University of North Dakota
UND Scholarly Commons

Theses and Dissertations | Theses, Dissertations, and Senior Projects

5-1999

The DNA Sequence From A Cloned 15 Kilobase Fragment Of The Chlamydomonas Acidophila Mitochondrial Genome And RNA Transcript Production In Response To Cadmium

Scott A. Hoffman

Follow this and additional works at: https://commons.und.edu/theses

Part of the Psychology Commons

Recommended Citation

Hoffman, Scott A., "The DNA Sequence From A Cloned 15 Kilobase Fragment Of The Chlamydomonas Acidophila Mitochondrial Genome And RNA Transcript Production In Response To Cadmium" (1999). *Theses and Dissertations*. 759.
https://commons.und.edu/theses/759

This Dissertation is brought to you for free and open access by the Theses, Dissertations, and Senior Projects at UND Scholarly Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UND Scholarly Commons. For more information, please contact zeineb.yousif@library.und.edu.

THE DNA SEQUENCE FROM A CLONED 15 KILOBASE FRAGMENT OF THE *CHLAMYDOMONAS ACIDOPHILA* MITOCHONDRIAL GENOME AND RNA TRANSCRIPT PRODUCTION IN RESPONSE TO CADMIUM

by

Scott A. Hoffman
Bachelor of Science, Kentucky State University, 1990

A Dissertation

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota
May
1999

T1999
H7112

This Dissertation, submitted by Scott A. Hoffman in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

T. M. Hill
(Chairperson)

Toni S. Spanier

Kevin D. Young

Ann M. Flower

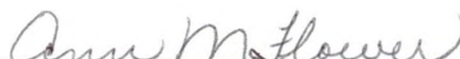
This Dissertation meets the standards for appearance, conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.

Harvey Knull
Dean of Graduate School

6-16-99
Date

This Dissertation, submitted by Scott A. Hoffman in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

(Chairperson)



This Dissertation meets the standards for appearance, conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.

____
Dean of Graduate School

____
Date

PERMISSION

Title — The DNA Sequence From a Cloned 15 Kilobase Fragment of the *Chlamydomonas acidophila* Mitochondrial Genome and RNA Transcript Production in Response to Cadmium

Department — Microbiology and Immunology

Degree — Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Signature Scott A. Hoffman

Date 6/4/99

TABLE OF CONTENTS

LIST OF FIGURES vi

LIST OF TABLES vii

ACKNOWLEDGEMENTS viii

ABSTRACT x

INTRODUCTION 1

- Reductive Genome Evolution 2
- Genome Structure 6
- Translation 9
- Electron Transport Genes 13
- Introns 14
- Mitochondrial Genomes of Alga 20
- Cadmium Tolerance 22
- Heavy Metal Tolerance 23
- Summary Statement 29

MATERIALS AND METHODS 30

- Cadmium Tolerance 31
- Mitochondrial Genome Analysis 45

RESULTS 49

- Mitochondrial DNA Isolation 49

Size and Genomic Organization of *C. acidophila* 51

Protein Coding Genes 69

Codon Usage 76

Introns and Intron Encoded Reading Frames 78

Tranfer RNA Genes 93

Ribosomal RNA 95

Intergenic Regions 98

Cadmium Challenge of *C. acidophila* Strain 122 100

DISCUSSION 110

Mitochondrial Genome 110

Cadmium Tolerance 120

APPENDICES 124

APPENDIX A. Media 125

APPENDIX B. Oligomers 126

APPENDIX C. Amino Acid Single Letter Code 127

APPENDIX D. List of Abbreviations 128

REFERENCES 130

LIST OF FIGURES

Figure Page

1. Ribosomal RNA coding regions from *E. coli*, *P. aurelia* and *C. eugametos* 12

2. Size classification of LambdaGEM-11® *C. acidophila* mtDNA insert 50

3. Linear map of mitochondrial DNA from *C. acidophila* 53

4. Nucleotide sequence of *C. acidophila* F3 clone 55

5. Predicted secondary structure of *C. acidophila cox1i2* 82

6. Conserved open reading frames within the *C. acidophila cobi1* intron 86

7. Conserved open reading frames within the *C. acidophila cox1i1* and *cox1i2* introns 88

8. Conserved open reading frames within the *C. acidophila nad5i1*, *nad5i2* and *nad5i3* introns 91

9. Deduced secondary structure of *C. acidophila's* mtDNA tRNAs 94

10. Sequence elements of *C. acidophila's* intergenic regions 99

11. Derivation of pJb2.5 101

12. RNA induction of *C. acidophila* strain 122 by cadmium 103

13. Restriction map and probe construction from pJb2.5 106

14. Localization of cadmium induced transcript 108

15. Direct comparison of the *C. acidophila* and *C. eugametos* mitochondrial genomes 112

LIST OF TABLES

Table Page

1. Partial digest of pJB1 32

2. Coding regions identified in *C. acidophila* mtDNA 51

3. Comparison of deduced NADH dehydrogenase subunit 2 (Nad 2) protein 70

4. Comparison of deduced apocytochrome b (Cob) protein 70

5. Comparison of deduced cytochrome oxidase subunit 1 (Cox 1) protein 73

6. Comparison of deduced NADH dehydrogenase subunit 1 (Nad 1) protein 73

7. Comparison of deduced NADH dehydrogenase subunit 5 (Nad 5) protein 75

8. *C. acidophila* F3 clone codon usage table 77

9. Introns from other species with identical insertion sites 84

10. Identity of rRNA coding regions between *C. acidophila* and *C. eugametos* 96

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Jonathan Spanier, for his friendship, patience, guidance and understanding during my graduate years. I would also like to thank my committee members Dr. Tom Hill, Dr. Ann Flower, Dr. Kevin Young, and Dr. Barry Milavetz for their guidance during my graduate career.

A special thanks is extended to Dr. Hill who went beyond the call of duty as an acting advisor and for editing this Dissertation. Dr. Hill and Dr. Flower for their friendship, and entrusting their home and dogs to me. Dr. Young who rescued me by challenging me. Dr. James and Sue Kelleher for actually paying me to drive their tractor, home cooked meals and good conversation.

I would like to thank the entire department of Microbiology and Immunology, both old and new members, for all their support and friendship. I especially wish to express my gratitude to those close departmental friends with whom I've shared countless gripe sessions, games of golf, and one too many barbecues. I hope we keep in touch.

I would also like to thank John, Jennifer, Alex and Zach Okemah, as well as Alice Casner. They are those most rare of friends who continue to bridge the passage of time and difference in geography.

Finally, I would like to thank my family. My parents John and Barbara Hoffman, my brother Steve and sister Sarah whose unconditional love and support have always been an inspiration. You are not only my family but also my friends.

…and this time it vanished quite
slowly, beginning with the end of the
tail, and ending with the grin, which
remained some time after the rest of it
had gone.

"Well! I've often seen a cat without a
grin, "thought Alice; "but a grin
without a cat! It's the most curious
thing I ever saw in all my life!"

"Cheshire-Puss,… would you tell me,
please, which way I ought to go from here?"
"That depends a good deal on where
you want to get to", said the Cat.
"I don't much care where"--- said Alice.
"Then it doesn't matter which way
you go," said the Cat.
"---so long as I get *somewhere*,"
Alice added as an explanation.
"Oh, you're sure to do that", said the
Cat, "if you only walk long enough."

-Lewis Carroll
-Alice's Adventures in Wonderland

ABSTRACT

Chlamydomonas acidophila is a unicellular green alga of the order Chlamydomonadales. Our research efforts were allied along two lines: (1) Characterization of the *C. acidophila* mitochondrial genome (mtDNA) and (2) Elucidation of any molecular events responsible for *C. acidophila's* heavy metal tolerance. The mitochondrial genomes of the protists have been underrepresented in the sequence databases. Among the protists, the alga genera *Chlamydomonas* shows a reduced mtDNA content with a highly rearranged gene structure. It was decided to sequence *C. acidophila's* mtDNA to further elucidate the evolutionary paths among the Chlamydomonads and add to the protist sequence database. A 15 kb fragment of *C. acidophila's* mtDNA was cloned and sequenced. The genes identified included apocytochrome b; partial sequences of subunits 2 and 5 and a complete subunit 1 of the NADH dehydrogenase complex; subunit 1 of the cytochrome oxidase complex; discontinuous and scrambled large and small subunit ribosomal rRNA; and four tRNAs whose anticodons specify tryptophan, glutamine, and 2 methionines (one of which appears to be a pseudogene). The mtDNA of *C. acidophila*, therefore, probably encodes a reduced gene coding capacity common among the Chlamydomonadales. In fact the basic gene order is colinear with that of *C. eugametos*. However, *C. acidophila* appears to have two distinctive features: (1) The reduced size of intergenic spacers, and (2) Non-synonymous insertion of a number of group I introns within the partial sequence.

These differences suggest a recent divergence between *C. acidophila* and *C. eugametos*, and place them very close phylogenetically. It was also noticed that *C. acidophila* exhibits a higher tolerance for cadmium than do other *Chlamydomonas* species. Cadmium is a potent environmental toxin and carcinogen that is accumulating in the environment through anthropogenic and natural means. Knowledge of the characteristics of metal tolerant species has yielded valuable insights into the nature of cadmium tolerance, and may one day aid in the safe disposal of this metal. In an attempt to understand the role of mtDNA during cadmium exposure, a 5 kb Hind III fragment of mtDNA was cloned onto a pGem vector (pJB2). That fragment was hybridized to Northern blots of cadmium challenged *C. acidophila* cells, and a transcript of ~300 bp in size was shown to increase during cadmium challenge. Restriction studies and DNA sequencing has revealed that the transcript was produced from a 1500 bp region and appears to be rRNA.

INTRODUCTION

One of the distinguishing characteristics of eukaryotes is the presence of subcellular organelles. Of those organelles, the chloroplast and mitochondria are distinctive in that they contain their own genetic systems. Mitochondria have come under further scrutiny because of the diversity of their genomes. The recent explosion of research in this area has revealed that mitochondrial genomes (mtDNA) vary almost as widely as the eukaryotic cells that house them. Size variability is enormous (~400 fold), ranging from the plant angiosperm *Cucumis* (muskmelon) at 2.4 Mb to the protist *Plasmodium* at 6 kb (Gillham, 1994; Gray, 1992). At first glance there appears to be a dichotomy of evolutionary paths taken by different mtDNAs. On the one hand are the 'animal like' mtDNAs that appear to be geared toward size minimization, both by loss of genes and by the elimination of intragenic spacers. Animals typically possess small circular molecules ranging in size from about 14 to 39 kb (Gray, 1992). By contrast, 'plant like' mitochondrial genomes appear to be evolving in the opposite direction; they are the largest organelle genomes known (up to 2400 kb). Their trend seems to be toward increased size, primarily by the accumulation of a large amount of non-coding DNA whose origin and function remain unknown (Gray *et al.*, 1998). As an aside, their structures appear to be circular and, interestingly, they are known to recombine actively to yield smaller circular molecules carrying only a subset of genes from the 'master' genome (Gillham, 1994).

Why should one be on the path of increasing mtDNA size and the other towards reduction? While this remains an interesting question, a more important aspect of organelle genomes is that of gene content. Mitochondrial genomes may grow larger, encoding

various proteins and RNAs, but in all cases the full complement of genetic material required for their biogenesis is not located in the mitochondria. Indeed, genome size is not indicative of gene content. The largest gene repertoire so far investigated is that of *Reclinomonas americana*, whose mtDNA size is average at around 69 kb (Gray *et al.*, 1998; Lang *et al.*, 1997). All other sequenced mtDNAs contain only subsets of *R. americana's* genes, implying that this protist is closest to the ancestral proto-mitochondrial genome (Lang et al., 1997). Further studies indicate that gene loss has occurred to different extents in different lineages (Palmer, 1997).

Reductive Genome Evolution in Mitochondria

Organelles represent a special class of symbiotic relationship - they have lost their identity as individual organisms. Recent investigations have revealed that the closest known relative to modern mitochondria are the α-proteobacteria, specifically the *Rickettsia* (Andersson & Kurland, 1998; Andersson *et al.*, 1998). Studies with the *Rickettsia*, as well as other obligate intracellular parasites in conjunction with organelles, have yielded valuable insight into the evolutionary mechanisms that influence organelle development. The loss of identity seems to involve two main themes. The first appears to be the loss of gene function either by the accumulation of mutations and/or the loss of DNA repair mechanisms. The other lies in the conservation of symbiotically useful genes through genetic redundancy.

The idea of an endosymbiotic origin of eukaryotic organelles was defended by Margulis (Margulis, 1975; Margulis, 1976) and others some 30 years ago. Since then, much information has been gathered, not only in defense of this hypothesis, but of the mechanisms by which endosymbiosis might proceed. An organism that leaves its free-living lifestyle faces a radically different environment once it adopts an intracellular existence conditioned by the host genome. Adaptation to this environment usually involves

one of two different evolutionary routes (Andersson & Kurland, 1998). The organism may become an obligate parasite to its new environment at the expense of the host. Such relationships are common, with the *Rickettsia*, *Chlamydia*, *Coxiella*, etc. serving as examples (Andersson & Kurland, 1998). Alternately, the host may become dependent on some product produced by the activities of the symbiont's genome. For example, the genus *Buchnera* has become an indispensable symbiont of aphids (Baumann *et al*., 1995). These bacteria reside in specialized cells called bacteriocytes that are maternally inherited from one generation to the next. *Buchnera* appears to supply the host with vital amino acids. If the aphids are treated with antibiotics, the result is either sterility or death. This relationship appears to be mutual in that cultivation of *Buchnera* outside its aphid host has not yet been achieved (Andersson & Kurland, 1998).

Regardless of the path of adaptation, a common theme emerges. Over the course of evolutionary time resident genomes are frequently involved in bottlenecks, manifesting in little opportunity for recombination between variants. Mutations under such conditions accumulate at a higher rate than in free-living organisms. The tendency of small asexual populations to accumulate deleterious mutations is known as Muller's ratchet (Felsenstein, 1974; Muller, 1964). This mechanism becomes an irreversible process within populations experiencing high mutation rates, lack of recombination and/or small population sizes. Accordingly, with the gradual accumulation of deleterious mutations, the most fit class of the population can be lost due to genetic drift. Andersson et al. (Andersson & Kurland, 1998) have raised the question whether such obligate intracellular parasites have been driven to extinction because of the effects of Muller's ratchet. Certainly Muller's ratchet is a powerful concept to invoke in explaining the formation of organelles, but it constitutes only a partial story. If a mutually beneficial relationship is to be established, then there must be a way to conserve those aspects that confer increased fitness before gene extinction while removing those that decrease fitness or are superfluous.

One possible way of conserving useful genes is to locate them in a stable genetic environment. Indeed this seems to be the strategy of eukaryotes. Much of the genetic material necessary for mitochondrial biogenesis and function is located within the nucleus. The process of mitochondrial gene migration to the nucleus has been defined differently by many authors (Brennicke *et al.*, 1993; Gillham, 1994; Thorsness & Weber, 1996). For the process to occur several criteria must be met. First, genes in the mitochondria must be present in multiple copies. This redundancy is necessary to ensure that the loss of some gene copies does not upset the cellular metabolism.

The second step is the escape of genetic material from the mitochondria and transfer to the nucleus. The frequency of this occurrence is much higher than previously thought. In a clever experiment by Thorsness and Fox (Thorsness & Fox, 1990) a plasmid was transformed into a yeast mitochondria. The yeast was a non-reverting uracil auxotroph and the plasmid contained a functional copy of the gene *ura3*, which was capable of repairing uracil synthesis in this yeast. In order for the yeast strain to become a uracil prototroph *ura3* would have to escape from the mitochondria and become localized in the nucleus. The transfer was shown to occur at a frequency of $2X10^{-5}$ Ura^{+} prototrophs per cell generation. Interestingly, the reverse does not appear to be favorable (DNA traveling from the nucleus to the mitochondria), apparently occurring at a rate 100,000 times less.

The last steps for the expression of mitochondrial genes in the nucleus probably represent the greatest barrier (Gillham, 1994). The gene must be capable of being transcribed by the nuclear machinery, translation must be able to occur on cytoplasmic ribosomes, and the product must be targeted back to the mitochondria. All of these events, according to Thorsness and Weber (Thorsness & Weber, 1996), should occur within a relatively short amount of time to prevent inactivation by random mutation.

Gene migration is evidenced from a number of different sources. In some cases of nuclear expression the mitochondrial gene counterpart was found intact, albeit inactive. In

the soybean the *cox II* gene is expressed from the nucleus, while its mitochondrial counterpart is clearly present and seemingly intact (Covello & Gray, 1992). In the flowering plant *Oenothera*, *rsp12* is expressed from the nucleus but about two-thirds of the gene can be found in the mitochondrial genome (Grohmann *et al.*, 1992). So far no example for the expression of two functional genes from both the nucleus and mitochondria exists, but as sequencing efforts progress this possibility may become realized. Taken together, these data imply that equilibrium for gene transfer has not been reached by many organelles, and gene migration is an ongoing process.

Is there a lower limit to the gene capacity of mitochondria? The answer may be contingent on what particular ecological niche an organism occupies (Thorsness & Weber, 1996). The study of trichomonads (largely parasitic and flagellated organisms) has yielded the existence of hydrogenosomes. Hydrogenosomes, like their mitochondrial ancestors, produce ATP by substrate level phosphorylation. However, they lack a genome and are entirely dependent on the nucleus for function and biogenesis (Palmer, 1997). Interestingly, they seem to have arisen from endosymbiotic origins. This argument has been strongly bolstered by finding that the *Trichomonas* nucleus contains one (Germot *et al.*, 1996; Horner *et al.*, 1996; Roger *et al.*, 1996) or all three (Bui *et al.*, 1996) of the mitochondrial heat-shock proteins Hsp10, Hsp60, and Hsp70. These genes have been shown to be the best tracers of the eubacterial origin of mitochondria (Palmer, 1997).

Hydrogenosomes serve as an extreme example of reductive evolution, but most other mitochondria do contain a genome of some sort. Is it to be expected that all mitochondria eventually lose their genome? Further, why are some genes incarcerated within the confines of the mitochondria while others are not? Again, the ultimate barrier for genome transfer to the nucleus may reside within the ecological niche filled by the organism in question (Thorsness & Weber, 1996). The overall picture of these conservation rules is, at this time, still vague. However, some barriers to nuclear localization have become

apparent. One good example is from a study in yeast by Claros and co-workers (Claros *et al.*, 1995). Previous sequence examinations revealed that most mitochondrially encoded proteins were intrinsic membrane proteins with a large number of hydrophobic, membrane spanning domains (Popot & de Vitry, 1990). *Saccharomyces cerevisiae* was engineered to express apocytochrome b (*cob*) in the cytoplasm. Cob is an integral membrane protein normally encoded in the mitochondria. Even with the help of a mitochondrial targeting protein, the Cob protein failed to translocate into the mitochondria. This study supported the hypothesis that large stretches of hydrophobic residues, corresponding to three or four transmembrane domains, can effectively inhibit import into the mitochondria.

Genome Structure

Mitochondrial genomes show a great variety in size and structure. The first completely sequenced mitochondrial genomes belonged to the mammals (human, cow, and mouse) (Gillham, 1994), which exhibited an invariant gene order. Animal mitochondrial genome sizes typically range from 13.8 to 39.3 kbp (Gray, 1992), and are typically circular with the exception of the *Cnidaria* which are linear (Bridge *et al.*, 1992). One of the hallmarks of animal genomes is their parsimony. Genetic information is packaged with a high information density, with coding regions directly adjacent to one another and in some cases overlapping (Gray, 1992). In most cases, over 90% of animal mtDNA have some coding function (Gillham, 1994). Until recently, it had been thought that animal mtDNA was devoid of intervening sequences. However, introns have been found in the mtDNA of a sea anemone (Gillham, 1994). Gene order is relatively constant within individual phyla but varies between phyla. Interestingly, although mammalian gene order is invariant, their mitochondrial genes diverge in sequence at an extremely rapid rate. This rate is 5-10 times the rate at which single-copy nuclear DNA diverges in the same species (Gray, 1992).

Fungal mitochondrial DNAs range more steeply in size from about 17 kb to 176 kb, are typically circular, but show few differences in gene content (the most notable being the complete absence of any *nad* genes) (Gillham, 1994; Gray, 1992; Zimmer *et al.*, 1984). In addition, gene order shows limited conservation between distantly related fungi (Clark-Walker, 1992). Most of the differences can be accounted for by two factors. The first is the expansion/contraction of AT-rich intragenic spacer regions, and the second is the presence or absence of introns. One of the most spectacular examples of this is intronic spacers within the *cox 1* gene of *Podospora anserina*. This gene spans 24.5 kb (1.5 times the size of the entire human mitochondrial genome), makes up about 26% of the 94.2 kb Podospora mtDNA, and contains 16 introns that account for 93% of the *cox I* gene (Cummings *et al.*, 1990).

The largest, most spacious and complex genomes belong to the plants, specifically the angiosperms (size ranges from 200 to 2400 kb) (Gillham, 1994). However, their size does not reflect their gene content. Indeed, much of the plant mtDNA contains introns and large intragenic spacers containing repeated blocks of sequence, some several kilobases long, present in both direct and inverted orientation, and appear to be recombinationally active (Gillham, 1994). Their genomes are typically circular, but they are known to recombine and form smaller genomes that contain only a fraction of the gene content from the 'master genome' (Gillham, 1994). Because of their large size, complete plant genome sequences are few. However sequences of individual genes are plentiful as well as other genetic analysis. Nonetheless our knowledge of plant mitochondrial genomes is lagging in comparison to the animal genomes. Apart from its large size, the most distinctive feature to date of plant mitochondrial genomes is their propensity for rapid structural change, while showing an extremely slow rate of sequence divergence (Palmer & Herbon, 1988). This is the complete opposite from what is observed in mammalian mtDNA. Studies with angiosperms reveals that mtDNA evolves about 10-fold less rapidly in sequence than single

copy nuclear DNA of the same species, and about 4-fold less rapidly than plastid DNA (Palmer & Herbon, 1988). These studies show that the plant mitochondrial genome is the most slowly evolving cellular genome so far characterized (Wolfe *et al.*, 1987).

Until recently, sequences of the mtDNAs of protists has been lacking. These organelles tend to be highly conserved. As of 1998, 63 complete mtDNA sequences were available through public domain databases. However, the sequences represented are both narrow and biased: 47 (75%) are from animal species; five (8%) from fungi; two (3%) from plants; and only nine (14%) from protists (Gray et al., 1998). To get some idea of the variability in gene content and even to catch a glimpse of intermediates in organelle evolution, it has become apparent that the protists and algae must be studied in greater detail. Indeed the recent explosion of protist sequences within sequence databases, in conjunction with formation of organelle genome sequencing consortiums (e.g. the Organelle Genome Megasequencing Program in 1992 (Gray et al., 1998)), has become apparent. Among all the studied mitochondrial DNA groups the protists are by far the most phylogenetically diverse (Gillham, 1994; Gray, 1993; Gray et al., 1998). The emerging data implies that the genome diversity of the protists is enormous and that unique characteristics found in other taxa are reflected within this group. An attempt to impart the various themes of mtDNA in protists would be counterproductive. The interested reader is directed to the excellent recent review by Gray et al. (Gray et al., 1998) on the diversity of this group. Suffice it to say that the study of protist mtDNAs is providing a fresh perspective on mtDNA evolution. Such questions as how the original mitochondrial genome may have been organized are becoming clearer as evidenced by the recent sequence of *Reclinomonas americana* (Lang et al., 1997), the most ancestral mtDNA resembling the eubacterial genome found to date.

Translation

One of the consequences of retaining a genomic complement is the need for translating the encoded products. Various species have approached the problem in a variety of ways. All mitochondrial genomes studied thus far encode both large (LSU) and small (SSU) rRNAs, albeit a minority of protist mtDNAs encode rRNAs that deviate from the typical 23S (LSU) and 16S (SSU) consensus. The trypanosomatid protozoa encode a 9S (SSU) and 12S (LSU) mitochondrial rRNA that are among the smallest and most structurally divergent known. In addition, their potential secondary structures contain only a few of the expected conserved structural motifs (Gutell, 1994; Gutell *et al.*, 1993).

From there, species begin to diverge in their mtDNA resident protein translation machineries. Like animal and fungal mtDNAs, most protist mtDNAs lack a 5S rRNA gene (Gray et al., 1998). The current exceptions to this trend are the chlorophyte algae *Prototheca wickeramii* (Wolff *et al.*, 1994) and *Nephroselmis olivacea* (Gray et al., 1998), the red algae *Chondrus crispus* (Gray *et al.*, 1998), and *Reclinomonas americana* (Lang et al., 1997). This sporadic phylogenetic distribution suggests that this gene has been lost independently a number of times over the course of evolutionary history (Gray et al., 1998).

Likewise, ribosomal protein genes have apparently undergone deletion in various lineages. Animal and fungal mtDNAs, with the exception of *rps3* in *Allomyces macrogynus* (Paquin & Lang, 1996), do not encode ribosomal proteins (Gray *et al.*, 1998). By contrast, plant mtDNAs typically encode their own set of ribosomal proteins (Gray, 1992). The protists show more divergence, with the most notable examples being *R. americana* encoding all ribosomal proteins, *P. wickeramii* encoding about half, and the Chlamydomonads encoding none of the ribosomal proteins (Gray *et al.*, 1998).

The presence of transfer RNAs (tRNA) show a high degree of divergence between taxa. Several protist mtDNAs do seem to contain a minimal tRNA set [*Monosiga*

brevicollis, *P. wickeramii*, etc. (Gray *et al.*, 1998)]. However, tRNAs recognizing one or more codons are absent from other mitochondrial genomes (e.g., *Chlamydomonas* sp.). In these cases, import from nuclear encoded DNA is usually invoked. In fact, import of tRNA into *Tetrahymena* mitochondria, long inferred from tRNA population studies (Suyama, 1986) has been shown experimentally (Rusconi & Cech, 1996). RNA editing must also be mentioned as an alternative possibility to tRNA variability. In this case opossum mitochondria serves as the archetype, in which tRNA-Gly undergoes partial C→U editing to generate tRNA-Asp (Borner *et al.*, 1996).

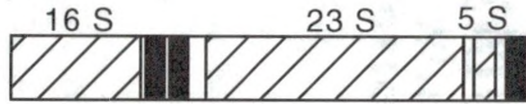
One of the more bizarre twists in the evolution of protein translation machines is that of discontinuous and scrambled rRNA genes. Bacteria and mitochondria typically organize their rRNA into operons structured as 16S-23S-5S. On occasion, tRNA genes are found in the spacer between the 16S and 23S genes (Srivastava & Schlessinger, 1990), and as mentioned, the 5S rRNA has been dropped from many mitochondrial genomes. Nevertheless the general structure is more or less conserved. A few mitochondrial lineages, as well as some bacterial ones, encode discontinuous rRNAs which are split into separate regions on the genome and are interspersed with other genes. Bacterial examples are usually confined to those with small genomes and low rRNA copy numbers such as *R. prowazekii* (Andersson *et al.*, 1995). Mitochondrial examples include *Paramecium aurelia* (Figure 1) (Seilhamer *et al.*, 1984), and *Tetrahymena pyriformis* (Heinonen *et al.*, 1987). The most extreme cases of rRNA rearrangement seem peculiar to certain lineages. *Chlamydomonas* is an example of extreme divergence with discontinuous and scrambled rRNA coding regions (Boer & Gray, 1988b; Denovan-Wright & Lee, 1994; Nedelcu, 1997). Other examples of discontinuous and scrambled rRNAs are portrayed by *Plasmodium* (Feagin *et al.*, 1992) as well as *Euglena* and *Crithidia* (Gray & Schnare, 1990).

A direct comparison of three rRNA coding regions in Figure 1 illustrates the concept of scrambled and discontinuous rRNA coding modules. *E. coli* encodes its rRNA operon as a continuous unit, and is transcribed as such. Post-transcriptional modifications result in functional 23s and 5s large (LSU) and 16s small subunits (SSU). *Paramecium aurelia* is an intermediate example of rRNA that is encoded discontinuously. Note that the rRNA is split into separate coding areas (divided by ~11 kb of sequence), and is interspersed with other genes.

