondrion founded? *Curr Opin Genet Dev* **7**, 792-9.

Spector, D. L. (1984a). Dinoflagellate Nuclei. In *Dinoflagellates* (Spector, D. L., ed.), pp. 107-47. Academic Press, Orlando.

Spector, D. L. (1984b). Dinoflagellates: An introduction. In *Dinoflagellates* (Spector, D. L., ed.), pp. 1-15. Academic Press, Orlando.

Spector, D. L. & Triemer, R. E. (1981). Chromosome structure and mitosis in the dinoflagellates: an ultrastructural approach to an evolutionary problem. *Biosystems* **14**, 289-98.

Spencer, D. F., Gray, M. W. & Schnare, M. N. (1992). The isolation of wheat mitochondria DNA and RNA. In *Modern methods of plant analysis new series* (Linskens, H. F. & Jackson, J. F., eds.), Vol. 14, Seed analysis, pp. 347-60. Springer-Verlag, New York.

Stoecker, D. K. (1999). Mixotrophy among dinoflagellates. *J Euk Micro* **46**, 397-401.

Suplick, K., Morrisey, J. & Vaidya, A. B. (1990). Complex transcription from the extrachromosomal DNA encoding mitochondrial functions of *Plasmodium yoelii*. *Mol Cell Biol* **10**, 6381-8.

Taylor, F. J. (1978). Problems in the development of an explicit hypothetical phylogeny of the lower eukaryotes. *Biosystems* **10**, 67-89.

Taylor, F. J. R. (1976). Flagellate phylogeny: a study of conflicts. *J Protozool* **23**, 28-40.

Tengs, T., Dahlberg, O. J., Shalchian-Tabrizi, K., Klaveness, D., Rudi, K., Delwiche, C. F. & Jakobsen, K. S. (2000). Phylogenetic analyses indicate that the 19'-hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol Biol Evol* **17**, 718-29.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80.

Trumpower, B. L. & Gennis, R. B. (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.

Turmel, M., Lemieux, C., Burger, G., Lang, B. F., Otis, C., Plante, I. & Gray, M. W. (1999). The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*. Two radically different evolutionary patterns within green algae. *Plant Cell* **11**, 1717-30.

Tuttle, R. C. & Loeblich, A. R. III. (1975). An optimal growth medium for the dinoflagellate *Crypthecodinium cohnii*. *Phycologia* **14**, 1-8.

Unseld, M., Marienfeld, J. R., Brandt, P. & Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat Genet* **15**, 57-61.

Vaidya, A. B., Akella, R. & Suplick, K. (1989). Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite. *Mol Biochem Parasitol* **35**, 97-107.

Vaidya, A. B. & Arasu, P. (1987). Tandemly arranged gene clusters of malarial parasites that are highly conserved and transcribed. *Mol Biochem Parasitol* **22**, 249-57.

Vaidya, A. B., Lashgari, M. S., Pologe, L. G. & Morrisey, J. (1993). Structural features of *Plasmodium* cytochrome *b* that may underlie susceptibility to 8-aminoquinolines and hydroxynaphthoquinones. *Mol Biochem Parasitol* **58**, 33-42.

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**, 201-10.

Vernet, G., Sala-Rovira, M., Maeder, M., Jacques, F. & Herzog, M. (1990). Basic nuclear proteins of the histone-less eukaryote *Crypthecodinium cohnii* (Pyrrhophyta): two-dimensional electrophoresis and DNA-binding properties. *Biochim Biophys Acta* **1048**, 281-9.

Watanabe, K. I., Bessho, Y., Kawasaki, M. & Hori, H. (1999). Mitochondrial genes are found on minicircle DNA molecules in the mesozoan animal *Dicyema*. *J Mol Biol* **286**, 645-50.

Watanabe, M. M., Suda, S., Inouye, I., Sawaguchi, T. & Chihara, M. (1990). *Lepidodinium viride* gen. et sp. nov. (Gymnodinilales, Dinophyta), a green dinoflagellate with a chlorophyll *a*- and *b*-containing endosymbiont. *J Phycol* **26**, 741-51.

Watanabe, M. M., Takeda, Y., Sasa, T., Inouye, I., Suda, S., Sawaguchi, T. & Chihara, M. (1987). A green dinoflagellate with chlorophylls *a*- and *b*-: morphology, fine structure of the chloroplast and chlorophyll composition. *J Phycol* **23**, 382-9.

Wilson, R. J., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D. J., Moore, P. W. & Williamson, D. H. (1996). Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J Mol Biol* **261**, 155-72.

Wilson, R. J. & Williamson, D. H. (1997). Extrachromosomal DNA in the Apicomplexa. *Microbiol Mol Biol Rev* **61**, 1-16.

Wilson, R. J. M. (1998). Plastid-like DNA in apicomplexans. In *Evolutionary relationships among protozoa* (Coombs, G. H., Vickerman, K., Sleigh, M. A. & Warren, A., eds.). Chapman and Hall, London.

Wissinger, B., Schuster, W. & Brennicke, A. (1991). Trans splicing in *Oenothera* mitochondria: *nad1* mRNAs are edited in exon and trans-splicing group II intron sequences. *Cell* **65**, 473-82.

Wolstenholme, D. R. (1992). Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol* **141**, 173-216.

Wolstenholme, D. R. & Fauron, C. M.-R. (1995). Mitochondrial genome organization. In *The Molecular Biology of Plant Mitochondria* (Levings III, C. S. & Vasil, I. K., eds.), pp. 1-59. Kluwer Academic Publ., Dordrecht, The Netherlands.

Yasuhira, S. & Simpson, L. (1997). Phylogenetic affinity of mitochondria of *Euglena gracilis* and kinetoplastids using cytochrome oxidase I and *hsp60*. *J Mol Evol* **44**, 341-7.

Zaug, A. J., Lingner, J. & Cech, T. R. (1996). Method for determining RNA 3' ends and application to human telomerase RNA. *Nucleic Acids Res* **24**, 532-3.

Zhang, Z., Green, B. R. & Cavalier-Smith, T. (1999). Single gene circles in dinoflagellate chloroplast genomes. *Nature* **400**, 155-9.

Zhang, Z., Green, B. R. & Cavalier-Smith, T. (2000). Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: A possible common origin for sporozoan and dinoflagellate plastids. *J Mol Evol* **51**, 26-40.

Ziaie, Z. & Suyama, Y. (1987). The cytochrome oxidase subunit I gene of *Tetrahymena*: a 57 amino acid NH_2-terminal extension and a 108 amino acid insert. *Curr Genet* **12**, 357-68.

Zurita, M., Bieber, D., Ringold, G. & Mansour, T. E. (1988). cDNA cloning and gene characterization of the mitochondrial large subunit (LSU) rRNA from the liver fluke *Fasciola hepatica*. Evidence of heterogeneity in the fluke mitochondrial genome. *Nucleic Acids Res* **16**, 7001-12.