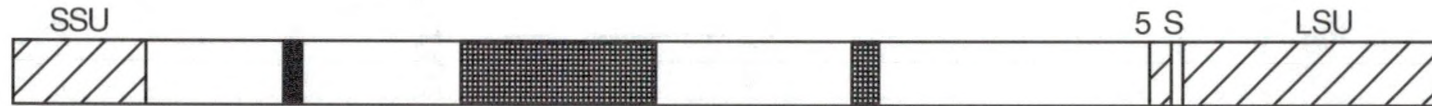
By contrast, *C. eugametos* encodes its rRNA modules discontinuously over a 14 kb stretch of DNA. These modules are interspersed with both protein and tRNA coding regions. Further, both LSUs and SSUs are scrambled. That is they do not obey the typical 5´-3´ architecture of other transcribed rRNAs. The *C. eugametos* rRNA has been labeled to denote how the subunits would piece together should they be continuously encoded and unscrambled. So the large subunit order would be L1-L2-L3-L4-L5-L6. Likewise, the small subunit would be S1-S2-S3. Region three of the *C. eugametos* map (Figure 1) has been enlarged to illustrate the scrambled coding within one of the modules. Note that not only are the L4 and L1 coding regions out of order 5´ to 3´, but the small subunit region S1 also separates them. There are a total of 4 separately encoded modules. The coding rRNAs within each module is as follows, Region 1: L3 and L2; Region 2: L6 and S2; Region 3: L4, S1 and L1; Region 4: L5 and S3.

One of the most distinctive features of gene loss and eventual migration to the nucleus is scrambling (Brennicke et al., 1993). Loss of coding integrity usually indicates the loss of a gene function that was either superfluous within the endosymbiotic environment or that has a backup redundancy. However, molecular modeling studies based on the secondary structure of *E. coli* rRNA have shown that the Chlamydomonads are capable of organizing their rRNA into a functional unit (Denovan-Wright & Lee, 1994). It is thought that the rRNA within *Chlamydomonas* sp. associates through hydrogen bonding associations.

E. coli



P. aurelia



C. eugametos

Figure 1. Ribosomal RNA coding regions from *Escherichia coli*, *Paramecium aurelia*, and *Chlamydomonas eugametos*. Genes coding for rRNAs, tRNAs, and proteins are represented by cross hatches, black and shaded regions respectively. Region 3 of *C. eugametos* is enlarged for illustrative purposes and is not drawn to scale. See text for details.

Indeed, though evidence for intermolecular associations is still lacking, research has provided evidence that these rRNAs do associate with ribosomes in mitochondria (Denovan-Wright & Lee, 1995).

Electron Transport Genes

While the outer membrane of mitochondria is rich in enzyme activity (e.g. phospholipid biosynthesis), only inner membrane proteins have been found encoded on the mitochondrial genome. The electron transport chain includes four multimeric complexes. The composition of each complex has been studied and the sequence of electron transfer has been established. Further, with the increased sequence information provided from a variety of different mitochondrial sources, an evolutionary picture is emerging. The description of the electron transport complexes below are listed from the initial electron entry from either complex I or II, to the electron's eventual transfer to molecular oxygen, resulting in the formation of water.

Complex I (NADH-ubiquinone reductase) catalyzes the oxidation of NADH (nicotinamide adenine dinucleotide) and is the entry point for electrons traveling from NADH into the electron transport chain. It has been described as one of the most complex enzymes ever characterized. For example, beef heart complex I contains around 41 separate polypeptides (Fearnley & Walker, 1992). However, no more than 12 of these have ever been shown to be encoded by any mitochondrial genome (Gray et al., 1998). The presence or absence of coding sequences for the *nad* genes shows no clear delineation along species lines. Interestingly, no *nad* sequences are present within the mtDNA of the yeasts *Schizosaccharomyces pombe* or *Saccharomyces cerevisiae* (Gray, 1992). Among the Chlamydomonads all species thus far characterized code for *nad 1, 2 ,4 ,5* , and *6.*

Complex II (succinate-ubiquinone oxidoreductase) is the second entry point of electrons through succinate from the citric acid cycle, and is the only citric acid cycle enzyme found on the mitochondrial inner membrane. Until recently, complex II had been

missing from characterized mitochondrial genomes. The current representatives containing *sdh 2, 3* and *4* of complex II are *R. americana* (Burger *et al.*, 1996), the red algal rhodophytes *Chondrus crispus* and *Porphyra purpurea* (Burger et al., 1996; Gray et al., 1998).

Complex III (ubiquinol-cytochrome C reductase) contains 8-11 subunits, depending on the species, and is in charge of reducing coenzyme Q by cytochrome C. Interestingly, one of the proteins is universally encoded by all species studied so far (Gray et al., 1998). That protein, apocytochome b, is highly hydrophobic with 8 to 9 membrane spanning domains (Popot & de Vitry, 1990). Conversely, none of the other proteins of this complex has ever been found encoded by any mitochondrial genome.

Complex IV (Cytochrome c oxidase) catalyzes the oxidation of reduced cytochrome c by molecular oxygen. Three of the polypeptides (Cox 1, 2, 3) of this complex are typically mitochondrial in origin and are the most hydrophobic (Gillham, 1994). While *cox 1* is universally encoded in all mitochondrial genome studies thus far, *cox 2* and *3* show some species variability. Most notable is the absence of *cox 2* and *3* in the Chlamydomonads (Denovan-Wright *et al.*, 1998) as well as *Chlorogonium elongatum* (Kroymann & Zetsche, 1998).

Introns

The discovery of introns was a major revolution in molecular biology. Not only had the model for the simple colinearity of the gene been shattered, but this intronic sequence was also capable of catalytic activity, a role that had been assigned only to proteins. Since their discovery, introns have been found in numerous DNA sequences. There are currently four recognized intronic types: (1) nuclear; (2) tRNA; (3) Group I and (4) Group II. Only group I and group II introns have been recognized in organellular DNA and will be the only ones discussed here.

Both group I and group II introns are capable of inserting themselves into intronless alleles as well as transporting to other locations (Lambowitz & Belfort, 1993). However, both groups differ in secondary structure, consensus sequences at their splice sites, and types of reactions involved in splicing. Therefore, group I and II introns can be described in two aspects of their life cycle: intron mobility and catalytic abilities.

Group I Introns

Group I introns have the widest distribution phylogenetically of any class (Lambowitz & Belfort, 1993), and have been found in eubacteria, eukaryotes and possibly archaebacteria (Lykke-Andersen *et al.*, 1997). One of the defining abilities of introns is their mobility. While the details of intronic horizontal transfer are still somewhat unclear (Gray, 1998), there is recent evidence that mass migrations may be more common than previously thought (Cho *et al.*, 1998). The details of intracellular genomic transfer are becoming well characterized. A number of group I introns achieve mobility via an intron-encoded site-specific endonuclease (Dujon, 1989; Gillham, 1994; Lambowitz & Belfort, 1993; Perlman & Butow, 1989). There are a number of intron-encoded site-specific endonucleases, each capable of cleaving an asymmetric target sequence. Four families of endonucleases have been defined for group I introns. The LAGLIDADG consensus sequence, which occurs as repeats, is present in the majority of endonucleases (Belfort & Perlman, 1995). The motif is phylogenetically widespread, occurring in all three kingdoms, and is present in archaeal introns, as well as in all four known group I intron-encoded maturases (Belfort & Perlman, 1995). The second group is the GIY-YIG proteins. The two components of the GIY-YIG motif are separated by 10-11 amino acids and occur upstream of a conserved sequence block of about 15 amino acids (Belfort & Perlman, 1995). The third motif, the H-N-H proteins, has the interesting property of being implicated in both group I and II introns (Lambowitz & Belfort, 1993). The H-N-H

proteins display a consensus sequence spanning 30-33 amino acids with four highly conserved histidine residues. This region is contained within 50-80 amino acids of less well-conserved residues (Shub *et al.*, 1994). The H-N-H motif also appears in a Zn^{2+} finger-like domain of group II introns. This reflects the coincidence of endonuclease functions in both group I and group II mobility pathways (Belfort & Perlman, 1995). Lastly, the Hys-Cys box motif contains 3 cysteine and 2 histidine residues in a conserved region about 30 amino acids long. The Hys-Cys box is likely to be a metal coordination site within the DNA binding domain (Johansen *et al.*, 1993).

These site-specific nucleases are capable of mobilizing their introns in two ways. The first is homing, a process by which an intron in one gene introduces itself into the same site in a homologous gene. Homing has been confirmed in a number of studies (Clyman & Belfort, 1992; Szostak *et al.*, 1983) and the mechanism has been extensively reviewed (Lambowitz & Belfort, 1993; Belfort & Perlman, 1995). Briefly, this mechanism occurs by creating a double stranded break in the intronless allele. The resulting DNA ends invade the intron containing allele to induce replicative transfer by a double-stranded break-repair process (Saldanha *et al.*, 1993; Belfort & Perlman, 1995). The second mode of mobility is transposition, a process which occurs by the addition of an intron to a specific site in a nonhomologus gene lacking the intron, or into a different site in the homologous gene. Group I intron transposition has not yet been observed in its entirety, but partial reactions have been observed both *in vitro* and *in vivo* (Mohr & Lambowitz, 1991; Roman & Woodson, 1995; Thompson & Herrin, 1994).

The secondary structure of group I introns was determined (Davies *et al.*, 1982; Michel *et al.*, 1982) based on comparative sequence analysis. However, the former assumption that the intronic RNA first folds into a secondary structure and then settles into tertiary interactions from the unpaired bases has recently been challenged. In NMR studies with a segment of the *Tetrahymena thermophila* group I intron it was found that upon

folding into the tertiary structure there was a rearrangement in secondary structure (Wu & Tinoco, 1998). This study indicates the need for investigation of intronic (indeed RNA in general) secondary and tertiary folding patterns. However, for this introduction, the accepted conventions will be observed in defining intronic secondary structure.

Group I introns have a characteristic conserved secondary structure of paired regions denoted P1 – P10, with elements of P3-P8 forming the intron's catalytic core (Lambowitz & Belfort, 1993). These paired regions are designated P, Q, R and S. At the 5′ and 3′ ends of the intron are one of the unusual features of group I introns, particularly in the context of mobility. These intronic ends, called the internal guide sequence (IGS), pair with the flanking exon sequences at the 5' and 3' ends to form helices P1 and P10 respectively (Lambowitz & Belfort, 1993). While the core secondary structure is conserved, DNA sequences can display extreme divergence (Lambowitz & Belfort, 1993). This sequence divergence, especially among those evolutionary distant from the current representatives in the DNA sequence databases, makes intronic group determinations problematic (Lisacek *et al.*, 1994). Nevertheless, the same secondary structure has been confirmed for numerous sequences.

In order to exist within the confines of a functional gene, introns must be capable of self-excision to impart no deleterious effects upon the gene transcript. Such interruptions would introduce sequences that would severely impair or destroy the function of the gene product. The catalytic nature of group I introns were first reported by Cech (Cech *et al.*, 1981) before it was even known to be a group I intron. The process of intronic excision has been extensively reviewed (Gillham, 1994; Lambowitz & Belfort, 1993). Briefly, group I introns operate via two guanosine-initiated transesterfication reactions. The first reaction is initiated by a guanosine residue, such as guanosine monophosphate, as the attacking nucleophile. This reaction releases the 5′ exon, leaving a free 3′-hydroxyl typically of a conserved uracil residue. The 5′ exon attacks the phosphorus atom at the 3′

splice site. Ligation of the exons and excision of the intron (which may either remain linear or circularize) is then completed.

Group II Introns

Group II introns have a wide distribution, have been found in fungal mitochondria (Michel & Ferat, 1995), are predominate in plant mitochondria and chloroplasts (Gillham, 1994; Oda *et al.*, 1992), and are abundant in *Euglena gracilis* chloroplasts (Hallick *et al.*, 1993). Group II introns, like their group I counterparts, initiate the removal of themselves from exonic RNA. However, group II introns differ in secondary structure as well as the mechanism for splicing. Like their group I counterparts the secondary structure of group II introns was determined by comparative sequence analysis (Michel et al., 1982; Schmelzer *et al.*, 1983). Several excellent reviews have been presented for the structure and catalysis of group II introns (Gillham, 1994; Lambowitz & Belfort, 1993; Michel & Ferat, 1995) from which the following description has been drawn. Structurally, group II introns are characterized as 6 helical domains radiating from a central wheel. This structure is essentially conserved, but there is little sequence similarity between group II introns. Among the most conserved sequences are the 5´ intron boundary sequence GUGYG (where Y is a pyrimidine) and the 3´ boundary AY. But even these sequences have their exceptions (Michel & Ferat, 1995). Group II introns are capable of the same forms of mobility as group I introns (i.e., homing, transposition and deletion), but achieve mobility in different ways. Group II mobility has been associated with reverse transcriptase -like proteins, maturases and endonucleases with a Zn^{2+} domain.

Catalytically, group II introns operate by a pair of transesterfication reactions. The initiating nucleophile is a 2´-hydroxyl of an adenosine residue in domain VI. This nucleophilic attack centers on the 5´ splice junction and releases the 5´ exon with a free hydroxyl end. A characteristic group II lariat is formed by the attachment of the 5´ intron end to the above mentioned adenosine residue responsible for the initial excision reaction.

The free 5´ exon then attacks the 3´ splice junction to produce the spliced exons plus the excised lariat structure.

The distribution of group I and group II introns has been identified in a number of different algal genera. Both group I and group II introns have been found in *Scenedesmus obliquus* (Kuck *et al.*, 1990). Only group I introns have been found in *Chlamydomonas eugametos* (Denovan-Wright & Lee, 1994; Denovan-Wright et al., 1998), *Chlamydomonas moewusii* (Turmel *et al.*, 1993), *Chlamydomonas smithii* (Colleaux *et al.*, 1990), *Chlorogonium elongatum* (Kroymann & Zetsche, 1998), and *Prototheca wickerhamii* (Wolff *et al.*, 1994). *Chlamydomonas reinhardtii* is postulated to contain a degenerate group II intron (Nedelcu & Lee, 1998) but is otherwise intronless (Michaelis *et al.*, 1990). The current data is sketchy at best but a few generalizations have been forwarded. Positionally equivalent and structurally homologous *cox1* introns have been found in *Monosiga brevicollis*, *Marchantia polymorpha* and *P. wickeramii*, suggesting a vertical inheritance from a mitochondrial ancestor of fungi, green algae and plants (Wolff *et al.*, 1993). It is important to point out that while these introns are structurally conserved, their sequence is not. One of the characteristics of vertical transmission is thought to be an extreme sequence divergence over evolutionary time from the common ancestor. For this reason authors tend not to speculate much about the evolutionary origins of sequences that are highly divergent from those in the database. In contrast, horizontal transmission (especially recent transmission events) is thought to show a greater degree of similarity (implying homology) between the conserved core and intronic open reading frames. For example, *Peperomia* is distantly related to the angiosperm *Veronica.* Yet they both contain an intron of 92% identity inserted in the same position of the *cox 1* gene (Cho *et al.*, 1998). After the investigation of 335 diverse genera and inquiries into the sequence databases, Cho and colleagues (Cho *et al.*, 1998) found that this intron showed a patchy phylogenic distribution. Their conclusion was that this intron had horizontally transferred over 1,000

times during angiosperm evolution. Further, this massive wave of intron transfer is a recent occurrence. In summary, the origin and transfer of introns remains a hotly debated subject, and will only resolve with research efforts.

Mitochondrial Genomes of Alga

The relationship of the mitochondrial genomes in chlamydomondales to other algal taxa and embryophytes (land plants) raises certain fundamental evolutionary questions. Mitochondrial rRNA sequences of the green alga *Chlamydomonas reinhardtii* do not branch with the higher plants although nuclear and chloroplast rRNA phylogenies place *Chlamydomonas* and higher plants in the same branch (Cedergren *et al.*, 1988). In fact, the question of whether mitochondria were of monophyletic or polyphyletic origin was for a time considered (Cavalier-Smith, 1992; Gray, 1988). However the recent sequencing of *Prototheca wickeramii* (Wolff et al., 1994) and other protists has demonstrated that plant mtDNA has an ancestral pattern that has been lost in the rapidly evolving and highly derived *Chlamydomonas* sp. (Gray et al., 1998). The division has been described by Denovan-Wright and co-workers (Denovan-Wright et al., 1998) as either 'animal like' in the chlamydomonadalean taxa or 'plant like' in *Prototheca wickeramii* and *Platymonas subcordiformis*. This borrowed analogy is used to reflect the differences in genome structure and organization. 'Plant like' algal taxa contain large mitochondrial genomes. By contrast, the 'animal like' division has a small mitochondrial genome size. However, this analogy fails to recognize the gene content divergence. The *Homo sapiens* mitochondrial mtDNA, for example, is a study in economy. Nearly all of its 16 kb of DNA codes for some gene (Gray, 1992), while *Chlamydomonas reinhardtii* encodes one of the lowest content of genes known (Gillham, 1994). Detailed genome analysis for a few chlorophytes including *Chlamydomonas reinhardtii* (Gray & Boer, 1988), *Chlamydomonas eugametos* (Denovan-Wright et al., 1998), *Prototheca wickerhamii* (Wolff et al., 1994), and

Chlorogonium elongatum (Kroymann & Zetsche, 1998) has been completed and their sequences deposited in GenBank.

The genetic coding is strikingly similar among *C. elongatum* and the two *Chlamydomonas* species. All three encode seven subunits of the mitochondrial respiratory chain (*nad 1, 2, 4, 5, 6, cob, cox 1*), three tRNAs ($tRNA^{met}$, $tRNA^{trp}$, $tRNA^{gln}$), and the large (LSU) subunit and small (SSU) ribosomal RNAs. The latter are discontinuous and scrambled, interspersed with tRNA and protein coding regions.

The differences among the three taxa are equally as interesting. None of the three are colinear with respect to one another, implicating different evolutionary routes after splitting from some common ancestor. While *C. eugametos* and *C. elongatum* are circular and roughly equal in size (22.9 kb and 22.7 kb respectively), *C. reinhardtii* has a linear (Michaelis et al., 1990; Ryan *et al.*, 1978) genome of about 15.8 kb. Further, this genome is flanked by characteristic inverted repeats with 3' single-stranded, noncomplementary extensions of 39 to 41 nucleotides (Ma *et al.*, 1992; Vahrenholz *et al.*, 1993). Both *C. eugametos* and *C. elongatum* encode all their genes on a single strand of the mtDNA genome whereas *C. reinhardtii* transcribes bi-directionally. Additionally, *C. reinhardtii* encodes an *rtl* gene coding for a reverse transcriptase-like protein (Boer & Gray, 1988a). This gene is not present in either *C. eugametos* or *C. elongatum.* Further, only *C. eugametos* encodes an additional $tRNA^{met}$ that is thought to be a pseudogene. No introns have been found in *C. reinhardtii* mtDNA, however one group I intron has been detected in the closely related, colinear but infertile strain *C. smithii*. By contrast, *C. eugametos*, *C. elongatum* and *P. wickeramii* have been shown to contain 9, 6, and 5 introns respectively.

The study of the evolution of various mitochondrial genome lineages is of increasing interest. In particular, the study of the highly derived Chlamydomonads has come under scrutiny because of their unique features. This laboratory chose to investigate the mitochondrial genome of *C. acidophila* in an attempt to further elucidate the various

evolutionary pathways these alga have undertaken. To that end, a partial segment of *C. acidophila's* mitochondrial DNA (~15 kb) was cloned and sequenced. The sequence was then subjected to various computer analyses. It is hoped that the knowledge of this sequence will contribute to the understanding of the evolution of mitochondrial genomes within the Chlamydomonads.

Cadmium Tolerance

Many metals, such as copper, zinc and magnesium, are essential for biological processes. They participate in a number of different cellular homeostatic functions such as transcription, electron transport, and osmotic regulation. While any metal can be toxic above certain threshold values, some metals such as cadmium, mercury, and lead serve no known biologic function and can be cytotoxic even at low concentrations. Further, unlike other environmental pollutants such as pesticides, heavy metals tend to persist for long periods of time. One of the metals, cadmium, is a potent environmental pollutant and carcinogen and has come under recent scrutiny (Waalkes *et al.*, 1992). It has a low crustal abundance but is present in all living organisms, albeit at concentrations that are extremely low and do not seem to hinder biologic function (Webb, 1979). Modern industrial operations have redistributed and concentrated cadmium into ecosystems not competent for high levels of this metal. The amount of cadmium present in soils is variable depending on location, but seems to be on the increase (Grant *et al.*, 1998; Jones *et al.*, 1992). Previous investigations have revealed that organisms exposed to cadmium contaminated sites contain high levels of this toxic metal (Grant et al., 1998; Webb, 1979). Some plants can accumulate relatively high levels of cadmium without adverse effects on growth (Bingham, 1979; Kuboi *et al.*, 1986). Strikingly, although plants do not require cadmium for growth or reproduction, the accumulation index of cadmium in many green plants exceeds that of all other trace elements (Kabata-Pendias & Pendias, 1992). The ability to accumulate

cadmium in food plants and its subsequent mobility through the food chain has put humans and other biota at a tangible risk.

The sites for cadmium damage are long and varied. Cataloging whether an effect is a primary cause for cell toxicity or just another in a long list of secondary consequences remains a daunting chore. It seems that cadmium is responsible for a multi-system breakdown of the cell. In mammalian cells, for example, cadmium at cytotoxic concentrations inhibits the biosynthesis of DNA, RNA, protein and induces lipid peroxidation (Beyersmann & Hechtenberg, 1997). At non-cytotoxic concentrations cadmium is weakly mutagenic, probably owing to the disruption of DNA repair enzymes (Beyersmann & Hechtenberg, 1997), and not due to direct DNA chemical damage. Further, at concentrations as low as 1 μM, cadmium has been shown to activate the proto-oncogenes *c-jun*, *c-fos*, and *c-myc* (Beyersmann & Hechtenberg, 1997). This metal/proto-oncogene interplay has been proposed to be the cause of carcinogenesis associated with cadmium (Beyersmann & Hechtenberg, 1997). In plants, photosynthetic inhibition is a common theme (Straton & Corke, 1979). Recently, Nagel and colleagues (Nagel *et al.*, 1996) challenged *C. reinhardtii* with Cd^{109} and found that the metal localizes primarily in the chloroplast with a specific inhibition of photosystem II (Voigt *et al.*, 1998). However, cadmium has also been shown to disrupt oxidative phosphorylation through inhibition of substrate oxidation and increasing proton leak in the mitochondria of potato tuber (Kesseler & Brand, 1994).

Heavy Metal Tolerance

A metal toxic environment, either natural or anthropogenic provides an interesting problem for organisms. On the one hand an established, metal naïve population suffering metal insult through pollution or other means will undoubtedly be forced through a strong selection sieve from which tolerant populations frequently arise (Macnair, 1993). On the

other hand, a metal competent population is forced to contend with a substance that is both unusable and toxic to the cellular environment. However, it must be stressed that metal contaminated environments may be viewed as an exploitable niche. An organism capable of survival within a metal toxic environment may have a selective advantage. Some organisms may have evolved to use these environments efficiently, to the exclusion of competing species. Indeed the conventional paradigms for certain metal metabolites may blur with increased study of biologically diverse niches. For example, cadmium is thought to be a universally unusable metal ion by biotic life. However, one study suggests that cadmium is able to replace zinc in marine phytoplankton (Lee & Morel, 1995). Still, as a general rule, certain metals (e.g. cadmium, mercury, arsenic, etc.) are toxic to most forms of life. In order to grow and reproduce in metal toxic environments organisms must be capable of negating the effects of metal toxicity.

Biotic life may employ a number of different defenses against metal toxicity. Perusal of the literature can be confusing especially in the categorization of metal defenses. Terms such as 'resistance' and 'tolerance' are used seemingly according to the author's preference. These definitions are usually defensible within the context of the author's field. However, crossing over to a different discipline may yield slightly different definitions or on different emphasis within the definitions. For example, Gadd (Gadd, 1992) clearly implicates metallothioneins and phytochelatins in metal 'resistance'. By contrast, Macnair (Macnair, 1993) defines these as 'tolerance' mechanisms (indeed refusing to use the term 'resistance' altogether), indicating the difficulty of agreement between fields. There are further definition lines among organisms and molecular mechanisms. Prudence dictates accepting those definitions that seem to have a consensus within a discipline.

The prokaryotes are replete with mechanisms capable of the specific elimination of toxic metals. Gadd (Gadd, 1992) defines 'resistance' as a microorganism's ability to survive toxic effects of metal exposure via the direct response of some detoxification

mechanism against the metal insult. It is important to note that Gadd never commits his terminology to a specific response mechanism for a specific metal or class of metals, though such mechanisms do seem to exist. Rather, this definition implies an 'active' response mounted against either the offending metal or the effects of that metal. Such mechanisms further imply an evolutionary history with metals. Indeed, Silver (Silver, 1996) suggests that the early earth was metal polluted and it was then, shortly after the appearance of prokaryotes, that metal resistance arose.

Of the bacterial resistance mechanisms, Silver (Silver, 1996) makes three generalizations: (1) Heavy metal resistance is very specific, there is no general mechanism capable of producing resistance to all heavy metal ions. In this sense metal resistance is analogous to plasmid-mediated antibiotic resistance and sugar or amino acid metabolism. (2) Metal-ion resistance has been found on plasmids of every bacterial group studied so far. Further investigations have revealed homologous genes on certain bacterial chromosomes. For example, the genome sequence of *Haemophilus influenza* (Fleischmann *et al.*, 1995) includes genes predicted for arsenite reductase and mercury transport similar to those previously sequenced on bacterial plasmids (Silver & Keach, 1982; Summers & Silver, 1972). (3) The mechanisms of resistance are generally efflux pumping and enzymatic detoxification converting a more toxic substance to a less toxic and less available metal. A clear and well-reviewed example of modification would be the bacterial reduction of Hg^{2+} to Hg^0 through a plasmid reductase gene (Silver & Misra, 1988; Summers, 1986). Transport examples include the *CadA* ATPase of gram positive bacteria, and the bacterial *Czc* antiporter system (Silver & Walderhaug, 1992).

These specific mechanisms are by no means the only defenses within a microorganism's repertoire. In contrast to 'resistance', 'tolerance' has been defined as the ability of a microorganism to survive metal insult by the means of some intrinsic mechanism (Gadd, 1992). Intrinsic properties can be seen as some part of an organism's

life cycle, not intended for specific metal resistance, that is nonetheless capable of modifying the toxicity of metals. Examples include impermeable cell walls, extracellular polysaccharide, and metabolite excretion.

Eukaryotes seem to have adopted a different strategy. In contrast to the bacterial theme of efflux and toxicity modification, eukaryotes typically mount a defense based on metal binding peptides, in effect inactivating harmful metal species through complex formation. The major metal complexing enzymes, which may play a dual role in metal homeostasis and detoxification, are metallothioneins and phytochelatins. Metallothioneins comprise a family of low molecular weight cysteine-rich, ribosomally translated proteins, that may have a multifunctional role consisting of metal ion homeostasis, metal ion detoxification, and detoxification of oxygen free radical species. Metallothioneins contain a number of Cys-Xaa-Cys stretches thought to bind metal ions (where Xaa is any amino acid other than cysteine). There is much evidence for the protective role of metallothioneins against metal ions. Rat hepatocytes exposed to cadmium, zinc, arsenic, mercury, and nickel were shown to induce the production of metallothioneins many fold over the controls (Bauman *et al.*, 1993). Resistance to cadmium toxicity in mammalian cell lines has been correlated with gene amplification (Crawford *et al.*, 1985). Sensitivity to cadmium toxicity has been associated with metallothionein gene deletions in transgenic mice (Masters *et al.*, 1994; Michalska & Choo, 1993). However, the role of metallothionein-like proteins in plants remains controversial. Many of these metallothionein-like genes have been found in plants (Robinson *et al.*, 1993). Zhou and Goldsbrough were even capable of restoring the cadmium and copper tolerance of yeast by complementation with genes coding for metallothionein from *Arabidopsis thaliana* (Zhou & Goldsbrough, 1994). But researchers such as Zenk have contended that there is no experimental evidence that these plant metallothioneins are involved in the detoxification of heavy metals in higher plants (Zenk, 1996).

Phytochelatins (class III metallothioneins, PC) are sulfur-rich peptides that are produced in plants, algae, and fungi. In contrast to Class I and II metallothioneins, phytochelatins are constructed not on ribosomes, but as a product of enzymatic reactions via a γ-peptide linkage, rather than the α-amino and carboxyl linkage seen in polypeptides. Their structure consists of a variable number (2-7) of glu-cys dimers and a carboxyterminal glycine. There are at least three distinct enzymes in the formation of phytochelatins (Zenk, 1996). The enzymes responsible for catalyzing the initial steps are γ-glutamylcysteine synthetase and glutathione synthetase, the product of which is glutathione. Glutathione forms a useable pool for phytochelatin synthesis via phytochelatin synthase (Zenk, 1996).

Early investigations showed that phytochelatins might be protective against metals. Pulse chase experiments, where the cellular glutathione pool was tagged with ^{35}S, show a loss of radiolabel from glutathione and a concomitant increase in radiolabeled phytochelatin when the organism was treated with cadmium (Robinson et al., 1993). While phytochelatin synthesis in response to metals has been shown in numerous organisms, their precise role in metal detoxification is still controversial. In an *in vitro* study by Kneer (Kneer & Zenk, 1992) plant enzymes tolerate 10 to 1000-fold the amount of cadmium in the presence of phytochelatins as compared to the free metal ion. In *C. reinhardtii* phytochelatins were shown to be protective against cadmium toxicity (Howe & Merchant, 1992). The same study speculates that phytochelatin synthesis is an evolutionary adaptation because of the inability of glutathione to effectively bind cadmium ions with the stability required for cadmium detoxification (Howe & Merchant, 1992). In support of this idea, the affinity for cadmium increases with increasing phytochelatin peptide length. One of the most convincing results of the role of phytochelatins in metal detoxification is with a cadmium sensitive mutant of *Arabidopsis thaliana* (Howden *et al.*, 1995). This mutant was sensitive to cadmium and deficient in its ability to form Cd-PC complexes while glutathione synthesis proceeded at the same rate as the wild type. An enzyme assay demonstrated that

phytochelatin synthase was the defective enzyme. Interestingly, the role of phytochelatins is somewhat murky in other organisms. For example, cadmium sensitive clones of *Silene vulgaris* produced more phytochelatin upon exposure to cadmium than cadmium tolerant lines (De Knecht *et al.*, 1994). This result is confirmed with similar results on zinc (Harmens *et al.*, 1993).

Previous investigators have studied the genetic and physiologic basis of metal tolerance in the genera *Chlamydomonas*. For example, Collard and Matagne (Collard & Matagne, 1990) selected clones of *C. reinhardtii* tolerant to Cd. In addition to confirming the role of the cell wall in cadmium tolerance, they found two independent major genes, each of which was capable of imparting tolerance on its own and that acted additively. Previous investigations by this group provided evidence that these mutations were cross protective against copper and zinc but not nickel and cobalt (Collard & Matagne, 1990). Nagel and Voigt isolated cadmium tolerant clones after 9 months of selection in cadmium infused medium (Nagel & Voigt, 1989). Later investigations by this group showed that the tolerance mutation occurred within the photosynthetic metabolic pathway (Nagel et al., 1996; Voigt et al., 1998). These results indicated that adaptive mutations could provide additional protection against cadmium toxicity by altering photosystem II. Two independent laboratories isolated cadmium sensitive mutants via transformation by insertional mutagenesis (McHugh & Spanier, 1994; Tang *et al.*, 1995). However, localization of these insertions has yet to be determined.

In an attempt to add to this data and elucidate mechanisms of metal ion tolerance, this laboratory investigated heavy metal tolerance in *C. acidophila*. Previous investigations had shown that this alga was capable of withstanding algastatic concentrations of copper 20-125 times greater than those of the laboratory strains of *C. reinhardtii*. Further studies revealed that *C. acidophila* amplified a ~20 kb segment of DNA in response to cadmium exposure (Spanier, unpublished data). In an attempt to understand the role of this DNA

during cadmium challenge we cloned a segment of the DNA and probed northern blots to obtain a profile of transcript production.

Summary Statement

The research reported in this dissertation reflects two major projects. Previous investigations revealed a high degree of copper tolerance in the chlorophyte *Chlamydomonas acidophila* (Twiss, 1990). By inference, since many mechanisms of eukaryotes typically show cross protection to other metals, it was decided to test *C. acidophila* for cadmium tolerance. The studies in this laboratory focused on potential mechanisms for that tolerance. *C. acidophila* was shown to amplify DNA in response to cadmium challenge (Spanier, unpublished results). In order to ascertain transcript production from this DNA, total cellular RNA from cadmium exposed *C. acidophila* was isolated and investigated. This process involved the cloning of the amplified DNA and using it as a probe against northern blots to obtain a profile of transcripts being produced from this region. During this process one of the cloned segments of DNA was sequenced. Investigations of this sequence revealed some unique features. It was determined that this DNA may be the mitochondrial genome. In an effort to further characterize this DNA, an attempt was made to clone the entire molecule. The cloned DNA was sequenced and analyzed by various computer-aided procedures.

MATERIALS AND METHODS

Strains

Chlamydomonas acidophila strain #122 was supplied by Dr. Judy Acreman, University of Toronto Culture Collection, Toronto, Canada. Chlamydomonas cultures were maintained at 22°C under continuous light. For sub-cloning procedures *E. coli* SURE® cells (Stratagene) and XL1-Blue (Stratagene) were used. For genomic cloning procedures *E. coli* strain LE392 (Promega) was used.

Plasmids and Cloning Vectors

Small DNA fragments were cloned onto pGEM®-3Zf+ vectors (Promega). This plasmid contains a gene for ampicillin resistance and also a multi-cloning site within the *lacZ* α-peptide gene that allows for blue/white screening of colonies. For the cloning of large genomic fragments LambdaGEM®-11 (Promega) was used. LambdaGEM®-11 is a derivative of EMBL3 that contains a multi-cloning site and is capable of holding large DNA fragments of 9kb to 23kb in size.

Growth Media

Minimal medium I (Sager & Granick, 1954) was used for all *Chlamydomonas* stock cultures. Two forms were used: (1) liquid and (2) solid media, in which washed

(Spanier *et al.*, 1992) Gibco Select Agar was used at a concentration of 1.5%. *Escherichia coli* strains were typically grown on Luria-Bertani (LB) (Maniatis *et al.*, 1982) media (broth or agar plates) with or without ampicillin (50 μg/ml).

Cadmium Tolerance Studies

Subcloning and Partial Digests

Two procedures were employed to clone DNA fragments into pGEM®. The first involved ligation of foreign DNA into pGEM®. Insert DNA was digested and fragments were separated by electrophoresis. The DNA was then quantitated by visualization of an ethidium bromide (EtBr) stained gel, containing standards of known amount. Based on this data the amounts of insert and vector DNA were estimated. Then both the insert and vector (cleaved with the appropriate enzyme) were mixed at a ratio of 1:3 of vector to insert ends, respectively. The two DNAs were then placed in 10 μl 1X T4 DNA ligase buffer (supplied by the vendor). After mixing, 1 μl was removed (pre-ligation mix) and 1 μl T4 DNA ligase (New England Biolabs, Inc.) was added. The mixture was then incubated overnight at 15-22°C. After ligation, 1 μl was removed (post ligation mixture) and both pre- and post-ligation mixtures were inspected on an agarose gel for confirmation of ligation.

Plasmid pJB1 was isolated by inserting *Hin*dIII fragments of mtDNA (Spanier, unpublished results) and was initially used to probe *C. acidophila* total cellular RNA. After restriction analysis it was found that pJB1 contained three *Hin*dIII fragments.

Instead of ligating individual Hind III fragments from pJB1 into new vectors, pJB1 was partially digested in an attempt to remove two of the fragments while retaining a third. Said another way, it was hoped that the partial digestion of pJB1 with *Hin*dIII would result in the enzyme cutting each plasmid an average of once or twice per unit time depending on the enzyme concentration. An effective enzyme concentration would cause two of the

fragments to 'drop out' of the plasmid, after which the DNA could be religated and transformed into *E. coli*. The components of the digestion can be seen in Table 1. Following digestion, an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added to the reaction and the sample was mixed by hand for 30 seconds. After mixing, samples were centrifuged for 10 min at 14,000 rpm in an Eppendorf 5414 microfuge. The upper aqueous phase was removed and placed in a fresh 1.5 ml microfuge tube. The aqueous DNA was precipitated by adding 2 volumes of ethanol and 1/10th volume 3 M sodium acetate to the aqueous phase. The nucleic acids were allowed to precipitate for 1 hour at –70°C. These samples were again centrifuged at 14,000 rpm for 30 minutes at 4°C. After centrifugation a white pellet could be visualized at the bottom of the tube. The aqueous phase was aspirated (taking care not to disturb the pellet), and the pellet was washed 3 times with 70% ethanol. After each wash the tubes were centrifuged at 14,000 rpm for 5 minutes. After the final spin, the pellets were air dried until the DNA became translucent, then resuspended in 10 µl sterile H_2O. A 1 µl aliquot of linear DNA was removed from each tube and ligated with T4 DNA ligase (New England Biolabs, Inc.). In addition, the concentration of the DNA was kept intentionally low to prevent the formation of dimers. After ligation, resulting plasmids were inspected by gel electrophoresis and transformed into *E. coli* SURE® cells. White colonies were selected and grown for plasmid miniprep analysis.

Table 1. Partial digest of pJB1.

	Tube 1	Tube 2	Tube 3	Tube 4
Hind III	1u/µl	0.2u/µl	0.04u/µl	0.01u/µl
pJB1 (305ng/µl)	10 µl	10 µl	10 µl	10 µl
Total volume	100 µl	100 µl	100 µl	100 µl

-Digests were conducted with 10X buffer supplied by New England Biolabs, Inc.
-Solution brought to a total volume of 100 µl with H_2O in a 1.5 ml microfuge tube.
-Hind III enzyme supplied from New England Biolabs, Inc. in aliquots of 20,000 u/µl.

DNA Fragment Isolation

On occasion it was necessary to isolate DNA fragments directly from agarose gels. Generally, the Geneclean® kit (BIO 101, Inc.) was used to isolate those fragments. Geneclean® employs 'glassmilk®', a silica matrix that binds DNA above a certain salt concentration. Size fractionated, EtBr stained DNA was visualized with UV light and desired bands were cut from agarose gels with a sterile razor blade. Gel slices were weighed, and gel slices totaling no more than 0.25 g (0.1 g equals approximately 100 µl) were placed into separate 1.5 ml microfuge tubes. To each tube 4.5 volumes of 6 M NaI was added followed by 1/2 volume of TBE modifier. These tubes were placed in a water bath between 45-55°C for 5 min. Usually 5 µl of 'glassmilk' was added (5 µl of glassmilk can bind ~5 µg of DNA; for each additional 0.5 µg of DNA, 1.0 µl of 'glassmilk' was added) and mixed into solution. The tubes were placed on ice for 5 minutes and mixed by inversion every 1-2 minutes during that time. Afterwards tubes were centrifuged and 'glassmilk' pelleted at 14,000 x g for 5 minutes. The NaI supernatant was removed and the pellets were resuspended in 500 µl NEW Wash®. The tubes were again centrifuged at 14,000 x g for 5 seconds. NEW Wash™ was removed and the washing procedure was repeated twice. After the final wash and aspiration of the supernatant, the tubes were again centrifuged to remove excess NEW Wash™. Finally, DNA was eluted by resuspending the pellet in sterile H2O in a volume equal to the volume of 'glassmilk'. The tubes were incubated at 55°C for 3 min and spun at 14,000xg for 20 sec. The DNA-containing supernatant was removed to a fresh microfuge tube. This process could be repeated and typical DNA yields were 80% recovery after the first elution and an additional 10-20% after the second.

Transformation

To produce competent cells, *E. coli* SURE® was inoculated into 5 ml LB broth (no ampicillin) and agitated overnight at 37°C. The following morning, 1 ml of the culture was removed and inoculated into 100 ml LB and grown to early log phase (3-4 hrs; 30-60 min after turbidity is detectable). The cells were centrifuged and the pellet was resuspended in 5 mls TSS and stored on ice. The cells could either be used immediately or divided into aliquots (0.2 ml or 0.6 ml) and frozen at –70°C.

The transformation protocol was based on a calcium chloride method. Plasmid DNA was added to 300 µl TCM (10 mM each of Tris pH 8, $CaCl_2$ and $MgCl_2$). Then 300 µl of competent cells were added, mixed gently and incubated at 4°C for 30 min. The cells were then transferred to 37°C for 10 min. For each transformation both negative (no transformed DNA) and positive controls (pGEM®-3Zf+ uncut vector plasmid) were included. LB plates containing ampicillin (50 µg/ml), 50 µl X-gal (20 mg/ml in dimethylformamide), and 20 µl IPTG (24 mg/ml in QH_2O) were inoculated with 10, 50, or 100 µl of transformed cells. Plates were grown at 37°C for 14-18 hrs then refrigerated (before the onset of satellite colonies).

Large Scale Preparations of Plasmid and Genomic DNA

The procedure for isolating large amounts of plasmids (or other DNA forms) was as follows. An overnight culture of *E. coli* SURE® cells, containing the plasmid of interest, was inoculated into 250 ml LB broth and grown overnight, under the selective pressure of ampicillin (50 µg/ml). The cells were concentrated in a GSA rotor using a DuPont Sorvall RC-5B centrifuge at 5000 x g rpm for 5 min. The supernatant was removed, the pellet resuspended in 20 ml TE, and the contents moved to a 50 ml Oak Ridge tube. The 5000 x g spin was repeated for 5 min in an SS-34 fixed angle rotor. The supernatant was discarded and the pellet was resuspended in 7 ml of solution I (15% sucrose, 50 mM Tris

pH 8.0, 50 mM EDTA) and 1/10th volume lysozyme (10 mg/ml in solution 1). The tube was mixed by gentle inversion and incubated at 25°C for 30 min. An equal volume of solution II (0.1% Triton, 50 mM Tris pH 8.0, 50 mM EDTA) was added, contents were mixed, and again incubated at 25°C for 30 min. The solution was centrifuged in the SS-34 fixed angle rotor for 60 min at 16,000 rpm. The supernatant was transferred to a fresh 50ml conical tube. Dry cesium chloride (CsCl) was added to 85.5% w/v and the solution was mixed until CsCl had dissolved. To the tubes 1.6 ml EtBr (10 mg/ml) was added and mixed. This solution was used to fill Beckman polyallomer quick-seal centrifuge tubes (16 x 76mm). The tubes were balanced, heat sealed, and placed in a Beckman 70.1Ti rotor and centrifuged in a Beckman L8-70M ultracentrifuge. The samples were spun for 48-72 hrs at 38,000 rpm at 20°C. After centrifugation, DNA bands were visualized with an UV light source and the plasmid band removed. To remove EtBr, plasmid DNA was washed 5 times (or once after all color was gone from both phases) with H_2O saturated n-butanol. Since n-butanol is less dense than water, the top (non plasmid containing) phase was discarded each time. The plasmid DNA was ethanol precipatated in 30 ml Corex tubes. After precipitation the plasmid pellet was washed three times with 70% ethanol and resuspended in sterile H_2O.

Miniprep

E. coli cells containing recombinant plasmids were inoculated into 5 mls LB broth with ampicillin in glass test tubes and incubated in a Bellco roller drum overnight at 37°C. A microfuge tube was filled with 1.5 ml of the overnight culture and centrifuged at 14,000 rpm for 45 seconds (all centrifugations were performed in an Eppendorf microfuge for this protocol). The supernatant was removed with a pasteur pipette and discarded. The pellet was resuspended in 150 µl STE (100 mM NaCl, 10 mM Tris-HCl pH 8, 1 mM EDTA, pH 8). To this suspension, 20 µl of fresh lysozyme (5 mg/ml) was added and the mix was

incubated at 22°C for 15-20 min. An equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added and gently mixed for 1 min. The solution was centrifuged at 14,000 rpm for 10 minutes. The upper, plasmid containing, phase was removed and placed in a fresh 1.5 ml microfuge tube. An equal volume of chloroform/isoamyl alcohol (24:1) was added and vortexed for 45 seconds. The solution was centrifuged for 3 minutes and the upper phase was again removed to a fresh microfuge tube. An equal volume of isopropanol was added and the solution was mixed, followed by a 10 minute incubation period at 22°C. After precipitation, the tube was again centrifuged for 10 minutes. The supernatant was removed (taking care not to disturb the nucleic acid pellet) and the tube inverted to allow the nucleic acids to air-dry (10-15 min). Nucleic acids were resuspended in 40 µl sterile QH_2O. The DNA could now be digested with the desired restriction endonucleases. Prior to agarose gel inspection, the sample was treated with RNAse (10 mg/ml) at 22°C for 10 min to remove endogenous RNA.

DNA Gel Electrophoresis and Transfer to Nylon Membranes

Plasmid DNA was digested with restriction enzymes and loaded with a gel running buffer (0.25% bromophenol blue, 0.25% xylene cyanol, and 30% glycerol diluted in sterile water) into EtBr (0.5 µg/ml) stained agarose gels (SeaKem LE or NuSieve GTG, both supplied by FMC Bioproducts). DNA was separated by gel electrophoresis in either an IBI Multi-Purpose Horizontal Electrophoresis Apparatus (model MPH) or Hoefer Scientific instruments horizontal mini-gel model HE33. In each case the buffer of choice was 0.5X TBE (10X stock: 890mM Tris Base, 890 mM boric acid, 20 mM EDTA). Both gel units used an EC 500 power supply (E-C Apparatus Corporation) as a power source. After size fractionation, gels were inspected by UV light fluorescence of stained DNA on a Fotodyne, Inc. UV light box. As needed, the DNA was transferred to Nylon membranes (Hybond-N^{+}™, Amersham Co.) by a capillary transfer procedure described by Southern (Southern,

1975). Briefly, gels were soaked in several volumes of 0.25 N HCl to 'nick' the DNA by depurination for 30 minutes. The gel was washed 3 times in several volumes of H_2O. Gels were washed in several volumes of a base wash (0.5 M NaOH, 1.5 M NaCl) for 30 min to denature the DNA. The gels were again rinsed with H_2O and subjected to neutralizing wash (0.5 M Tris pH 7.2, 3.0 M NaCl) for 30 minuntes. A large sponge was placed in a glass tray and the tray was filled with 20 X SSC (20X stock: 3 M NaCl, 0.3 M Na_3 citrate, pH 7.4 with 2 N HCL). On top of the sponge was stacked four sheets of Whatman 3MM paper (equal size to the gel), the DNA gel, the nylon membrane, and two more sheets of blotting paper. During the placement of these layers care was taken to remove any bubbles between the layers. These layers were topped with a 5-8 cm high stack of paper towels. The capillary action was allowed to continue overnight and the following morning DNA was UV cross-linked to the nylon membrane by a UV Stratalinker 1800.

Probe Construction

Large DNA templates were labeled by random primer labeling. Two microfuge tubes were used for each reaction. The first tube, containing probe DNA (20 to 200 ng), was brought to a volume of 9.4 µl with sterile dH_2O (template/primer mix). To that was added a 6-mer random primer ($d\{N\}_6$ where N=A,C,G,T from New England Biolabs, Inc.) to a concentration of 25 ng/µl. To the second tube (labeling mix) the buffer (10X = 75 mM Tris-HCl pH7.6, 55 mM dTT, 50 mM $MgCl_2$) was added to a 1X concentration. Next, all dNTPs were added to a 500 µM concentration (with the exception of dCTP). 2 µl of Radioactive α-^{32}P dCTP (10 µCi/µl) was then added, followed by 2.0 µl of the Klenow fragment of DNA polymerase (5 units/µl). Tube 1 was boiled for 2-5 min, centrifuged and briefly placed on ice. After the template/primer mix had cooled it was added to the labeling mix. The solution was incubated at 22°C for 1-2 hours. Afterwards 80 µL of TE was

added to halt the reaction. Unincorporated nucleotides were then separated from the labeled DNA by a spin column.

Spin columns were constructed by loading a 1-cc syringe (Becton Dickson and Company) with sterile glass wool to prevent loss of glass beads. Next, hydrated G50 beads (Sigma) were added, and the columns were spun in a tabletop IEC (model HN) swinging bucket centrifuge for 20 sec. This process continued until the packed volume of G50 beads reached ~0.8 cc. The probes were added and the columns were spun for 45 sec. The probe was collected in a catch tube and was ready for hybridization protocols.

Alternately, smaller fragments of DNA (e.g. oligomers under 100 bases in size) were end labeled. In this process, dephosphorylated 5' ends (1-50 pmol) were added to a 1X kinase buffer solution. To this mix $\gamma^{32}P$ ATP was added to a final concentration of 3000 Ci/mmol, 10-20 units of T4 polynucleotide kinase (New England Biolabs, Inc.) was added, and the solution brought to 50 µl with H_2O. The reaction was allowed to continue at 37°C for 10 min. During this process T4 polynucleotide kinase catalyzes the transfer of the radioactive gamma phosphate from ATP to the 5' terminus of the single stranded oligomers. Upon completion, a chelating agent (2 µl, 0.5 M EDTA) was added to stop the reaction and the radioactive oligomers were cleaned by two ethanol precipitations.

DNA and RNA Hybridizations

Immobilized DNA and RNA was inspected by two different methods depending on the probe used. For long oligomers (those labeled by random primer labeling) stringent aqueous hybridization was used. Nylon membranes containing immobilized DNA or RNA were placed in heat sealable bags with 10 ml hybridization buffer (0.5 M NaCl, 0.1 M $NaPO_4$ pH 7.0, 6 mM EDTA pH 7.0, 1% SDS and 100 µg/ml sonicated and denatured salmon sperm DNA). The bag was sealed and the membrane was allowed to agitate at 65°C for at least 1 hr. The bag was opened, 5 mls hybridization buffer was removed and 4

mls sterile H_2O was added. Probe DNA was boiled and added to the bag, which was resealed and again placed into the 65°C water bath for 18-24 hrs. The following day the nylon membrane was removed from the bag and placed into a glass dish. Roughly 300-500 ml wash buffer (2X SSC, 25 mM Na_2HPO_4, 0.1% NaP pH 7.0, 6 mM EDTA pH 7.0, 1%SDS), equilibrated to 65°C, was added to the membrane-containing dish and allowed to agitate at 65°C for 20 min. This process was repeated twice, followed by two washes with buffer at 1:2 and 1:5 dilutions for 5 min and 2 min respectively. Membranes were blotted with Whatman 3MM paper to remove excess liquid but allowed to remain damp. The membranes were then wrapped in plastic wrap and exposed to Amersham Hyperfilm™-MP autoradiograph film. The film was developed and visually inspected.

Hybridizations performed with short oligomers (those 5' end labeled with γ-ATP) were performed much like those with long oligomers except the hybridization buffer consisted of 10X Denhardt's solution (0.02% bovine serum albumin, 0.02% polyvinylpyrrolidone, 0.02% Ficoll) and 6X SSC. The membrane was sealed in a plastic bag and incubated at 25°C overnight. The membranes were washed in 6X SSC at 25°C twice. If membranes still contained a high amount of nonspecific binding, the temperature of the wash buffer was increased in 5°C increments. Finally the membranes were autoradiographed, as above, and visually inspected.

RNA Extraction

C. acidophila cells were grown on minimal agar plates at 22°C under continuous light until the cell lawn had reached confluence and produced a deep green color. Cells were then scraped from the plates with sterile razor blades, and suspended in M broth to an optical density of ~0.110 at A_{750} (7 x 10^5 cells/ml). Cells were agitated in broth under continuous light overnight. The following morning, cells were treated with various

concentrations (0 μM–200 μM) of $Cd(NO_3)_2$ over an increasing amount of time (0 hr-8 hr) at 22°C under continuous light.

RNA is noted for its susceptibility to ribonucleases (RNAses), due to the ability of the 2' hydroxyl group to act as an intra-molecular nucleophile. Whereas deoxyribonucleases (DNAse) require metal ions for activity and can be inactivated by chelating agents (e.g. EDTA), many RNAses bypass the need for metal ions by taking advantage of the 2′ hydroxyl group as the reactive agent. RNAse contamination is one of the primary reasons for the failure to isolate good quality RNA. Therefore the following precautions were taken to avoid RNAse contamination of naked RNA: (1) All glassware, spatulas, etc. were baked at 200°C for at least 8hrs; (2) Sterile, disposable plasticware was used where possible; and (3) Solutions were treated with 0.1% diethyl pyrocarbonate (DEPC) as previously described (Sambrook *et al.*, 1989), which operates by alkylating histidine residues that serve as the active site for most RNAses. Compounds with primary amine groups (e.g. Tris) will react with DEPC. Consequently, these materials were prepared from fresh sources and reserved specifically for RNA use. In addition, RNA was kept on ice during any experimental procedures, and frozen at –70°C for long term storage.

The RNA extraction was performed essentially as described by Keller (Keller, 1995) but with modifications. Flasks containing 250 mls of cells were placed into 250-ml bottles and pelleted for five minutes at 3000 x g with a Sorval GSA rotor. The liquid was aspirated and the pellet was resuspended in 5 mls of lysis buffer (50 mM Tris HCl pH 8, 0.3 M NaCl, 5 mM ethylene glycol bis(β-aminoethyl ether)-N,N′-tetraacetic acid [EGTA], and 2% sodium dodecyl sulfate [SDS]) with Proteinase K (40 μg/ml). The mixture was then transferred to a polypropylene Oak Ridge tube and mixed gently by inversion for 5 minutes. Phenol/chloroform/isoamyl alcohol (25:24:1) was added at a 1:1 volume to separate the proteinaceous material from the RNA. The tubes were again mixed by inversion for 30 seconds and placed on ice for 10 minutes (mixing/vortexing every two

minutes). Tubes were balanced and centrifuged in a Sorvall SS-34 rotor at 8000 x g for 20 minutes. After centrifugation the upper aqueous layer was removed and transferred to another Oak Ridge tube. Chloroform/isoamyl alcohol (24:1) was added at a 1:1 volume, mixed by vortexing, and centrifuged again at 8000 x g for 15 minutes. The upper aqueous layer was removed and placed in a sterile 50 ml polypropylene conical tube. RNAse free CsCl (biotechnology grade, Fisher Scientific) was added at 1 g/ml of aqueous phase, and mixed until the CsCl had dissolved. This solution now contained all genetic material formerly within the *Chlamydomonas* cell. RNA would be separated from the rest of the genetic components by ultracentrifugation. The procedure was as follows: First 1 ml of CsCl cushion (5.99 M CsCl) was pipetted into a Beckman polyallomer quick-seal centrifuge tube (16 x 76 mm). The RNA sample (up to 5 mls) was gently laid over the CsCl cushion without disturbing the cushion/sample interphase. The volume of the sample was brought to within 3 mm of the centrifuge tube top with TE (10 mM Tris-HCl, 1 mM EDTA) or mineral oil. The tubes were balanced and sealed with a Beckman tube sealer. The sealed samples were placed into a Beckman 70.1 Ti rotor and spun at 33,000 rpm for 18.3 hours at 20°C. After centrifugation the tubes were removed and the whitish RNA pellet was located and marked at the bottom of the tube. The sealed tubes were decapitated with a hot razor and the liquid was aspirated, care being taken not to disturb the RNA pellet. The pellets were dissolved with approximately 200 µl TE and transferred to a fresh Corex 30 ml tube containing 2.8 mls of TE. Ammonium acetate was added to a final concentration of 0.4 M, 2 volumes of ethanol were added and the solution was placed at -20°C overnight (or alternatively at -80°C for one hour) to precipitate the RNA. The RNA was collected by centrifugation at 8000 x g for 20 min in a Sorvall HB-6 rotor. The precipitation procedure was repeated and the pellet was resuspended in ~300 µl of RNAse free water. A 5 µl aliquot was removed and diluted to a 1:100 concentration in RNAse free water. The samples were placed in quartz cuvettes and the UV absorbance measured

spectrophotometrically at a wavelength of 260 nm. RNA concentration (in µg/ml) was determined by the equation: Total RNA=A_{260} x 40 x df (where df = the dilution factor). The purity of the sample could also be determined by reading the absorbance of the sample at A_{280}, then calculating the ratio of A_{260}/A_{280}. RNA samples of good purity were usually between the range of 1.8 to 2.0.

Northern Analysis

RNA was analyzed by one of two methods: (1) Formaldehyde gel electrophoresis and (2) Dot blot analysis. Formaldehyde gels were essentially as described by Sambrook (Sambrook et al., 1989) but with modifications. Gels were prepared by melting the appropriate amount of agarose in water. For formaldehyde gels, Seakem LE agarose (FMC, Bioproducts) was used at 1.0 to 2.5% or NuSieve® GTG agarose (FMC, Bioproducts) at 2.0 to 3.5% . Because of the frothing usually associated with a high concentration of agarose, water was added to excess and then reduced during the melting process. Once the agarose was melted and cooled to ~60°C, formaldehyde (12.3 M stock concentration) and 10X MOPS [3-(N-morpholino) propane-sulfonic acid] gel running buffer (0.4 M MOPS, 100 mM sodium acetate, 5 mM EDTA) were added to yield their final concentrations of 2.2 M and 1X respectively. The gel was then immediately cast into 10 x 14 centimeter trays and allowed to set at least 30 minutes under a chemical hood. During the setting process, total *C. acidophila* cellular RNA was prepared by adding 10 µg RNA, 1 µl 10X MOPS, 3.5 µl formaldehyde, and 10 µl formamide to a fresh microcentrifuge tube. These samples were brought to a total volume of 20 µl with sterile DEPC treated H_2O, and incubated at 65°C for 15 min. The samples were cooled on ice and centrifuged to collect all the liquid on the bottom of microfuge tubes. To each RNA sample 2 µl of gel loading buffer (50% Glycerol, 1 mM EDTA, 0.4% bromophenol blue, 0.4% xylene cyanol) and 1 µl of EtBr (1mg/ml) was added. Before loading these samples the gel

was submerged in a 1X MOPS gel running buffer and prerun at 20 volts for 5 minutes. The samples were immediately loaded onto the gel and RNA was separated by gel electrophoresis at 20 volts for 18-24 hrs. After electrophoresis, gels were photographed under UV light. RNA gels were then soaked in sterile H_2O for 30 min and the RNA was transferred to nylon membranes (HyBond™-N+, Amersham) through capillary action much like the Southern procedure described above. Gels were placed on top of glass plates and 20X SSC was wicked up from the glass dish by Whatmann 3MM paper.

DNA Sequencing

RNA transcript production was found to originate from a 1500 bp *Hae* III fragment. This blunt end fragment was cloned into the *Sma* I site of pGEM® (pJB5) and sequenced. Although some of the sequencing was performed with Sequenase® (Amersham Life Sciences), the majority of sequence analyses was performed with fmol® (Promega), which is described here. It should also be noted that some sequence information, as well as confirmation of our previous efforts, was kindly provided on an ABI Model 377 automated sequencer by Scott Bingham of Arizona State University.

Because the 1500 bp fragment was cloned into pGEM® the initial sequencing efforts could be started at the ends of the insert. The plasmid pGEM has the T7 and SP6 promoter sites flanking the multi-cloning site, which contained the 1500 bp insert. Further, primers (T7 and SP6, Promega) for those regions are readily available. The sequencing efforts proceeded by 'walking' the insert and constructing primers from the derived sequence. Sequencing primers were constructed by Operon (Alameda, CA) and are listed in Appendix B.

The sequencing protocol utilizing direct incorporation of ^{35}S was as follows. Four microcentrifuge tubes (G,A,T,C) were labeled and to each tube the appropriate d/ddNTP mix was added (the fmol® system substitutes 7-deaza dGTP for dGTP, which resolves

band compressions associated with GC-rich regions). For example, the G tube contains a mix of all four dNTPs and a limiting amount of ddGTP. The rationale is that on occasion Taq polymerase incorporates one of these dideoxyribonucleoside triphospates. Because ddNTPs lack the terminal 3´-OH group necessary for chain elongation, the growing chain is terminated, in this case with a ddGTP. The following reagents were then added to a fresh tube (denoted primer/template mix): 500 fmol of the plasmid pJB5, 3.0 pmol of the appropriate primer, 0.5 µl α-^{35}S dATP (10 µCi/µl) and 5 µl fmol® 5X sequencing buffer. The contents of the tube were brought to a volume of 16 µl with sterile H_2O followed by the addition of 1.0 µl of sequencing grade Taq (5u/ml) and gently mixed. To each of the four d/ddNTP tubes 4.0 µl of the template/primer/enzyme mix was added and the entire contents covered with 20 µl mineral oil. The four d/ddNTP tubes were centrifuged briefly and transferred to a Coy thermocycler that had been preheated to 95°C. The typical thermocycler profile for sequencing was as follows. Tubes were initially heated to 95°C for 2 min. A programmed profile of 95°C for 30 sec, 42°C for 30 sec, and 70°C for 1 min, was repeated for 30 cycles and then brought to a final temperature of 4°C at the end of the run. After completion of the program, a stop solution containing EDTA and a sequencing gel dye was added.

The products of the fmol® sequencing reactions were separated initially on acrylamide gels, but later we utilized the Long Ranger™ gel solution from FMC Bioproducts because of this gel's capacity for longer, cleaner reads. Usually a total of 4µl of sequencing samples were heated to 75°C, loaded and run on a 6% Long Ranger™ gel solution (42 g urea, 5 ml 10X TBE, 12 ml 50% stock Long Ranger gel solution, H_2O to 100 ml, solution was degased and 50 µl TEMED and 500 µl 10% APS added prior to casting) through 0.5X TBE gel running buffer solution. Sequencing gels were typically run at 55 Watts until the dye front had reached the bottom of the gel, after which more samples could be loaded for extended reads into the DNA sequence. After the completion of electrophoresis, gels were

mounted onto Whatman 3MM filter paper, wrapped in plastic wrap, dried at 70°-80°C for 60 min and exposed to Amersham Hyperfilm-MP™ (with or without intensifying screen). From the autoradiograph, the DNA sequence was inspected and recorded manually. The sequence was compared to the DNA database at the National Center for Biotechnology Information website (NCBI, http://www.ncbi.nlm.nih.gov/) using the available suite of BLAST algorithms.

Mitochondrial Genome Analysis

Preparation of Mitochondrial DNA

Previous investigations led us to believe that the mitochondrion of *Chlamydomonas acidophila* was circular. Based on this information we attempted to retrieve mtDNA via alkaline lysis. Cells were spun down in a microfuge tube, and the supernatant was discarded. The cells were resuspended in 500 µl solution I (50 mM glucose, 50 mM tris, and 50 mM EDTA) and the resulting solution was split into 2 microfuge tubes (each containing ~250 µl of sample). To each of the tubes 500 µl of solution II was added (0.2 N NaOH, 1% SDS), the samples were vigorously mixed (but not vortexed) and stored on ice for 10 minutes. After mixing, 375 µl solution III (3 M KCl and 5 M Na0Ac) was added to each tube, samples were mixed vigorously and stored on ice for 5 minutes. The microfuge tubes were then spun at 14,000 rpm for 5 minutes and the supernatant was harvested. This last step was repeated once.

The supernatant was extracted with an equal volume of phenol/chloroform/isoamyl (25:24:1). The solution was mixed by inversion for 5 minutes and spun at 14,000 rpm for 15 minutes. After centrifugation, the upper aqueous phase was removed to a new tube and an equal volume of chloroform/isoamyl alcohol (24:1) was added, mixed, and again centrifuged at 14,000 rpm for 10 minutes. Chloroform extraction was repeated until the

entire interface became clear. To the aqueous phase was added 2 volumes of ethanol, 0.1 volumes of NaOAc, and the solution was mixed and allowed to precipitate at –20°C overnight or at –80°C for 1 hour. Microfuge tubes were spun for 30 minutes, the supernatant removed and the pellet washed three times with 70% ethanol. The pellets were dried under vacuum desiccation and the DNA was resuspended in a total of 500 μl sterile H_20. To the tube was added RNAse and incubated for 1 hour at room temperature. Phenol/chloroform extractions were performed, as above, after which the DNA was ethanol precipitated and resuspended in 50 μl H_2O.

Cloning of *C. acidophila* mtDNA

Earlier investigations had suggested that *C. acidophila's* mtDNA contained a single *Bam* HI site (Spanier, unpublished results). Mitochondrial DNA, isolated by alkaline lysis, was cleaved with *Bam* HI and cloned into LambdaGEM-11® from Promega (Madison, WI). Cloning procedures were performed according to manufacturer protocols. Briefly, mitochondrial DNA was ligated at a *Bam* HI site between LambdaGEM-11® arms (left arm: 20 kb; right arm 9 kb). Recombinant DNA was packaged into Packagene® Lambda DNA packaging extract according to manufacturer specifications. Phage was transfected into *E. coli* cells strain LE392, and titered. After determination of phage numbers, phage was added to *E. coli* and allowed to adsorb for 30 minutes at 37°C. To this mix molten LB top agar was added and immediately poured onto LB plates. Plates were incubated at 37°C until plaques were pinpoint in size (~3.5 hours), then removed to 4°C for at least 1 hour to harden agar. Nitrocellulose filters were labeled, marked for orientation, and placed on agar. Filters were left on plates for 1-10 minutes at room temperature, removed and allowed to air dry at room temperature for 10-20 minutes. The filter was then moved to three trays each containing Whatman® 3MM saturated with the

following solutions for each tray: (1) 0.2 M NaOH, 1.5 M NaCl; (2) 0.4 M Tris-HCl, pH 7.6, 2X SSC; (3) 2X SSC. Incubation times for each treatment were 1-2 min. Nitrocellulose filters were probed with pJB1 and phage that appeared positive were isolated and amplified for further screening procedures.

Characterization of Recombinant Phage Clones

DNA isolated from phage particles, was screened according to two procedures. The first involved creating a restriction map of the isolated mtDNA. The second confirmed the insert size by the polymerase chain reaction, by using the GeneAmp® XL PCR kit from Perkin Elmer Applied Biosystems (Foster City, CA). Insert DNA amplifications were performed with the primers LLA20015 and LRA97 (Appendix B), both of which bound to DNA in the vector arm regions and amplified the inserted mitochondrial DNA between the arms. The reaction mix was in accordance with the manufacturer recommendations. Reaction mixes were placed in a Perkin Elmer GeneAmp® PCR System 2400 thermocycler. The PCR cycling parameters were as follows: (1) Pre-PCR Hold: 94°C for 1 minute. (2) 94°C for 15 seconds, 68°C for 10 minutes; 16 cycles. (3) 94°C for 15 seconds, 68°C for 10 minutes (and increasing by 15 seconds each cycle) for 12 cycles. (4) 72° for 10 minutes, 4°C until post-PCR analysis. After amplification, PCR products were inspected on EtBr stained agarose gels. After analyses, the recombinant clone F3 was selected for DNA sequencing. Both strands of the mtDNA insert were sequenced by Commonwealth Biotechnologies, Inc. (Richmond, VA).

DNA Sequence Analysis

DNA sequence analysis was accomplished by a number of different methods. Initial determination of coding regions was provided by both BLAST and FASTA search algorithms at the NCBI (http://www.ncbi.nlm.nih.gov/) and EBI

(http://www.ebi.ac.uk/ebi_home.html) websites respectively (Altschul *et al.*, 1990; Pearson, 1994b; Pearson, 1994a; Pearson & Lipman, 1988). Multiple sequence alignments for protein comparisons, the determination of intron insertion, and the search for group I intron conserved sequences was prepared by the CLUSTAL X program (Jeanmougin *et al.*, 1998; Thompson *et al.*, 1997). Determination of protein hydrophobic regions was provided by Gene Inspector™ (Textco, Inc.) The sequence in Figure 4 was constructed primarily by the Gene Construction Kit™ (Textco, Inc.). Codon bias tables were constructed with MacVector (Oxford Molecular Group). RNA folding and secondary structures were provided by the MULFOLD program (Jaeger *et al.*, 1989; Jaeger *et al.*, 1990; Zuker, 1989). Visualization of RNA structures was provided by the LoopDLoop program written by D. G. Gilbert (Indiana University).

RESULTS

Mitochondrial DNA Isolation

Previous investigations have suggested that the mtDNA of *C. acidophila* strain 122 is circular and contains a single *Bam*HI site (Spanier, unpublished data). Based on these data the decision was made to clone *C. acidophila* mtDNA by cutting with *Bam*HI and ligating the linearized DNA into the cloning vector LambdaGEM-11®. Prospective clones containing mtDNA were identified by hybridizing a radioactive subfragment of the genome, called pJB1 (Figure 11 A, page 99), to phage DNA attached to nitrocellulose filters. Of the candidates that showed hybridization to the probe, three clones were selected for further analysis.

Two methods were employed to determine which clone contained the largest insert. The first involved restriction digestion of each clone, size analysis of the restriction fragments, and construction of a map to estimate the total insert size. The second method used PCR amplification of the entire insert to estimate the size of the mtDNA fragments. Two primers, LLA20015 and LRA97, were designed to amplify DNA inserted within the lambda DNA vector arms. Figure 2 shows the results of this amplification. Lane 1 contains lambda DNA cleaved with Hind III and serves as size markers. Lane 2 contains control insert DNA. Lanes 3 – 5 contain DNA amplified from *C. acidophila* mtDNA inserted within LambdaGEM-11. Lane 5 contained the slowest migrating DNA,

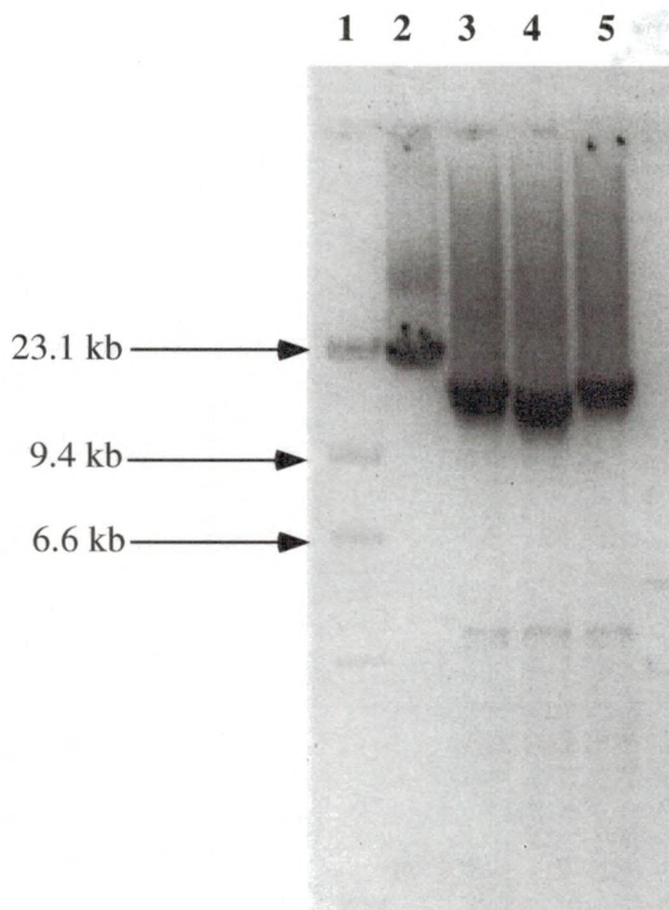


Figure 2. Size classification of LambdaGEM-11® *C. acidophila* mtDNA insert. Negative image of 0.6% Agarose gel stained with EtBr. DNA products were generated with GeneAmp® XL PCR. Lane 1: Lambda DNA digested *Hind*III, size of products are listed. Lane 2: Control amplification of lambda template DNA (~20 kb in size). Lanes 3, 4, and 5 contains *C. acidophila* mtDNA cloned into LambdaGEM-11®. Isolates were designated B2, E2, and F3 respectively.

migrating DNA, designated F3. From both restriction mapping analysis and PCR amplification it was determined that the F3 clone contained the longest DNA insert. Therefore, F3 was selected for sequence analysis.

Size and Genomic Organization of *C. acidophila*

Sequence analysis of the F3 clone indicated that it contained only part of the entire mitochondrial genome rather than the full compliment of genetic information. Consequently, the F3 fragment is represented as a linear entity in Figure 3 instead of a circular entity, which is believed to be the true configuration of the mitochondrial genome (Spanier, unpublished results).

The coding regions identified in the mitochondrial genome of *C. acidophila* were identified by sequence similarities with other organisms. These regions are indicated as a linear map in Figure 3 and are defined in Table 2 below. The sequence of the F3 clone,

Table 2. Coding regions identified in *C. acidophila* mtDNA.

A. Ribosomal RNA gene pieces (5) Small subunit rRNA in pieces *rns a*, *rns c* Large subunit rRNA in pieces *rnl a*, *rnl d*, *rnl e*
B. Transfer RNA genes (4) $tRNA^{met1}$ (cau), $tRNA^{met2}$ (cau), $tRNA^{trp}$ (cca), $tRNA^{gln}$ (uug)
C. Respiratory chain genes (5) NADH dehydrogenase (*nad 1*, *nad 2* [incomplete], *nad 5* [incomplete]) Apocytochrome b (*cob*) Cytochrome oxidase (*cox 1*)
D. Group I intronic ORF's (7) *cob i1 orf149* and *orf112*, *cox 1i1 orf306*, *cox 1i2 orf358*, *nad 5i1 orf229*, *nad5i2 orf 296*, *nad5i3 orf 267*

including defined regions, can be seen in Figure 4. The sequence spans 15,010 bp and has a G + C content of 32%. Based on the sequence data, the F3 clone contains five

Figure 3. Linear map of mitochondrial DNA from *Chlamydomonas acidophila*.
Protein coding regions are in red. Ribosomal RNA and transfer RNA are shaded in dark blue and light blue respectively. Intronic regions are yellow. Abbreviations: *nad2* - NADH dehydrogenase subunit 2; *cob* -apocytochrome b; *cobi1* - intronic region of *cob*; *rnl d* – ribosomal RNA, large subunit fragment d; *rns a* - rRNA, small subunit fragment a; *rnl a* - rRNA, large subunit fragment a; tRNA met1 - transfer RNA for methionine; tRNA met2 - tRNA for methionine (possible pseudogene); *rnl e* - rRNA for large subunit fragment e; *rns c* - rRNA for small subunit fragment c; *cox1* - cytochrome oxidase subunit 1; *cox1i1* and *cox1i2* - intronic regions of cox gene; tRNA trp - tRNA for tryptophan; tRNA gln - tRNA for glutamine; *nad1* - NADH dehydrogenase subunit 1; *nad5* - NADH dehydrogenase subunit 5; *nad5i1*, *nad5i2* and *nad5i3* - introns within *nad5*.

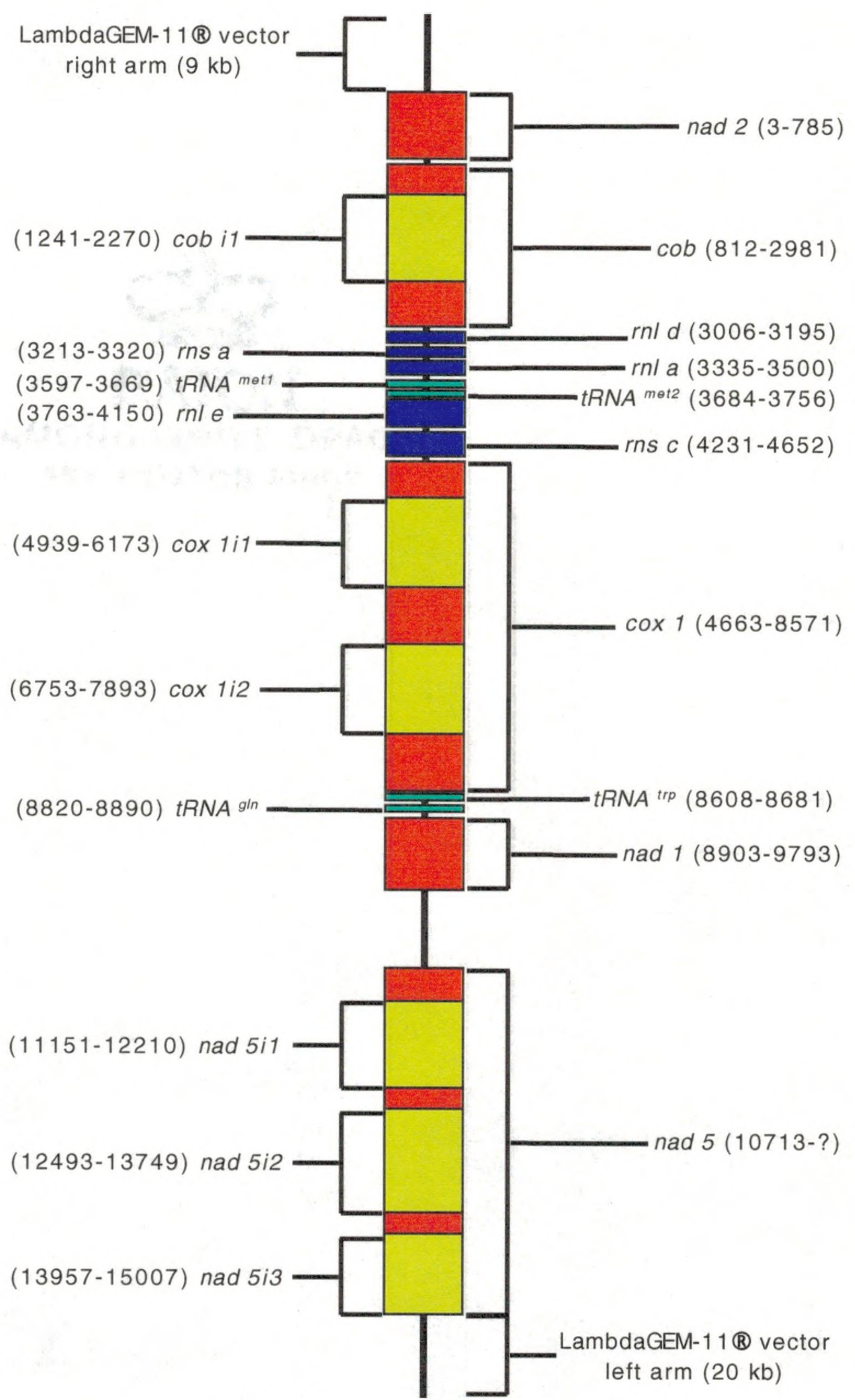

LambdaGEM-11® vector
right arm (9 kb)
nad 2 (3-785)
(1241-2270) cob i1
cob (812-2981)
rnl d (3006-3195)
(3213-3320) rns a
rnl a (3335-3500)
(3597-3669) tRNA met1
tRNA met2 (3684-3756)
(3763-4150) rnl e
rns c (4231-4652)
(4939-6173) cox 1i1
cox 1 (4663-8571)
(6753-7893) cox 1i2
tRNA trp (8608-8681)
(8820-8890) tRNA gln
nad 1 (8903-9793)
(11151-12210) nad 5i1
nad 5 (10713-?)
(12493-13749) nad 5i2
(13957-15007) nad 5i3
LambdaGEM-11® vector
left arm (20 kb)

Figure 4. Nucleotide sequence of the *C. acidophila* F3 clone
Since *C. acidophila* codes all of its genetic elements on one strand, only the sense strand is shown. The deduced amino acid sequences are located below their coding DNA sequences. Nucleotide numbers are in regular typeface and located to the left of the DNA sequence. Amino acid numbers are also located to the left of the protein sequence but are in boldface type. An asterisk (*) indicates a termination codon. Ribosomal RNA, transfer RNA and intronic sequences are underlined, including the open reading frames contained therein. Important restriction enzymes are shaded and labeled. Repetitive elements are also shaded and labeled. The boundaries of all genetic elements are numbered.

```
      nad2
   1 ATATCATTCCTCAAATGATGATATTAATAGCACTTATGTTTAAGTTAGGGGGTGCTCCTT
   1> I  I  P  Q  M  M  I  L  I  A  L  M  F  K  L  G  G  A  P

  61 TACATATATGGATGGTAGATATTTATAGTGGTGTAAAACGTCAATTATTAATGTATTTGT
 20>L  H  I  W  M  V  D  I  Y  S  G  V  K  R  Q  L  L  M  Y  L

 121 CTACAGCTCCTAAATTAAGCTTATTTGGTTTTTGGGTATCTACTTGGCATTCAGTATGGA
 40>S  T  A  P  K  L  S  L  F  G  F  W  V  S  T  W  H  S  V  W

 181 CTGATTTTACATTATTTTTATTTGTAGCTTTATCTATGATTATTGGTTGTTTCGGTGCTT
 60>T  D  F  T  L  F  L  F  V  A  L  S  M  I  I  G  C  F  G  A

 241 ATAATCAACCAACATTACGAGCGTTATTTGCATATAGTACAATAAATGAAATAGGGTTAA
 80>Y  N  Q  P  T  L  R  A  L  F  A  Y  S  T  I  N  E  I  G  L

 301 TGTTAATGGCTATTGAAACAGCTGGGTTTCATTCAATGTTTCAACATTTAAGTATATATA
100>M  L  M  A  I  E  T  A  G  F  H  S  M  F  Q  H  L  S  I  Y

 361 TAGTAACCATGTTGTTACTTTGGAATATAACAGATCATCGCTTTTTTTCTATTTTAGCTG
120>I  V  T  M  L  L  L  W  N  I  T  D  H  R  F  F  S  I  L  A

                                                  HindIII
 421 TTAGTTTAGCAGGATTGCCACCATTAGCTGGCTTTTTTGGTAAAGCTTGGATTTTTAATA
140>V  S  L  A  G  L  P  P  L  A  G  F  F  G  K  A  W  I  F  N

 481 GTGTAGCTATAGGTTCAATGGCTGGTCCCTACCTAGGGCTGTTATTAATCTCATTATTCT
160>S  V  A  I  G  S  M  A  G  P  Y  L  G  L  L  L  I  S  L  F

 541 GTACAGGATTGTCTTTAGTATATTATTTACGTGTTTTTCGCTTATTTACGATGAGTAATC
180>C  T  G  L  S  L  V  Y  Y  L  R  V  F  R  L  F  T  M  S  N

 601 AAGTGAGTCGCAATAATATAATTTATCCTGTAGGTGTTGATGGTAGTGTAGGTAATCCTT
200>Q  V  S  R  N  N  I  I  Y  P  V  G  V  D  G  S  V  G  N  P

 661 ACAATATGTCAGTGGTTAGTCGTACTTATAATACTCATTTGGTTGATTTTAATATAAAAA
220>Y  N  M  S  V  V  S  R  T  Y  N  T  H  L  V  D  F  N  I  K

 721 TGACTTCTTTTTGTGTTATTTTCTTAATGTTTGCACCTTTATTCTATATTAAGCCTTTTG
240>M  T  S  F  C  V  I  F  L  M  F  A  P  L  F  Y  I  K  P  F

               788                   812  cob exon 1
 781 TGCTATAAAAAAATTTACAAAAAAAATTAAAAATGCGTTTACATAATAAAATTCAAGTATT
260>V  L  *                          1> M  R  L  H  N  K  I  Q  V  L

 841 AAATTTATTAAATCATCATATTGGTGTTTATCCAACACCTATGAATATTAATTGGAATTG
 11>  N  L  L  N  H  H  I  G  V  Y  P  T  P  M  N  I  N  W  N  W

 901 GAGTTGGGGATCATTATCAGGTTTGGTTTTAGCCAGTCAAATAGTAACTGGTATATTGTT
 31>  S  W  G  S  L  S  G  L  V  L  A  S  Q  I  V  T  G  I  L  L

 961 GGCAATGCATTATGTTGGTCATGTTGATCATGCTTTTTCTAGTGTACAACATTTAATGGT
 51>  A  M  H  Y  V  G  H  V  D  H  A  F  S  S  V  Q  H  L  M  V

1021 TGATGTACCTTCTGGTGTAATATTACGTTATGCTCATGCAAATGGTGCTAGTTTATTTTT
 71>  D  V  P  S  G  V  I  L  R  Y  A  H  A  N  G  A  S  L  F  F
```

```
1081 TACCGTAGTTTATTTGCACGTGTTGCGTGGTTTATATTATAGTAGTGGTAATCAGCCTCG
 91>  T  V  V  Y  L  H  V  L  R  G  L  Y  Y  S  S  G  N  Q  P  R

1141 TGAAATTGTATGGATTTCCGGTGTTGTTATTTTATTATTGATGGTAATTACAGCCTTTAT
111>  E  I  V  W  I  S  G  V  V  I  L  L  L  M  V  I  T  A  F  I

                                       1241  cobi1 orf149
1201 TGGTTATGTGCTCCCTTGGGGTCAAATGAGCTTTTGGGGTATTCTATATTGCCCTAAATA
131>  G  Y  V  L  P  W  G  Q  M  S  F  W  G
                                           1> I  L  Y  C  P  K  Y

1261 TGATTTTATATTTGTTTTTGTTGTTTTTATTCATCCACCAAAACGTTTGTTAGCTAAACA
  8>  D  F  I  F  V  F  V  V  F  I  H  P  P  K  R  L  L  A  K  Q

1321 ACGTATAGGTCCACATAATATTGATATATTGTCAATAATTGTTGGTTCTTTGTTGGGAGA
 28>  R  I  G  P  H  N  I  D  I  L  S  I  I  V  G  S  L  L  G  D

1381 CAGTTATGCTGAAAAACGCAATGGATCTACACGTATACATTTTCAACAAGAAAGTTCAAA
 48>  S  Y  A  E  K  R  N  G  S  T  R  I  H  F  Q  Q  E  S  S  N

1441 TCGAGAGCATTTATTGAATTGTTGGAAAATATTAAATAAAGGTAGTTATTGTTCTGATAT
 68>  R  E  H  L  L  N  C  W  K  I  L  N  K  G  S  Y  C  S  D  I

1501 TGAACCAAAAATAGAAGAGCGTCTTGGAAAAGAAGGTAAAATACGTTTTTTATCTCGTTT
 88>  E  P  K  I  E  E  R  L  G  K  E  G  K  I  R  F  L  S  R  F

                                 1596  orf112
1561 TAAAACATATTCTTTTTCAAGTTTCAATTGGATTCATGATGCTTTTTTATTTGAAACGCA
108>  K  T  Y  S  F  S  S  F  N  W  I  H  D  A  F  L  F  E  T  Q
                                      1> M  M  L  F  Y  L  K  R

1621 AAAAGGTAGTTCCTTTAGATTTAATAAATCTTTTAACTCCTTTAGCATTAGCTATTTGGA
128>  K  G  S  S  F  R  F  N  K  S  F  N  S  F  S  I  S  Y  L  D
   9>K  K  V  V  P  L  D  L  I  N  L  L  T  P  L  A  L  A  I  W

             1691
1681 TAATGGATGATGGTACGTGGCAAGGATCAGGCGTTCGTATAGCTACTAATTGTTTTTCAT
148>  N  G  *
  29>I  M  D  D  G  T  W  Q  G  S  G  V  R  I  A  T  N  C  F  S

1741 TCAATGAAAATAAATTATTGTGTTCATTATTAAATAAAAAATATAATTTATTTTGTACGG
  49>F  N  E  N  K  L  L  C  S  L  L  N  K  K  Y  N  L  F  C  T

1801 TGGTAAAAAATGGCAAAAACAAAGATCAAACGATAGCATATAATATATATATACATAAAC
  69>V  V  K  N  G  K  N  K  D  Q  T  I  A  Y  N  I  Y  I  H  K

1861 AATCTATAATAGAATTACAAAAAATAGTAAAACCTTTTTTTGTTAAAAGTATGTTGTATA
  89>Q  S  I  I  E  L  Q  K  I  V  K  P  F  F  V  K  S  M  L  Y

                1934
1921 AGATAGGCTTATAGTTAAATAAAATCATATAATTTATGTATGTATATATCTAATAACTTG
 109>K  I  G  L  *

1981 TACAAGAAAAGATATAATTTATATATAAGTAATGCAAGTGAAAACGGTTAAATTATGTAT
```

```
2041 CTTTAAATAAAAGACCGTCAGTTTTGTACAAAACTGCTACAGATTAGAAATCTTCTGTTC

2101 AAGAAGCAGCACTTGTGGGTATCTGAAATGATGCTTAATGTATAATCGGAGTTTCAGTTT

                                                  ScaI
2161 ATAATATATATTATAAGCAATAAGGCTTTAAGTAAATCCAGTACTGGTATTAAAGTACAT

                                                           2271
2221 TTAAAATAAAACAAAAAGAATTTAATTATATTTTGTTTTATTGAAACTTGGCTACAGTAA
                                                     144> A  T  V

     cob exon 2
2281 TTACAAGTTTAGTAACAACAGTACCCATCGTAGGAAAACAAATTGTTTTTGGTTATGGG
 147>I  T  S  L  V  T  T  V  P  I  V  G  K  Q  I  V  F  W  L  W

2341 GTGGATTTAGCATTGATCATCCAACATTAAATCGTTTTTACAGTTTGCATTATACATTAC
 167>G  G  F  S  I  D  H  P  T  L  N  R  F  Y  S  L  H  Y  T  L

                                                            ScaI
2401 CTTTTGTGTTAGCTGCTTTAAGCATCTTTCATATAGCAGCATTACATCAATATGGTAGTA
 187>P  F  V  L  A  A  L  S  I  F  H  I  A  A  L  H  Q  Y  G  S

2461 CTAATCCGTTAGGTATCAATACACAAAGCAGCACAATACATTTTGGAATTTACTTTTTAA
 207>T  N  P  L  G  I  N  T  Q  S  S  T  I  H  F  G  I  Y  F  L

2521 GTAAAGATTTATTAGCTCTATTATTCTTTATATTGGTTTTTGCCGTTTTAGTGTTTTTTT
 227>S  K  D  L  L  A  L  L  F  F  I  L  V  F  A  V  L  V  F  F

2581 ATCCTGAATGGTTAGGTCATCCTGACAATTTAATCCCTGCAAATCCATATTCAACTCCAC
 247>Y  P  E  W  L  G  H  P  D  N  L  I  P  A  N  P  Y  S  T  P

2641 AACACATTGTACCTGAATGGTATTTCTTATGGGTTTATGCTATTTTACGTAGCATTCCTA
 267>Q  H  I  V  P  E  W  Y  F  L  W  V  Y  A  I  L  R  S  I  P

2701 ACAAAGCCATGGGTTTTGTGGGTGTATTGTTGGTATTTGCGTGTTTAATAGCATTACCTT
 287>N  K  A  M  G  F  V  G  V  L  L  V  F  A  C  L  I  A  L  P

2761 TCATTAGTGTAGTACAAGTAGGGTCTCCTCGTTTTCGTATAATATATGAACGTTTATTTT
 307>F  I  S  V  V  Q  V  G  S  P  R  F  R  I  I  Y  E  R  L  F

2821 GGGTATTAGTAGCTGATTTATTCTTATTAACTTGGGTAGGTGCTCAAGAAATTATGCCAG
 327>W  V  L  V  A  D  L  F  L  L  T  W  V  G  A  Q  E  I  M  P

2881 CTACCGTATTATTAGGACAAATTTGTACCGTTGTATTGTTTGTCTATTTATTAGTCATAT
 347>A  T  V  L  L  G  Q  I  C  T  V  V  L  F  V  Y  L  L  V  I

                                                  2985
2941 TACCTTTTTTAGGTTGGTTAGAAACTGCTTTAGTATTAGCTTAACATACTAAAAATAAAA
 367>L  P  F  L  G  W  L  E  T  A  L  V  L  A  *

         3004  rnl d
3001 TTTAAAATCTTTCTTTTGGTAGCTAGGCGTTCTATTTTTTAAATTAATTTTTAATAAAAG

3061 ATTATTATTAAATTGAATTTAGTTCAATTTTTTATGGAAACAATAAATAGAATTGAGACT

                                        HaeIII
3121 GCTGACATTAGTAACTTTTAGTTTGAAAACGGCCGAAAGTCCGAGGGTTTTTTAGTGAA
```

```
                3192                     3213 rns a
3181 TCGGTTAAATATTAGTTAATTTATTAAAATTAAAAATCAATTTATGATGAGTTTGATGCT

3241 GGCTCCGCAAAAATGCTTTTATTAAGGTTAATACATGCGAGTGATATAGAGTAACGTACG

                     3320           3335  rnl a
3301 GGTGCGTAATATGCGAGTTTAAAAAAAAAATTAAAAATGAATTTAAGGTAACTGATAGTG

3361 AATTAATTAGGATTTTTATTCTTAATCAGTACAATGATGGAAACAATGAAATTATAACTT

3421 AGCTAATACAATAATAAAAATTATGTACCTTTTGCATCATGGGTCAGTCACTTAATAAAC

                     3500
3481 ATAAACTAAAGTTATTATAACTCTTCTAAAAAAATCGAACCTTTTTTCTTTTTTAAAGAA

                                                           3597
3541 AAACCGTGAGTAGCTTTCACATAAATTATAAACTTTTTTTTTCAACTTTATAATATTGGT

     tRNA^met1
3601 GATTAGCTCAATGGTTAGAGCATAGGTCTCATAAACCTATGGTTACGAGTTCAAGTCTTG

             3669           3684  tRNA^met2
3661 TATCACCAAAATAAGATGTTAATTGGTATGATACCGAAGGGTAGAGGTAAGGGGTTCATG

                                             3763  rnl e
       BstN1                              3756
3721 CCCCCTGGTAAGCAGGTTCGAGTCCTGAGATACCAAAAATAACCGTACCACAAACCAACG

3781 CAGGTGGACTACAATCATATTGTTAGGCGTAGAATTAACTATATATAGGGAACTCGGCAA

3841 AATGTTTTATAGACTTAGGTTTAATAAAATCCCTTTTTCAAGGGAATAATATAAAAGATA

3901 GCTGCGACTGTTTACCAAAAACACATGACTATGCAAAGAAAAACCAAGTATATAGTCTGA

3961 CACCTGCCCAAAGGCTATAGGCAAACGGCAGCCGTAACTCTAACGGTTCAAAGGTAGCAA

4021 AATTCCTTGACGTTTAATTGGCGTCCTGCATGAAGGGTGTAACGATGGCTATGCTGTCCC

4081 ATATATAGATTCAGTGAATTTGAATTACCCGTGCAGATGCGGGTTTTTAAGCACCGGACG

              4150
4141 AAGAGACCCTGTGCACCTTTACATGTTGTTACAACGTAAAACACAATAGATCTTCAATGA

                                  4230  rns c
4201 ATAGGTGGAATAAATAAAAAAATTTAGAAAAGCATTGCATGGCTGACTAGCTGTTTTAT

4261 ATATTAGTAATAAATAAAGTATAACCAAAAAAGCTGCACAAGTCCGCATGGTCTTTATAA

4321 AGTGGGCTACACGTTTGCTACAATGGATGGTATAACATAAAAAATCATTCGTAGTCCAGA

4381 TTAAAAACCTGAAATTGGTTTTATTAAGGAGGAATCGCGAGTAATCGAAAATCAGACAAG

4441 TTTCGGTGAAGTTTTTGTTAATTTTGTAATCTTTTAATTATGAAATTATTAACTAATCTA

4501 GTTGATCTCGTACTCACTGCCCGTCAAGGGCTTATAGAATTTAAGACGTCAAATTTTGAT
```

```
                                                       HaeIII
4561 ATTCTACTTTAGAAAAATTAAAAATAACGTCTTACGTTTGATAGGCCTTAAGTCGTAACA

                            4652            4663  cox1 exon 1
4621 AGGTAGGACTAGGGGAACCTGGTCCTGTGATATTTTAAACACATGGCAATTCGTTGGTTA
                                             1> M   A   I   R   W   L

4681 TATTCAACAAATCATAAAGATATTGGAATTTTATATTTATTATTAGCCTTATTTGCAGGT
  7> Y   S   T   N   H   K   D   I   G   I   L   Y   L   L   L   A   L   F   A   G

4741 ATAATTGGTACTACTTTATCAATGTTTATTCGTTTAGAATTAGGTTTACCTGGTGAAGGT
 27> I   I   G   T   T   L   S   M   F   I   R   L   E   L   G   L   P   G   E   G

4801 TTATTAAATGGAAATGGACAATTATATAATGTTATTATTACTGGACATGGTATTATTATG
 47> L   L   N   G   N   G   Q   L   Y   N   V   I   I   T   G   H   G   I   I   M

4861 TTATTATTCATGGTAATGCCTGCTTTGTTTGGCGGTTTTGGTAACTGGTTAGTTCCTATT
 67> L   L   F   M   V   M   P   A   L   F   G   G   F   G   N   W   L   V   P   I

                        XbaI
                   4939  cox1i1
4921 TTAATTGGTGCTCCTGATAATCTAGAGCTTTTAAATTATTATTCAATATTATTTACTACT
 87> L   I   G   A   P   D
                      >1 N   L   E   L   L   N   Y   Y   S   I   L   F   T   T

     orf306
4981 TCATCAAATCTTTTAATGCACAATAACTTGCATTCAAACACTAAAATGGCTAGTTATTTA
 15> S   S   N   L   L   M   H   N   N   L   H   S   N   T   K   M   A   S   Y   L

5041 GCAGGTTTGTGGGAAGGTGACGGTCATATTGTTTTACCGACACACAATAATACACCCTGT
 35> A   G   L   W   E   G   D   G   H   I   V   L   P   T   H   N   N   T   P   C

5101 ATAGCTATTACTTTTTCTGATAAAAATGCACCTTTAGTTGATTTTTTAATAAAAAATTAT
 55> I   A   I   T   F   S   D   K   N   A   P   L   V   D   F   L   I   K   N   Y

5161 GGAGGTTGGGTTCGTATTAAAAAAAAGGAATCTTCACTTGTTTGGACAATAACAAAGCAA
 75> G   G   W   V   R   I   K   K   K   E   S   S   L   V   W   T   I   T   K   Q

5221 ATTGATTTATTAAAGATCGTATGTTTACTTAATGGATATTTACGTACCCCTAAAATCCAT
 95> I   D   L   L   K   I   V   C   L   L   N   G   Y   L   R   T   P   K   I   H

5281 CAATTTAACATTCTTTTAAATTATTTGAAAACAAAATATTCGGATATATCCTTAAATATA
115> Q   F   N   I   L   L   N   Y   L   K   T   K   Y   S   D   I   S   L   N   I

5341 CAAAAAGTAGATACTTCACCACTCTCTGAAAATGCCTGGTTAGCTGGTTTTATTGACGCT
135> Q   K   V   D   T   S   P   L   S   E   N   A   W   L   A   G   F   I   D   A

5401 GATGGCTGTTTCAAAATACGTTATACAAAAGCAAAACATTGCGTAAACACTGGTAAGTGT
155> D   G   C   F   K   I   R   Y   T   K   A   K   H   C   V   N   T   G   K   C

5461 ATTACAAAAGAACGTCTTGGGTTATCATTTACTATTGAGCAACAAATGATACATCTAAAA
175> I   T   K   E   R   L   G   L   S   F   T   I   E   Q   Q   M   I   H   L   K

5521 ACACAAGAATCTTTTGAACCTATAATGAGTGAAATTGCTAAATTTTTAGATGTTAATTTA
195> T   Q   E   S   F   E   P   I   M   S   E   I   A   K   F   L   D   V   N   L
```

```
5581 AGAATATGTAAACATTTAAAAAAAGGAAGTGTTGTAAACTTTTGGTGTATCGAATTAAGT
215> R  I  C  K  H  L  K  K  G  S  V  V  N  F  W  C  I  E  L  S

5641 AGTTTTCAAAAAATAGATAAGTTAATAAAATATTTAGAAATTAATCACTTATTAACTGTA
235> S  F  Q  K  I  D  K  L  I  K  Y  L  E  I  N  H  L  L  T  V

5701 AAACGAATGAATTATATGGATTGGGTTAAGGCTTATGATATATTTAAACACAATTTGCAT
255> K  R  M  N  Y  M  D  W  V  K  A  Y  D  I  F  K  H  N  L  H

5761 TTAACCGAAAAAGGTAAGAATTCTATTATTAATATTAAAACTCAAATGAATAGTAAACGT
275> L  T  E  K  G  K  N  S  I  I  N  I  K  T  Q  M  N  S  K  R

                         XbaI              5859
5821 ATTGAATATGATTGGAGTTTTCTAGAATGTCCCTATTAAAATAACCTTTTAATAGGAAAT
295> I  E  Y  D  W  S  F  L  E  C  P  Y  *

                                                       HindIII
5881 TGGGTTAATTGTCAAGAATCTCTTATTTAATAAATAAGACAACTTGCAGCGAAGCTTAGC

5941 TTAACACTATAAATGTTATTTATTTAATTATTAACATTTATGACCAATAAAAAGTAAGGT

6001 TTTAATACTTATTTTGTTGGAAAGCTAAGAACGTTCAACGACTAGGAAGTGAGTAGTGTT

6061 AACAATAATCTTCCCACGAACGCCCAACTGTTTTATAAAAACAGATGACATAGTCTAGAC

                                                         6174
6121 TTACTAGTGATAGTAAGAACTGAAATATAAACAATTTCAGGTTAATAATCTCGATGGCTT
                                                      93> M  A

     cox1 exon2
6181 TCCCTCGTTTAAATAATATTAGTTTTTGGTTAAATCCATCTGCTTTAGGCTTATTATTAT
 95>F  P  R  L  N  N  I  S  F  W  L  N  P  S  A  L  G  L  L  L

6241 TGTCTACTATGGTAGAACAAGGTGCTGGTACTGGATGGACTGCATACCCACCATTAAGTA
115>L  S  T  M  V  E  Q  G  A  G  T  G  W  T  A  Y  P  P  L  S

6301 TACAATCAACAGGAGCTTCTGTTGATTTAGCTATATTAAGTTTGCACTTAAATGGTTTAA
135>I  Q  S  T  G  A  S  V  D  L  A  I  L  S  L  H  L  N  G  L

6361 GTTCCATACTAGGAAGCATAAATATTTTAGTAACAATAGCAGGAATGCGTGCTGTTGGTA
155>S  S  I  L  G  S  I  N  I  L  V  T  I  A  G  M  R  A  V  G

6421 TGAAATTGTCTCAAATGCCCTTATTTGTATGGTCCATAGCTTTTACTGCTATTTTAGTAA
175>M  K  L  S  Q  M  P  L  F  V  W  S  I  A  F  T  A  I  L  V

6481 TATTAGCCGTACCTGTATTAGCAGCTGCTTTAGTTATGTTATTAACAGATCGTAATTTAA
195>I  L  A  V  P  V  L  A  A  A  L  V  M  L  L  T  D  R  N  L

6541 ATACTGCATATTTCTGTGAAAGCGGTGACTTAATATTGTATCAACATCTTTTCTGGTTCT
215>N  T  A  Y  F  C  E  S  G  D  L  I  L  Y  Q  H  L  F  W  F

6601 TCGGACACCCTGAGGTTTATATTTTAGTATTACCAGCTTTCGGAATCGTTAGTCATGTTA
235>F  G  H  P  E  V  Y  I  L  V  L  P  A  F  G  I  V  S  H  V

6661 TTAGTTTTTTCAGCCAAAAACCCATTTTTGGTAACATGGGTATGATTTGTGCTATGGGTG
255>I  S  F  F  S  Q  K  P  I  F  G  N  M  G  M  I  C  A  M  G
```

```
                                       6753  cox1i2 orf 358
6721 CCATTAGTATTTTAGGTTTCATTGTATGGGCTCAATTGGGTCTCCTGTCATGTGAATGTC
 275>A  I  S  I  L  G  F  I  V  W  A
                                   1> Q  L  G  L  L  S  C  E  C

6781 AGGCAACATATCTCTGCCATATGCTGGAAACATCTTTAATTTTTAACTCAAATAAAATAT
  10>Q  A  T  Y  L  C  H  M  L  E  T  S  L  I  F  N  S  N  K  I

6841 ATCAAACCACAATCTACTTAGTATTCTATATATATTGTTTAAGTTTTATTGGAAAAATGT
  30>Y  Q  T  T  I  Y  L  V  F  Y  I  Y  C  L  S  F  I  G  K  M

6901 TAAAAATAGATCAATCAGCAGGAAACGGTGTTATATACTGTATTCAAGGTTTAAACACTA
  50>L  K  I  D  Q  S  A  G  N  G  V  I  Y  C  I  Q  G  L  N  T

6961 CAGTAGATTCTCTTAAAGAGTCCTTCGGGTTTGCCGGTTCCTCAGAGACTAGACGCAGAG
  70>T  V  D  S  L  K  E  S  F  G  F  A  G  S  S  E  T  R  R  R

7021 TATCTTTTCTTAGCAACAATTTATCAAAATATGACCCTATTTTTCTTGATTGGTTCATTG
  90>V  S  F  L  S  N  N  L  S  K  Y  D  P  I  F  L  D  W  F  I

7081 GTTTTACAGAAGGCGATGGAGGTTTTTATCATAATATTAAAGACGGACGTTTCTATTATA
 110>G  F  T  E  G  D  G  G  F  Y  H  N  I  K  D  G  R  F  Y  Y

7141 AAATACGTCAAAAAAATCCTAAAGTGTTACTTTATATAAAAAAAAATTTAGGTATAGGAA
 130>K  I  R  Q  K  N  P  K  V  L  L  Y  I  K  K  N  L  G  I  G

7201 CCCTTAAACTAGCTAAAGATAATTATTGGACTTATACAGTAACAGCCATTTCTGATATTG
 150>T  L  K  L  A  K  D  N  Y  W  T  Y  T  V  T  A  I  S  D  I

7261 AAATATTAATAAATATTTTTAATGGCAATCTTCTATTAGAAAAAACTAATTATCGTTTTG
 170>E  I  L  I  N  I  F  N  G  N  L  L  L  E  K  T  N  Y  R  F

7321 TATCAGAGTGGCTTACTCCTTATAACAAAATGTATACAGATAAAGCTATAAAATATTTAG
 190>V  S  E  W  L  T  P  Y  N  K  M  Y  T  D  K  A  I  K  Y  L

7381 GCCCTGGTACCTTTGTAGGGTTAAAGAATGCTTGGTTATGCGGATTTTCTGATGCAGAAG
 210>G  P  G  T  F  V  G  L  K  N  A  W  L  C  G  F  S  D  A  E

7441 GCAGTTGTGGGTTTAAGTTAGTAGCTGATAAAACCCGTAAAAATGGTTATCGTTTACGCC
 230>G  S  C  G  F  K  L  V  A  D  K  T  R  K  N  G  Y  R  L  R

7501 TTTTTTGGTACATTGATCAAACCGATGAAAAAGCTTTTTTTGATAAAATGAAACTGGTTT
 250>L  F  W  Y  I  D  Q  T  D  E  K  A  F  F  D  K  M  K  L  V

7561 TAGGATGGGGTTATATTGAAAAAAAACTTGCTAATGATACATCTTTTAAAGCAGATCCTA
 270>L  G  W  G  Y  I  E  K  K  L  A  N  D  T  S  F  K  A  D  P

7621 ATAAAAAAGCTTGGCGCTTTAAAACAGAAAGTAATCATATTGTTCAACAAATAGTTACCT
 290>N  K  K  A  W  R  F  K  T  E  S  N  H  I  V  Q  Q  I  V  T

7681 ATTTTGATCAATATAATCCACATACTACCAAGCTTTATGTACGTTATATTCGATTACGAC
 310>Y  F  D  Q  Y  N  P  H  T  T  K  L  Y  V  R  Y  I  R  L  R

7741 GTGTATTGAATTGGATAGTTAAAGATGGGTGGCATAGTCGATTGAAAGATATCAGCCATT
 330>R  V  L  N  W  I  V  K  D  G  W  H  S  R  L  K  D  I  S  H
```

7829
7801 TGATTCAGTTAAATAAGCGTTTAAGATAGTAGAATCTCTTTTTAAGAAAAAGATAAAGGT
350>L I Q L N K R L R *

7894 *cox1* exon 3
7861 ATAGTCCATTTAAGTGAATAATTTCACTTATTGCATCATATGTTTACTGTTGGTTTAGAT
286> H H M F T V G L D

7921 TTAGATACAATTGCATATTTTACCTCAGCTACTATGATTATTGCAGTACCCACTGGTATG
295> L D T I A Y F T S A T M I I A V P T G M

7981 AAAATTTTCAGTTGGTTAGCCACTATTTACGGTGGTAGCGTATGGATGACAACACCTATG
315> K I F S W L A T I Y G G S V W M T T P M

8041 TGGTTTGCTGTTGGTTTCATTTGCTTATTTACTATTGGAGGTGTTACCGGTGTCGTGCTA
335> W F A V G F I C L F T I G G V T G V V L

8101 GCTAACGCTGGTATTGACATGTTAGTACATGACACTTATTACGTAGTAGGTCACTTCCAT
355> A N A G I D M L V H D T Y Y V V G H F H

8161 TACGTATTAAGCATGGGAGCTTCCTTTGGTATATTTGCAGGTATTTACTTCTGGTTTGGT
375> Y V L S M G A S F G I F A G I Y F W F G

8221 TTAATGACTGGATTAAGTTACATAGAAAGTCGTGGTCAAGTTCAATTTTGGACCTTATTT
395> L M T G L S Y I E S R G Q V Q F W T L F

8281 ATTGGCGTTAACTTAACTTTCTTCCCTATGCATATGTTAGGTTTGGGCGGGATGCCTCGT
415> I G V N L T F F P M H M L G L G G M P R

8341 CGAATGTTTGATTATGCTGATTGCTTTTATGGATGGAATGCTATTGCCAGTTTTGGTGCT
435> R M F D Y A D C F Y G W N A I A S F G A

8401 TTAATTTCATTCCTATCCATTTTAATGTTAGCAGGCCCAATAAACTTTGTTCCAGAACAT
455> L I S F L S I L M L A G P I N F V P E H

8461 GACACAAAAGCGGCTAATTACCCACGCACTGCTACTACATTAGAATGGTTACAACCATGT
475> D T K A A N Y P R T A T T L E W L Q P C

8572
8521 ACACCAGCAAGTCACGTCTTTACACAATTACCTGTAATACGTAGCTACTAATCATTTTTT
495> T P A S H V F T Q L P V I R S Y *

8608 $tRNA^{trp}$
8581 TTATCTCTTTTTTTAAAAAAAGGATATAGAAAGGTAGCTCAATTAGGTAGAGCATAGGAT

8681
8641 TCCAAATCCTAAGGTTGCAAGTTCAATTCTTGTTCTTTCTGTTCTAATATTCTTAATCTT

8701 GTATATTGGTTTCTATTTGAAATAATTTCAGCAATAAGAATGTAAAACATATTTATTTTG

8820
8761 CTATTATTTTAAATTAATTGTTTTTATACCTAAATAACACATTTTTACGCCATTACTTTT

$tRNA^{gln}$
8821 GGGCTATAGCCAAGCGGTAAGGCACTGGGTTTTGGTCCCAATATCACAAGTTCGAATCTT

```
             8890           8903  nad1
8881 GTTAGCCCAGCTATTATTAATTATGATTATTTTATCTGTTTTAACAATTACAGTACCTGT
                             1> M  I  I  L  S  V  L  T  I  T  V  P  V

8941 ATTATTATCTGTAGCTTTTTTTACCTTAGCTGAACGTCAAATAATGGCAAGTATGCAACG
 14>  L  L  S  V  A  F  F  T  L  A  E  R  Q  I  M  A  S  M  Q  R

9001 TCGTTTTGGACCCCATGTAAGCGGTATTGGTGGTGTTTTACAACCTTTTTGGGACGGTTT
 34>  R  F  G  P  H  V  S  G  I  G  G  V  L  Q  P  F  W  D  G  L

9061 AAAATTAGGGGTAAAAGAACCAATACTACCTTCATTAAGTTCTTATGGCGCTTTTAGTGC
 54>  K  L  G  V  K  E  P  I  L  P  S  L  S  S  Y  G  A  F  S  A

9121 TGCCCCAATGATTAGTTTTATATTAAGCCAAATTTCCTGGTGTGGTATCTTTATATCAGA
 74>  A  P  M  I  S  F  I  L  S  Q  I  S  W  C  G  I  F  I  S  D

9181 TGCTTCTTTTCAGGGTCTAGTTTTAATGGCTTTGAGTTCTTTAGCGGTTTATGGTGTGTT
 94>  A  S  F  Q  G  L  V  L  M  A  L  S  S  L  A  V  Y  G  V  L

9241 ACTTGCTGGTTGGGCTAGCAACAGCAAATATGCTTTTTTAGGGTGTTTGCGCTCAGTTGC
114>  L  A  G  W  A  S  N  S  K  Y  A  F  L  G  C  L  R  S  V  A

9301 TCTAATGGTTTCATATGAGTTGAGCTTAGGAGCTGCTTTATTATCTATTGGCTTATTTCT
134>  L  M  V  S  Y  E  L  S  L  G  A  A  L  L  S  I  G  L  F  L

9361 AACGGACAGTACTGGTATGAAATGTTTATCTTTTTATGATGCGCCGTCTACTGTTCAATT
154>  T  D  S  T  G  M  K  C  L  S  F  Y  D  A  P  S  T  V  Q  F

9421 TGCTTTATTACCTTTATGTCATATTTTTTTGATTTGTATATTAGCTGAAACTAAACGTAT
174>  A  L  L  P  L  C  H  I  F  L  I  C  I  L  A  E  T  K  R  I

9481 ACCTTTCGATTTACCAGAAGCCGAAGCCGAATTAGTAGCTGGTTACAATGTAGAATTTTC
194>  P  F  D  L  P  E  A  E  A  E  L  V  A  G  Y  N  V  E  F  S

9541 ATCCTTAGGATTTGCCTTATTTTTCATAGCTGAGTATGCAAATATGGCCGTAATGAGTGC
214>  S  L  G  F  A  L  F  F  I  A  E  Y  A  N  M  A  V  M  S  A

9601 TTTAGCCTCCATTTACTTTTTAGGTGGTTTTTCTGCTTTAAAAATAACAGCTATATTTTT
234>  L  A  S  I  Y  F  L  G  G  F  S  A  L  K  I  T  A  I  F  F

9661 TGCGTTTGTCTGGACTCGAGGAACCTTACCACGCTATCGTTATGACCAATTTATGCGTTT
254>  A  F  V  W  T  R  G  T  L  P  R  Y  R  Y  D  Q  F  M  R  L

9721 AGGTTGGAAAGCTTATTTACCACTTACACTTGCTATTTTTGCCATAAACGCTTGTTTTGA
274>  G  W  K  A  Y  L  P  L  T  L  A  I  F  A  I  N  A  C  F  D

                        9796
9781 CGTTTTTGTTATTTAATTTAATCTTTTTAAACGGCGTTAATTTAATAATCTTTTATGTTT
294>  V  F  V  I  *

             Repeat
9841 TTATAAACGAAAAAACCAAACTTTAAAATTTTTCACCTTAATGAAAAAGTATAGAAGCGG

9901 AGGGTGAACCCCCTCCTTTCTATTTATATTCAATGGTTGACGTAGACTTTATATTCTCAT

                                                  Repeat
9961 TTAATAAGTATTCTAGCCGCTTGAGCGTCATGGAAAAAATGCATAGAAAAAACCAAACTT
```

10021 TAAAATTTTTCACCTTAATGAAAAAGTATAGAAGCGGAGGGTGAACCCCTCCTTTCTATT

10081 GAGAAAAAATTAAGAATCTATAAAAACAGATGAACTTATAAGCACTGTATACTTGTGTTT

Repeat

10141 CTTATGATTACGCATCAGCTCTGCGTAATCTAATACCAAGTACACGGAGCTGACGCGGAG

10201 CTGACGCGGAGCTGACGCGGAGCTGACGCGGAGCTGACGCGGAGCTGACGCGGAAGTAAA

10261 GTCACAGGTTTTACCATAACTTAATAAGCATATTAGATTTTTATAATAAAAAAAGCAAAG

10321 GTGGGTGAGATAGAAACTAAGCAACGCTGGGTGGGGGAGAGGGGGGGAGGGGGAGAGCGG

10381 GTGTGGGGCTCTCCCCCCCTCTTTAATGCTATCAATAGATTATGCCATTATATACATATA

10441 TGCACCTAGTTTGAACTTGAATTTACAATGAATATGATCCGATAGTAAACGGTTAGAATG

10501 GTATTTTACTTAAAGGGCGGGCGGGGGTTAGGGGGTGGTGGTTCCGTCCCGGAAGGGAAA

10561 CCCCCTAGTAGAACATAGAATAGAGTTACTTTAGGCATTCAAGGTGGCTTAAAACTAAAG

10621 TAGCAACTAAATTGATAAATCAAAATGTTTTTGCACTACAAAAGCCATTGAATATACAAG

10713 *nad5* exon1

10681 GTTTTGATGAACAAAGTCTATAAGAATTTAAAATGTATTTAATTCCTTTGTTGGCTACAT
1> M Y L I P L L A T

10741 TGATGAGTAGTATATTCGCAGGTGCTTTACCTATAACAGCTCGAAATTTAGGGCATCGTG
10> L M S S I F A G A L P I T A R N L G H R

10801 GAGTTGCCGTCTTTTCCATAATAAGCCTAGTAATTGCTTTTCTAAGCAGTGCTTTAATTT
30> G V A V F S I I S L V I A F L S S A L I

10861 GGATTGATTTGTATATAGGGTCTTCACCTGTCTGGTTGGATTTATTTGGTGCTTGGTTTG
50> W I D L Y I G S S P V W L D L F G A W F

10921 AAGTAGGAACTGTAACAGTTTCATGGGTGTTCTATTATGATTTGTTAACGGCCCATATGT
70> E V G T V T V S W V F Y Y D L L T A H M

10981 TGTTTACGGTGACTAGTGTAAGCTTAGCAGTACACATTTATGCTGTTGTATATATGCGTA
90> L F T V T S V S L A V H I Y A V V Y M R

11041 GTGATCCACATCTGACATTATTTATGTCATATCTTTCATTATTTACGTTTTTTATGTTAG
110> S D P H L T L F M S Y L S L F T F F M L

11151

11101 TATATGTTTGTGGTGATAATTTGTTAGTAATGTTAGTCGGTTGGGAAGGTAAAATACATT
130> V Y V C G D N L L V M L V G W E G
1> K I H

nad5i1 orf229

11161 GCCTTAATGGTTTTGACAAAGAAATAATTATTTATGCAGGTGTTAGTATTCCAGGTAATA
4> C L N G F D K E I I I Y A G V S I P G N

11221 GTCGAAAAGGGCCACATAGTAGTTTATTTAAACAAATAATGGTGGGAGGTTTATTAGGAG
24> S R K G P H S S L F K Q I M V G G L L G

```
11281 ATGGTTGGTTGGAAAAACATGGAGCAGGAGCACGTTTTGGAATGTCATTTAAACATACTT
  44>D  G  W  L  E  K  H  G  A  G  A  R  F  G  M  S  F  K  H  T

11341 ATAAAGATGTTGCTAATTGGTATCAGTTTATGTTGTATGCGTTAGGTTATCATCATAAAC
  64>Y  K  D  V  A  N  W  Y  Q  F  M  L  Y  A  L  G  Y  H  H  K

11401 TAAGTGTTGATGAGCCGTTGGAACGTATAACGAAACAAGGTAAAATAAGTAATTATTACC
  84>L  S  V  D  E  P  L  E  R  I  T  K  Q  G  K  I  S  N  Y  Y

11461 AAGTGCGCACATTGACTTTTAACAGTCTATTGAAATATTATAACCTTTGGTATGTTAAAG
 104>Q  V  R  T  L  T  F  N  S  L  L  K  Y  Y  N  L  W  Y  V  K

11521 TAGATGGACGTCGTCAAAAAGTAATACCACGTAATTTAGAAAATGATTTGACACCAATAA
 124>V  D  G  R  R  Q  K  V  I  P  R  N  L  E  N  D  L  T  P  I

11581 GCTTAGCTTTGTGGTTGATGGGCGACGGTTCAGGTATGCGAGACGGTGGTTTTAAAATAG
 144>S  L  A  L  W  L  M  G  D  G  S  G  M  R  D  G  G  F  K  I

11641 CGACACATTCATTTTCAATAGAAGACAATTTATATTTAATTGATTTATTAAAAGAAAAGT
 164>A  T  H  S  F  S  I  E  D  N  L  Y  L  I  D  L  L  K  E  K

11701 ATGGATTAAAAGCAAGCTTGCATAAAGATGGCAATAAAGTATGTATTTATATATGGAAAC
 184>Y  G  L  K  A  S  L  H  K  D  G  N  K  V  C  I  Y  I  W  K

11761 AATCGGTTCCTAAATTAAAGGCTATAGTATTACCATTTTTTCAAGAGTCATGTTTGTATA
 204>Q  S  V  P  K  L  K  A  I  V  L  P  F  F  Q  E  S  C  L  Y

                          11840
11821 AATGGCGTCATGTAAAATAATAGTTCTTTTTGTCAAAAATCAAGTCTTATTTGTAAAAAC
 224>K  W  R  H  V  K  *

11881 GGTATGAAAACACATAAAATAAAAAACGTTAAAAATTTTAACAAAATATATAATTTTGTT

11941 AAATAATACAAATATTAAAACCTGACGGGGTTTTAACTTAAATTCTTTAAAATGTTATGT

12001 TTTGTGAGTTTAAGAGGCGACATTGGTGAAAACGATTAAAGTCTTTAAGACAAGATCGTC

12061 GGTTTTTTTTATGAAACCGCGACAGACTGGGTCACTGATGTGTGTCTGAAATGATGCATA

12121 ATGTACAGTCGATATATCTTTTTCAAAAGTCGCGTAAGCGGCTAGCTTTTTTTTTAAAAA

                                  12211  nad5 exon 2
12181 AAATAAAAGCTTTTTCAAAACAAGATATGGATTGGTGTTTGTTCCTATTTATTAATAGGT
                             147> I  G  V  C  S  Y  L  L  I  G

12241 TATTATTCACATCGTTTAGCGGCTGTTAAAAGTGCTCAGAAAGCTATTTTAGTAAATCGT
 157> Y  Y  S  H  R  L  A  A  V  K  S  A  Q  K  A  I  L  V  N  R

12301 GTTAGTGATGGCATGTTACTTTGGGGTGTGTTGTGGATTTGGTATTATGCCGGTAGTTTA
 177> V  S  D  G  M  L  L  W  G  V  L  W  I  W  Y  Y  A  G  S  L

12361 GAGTATGACTTAGTTTTGTTAAATCAAACATCAAGTATTAGTATGTTTATTGTCTTAAGT
 197> E  Y  D  L  V  L  L  N  Q  T  S  S  I  S  M  F  I  V  L  S

12421 GTATTAGTAGGTGCTATGGGTAAAAGTGCACAGATTTTGTTCCATGTATGGTTAGCAGAT
 217> V  L  V  G  A  M  G  K  S  A  Q  I  L  F  H  V  W  L  A  D
```

```
                   12493  nad5i2  orf 296
12481 GCAATGGAGGGTTTGAATTACATAAAAAAGTTTCTAGTTATGTTATCAGGTTGGGTTATG
 237> A  M  E  G
                 1> L  N  Y  I  K  K  F  L  V  M  L  S  G  W  V  M

12541 TCCGTCTGGTCTAAAATTATTTTTGATTTGATCTTGTGTTTTATTTGGGTGCTGTATTGG
  17> S  V  W  S  K  I  I  F  D  L  I  L  C  F  I  W  V  L  Y  W

12601 GAGTGGGCTTGCTACCCCTGGTTAAATTTTTGTCAGGATTTTGTAGTCTGGGTTATTATG
  37> E  W  A  C  Y  P  W  L  N  F  C  Q  D  F  V  V  W  V  I  M

12661 GATTCTCGTGATTTTTCTTATCAAACACTTGACTTTAGTAATAATTTAGTACTATTTGCT
  57> D  S  R  D  F  S  Y  Q  T  L  D  F  S  N  N  L  V  L  F  A

12721 GCCTTTACAGGTCGAAAGAATGGTACTCAGGTAGCCACCCCTTATCAGTTAGAGGTTATT
  77> A  F  T  G  R  K  N  G  T  Q  V  A  T  P  Y  Q  L  E  V  I

12781 GCGGGTTTGTTACTATCAGATGGTTGCTTGCGTAATCCTAATAGTAATAAGCGTAGTACT
  97> A  G  L  L  S  D  G  C  L  R  N  P  N  S  N  K  R  S  T

12841 GGAAATTACCGCTTAGAGTTTACTTTTAAGTCTCCAGTATACGATTATATTACCTGGCTT
 117> G  N  Y  R  L  E  F  T  F  K  S  P  V  Y  D  Y  I  T  W  L

12901 AAATTTGATGTTTTGGGTAGTTTGTGTACAGATTCTTTGCCTACCCCTTATCCAAAGATC
 137> K  F  D  V  L  G  S  L  C  T  D  S  L  P  T  P  Y  P  K  I

12961 AATCCAAACCAATATTGGTTTGCCAGTCGAAGTATGTCCTTATTTACAGAATTGAACGAG
 157> N  P  N  Q  Y  W  F  A  S  R  S  M  S  L  F  T  E  L  N  E

13021 GTTTGGTATACCGTTATTGATAAAAAACGCGTTAAGGTTGTTCCAAGTAATCAAGTACTT
 177> V  W  Y  T  V  I  D  K  K  R  V  K  V  V  P  S  N  Q  V  L

13081 TCACCTTTGTTTACACCTATTTGTTTGGCTCACATGATAATGGGCGACGGTTATTGGGAT
 197> S  P  L  F  T  P  I  C  L  A  H  M  I  M  G  D  G  Y  W  D

13141 AACGATAGTAACACTATTCTTCTTTGTACTGAATGTTATACCAAGGAAGAAGTATTGCGT
 217> N  D  S  N  T  I  L  L  C  T  E  C  Y  T  K  E  E  V  L  R

13201 TTGATTGTTCTCTTGGACACTTGTCTAGGTATAAAAGCAACTTTAAAACGTCGTGTTTCT
 237> L  I  V  L  L  D  T  C  L  G  I  K  A  T  L  K  R  R  V  S

13261 GATAAAGGTATTATTAACTATCGTATTCGTATTAGCGGTGCTGCTGCTAATTTAGCGCTT
 257> D  K  G  I  I  N  Y  R  I  R  I  S  G  A  A  A  N  L  A  L

13321 ATTCGAGCTTTGGTAAAGCCACATATGCACCCTAGTATGTTCTATAAATTGGGAATAATT
 277> I  R  A  L  V  K  P  H  M  H  P  S  M  F  Y  K  L  G  I  I

         13383
13381 TAAGCTAATTAGCTTTCTTATAAAATTTTGGTTAAAAATAATTTTTATTATCTAGTTACT
     > *

13441 ATGTATAATAAATAGGGCCCTCCCTAAATTTCGCTGTATGCTGGAACACCCTAAAGCTTG

13501 GATTACGTATATTTTAATATACCCGTGAAAACATTCAAGATATAACAATGGGCAATCAGC

13561 AAGAAACCAAACTATCTACTCATTTCAATCAAGGCTTTTTAGCCTTAATTTTTGTAGGTA
```

```
13621 GATAGAAGTAGGATCTTCAGAGACTATGCGCGAAAATACCCTAAAAATGACGAAGTGATT

13681 GCTATTTATAGCATTAAGATTAGGGTATATGATATAGTACGACTCTTTACGAAAGTTTAG

               13750  nad5 exon 3
13741 AGATTATTGCCTACCCCTGTGTCTGCTTTAATACATGCAGCTACATTGGTTACTGCAGGA
       241> P  T  P  V  S  A  L  I  H  A  A  T  L  V  T  A  G

13801 GTTTATTTAATGGTGCGCTTAGGGCCTTTTATGGCTGGATCTGATTTGGTGATTTTAATT
 258> V  Y  L  M  V  R  L  G  P  F  M  A  G  S  D  L  V  I  L  I

13861 GGTAGTTTAACTGCTTTTATGGCTGGAATTTTTGGTTTTTTTCAAGCCGATTTAAAACGT
 278> G  S  L  T  A  F  M  A  G  I  F  G  F  F  Q  A  D  L  K  R

                                        13957  nad5i3 orf267
13921 GTAATTGCTTTTAGTACTTGCAGTCAATTAGGGTGGAATAGTCAAAAAAATACATTAAAT
 298> V  I  A  F  S  T  C  S  Q  L  G  W
                                     1> N  S  Q  K  N  T  L  N

13981 TATAATAAAATGATTTCTTCAAATAATCTTTCTTATTTCAATGAAATAAATGTACGTAAC
   9> Y  N  K  M  I  S  S  N  N  L  S  Y  F  N  E  I  N  V  R  N

14041 TATTCAACAAAAATTGAAATAGAGTGCGATATATTACCAATAACATTTGTAGATAAGTTT
  29> Y  S  T  K  I  E  I  E  C  D  I  L  P  I  T  F  V  D  K  F

14101 GAGACTATTGTGGATGATCAAAAATATAATATAAAAAGTAAATATAAAAAAGTGGCCGTA
  49> E  T  I  V  D  D  Q  K  Y  N  I  K  S  K  Y  K  K  V  A  V

14161 ATTTATTTATGGTATAATAAAGTTAACAACAAGTGTTATGTAGGTCGTTCAACAAATTTA
  69> I  Y  L  W  Y  N  K  V  N  N  K  C  Y  V  G  R  S  T  N  L

14221 GCCTCACGTTTGGAAAATTATTTTCGTGTTAAATATTTGAATGATATGAAAAATAAGATG
  89> A  S  R  L  E  N  Y  F  R  V  K  Y  L  N  D  M  K  N  K  M

14281 CCAATATGTAGCGCTTTATTGAAATATGGTTTAGATAATTTTATTTTATATGTACTTGAA
 109> P  I  C  S  A  L  L  K  Y  G  L  D  N  F  I  L  Y  V  L  E

14341 ATAATACCAACGGAGAATATAACAAAGCTACCAGAACGCGAGGATTATTATGTTTCAATA
 129> I  I  P  T  E  N  I  T  K  L  P  E  R  E  D  Y  Y  V  S  I

14401 GTTAAACCTAGTTATAACATAGCTAAAATAATAGACCAGTTTGTGGGTGCAAATCATCCC
 149> V  K  P  S  Y  N  I  A  K  I  I  D  Q  F  V  G  A  N  H  P

14461 CGTTATGGTAAAGTTATTTCACAAGAAGTGCGTGAAAAAATAAGTAAAGCATTAACTGGG
 169> R  Y  G  K  V  I  S  Q  E  V  R  E  K  I  S  K  A  L  T  G

14521 CGTACACTAACAAAGTTGGAAATAGAAAATCATCGTAAAGGTGCTCGAAAAAAAGTAGTA
 189> R  T  L  T  K  L  E  I  E  N  H  R  K  G  A  R  K  K  V  V

14581 TATTGTTATGATGTTACGTCCAAAAAGTTAGTAACCACTTTTGAATCAATGCGCGCTTTA
 209> Y  C  Y  D  V  T  S  K  K  L  V  T  T  F  E  S  M  R  A  L

14641 AGTCGTCAAATGAATATAAATCGAGGTATACTTTATCGAACTATTGATAAAAATAAACCA
 229> S  R  Q  M  N  I  N  R  G  I  L  Y  R  T  I  D  K  N  K  P

14701 ATTTATGTGAAATTTCAAGATAAAGAGTGTGCTTGGCTTTTATATTATAAACCTATTTAA
 249> I  Y  V  K  F  Q  D  K  E  C  A  W  L  L  Y  Y  K  P  I  *
```

14761 TTTAATGTATAAATAAATTCATATGCTCTACTTTAGTGTAAACTAAAGATATAAAATCTC

14821 ACTGTATGCTGGAAAATCCTAAAGCTTATTTTGTTCTTGTTTTTCATAATAAAAGGTGGG

14881 AAATAAAGTAAGATATTTTTTTTTAAAGGTCGCACAAGCGGCTTTACTGGAGGTCTTCA

14941 GTCCCTCAGGTAAGAAAATAATGGACAATCAGCAGGAAACCAATTTAAATGTTTAATAGT

15010

15001 AGGATCCTCG

*Bam*HI boundary of 20 kb LambdaGem-11® left arm

respiratory chain proteins (*cob*, *cox 1*, *nad1* and partial fragments of *nad2* and *nad5*), four tRNAs (one, *tRNA met2*, is a possible pseudogene), and discontinuous LSU and SSU rRNAs. In addition, seven open reading frames (ORFs) were identified within six introns interrupting the coding regions of *C. acidophila* mtDNA. In Addition, an intergenic region containing a number of repetitive elements and an elevated G + C content maps between positions 9794-10712. Finally, all genetic elements are encoded on one strand of *C. acidophila's* mtDNA. The specific details of each element are discussed below.

Protein Coding Genes

Nad 2

The first gene in the F3 molecule is a portion of the *nad2* gene. The rest of *nad 2* was probably lost during the cloning procedure. Interestingly, the intended *Bam*HI ligation site at this end has been lost (data not shown). Similarity searches with BLASTX revealed strong sequence homology with other Nad2 proteins. The region of DNA that corresponded to the *nad 2* gene mapped between nucleotide positions 3-788 (Figure 4) with the last three forming the ochre (TAA) stop codon. The DNA sequence of this region was converted to an amino acid sequence, albeit incomplete because of the amino-terminal truncation of the DNA molecule during the cloning procedure. The amino acid sequence was compared against the NCBI database with the program BLASTP. The sequence, 261 amino acids in length, shows a high degree of similarity with the carboxy terminal end of other Nad 2 proteins (Table 3), most notably to the *C. eugametos* Nad 2 counterpart. Hydrophobicity studies revealed this protein to contain at least five transmembrane segments. The five membrane spanning segments correspond to amino

Table 3. Comparison of deduced NADH dehydrogenase subunit 2 (Nad 2) protein.

Organism	Length in amino acids	Identity (%) with *C. acidophila*	Similarity (%) with *C. acidophila*
Chlamydomonas acidophila	261*		
Chlamydomonas eugametos	496	79%	91%
Chlorogonium elongatum	454	50%	67%
Chlamydomonas reinhardtii	382	52%	65%
Prototheca wickeramii	510	31%	48%
Chondrus crispus	497	26%	47%
Marchantia polymorpha	489	28%	44%
Reclinomonas americana	498	26%	48%
Rickettsia prowazekii (NuoN1)	458	26%	45%

**nad 2* gene of *C. acidophila* is a partial sequence.

Table 4. Comparison of deduced apocytochrome b (Cob) protein.

Organism	Length in amino acids	Identity (%) with *C. acidophila*	Similarity (%) with *C. acidophila*
Chlamydomonas acidophila	380		
Chlamydomonas eugametos	380	94%	98%
Chlorogonium elongatum	385	76%	87%
Chlamydomonas reinhardtii	381	73%	85%
Chlamydomonas smithii	381	73%	85%
Prototheca wickeramii	384	58%	73%
Marchantia polymorpha	404	59%	71%
Chondrus crispus	381	53%	72%
Reclinomonas americana	390	54%	69%
Rickettsia prowazekii (PetB)	398	48%	67%

acid residues 60-80, 100-130, 134-154, 169-189, and 238-358; other possible transmembrane segments map to amino acid residues 35-55 and 87-107.

Cob

Downstream of the *nad2*-coding region is a putative apocytochrome b (*cob*) coding gene. The gene is 2170 nucleotides (nucleotides 812-2981, Figure 4) and terminates with an ochre (TAA) codon, encoding a protein 380 amino acids in length. Apocytochrome b is one of the more conserved proteins across taxa as is evident from Table 4, with the closest match to *C. acidophila* being *C. eugametos* (94% identity and 98% similarity). Analysis of the Cob protein reveals that it contains at least seven transmembrane spanning segments. The seven membrane spanning segments correspond to amino acid residues 33-53, 112-132, 138-158, 181-201, 229-249, 289-309, and 353-373; in addition, two potential transmembrane segments mapped to amino acid residues 35-55 and 87-107.

Histidine residues are known to coordinate iron molecules for electron transfer, an important function of the Cob protein, and are highly conserved (Esposti *et al*., 1993). Multiple sequence alignments of the *C. acidophila* Cob protein against other known sequences (data not shown) indicate that the histidine residues at amino acid positions 82 and 197 coordinate one iron molecule. A second iron molecule is bound by the two histidines at positions 96 and 183. The only other conserved region for apocytochrome b is the P-E-W triplet, which seems to be important for electron transfer at the ubiquinone redox site located outside the mitochondrial membrane (Esposti *et al*., 1993). This triplettypically exists in the loop that separates the fifth and sixth transmembrane segments (Esposti *et al*., 1993). In the *C. acidophila* Cob protein this triplet maps to

positions 271-273 of the amino acid sequence. This region lies between the proposed transmembrane domains at positions 229-249 and 289-309, indicating the putative transmembrane domain between amino acids 80-100 may not traverse the membrane.

Cox 1

The region of the F3 molecule mapping between positions 4663-8571 (Figure 4) appears to code for the cytochrome oxidase subunit one gene (*cox 1*). The *cox 1* exons are fragmented into three segments interspersed by two intronic regions (Figure 3 and 4), and encodes for a protein 510 amino acids in length ending in an Ochre (TAA) stop codon. The *C. acidophila* protein shows high similarity to other Cox 1 proteins (Table 5) with *C. eugametos* being the closest (93% identity, 96% similarity). The protein contains 12 transmembrane segments and one possible segment. The twelve transmembrane segments are located at amino acid residues 16-36, 60-80, 100-120, 150-170, 180-200, 240-260, 266-286, 297-317, 329-349, 380-400, 409-429, and 448-468; the possible transmembrane segment maps to amino acid positions 123-143.

Cytochrome oxidase is responsible for generating a transmembrane proton gradient. Cox 1 coordinates a heme (cytochrome a) and a bimetallic cytochrome a_3/Cu_B for its active site (Castresana *et al.*, 1994; Garcia-Horsman *et al.*, 1994; Saraste, 1999; Saraste & Castresana, 1994). Six conserved amino acid residues are involved in the binding of these three metals. Multiple sequence alignments (data not shown) with known Cox 1 proteins reveals that the cytochrome a heme is probably bound by two histidine residues at amino acid positions 62 and 374. A histidine residue at amino acid position 372 binds cytochrome a_3 of the bimetallic active site, while the other copper- containing

Table 5. Comparison of deduced cytochrome oxidase subunit 1 (Cox 1) protein.

Organism	Length in amino acids	Identity (%) with *C. acidophila*	Similarity (%) with *C. acidophila*
Chlamydomonas acidophila	510		
Chlamydomonas eugametos	509	93%	96%
Chlorogonium elongatum	510	82%	91%
Chlamydomonas reinhardtii	505	79%	87%
Prototheca wickeramii	515	61%	75%
Chondrus crispus	532	61%	76%
Marchantia polymorpha	522	62%	75%
Reclinomonas americana	531	62%	75%
Rickettsia prowazekii (CoxA)	534	60%	73%

Table 6. Comparison of deduced NADH dehydrogenase subunit 1 (NAD 1) protein.

Organism	Length in amino acids	Identity (%) with *C. acidophila*	Similarity (%) with *C. acidophila*
Chlamydomonas acidophila	297		
Chlamydomonas eugametos	295	92%	96%
Chlorogonium elongatum	296	78%	88%
Chlamydomonas reinhardtii	292	83%	92%
Prototheca wickeramii	328	51%	65%
Chondrus crispus	326	49%	67%
Marchantia polymorpha	328	50%	64%
Reclinomonas americana	333	47%	66%
Rickettsia prowazekii (NuoH)	339	48%	66%

cytochrome is bound by three histidines at amino acid positions 286, 287 and 372. Previous investigations have also shown that one of the ligands to this copper, a tyrosine residue, forms a covalent linkage (Tsukihara *et al.*, 1995). In *C. acidophila* this tyrosine residue appears to be at position 241.

Nad 1

Nad 1 is part of the NADH dehydrogenase complex and has a homolog in the *C. acidophila* mtDNA-coding region between nucleotides 8903-9796 (Figures 3 and 4). This region is 891 nt in length, ending with an ochre (TAA) stop codon, and codes for a protein 297 amino acids in length. This protein has a strong similarity with other Nad1 proteins in the database (Table 6), with the *C. eugametos* Nad 1 being the most notable (identity 92%, similarity 96%). Since Nad 1 is purported to be a transmembrane protein, hydrophobicity analyses were performed. The protein has six strong candidates for transmembrane domains and one potential segment. The transmembrane domains correspond to the amino acid residues 72-92, 98-188, 137-157, 169-189, 226-246, 277-297; and the potential transmembrane segment maps to positions 203-223.

Nad 5

The last protein coding region flanks the LambdaGEM-11® vector left arm. Sequence comparisons indicate that this region (Figures 3 and 4; nt 10713-15010) codes for the N-terminal domain of the *nad 5* gene. This 927 bp fragment encodes a partial Nad5 protein 309 bp in length, as determined by BLASTP, with *C. eugametos* showing the strongest similarity (identity=83%, similarity=93%; Table 7). Hydrophobicity studies

Table 7. Comparison of deduced NADH dehydrogenase subunit 5 (Nad 5) protein.

Organism	Length in amino acids	Identity (%) with *C. acidophila*	Similarity (%) with *C. acidophila*
Chlamydomonas acidophila	309*		
Chlamydomonas eugametos	576	83%	93%
Chlorogonium elongatum	544	69%	83%
Chlamydomonas reinhardtii	567	65%	77%
Prototheca wickeramii	689	44%	62%
Chondrus crispus	666	46%	61%
Marchantia polymorpha	669	40%	57%
Reclinomonas americana	670	44%	62%
Rickettsia prowazekii (NuoL1)	653	42%	60%

**nad 5* gene of *C. acidophila* is a partial sequence.

of *C. acidophila's* Nad 5 give strong evidence for eight transmembrane domains. The transmembrane segments are located at amino acid residues 31-51, 84-104, 113-133, 136-156, 177-197, 205-225, 242-262, 273-293.

Codon Usage

Table 8 compares the codon usage pattern of *C. acidophila* to three closely related algal species (*Chlamydomonas eugametos*, *Chlorogonium elongatum*, and *Chlamydomonas reinhardtii*). The table takes into account only those protein-coding regions that could be determined with the *C. acidophila* cloned mtDNA. Therefore, care was taken to survey only similar coding regions from the other species, which include apocytochrome b (*cob*), cytochrome oxidase subunit 1 (*cox 1*), NADH dehydrogenase subunit 1 (*nad 1*) and partial sequences of NADH dehydrogenase subunits 2 and 5 (*nad 2* and *nad 5*). In the case of *nad 2* and *nad 5* multiple sequence alignments were performed to obtain only those codons for which a counterpart had been sequenced in *C. acidophila*. Previous investigations have shown that the mitochondrial genome of both *C. eugametos* (Denovan-Wright *et al*., 1998), and *C. reinhardtii* (Boer & Gray, 1988; Michaelis *et al*., 1990) use the standard genetic code and the same was assumed for both *C. elongatum* and *C. acidophila*. While a detailed picture of codon usage cannot be obtained from this data, biases can be explored. Indeed, codon tendencies for this partial set are very similar to previous studies of the Chlamydomonads taking into account all protein coding units (Denovan-Wright et al., 1998). Among all codons, the only one not used among all four algae is the opal termination codon (TGA), which is in agreement with complete codon

Table 8. Codon usage table

Cod	AA	Cac	Ceu	Cel	Cre	Cod	AA	Cac	Ceu	Cel	Cre
UUU	Phe	106	109	63	46	UAU	Tyr	59	60	55	24
UUC	Phe	29	23	79	82	UAC	Tyr	14	13	27	46
UUA	Leu	193	204	212	0	UAA	ter	4	3	4	3
UUG	Leu	39	21	0	147	UAG	ter	0	1	0	1
CUU	Leu	8	8	22	16	CAU	His	32	34	20	13
CUC	Leu	1	2	11	1	CAC	His	7	8	22	37
CUA	Leu	12	16	16	71	CAA	Gln	35	34	34	29
CUG	Leu	2	2	0	10	CAG	Gln	4	5	0	9
AUU	Ile	89	57	92	91	AAU	Asn	43	45	18	2
AUC	Ile	8	11	28	36	AAC	Asn	8	11	32	41
AUA	Ile	51	89	1	0	AAA	Lys	24	24	31	25
AUG	Met	74	76	64	72	AAG	Lys	2	2	0	1
GUU	Val	52	31	58	42	GAU	Asp	30	27	29	22
GUC	Val	9	9	22	32	GAC	Asp	10	11	11	17
GUA	Val	71	91	80	74	GAA	Glu	24	26	29	0
GUG	Val	16	13	6	3	GAG	Glu	5	3	2	28
UCU	Ser	25	19	25	33	UGU	Cys	16	17	12	7
UCC	Ser	10	9	4	10	UGC	Cys	3	4	3	14
UCA	Ser	24	39	16	1	UGA	ter	0	0	0	0
UCG	Ser	0	1	2	1	UGG	Trp	49	45	41	45
CCU	Pro	42	22	23	8	CGU	Arg	31	21	33	27
CCC	Pro	6	7	2	3	CGC	Arg	7	3	10	12
CCA	Pro	23	40	45	55	CGA	Arg	4	7	2	1
CCG	Pro	2	2	1	4	CGG	Arg	0	1	0	0
ACU	Thr	37	19	53	60	AGU	Ser	54	69	66	42
ACC	Thr	10	13	21	33	AGC	Ser	21	11	32	39
ACA	Thr	38	49	23	0	AGA	Arg	0	11	0	0
ACG	Thr	5	3	0	0	AGG	Arg	0	0	0	0
GCU	Ala	88	50	83	110	GGU	Gly	91	66	101	118
GCC	Ala	20	22	31	53	GGC	Gly	9	8	17	26
GCA	Ala	30	54	25	1	GGA	Gly	30	53	13	3
GCG	Ala	7	12	3	2	GGG	Gly	12	9	13	0

Comparison of codon usage with *C. acidophila* sequenced coding regions. Codons represent those taken from *cob*, *cox 1*, *nad 1*, and partial sequences from *nad 2* and *nad 5* in *C. acidophila* (Cac), *C. eugametos* (Ceu), *C. elongatum* (Cel), and *C. reinhardtii* (Cre) mitochondria. Cod, codon. AA, amino acid.

tables for *C. reinhardtii* and *C. eugametos*. While most codons are used at similar frequencies in these algae, some show distinct biases across species lines. Some examples include UUA (Leu), UUG (Leu), AUA (Ile), ACA (Thr), GCA (Ala), and GAA (Glu).

The *C. acidophila* mtDNA sequenced fragment appears to only encode three functional tRNAs, despite the fact that from this partial sequence all but six (including two stop codons) of the amino acid specifying codons are used in protein-specifying genes. These data indicate that the remaining tRNAs may be imported from the nucleus.

Introns and Intron Encoded Reading Frames

Introns are mobile genetic elements that typically encode their own maturases and nucleases. Defining intronic groups is a daunting task because of the instability of their nucleotide sequences. While introns may show little nucleotide similarity, their secondary and tertiary structures are conserved (Lambowitz & Belfort, 1993). Unfortunately, secondary structures are rarely published, probably owing to the difficulty in determining the intronic core sequence base-pairing interactions. However, intronic families may be inferred by two criteria: (1) Intronic open reading frames encoding maturases and/or endonucleases; and (2) intron insertion sites.

Intronic open reading frames, the first criterion, have been shown to encode maturases and endonucleases. Maturases have been implicated in aiding introns in splicing reactions (Lambowitz & Belfort, 1993; Lambowitz & Perlman, 1990), but conserved consensus sequences have not been identified. Many maturases contain endonucleases that target intronless alleles, giving these proteins a dual function. In

contrast to maturases, endonucleases display a well defined motif. The two main families of nuclease motifs are the LAGLIDADG and GIY…YIG motifs. These enzymes are named by their acronyms that correspond to the genus and species from which they were isolated, preceded by the letter 'I' to denote an intron (Dujon *et al.*, 1989). I-*Sce*I, for example, corresponds to the first intron endonuclease to be discovered from a *S. cerevisiae* mitochondrial intron (Colleaux *et al.*, 1986). Because maturases and endonucleases serve specific functions in RNA splicing and mobility, their amino acids tend to be conserved.

In discussing intron open reading frames (ORF) in the results below it should be noted that intronic ORFs are listed as either free standing (FS) or in frame (IF) with the 5′ exon sequence. FS ORFs are not in frame with the 5′ exon coding sequence as typified by the *cobi1 orf112*, which is in a +1 reading frame with respect to the 5′ exon sequence. The IF ORFs are given in amino acid lengths commensurate with their starting position at the 5′ exonic juncture. The reason for displaying this protein fusion is because all of *C. acidophila's* IF ORFs show similarity with yeast maturases. Maturases of group I and group II introns are typically in frame with their upstream exons. Previous investigations have shown that an active maturase may be synthesized as an exon/intron fusion protein (Banroques *et al.*, 1987; Carignani *et al.*, 1983; Weiss-Brummer *et al.*, 1982). This mode of synthesis presumably results in feedback regulation, such that a slower rate of splicing leads to the production of more maturase, which in turn promotes splicing (Lazowska *et al.*, 1980).

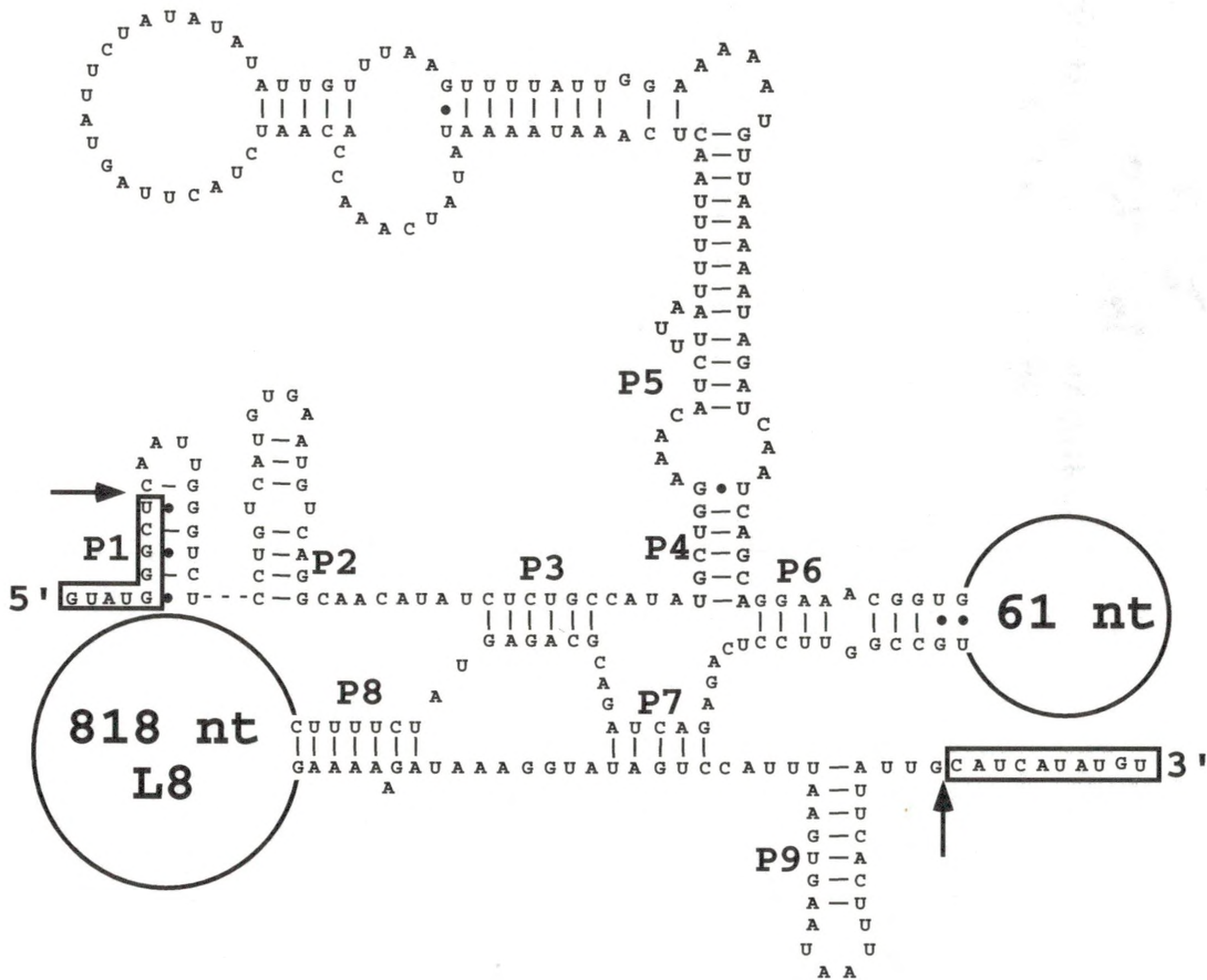
The second criterion for intron group inference is their site of insertion within the corresponding exon sequence. Figure 5 shows the putative secondary structure for the *C.*

acidophila cox1i2 intron, which is similar to the group IB2 class of introns. For the purpose of insertional criteria, note the P1 loop. The boxed region corresponds to the 5′ exon region. This sequence base pairs with a complementary region, also known as the internal guide sequence (IGS), of the intron. This interaction forms the P1 loop. It is noteworthy to this study that intron splice-site recognition relies on this P1 pairing (Lambowitz & Belfort, 1993). Therefore, the determination of insertion site is important because group I introns tend to insert into sequences that are capable of folding with the intron to form a P1 loop.

The structure in Figure 5 was constructed by multiple sequence alignments with similar intronic sequences and visual adjustment of the nucleotide sequence data. Once sequences corresponding to the conserved intronic core were determined (P3-P8) a secondary structure of the entire RNA molecule was predicted with the help of the MULFOLD RNA folding program (Jaeger *et al.*, 1989; Jaeger *et al.*, 1990; Zuker, 1989). In addition, all structural elements are labeled by standard conventions according to Burke et al. (Burke *et al.*, 1987) .

Finally, intron and intron encoded protein nomenclature is as follows. Introns are named according to their position 5′ – 3′ in the corresponding coding strand. For example, the second intron of the *cox 1* gene is referred to as *cox 1i2*. Internal intron open reading frames are named according to their amino acid length. The reading frame of the *cox 1i2* is *orf358*. Initial referrals to intron open reading frames in this text are listed as *cox 1i2 orf358*, which is then truncated to *orf358* for the remainder of the

Figure 5. Predicted secondary structure of *cox1i2* of *C. acidophila*.
Nomenclature of structure was assigned according to Burke et al. (Burke *et al.*, 1987). Catalytic core region was searched against the data provided by Michel and Westhoff (Michel & Westhof, 1990). Boxed sequences represent the adjacent exon. Arrows indicate splice sites. Internal guide sequence is located at P1. Catalytic core is P3-P8.


P1
P2
P3
P4
P5
P6
P7
P8
P9
5'
3'
818 nt
L8
61 nt

chapter heading. Additional references to this protein in the rest of this work will carry the full label *cox1i2 orf358*.

The partial sequence of *C. acidophila* reveals the presence of six introns. Based on sequence homology all six were classified as group I introns. One intron (*cobi1*) was present within the apocytochrome b (*cob*) coding region, two (*cox1i1* and *cox1i2*) in the cytochrome oxidase subunit one (*cox1*) gene and at least three introns (*nad5i1*, *nad5i2* and *nad5i3*) exist in the NADH dehydrogenase subunit five (*nad5*) region. A map of all introns is located in Figure 3 and complete sequences are marked in Figure 4.

Cob i1

The *cob i1* intron is 1030 bp in length and is inserted between the exon sequences of *cob* at positions 1240-2271. The insertion shows the highly conserved 'U' residue of the 5′ exon sequence immediately preceding the intron, and the conserved 'G' residue at the last 3′ position of the intron immediately preceding the 3′ exon nucleotides. A search for other organisms with identical insertional positions by multiple sequence alignment revealed that this insertion site is shared by the *cob i2* introns of *Chlorogonium elongatum* and *Saccharomyces cerevisiae* and the *cob i3* intron of *Allomyces macrogynus* (Table 9).

The *cobi1* intron encodes two reading frames. The first is in frame with the 5′ coding sequence of the *cob* gene. This region is 450 bp in length and terminates with a UGA stop codon producing a 149 amino acid product and is called *cobi1 orf149*. BLASTP searches against the protein database returned a number of similar sequences.

Table 9. Introns from other species with identical insertion sites.

C. acidophila	**Other species**			
Cob i1 ID/L	*Chlorogonium elongatum cob i2* ID/L (Y07814)	*Saccharomyces cerevisiae cob i2* ID/L (J01472)	*Allomyces macrogynus cob i3* ID/L (U41288)	
Cox 1i1 IB/L	*Podospora anserina cox1 i3* (X55026)[a] IB2/L	*Emericella nidulans nox2* (X00790)[a] IB4/L		
Cox 1i2 IB2/L	*C. elongatum cox 1i1* (Y13644) IB2/L	*Marchantia polymorpha cox 1i9* (M68929) ns/L	*A. macrogynus cox 1i12* (U41288) ns/L	*P. anserina cox 1i11* (X55026) IB2/L
Nad 5i1 ID/L	*C. elongatum nad 5i1* (Y13643) ID/L	*C. eugametos nad 5i1* (Af008237) ns/L	*A. macrogynus nad 5i1* (U41288) ns/L	
Nad 5i2 IB/L	*C. elongatum nad 5i2* (Y13643) IB/L	*C. eugametos nad 5i2* (Af008237) ns/L	*M. polymorpha nad 5i1* (M68929) Ns/-	*Neurospora crassa nad 5i2* (X05115) IB4/L
Nad 5i3 ns/G	?			

Intron subgroups and ORF types of group I intron mitochondrial genes with insertion sites shared by *C. acidophila* and other species.
G, an intronic open reading frame of the GIY-YIG type is encoded.
L, intronic open reading frame of the LAGLIDADG type.
-,no intronic open reading frame.
ns, intron subtype not specified
Accession numbers are in parentheses.
[a]*P. anserina* and *E. nidulans* do not share identical insertions but are located 6 bases downstream from *C. acidophila* insertion site.
Data for this table comes from Kroymann and Zetsche (Kroymann & Zetsche, 1998) and Michel and Westhof (Michel & Westhof, 1990).
? No introns with identical sites found.

In Figure 6A *Allomyces macrogynus* showed an identity of 59% and a similarity score of 73%. Interestingly, two of the highest scores belonged to *Chlamydomonas smithii* (which contains only one known intron) and *Chlorogatum elongatum*. In all cases the scores reflected the homing endonuclease LAGLIDADG motif. The LAGLIDADG consensus typically occurs as repeats (P1 and P2), which is evident in all returned matches of the motif in Figure 6A. However, *orf149* of *C. acidophila* contains only the P1 repeat. The second open reading frame (*cobi1 orf112*) within the *cobi1* intron begins at nucleotide position 1596 of the F3 clone with an AUG codon and is 339 bp in length terminating in a UAG codon, producing a protein of 112 amino acids. Interestingly, this sequence is in a +1 reading frame in respect to *orf149*, and overlaps *orf149* by 95 bp. The sequences returned by BLASTP in Figure 6B were extremely similar to those returned by *orf149* with only *Allomyces macrogynus* and *Saccharomyces capsensis* exchanging rank. Multiple sequence alignment showed that *orf112* was missing the P1 repeat while possessing the P2. From this data it may be inferred that *orf149* and *orf112* were, at one time, two parts of the same gene.

A comparison of the well-conserved LAGLIDADG P1 and P2 protein motifs indicates that these two reading frames most closely resemble the I-CsmI class of homing endonucleases. Indeed, *C. smithii* encodes the archetype conserved repeats: IAVGLLLSDAHA (P1) and ALAYWIAGDGCW (P2). Taken together, the insertional and ORF data imply that this intron belongs to the group ID class of introns.

A

Organism (pan)	Motif LAGLIDADG	Motif LAGLIDADG	ORF	Identity/ Positives	Intron* location
C. acidophila orf149	(39)IIVGSLLGDSYA	....	149		
A. macrogynus (2147528)	(44)IIFGSLLGDAFA	(147)ALAIWIQDDGGA	233	59%/73%	*cobi1*
S. capensis (260036)	(82)IIYGSMLGDGHA	(193)ALAIWIMDDGCK	280	47%/61%	
C. smithii (296719)	(43)IAVGLLLSDAHA	(146)ALAYWIAGDGCW	358	28%/43%	
C. elongatum (2193889)	(101)VAVGLLLSDAHA	(209)SLRHAICGDGSS	292	32%/48%	*cobi2*

B

Organism (pan)	Motif LAGLIDADG	Motif LAGLIDADG	ORF	Identity/ Positives	Intron* location
C. acidophila orf112	....	(24)ALAIWIMDDGTW	112		
S. capensis (260036)	(82)IIYGSMLGDGHA	(193)ALAIWIMDDGCK	280	50%/65%	
A. macrogynus (2147528)	(44)IIFGSLLGDAFA	(147)ALAIWIQDDGGA	233	39%/56%	*cobi1*
C. smithii (296719)	(43)IAVGLLLSDAHA	(146)ALAYWIAGDGCW	358	39%/59%	
C. elongatum (2193889)	(101)VAVGLLLSDAHA	(209)SLRHAICGDGSS	292	33%/51%	*cobi2*

C

Organism (pan)	Motif LAGLIDADG	Motif LAGLIDADG	ORF	Identity/ Positives	Intron* location
C. acidophila orf231	(39)IIVGSLLGDSYA	(142)ALAIWIMDDGTW	231		
S. capensis (260036)	(82)IIYGSMLGDGHA	(193)ALAIWIMDDGCK	280	45%/58%	
A. macrogynus (2147528)	(44)IIFGSLLGDAFA	(147)ALAIWIQDDGGA	233	45%/60%	*cobi1*
C. smithii (296719)	(43)IAVGLLLSDAHA	(146)ALAYWIAGDGCW	358	32%/48%	
C. elongatum (2193889)	(101)VAVGLLLSDAHA	(209)SLRHAICGDGSS	292	31%/46%	*cobi2*

Figure 6. Conserved open reading frames within the *C. acidophila cobi1* intron. Number in parentheses following organism name (pan) is the protein accession number from the NCBI database. Number in parentheses preceding the aligned sequences is the amino acid position in the protein. Highlighted amino acids above the aligned sequences correspond to proposed homing endonuclease class. Dots represent gaps. (A) Alignments against *C.acidophila's orf149*. (B) *C. acidophila's orf112*, a free standing open reading frame at +2 with respect to *orf149*. (C) *C. acidophila's orf149* and *orf112* brought into the same reading frame by a deletion of an 'A' residue at position 324 within the intron. *Where available. Organism names are as follows: *Chlamydomonas acidophila*, *Allomyces macrogynus*, *Saccharomyces capensis*, *Chlamydomonas smithii*, and *Chlorogonium elongatum*.

Cox 1i1

The cytochrome oxidase subunit one (*cox 1*) coding region contains two introns (Figure 3). The *cox 1i1* intron is 1235 bp in length and is inserted between the *cox 1* exon nucleotides 4938-6174. The conserved 'U' (nt 4938) and 'G' (nt 6173) residues as described above apply. A search for similar insertion sites among other species failed. However, it is significant that *Podospora anserina* and *Emericella nidulans* insert their introns *cox 1i3* and *nox 2* respectively, 6 bp upstream from the *C. acidophila* insertion site (Table 9).

Cox 1i1 encodes one large reading frame named *cox 1i1 orf306*. This coding region is in frame with the flanking *cox1* 5′ exon sequence, producing a protein 306 amino acids in length. Interestingly, the ORFs of *P. anserina cox 1i3* and *E. nidulans nox2* are among the most similar to *C. acidophila's cox 1i1* according to database searches with BLASTP. *P. anserina* had the top score with an identity of 34% and a similarity of 56% over a 285 amino acid overlap (Figure 7A), while *E. nidulans* showed a 31% identity and 50% similarity. Sequence alignments with Clustal X revealed a clear LAGLIDADG motif (Figure 7A). However, *C. acidophila's* LAGLIDADG sequences appear to be somewhat degenerate. The closest match appears to that of the I-*Sce*III homing endonuclease of *S. cerevisiae*.

Cox 1i2

The *cox1i2* intron is the second intronic region of the *cox1* gene. *Cox1i1* spans 1142 nucleotides, and is inserted between nucleotide positions 6752-7894. Again the conserved 'U' (nt 6752) and 'G' (nt 7893) residues are present. Sequence alignments

Organism (pan)	Motif		ORF	Identity/ Positives	Intron* location
A					
cox 1i1					
	LAGLIDADG	**LAGLIDADG**			
C. acidophila orf306	(33)YLAGLWEGDGHI	(147)WLAGFIDADGCF	306		
P. anserina (g483208)	(26)YLAGLYEGDGHI	(144)WLSGFIDSDGSF	296	34%/56%	*cox 1i3*
E. nidulans (g83726)	(65)YLAGLIEGDGTI	(184)WLSGFIEADGSF	334	31%/50%	*nox 2*
S. cerevisiae (g450080)	(18)YLAGLIEGDGSI	(150)WLAGMTDADGNF	321	26%/45%	
S. cerevisiae (g320873)	(29)WLAGLIDGDGYF	(135)WFVGFFDADGTI	258	25%/46%	
B					
cox 1i2					
	LAGLIDADG	**LAGLIDADG**			
C. acidophila orf358	(107)WFIGFTEGDGGF	(221)WLCGFSDAEGSC	358		
C. elongatum (g3413805)	(146)WVLGFIEGDGGF	(258)WFLGFIEADGSL	338	39%/56%	*cox 1i1*
S. pombe (g141034)	(89)YLAGLIDGDGHF	(200)WLAGFSDADASF	323	39%/56%	
P. anserina (g483194)	(60)WFIGFAEGDGAI	(179)WLSGFTDAEGCF	321	27%/43%	*cox 1i11*
A. macrogynus (g2147556)	(88)WFIGFAEGDGAI	(201)WVSGFTDAEGCF	342	26%/45%	*nad 5i1*

Figure 7. Conserved open reading frames of *C. acidophila* within the two introns of *cox1* coding gene. Number in parenthesis following organism name (pan) is the protein accession number in the NCBI database. Number in parenthesis preceding the aligned sequences is the amino acid position in the protein. Highlighted amino acids above the aligned sequences correspond to proposed homing endonuclease class. *Where available. (A) First intron (*cox 1i1*) in *C. acidophila's cox 1* coding region. (B) Second intron (*cox 1i2*) in *cox1*. Species names are: *Podospora anserina*, *Emericella nidulans*, *Saccharomyces cerevisiae*, *Chlorogonium elongatum*, *Schizosaccharomyces pombe*, *Allomyces macrogynus*.

show that a number of other species share identical insertion sites with this intron (Table 9). They include: *Chlorogonium elongatum* (*cox 1i1*), *Marchantia polymorpha* (*cox 1i9*), *Allomyces macrogynus* (*cox 1i2*) and *P. anserina* (*cox 1i11*).

Cox 1i2 contains a single large open reading frame called *cox 1i2 orf358* (Figure 7B). This open reading frame produces a protein 358 amino acids in length. A search with BLASTP using *C. acidophila's orf 358* returned matches for ORFs within introns that insert at the same site in *cox 1*. Among the highest scoring was that of *C. elongatum*, which showed an identity of 39% and a similarity score of 56% over a 331 amino acid overlap. Multiple alignments of the protein sequence data revealed a LAGLIDADG motif. However, while the sequence alignments showed some conservation, these comparisons failed to suggest a possible endonuclease designation without ambiguity. Information from the insertional position, ORF similarity and structural data (Figure 5) confirm that this intron belongs to intronic group IB2.

Nad 5i1

The NADH dehydrogenase subunit 5 coding region (*nad 5*) in the F3 clone contains at least 3 intronic regions. The first intron *nad 5i1* is 1060 bp in length, and is inserted between exon nucleotide positions 11150-12211. The conserved residues 'U' (nt 11150) and 'G' (nt 12210) are both apparent. Multiple sequence alignments with similar introns revealed that three other organisms share identical insertion sites. They include the *nad 5i1* of *C. elongatum*, the *nad 5i1* of *Chlamydomonas eugametos* and the *nad 5i1* of *A. macrogynus* (Table 9).

Nad 5i1 contains an in frame coding region, called *nad 5i1 orf229*, that is proposed to produce a protein product of 229 amino acids. BLASTP searches with *C. acidophila's* putative *orf229* returned results that were in agreement with the insertional data. Strong similarities (Figure 8A) were noted from *C. eugametos's nad 5i1* encoded protein (identity 61%; similarity 74%) and *C. elongatum's nad 5i1* (identity 48%; similarity 72%). Multiple sequence alignments of these proteins revealed a LAGLIDADG motif, which shows some similarity to the I-*Sce*IV homing endonucleases. Based on the intron insertion site and high similarity scores with the other LAGLIDADG motifs, this intron is proposed to belong to the group ID class of introns.

Nad 5i2

The *nad 5i2* intron of *C. acidophila* is 1275 bp in length and is inserted between nucleotides 12492-13750 of the F3 clone. Once again, conserved 'U' and 'G' residues reside at nucleotide positions 12492 and 13749 respectively. Identical exon insertions (Table 9) among other organisms include the *nad 5i2* of *C. elongatum, C. eugametos's nad 5i2, M. polymorpha's* and *Neurospora crassa's nad 5i1* and *nad 5i2* respectively.

Nad 5i2 contains an in frame sequence capable of producing a protein 296 amino acids in length (*nad 5i2 orf296*). BLASTP searches (Figure 8B) with *orf296* of *C. acidophila's nad5i2* showed the strongest identity (51%) and similarity (65%) with the protein product of *orf231* of *Allomyces macrogynus*. Strong similarities were also noted with the open reading frames of both *C. elongatum's* and *C. eugametos's nad 5i2*. Interestingly, the researchers listed the *nad 5i2* intron of *C. eugametos* (Denovan-Wright *et al*., 1998) as having two open reading frames. Multiple alignment of a number of

Organism (pan)	Motif		ORF	Identity/ Positives	Intron* location
A					
Nad 5i1	**LAGLIDADG**	**LAGLIDADG**			
C. acidophila *orf229*	(36)IMVGGLLGDGWL	(144)SLALWLMGDGSG	229		
C. eugametos (2865263)	(48)ILTGLLLGDGWL	(152)ALAIWLMGDGSG	283	61%/74%	*nad 5i1*
C. elongatum (3483054)	(4)IVVGCLLGDGHL	(115)ALALWCMGDGSA	180	48%/72%	*nad 5i1*
S. capensis (260036)	(82)IIYGSMLGDGHA	(193)ALAIWIMDDGCK	280	38%/55%	
A. macrogynus (2147537)	(66)VIFGSLLGDSHL	(177)ALAIWFMDDGSK	261	28%/47%	*nad 5i1*
B					
Nad 5i2	**LAGLIDADG**	**LAGLIDADG**			
C. acidophila *orf296*	(95)VIAGLLLSDGCL	(204)CLAHMIMGDGYW	296		
A. macrogynus (2147527)	(40)AINGLLLSDGHV	(145)TLAFWIMGDGFW	231	51%/65%	
C. elongatum (3646347)	(70)AIAGLMFSDGHI	(188)SLAFAIMGDGYW	283	41%/55%	*nad 5i2*
C. eugametos (2865262)	(59)AITGLMLSDGHL	....	145	52%/65%	*nad 5i2*
C. eugametos (2865265)	....	(35)ALAFMIMGDGYW	119	45%/63%	*nad 5i2*
C. humicola (3122034)	(10)LIFGSLLGDGNL	(121)ALAYFYIDDGAL	218	28%/45%	
C					
Nad 5i3	**GIY** **YIG**				
C. acidophila *orf267*	(67)AVIYLWYNKVNNKCYVGRSTN(88)		267		
C. eugametos (2865254)	(84)AGVYLIYDNLTHDFYVG..SA(103)		306	26%/43%	
P. anserina (478093)	(87)SGIYMIVNKVTKDYYIG..SA(106)		301	25%/43%	
N. crassa (478093)	(232)SGVYMIINKTTKDYYIG..SA(251)		448	23%/41%	
C. elongatum (2193888)	(72)SGVYLVRNDINGNCYVG..SA(91)		299	25%/43%	

Figure 8. Conserved open reading frames within *C. acidophila's* three *nad5* introns. Number in parentheses following organism name is the protein accession number from the NCBI database. Number in parentheses preceding the aligned sequences is the amino acid position in the protein. Highlighted amino acids above the aligned sequences correspond to proposed homing endonuclease class. Dots represent gaps. *Where available. (A) *C. acidophila's* open reading frame within the first intron of *nad 5* (*nad 5i1*). (B) *C. acidophila's* open reading frame within *nad 5i2*. (C) *C. acidophila's* open reading frame within *nad 5i3*. Organism names are as follows: *Chlamydomonas acidophila*, *Chlamydomonas eugametos*, *Chlorogonium elongatum*, *Saccharomyces capensis*, *Allomyces macrogynus*, *Chlamydomonas humicola*, *Podospora anserina*, *Neurospora crassa*.

similar sequences revealed that these two open reading frames probably once formed a single coding region (Figure 8B). Support for this is that the first and second open reading frames of *C. eugametos nad 5i2* shows strong similarity to the P1 and P2 region of LAGLIDADG motifs belonging to the I-*Chu*I (*Chlamydomonas humicola*) class of endonucleases. Based on the insertional and LAGLIDADG similarities, the *nad 5i2* intron appears to belong to the group IB class of introns.

Nad 5i3

The entire *nad 5i3* intron of *C. acidophila* was probably truncated in the cloning procedure, however, enough of the intron was present to analyze. What is present of the intron is 1051 bp in size, and is inserted at the exon nucleotide position 13956, and is assumed to separate amino acid positions 309-310 on the Nad5 protein. A search for similar insertion sites across taxa failed (Table 9). In Addition, the typically conserved 'U' residue at the 5´ splice site (nt 13956) is replaced with a 'G'.

C. acidophila's nad 5i3 contains an inframe coding sequence, called *nad 5i3 orf267*, that produces a protein predicted to be 267 amino acids in size. A search for ORF proteins with BLASTP searches of the *orf267* returned very low but distinguishable similarities with proteins from both the *C. eugametos* and *C. elongatum cob i1* intron (Figure 8C). Further, multiple sequence alignments revealed that the *C. acidophila orf267* had a highly degenerate GIY…YIG homing endonuclease motif. Unfortunately, because of the truncation of this sequence, inference of intronic core sequences is difficult. However, using the available data, this intron is speculated to be either a group IB or ID class of intron.

Transfer RNA Genes

The sequence of the F3 clone contains four tRNA coding regions. They are $tRNA^{met1}$(CAU) (nt 3597-3669), $tRNA^{met2}$ (CAU) (nt 3684-3756), $tRNA^{trp}$ (CCA) (nt 8608-8681) and $tRNA^{gln}$ (UUG) (nt 8820-8890). The structure of the four tRNAs was predicted using the MULFOLD RNA folding program (Jaeger *et al.*, 1989; Jaeger *et al.*, 1990; Zuker, 1989) (Figure 9), which predicts folding RNA based on energy minimization, and tRNAs were numbered according to standardized numbering (Sprinzl *et al.*, 1996). Folding procedures required the manipulation of certain tRNA sequences within the MULFOLD program. For example, in some instances certain nucleotides were forced to pair in order to achieve a conserved secondary structure. Only $tRNA^{gln}$ required no modifications and a final free energy state of –17.2 kcal/mole was obtained after folding. The final structure of $tRNA^{trp}$ required some modifications and had an initial folding structure energy of –15.4 kcal/mole. To obtain the final structure positions, 33(U) and 39(A) were prohibited from interacting within the anticodon loop. The free energy was raised to –13.7 kcal/mole which was still well within acceptable limits. The $tRNA^{met1}$ also required some modifications to achieve a final conserved structure. Initial folding yielded a structure with an available free energy of –16.7 kcal/mole. The final structure was obtained by forcing an interaction between position 7(U) and 66(A), and inhibiting an interaction between positions 26(A) and 44(U). The resulting structure had a free energy of –15.6 kcal/mole. The final structure for $tRNA^{met2}$ could not be forced into a final conserved secondary structure. However a close representation could be formed with the following modifications. Positions 6(U) - 67(A), 10(A) – 25(U), and

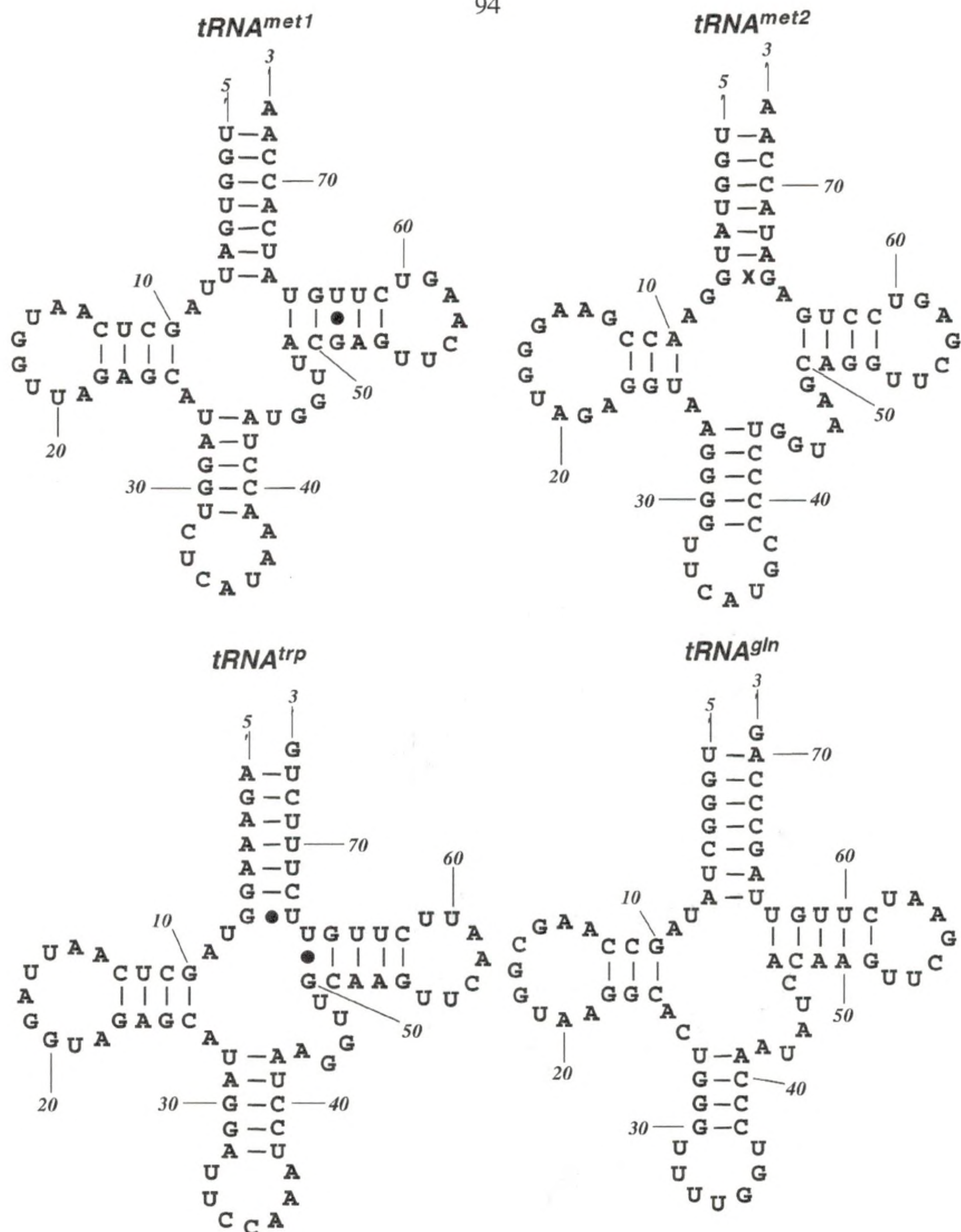


Figure 9. tRNA deduced secondary structure from tDNA sequences identified in the *C. acidophila* mtDNA. Nucleotides involved in Watson-Crick base-pairs are indicated by (-), G:U pairs are indicated by(□) and purine:purine base-pairs are indicated by (x). See text for details.

27(U) – 43(A) were forced to base pair, while positions 9(A) – 25(U) were prohibited from interacting.

Three of the *C. acidophila* mitochondrial encoded tRNAs ($tRNA^{met1}$, $tRNA^{gln}$, and $tRNA^{trp}$) show the normal pattern of invariant and semi-invariant nucleotides recognized in other conventional tRNAs (McClain, 1993). The *C. acidophila* $tRNA^{met2}$, however, contains a number of unusual features that include: (1) 8(G) rather than the invariant 8(U); (2) the lack of potential tertiary interactions between 8(A) and 14(A); (3) the normally present base pairs of 7(G):66(G) and 49(G):65(A) cannot occur as seen in Figure 9. Taken together, the inability to fold $tRNA^{met2}$ into a stable secondary structure with MULFOLD, and the unusual features of this tRNA with respect to other tRNAs seems to suggest that $tRNA^{met2}$ is a pseudogene. Finally, it appears that the 3' CCA terminus is added post-transcriptionally to the *C. acidophila* mitochondrial tRNAs as indicated earlier in both *C. eugametos* and *C. reinhardtii* (Denovan-Wright et al., 1998).

Ribosomal RNA

The rRNA genes were identified according to their similarity at the primary sequence level with their counterparts from *C. eugametos*. A secondary structure was not attempted due to the lack of a complete sequence set of either large subunit (LSU or *rnl*) and small subunit (SSU or *rns*) coding regions. The rRNA subunits that are absent from the sequence data are believed to be part of the *C. acidophila* mtDNA that was not cloned into the LambdaGEM-11® vector. Finally, the exact 5′ and 3′ termini could not be determined and can only be resolved with experimental data. It should be mentioned that BLASTN search and alignment parameters were employed using reduced gap penalty

costs. The rationale for this generous estimate was to increase BLASTN sensitivity to divergent sequences. While this approach worked in some cases, it overestimated sequence identities in others.

The rRNA sequence that is present shows an orientation characteristic of its Chlamydomonad counterparts. The coding sequences are scrambled and discontinuous with respect to each other and are divided into two regions. Region one contains *rnl d*, *rns a*, *rnl a* and is separated by 250 bp from region two, which contains *rnl e* and *rns c* (Figures 3 and 4). Table 10 shows identity between *C. acidophila* and *C. eugametos* rRNA coding regions. Coding pieces are listed in the same order as they appear on each respective genome (i.e., they are colinear).

Table 10. Identity of rRNA between *C. acidophila* and *C. eugametos*.

	rnl d	*rns a*	*rnl a*	*rnl e*	*rns c*
Chlamydomonas acidophila	189 bp	108 bp	166 bp	387 bp	422 bp
Chlamydomonas eugametos	175 bp	102 bp	170 bp	480 bp	508 bp
Identity	72%	79%	80%	86%	71%

bp = base pairs.

<u>**Ribosomal RNA Region 1 (*rnl d, rns a, rnl a*)**</u>

Ribosomal RNA coding region 1 of both species contains the same 3 rRNAs. The first, *rnl d*, was predicted to be 189 bp long from the 175 bp *C. eugametos rnl d*. This discrepancy in size is not due to the random insertion of nucleotides along the length of *C. acidophila rnl d*. Instead, a localized area between nt 3066-3095 of *C. acidophila* seems to contain more nucleotides than the *C. eugametos* counterpart. Whether *C. acidophila* gained these nucleotides or *C. eugametos* experienced a reduction within this

area is unclear. The *rns a* region in *C. acidophila* is located at positions 3213-3320. Both genes are roughly equivalent in size for *rns a* and no particular stretch of DNA has a pattern for areas of insertions or deletions. Likewise, *rnl a* genes of both species are roughly equivalent in length and show an 80% identity.

Ribosomal RNA Region 2 (*rnl e, rns c*)

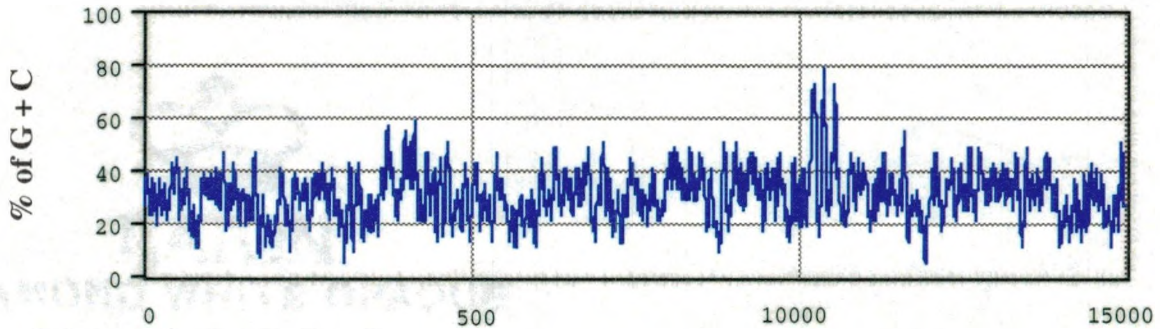
BLASTN searches revealed that *C. acidophila rnl e* is encoded at positions 3763-4229 and shows an 83% identity with its *C. eugametos* counterpart. However, empirical evidence in this laboratory refutes this alignment. During heavy metal induction experiments, radiolabeled oligomers were hybridized to total RNA extracts from *C. acidophila*. Two oligomers, Praer 998 and PFL5, hybridize in flanking positions to the coding strand of *C. acidophila* mtDNA. Praer 998 is a 15-mer capable of hybridizing to *C. acidophila* RNA produced from the template strand at 4136-4150. PFL5, a 15-mer, hybridizes to RNA produced from the template region located at 4151-4165. When both were independently hybridized to total *C. acidophila* RNA (see Figure 14) only Praer 998 showed hybridization with the rRNA. In contrast, PFL5 showed no hybridization, indicating that an RNA transcript was not produced from this region. These results show that *rnl e* is shorter than that represented by BLASTN alignments and probably stops coding around base pair 4150 of *C. acidophila* mtDNA. These data also indicate the danger in defining coding regions on identity data alone. Adjusting this coding region to better fit the hybridization data, the positions 3763-4150 were assigned for *rnl e*. Identity searches with this region yielded a slightly better identity (86%) over a shorter span of DNA.

Lastly, similarity searches place *rns c* at positions 4230-4652. Major differences between the two species include: (1) The 5′ coding region of *C. acidophila* is missing the equivalent 5′ *C. eugametos* 37 base pairs. (2) *C. acidophila* is missing 88 base pairs of *rnl c* that are present at the 3′ end of *C. eugametos.* (3) Internally, sequence alignments showed large gaps missing from *C. acidophila rnl c* as compared to the *C. eugametos* counterpart. The two major gaps in *C. acidophila* exist around 4345 (missing 19 base pairs) and 4481 (missing 16 base pairs). In addition, a small gap of 6 base pairs is missing between positions 4453-4454.

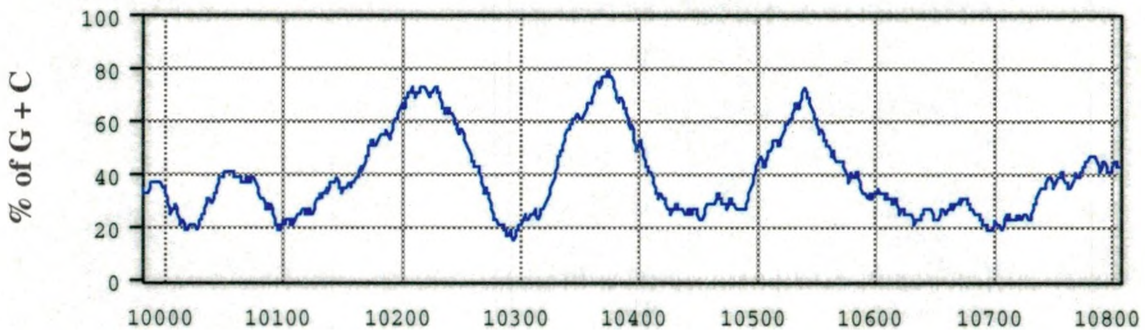
Intergenic Regions

C. acidophila contains one very large intergenic region (nt 9797-10712). Sequence analysis of this region failed to find any counterparts within DNA or protein databases. Since *C. eugametos* and *C. elongatum* (Denovan-Wright et al., 1998; Kroymann & Zetsche, 1998) were both shown to encode repetitive sequences within their intergenic regions, the same analysis was applied to *C. acidophila.* Contained within this region are a number of unique elements found only within this segment of DNA. Figure 10 A1 shows the G-C content of the entire F3 clone. Note the large peak centered between nucleotides 10,000 – 11,000. Closer inspection (Figure 10 A2) reveals that this region is composed of three peaks representing high (close to 80%) G-C content. The first peak appears at around nucleotides 10188-10249 and corresponds to the repeating element 'ACGCGGAGCTG' in Figure 10 B2. This element repeats 5 complete times and is flanked by portions of itself on the 5′ and 3′ ends. Peaks two and three in Figure 10 A2 map to around nucleotides 10338-10398 and nucleotides 10504-10560,

A1



A2



B1

```
9849  GAAAAAACCAAACTTTAAAATTTTTCACCTTAATGAAAAAGTATAGAAGCGGAGGGTGAACCCCCTCCTTTCTATT 9924
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
10006 GAAAAAACCAAACTTTAAAATTTTTCACCTTAATGAAAAAGTATAGAAGCGGAGGGTGAACCCC-TCCTTTCTATT 10080
```

B2

```
10185 CGGAGCTGACGCGGAGCTGACGCGGAGCTGACGCGGAGCTGACGCGGAGCTGACGCGGAGCTGACGCGGA 10254
```

Figure 10. Sequence elements of intergenic region of *C. acidophila* mtDNA nt 9797-10712. A1. G-C content of entire F3 clone. A2. G-C content of intergenic region enlarged. B1. 76 nt direct repeats. B2. 11 nt repetitive element is alternated in red and black for easier visualization. Numbers indicate the first and last bases of the elements and refer to their positions in the complete F3 genomic sequence.

respectively. These areas of high G-C content do not correspond to any large repetitive sequences in F3.

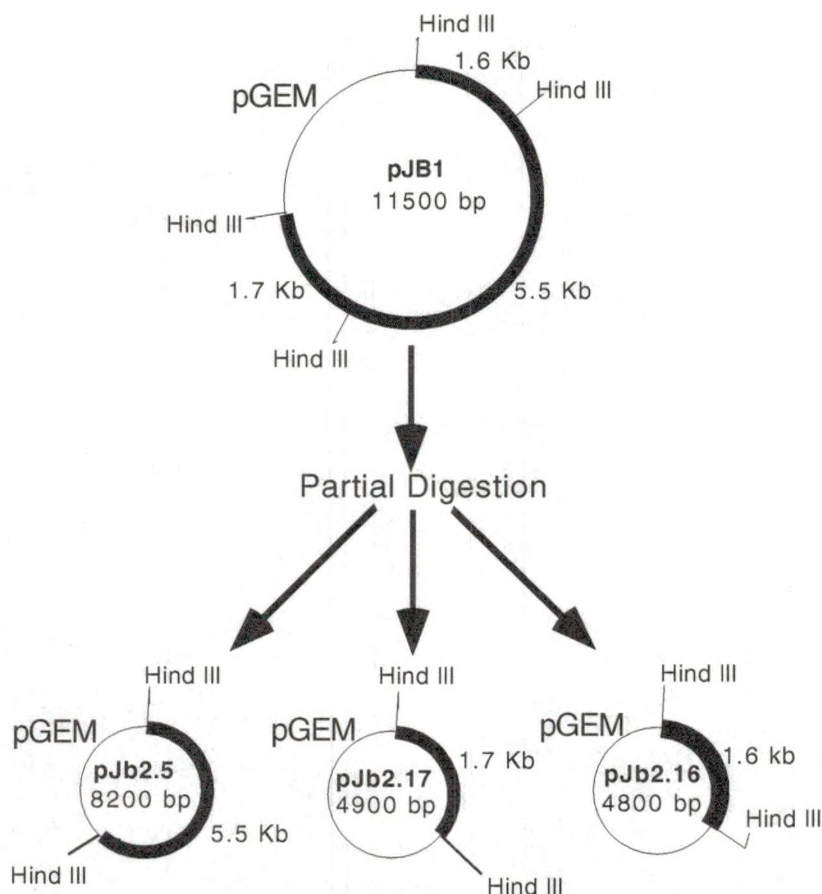
Figures 10 B1 and 10 B2 show repetitive elements in the *C. acidophila* intergenic region. Figure 10 B1 shows two large direct repeats, separated by 83 nucleotides. The repeat at position 9849-9924 is 76 nucleotides long, while the repeat at position 10006-10080 is 75 nucleotides long. This discrepancy can be accounted for by the 'C' residue at position 9913 of the longer molecule which is absent from the short one. Many other short direct repeat regions (8-17 mers) were present, however, none were analyzed further.

One of the most striking differences between the *C. acidophila* and *C. eugametos* mitochondrial genomes is the reduced amount of intergenic space in *C. acidophila* mtDNA. Examples include the large intergenic region around the $tRNA^{met}$ coding region of *C. eugametos*. This region is completely absent from the *C. acidophila* counterpart.

Cadmium Challenge of *C. acidophila* Strain 122

Previous investigations revealed that *C. acidophila* amplifies its mtDNA in response to cadmium exposure (Spanier, unpublished results). The amplified DNA was isolated and cleaved by the restriction enzyme *Hind*III into many subfragments. Three of those *Hind*III fragments were cloned into the vector pGEM-3zf® and the resulting plasmid was named pJB1. Figure 11 A shows pJB1 with the three cloned *Hind*III fragments of 1.6, 1.7 and 5.5 kb in size.

Broth cultures of *C. acidophila* cells (7 X 10^5 cells/ml) were exposed to $Cd(NO_3)_2$ at 0, 25, 50, and 100 μM concentrations over a span of four hours. After cadmium



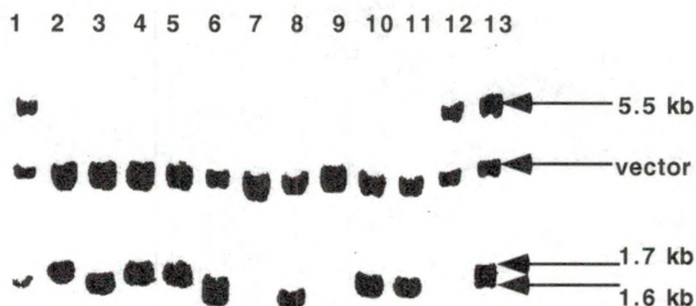


Figure 11. Derivation of pJb2.5. Three *Hind*III fragments were cloned into the vector pGEM-3Zf®. pJB1 was partially digested with *Hind*III which resulted in isolation of the the three *Hind*III fragments into separate vector molecules. (A) restriction maps of pJB1 and plasmids pJb2.5, pJb2.17, and pJb2.16. (B) Negative image of an EtBr stained gel showing results of partial digestion. Lane 13 is a control containing three *Hind*III fragments and vector DNA. Lanes 2, 3 and 12 contain the DNAs of interest 1.7, 1.6, and 5.5 kb respectively.

exposure total RNA was isolated, size fractionated on 1.2% agarose gels, and transferred to nitrocellulose. When ^{32}P labeled pJB1 was used as a probe (Figure 12), one small transcript (~350-400 bp) was strongly induced. Based on densitometry analysis, this transcript increased between fifteen and fifty fold compared to the control, depending on the cadmium treatment.

In order to localize the origin of this transcript, subfragments of pJB1 were used to probe Northern blots containing the amplified RNA. To obtain the subfragments, plasmid pJB1was partially digested with *Hind*III, religated and transformed into *E. coli.* White colonies, indicating that a fragment had been retained with the vector molecule, were selected and plasmid minipreps were prepared. Following isolation, the plasmids were completely digested with *Hind*III and size fractionated on 0.7% agarose gels by electrophoresis. Inspection of the gel in Figure 11 B reveals that three plasmids (lanes 2, 3 and 12) retained single *Hind*III insertions of the desired size classes (1.7, 1.6 and 5.5 kb). These clones were isolated and tested by Southern hybridization. Probes were constructed from the 1.6 kb, 1.7 kb, 5.5 kb, and vector DNA by random primer labeling with ^{32}P. DNA was transferred to nylon membranes and hybridized with each of the constructed probes in succession. As expected, each probe hybridized only to those regions corresponding to the fragments from which they were constructed (data not shown). The plasmids containing the appropriate inserts were named pJb2.5, pJb2.17 and pJb2.16.

Hybridization of these fragments to RNA gels revealed that pJb2.5 contained the DNA coding for the amplified transcript of interest (data not shown). Further localization of the coding region for this transcript was accomplished by restriction mapping of the

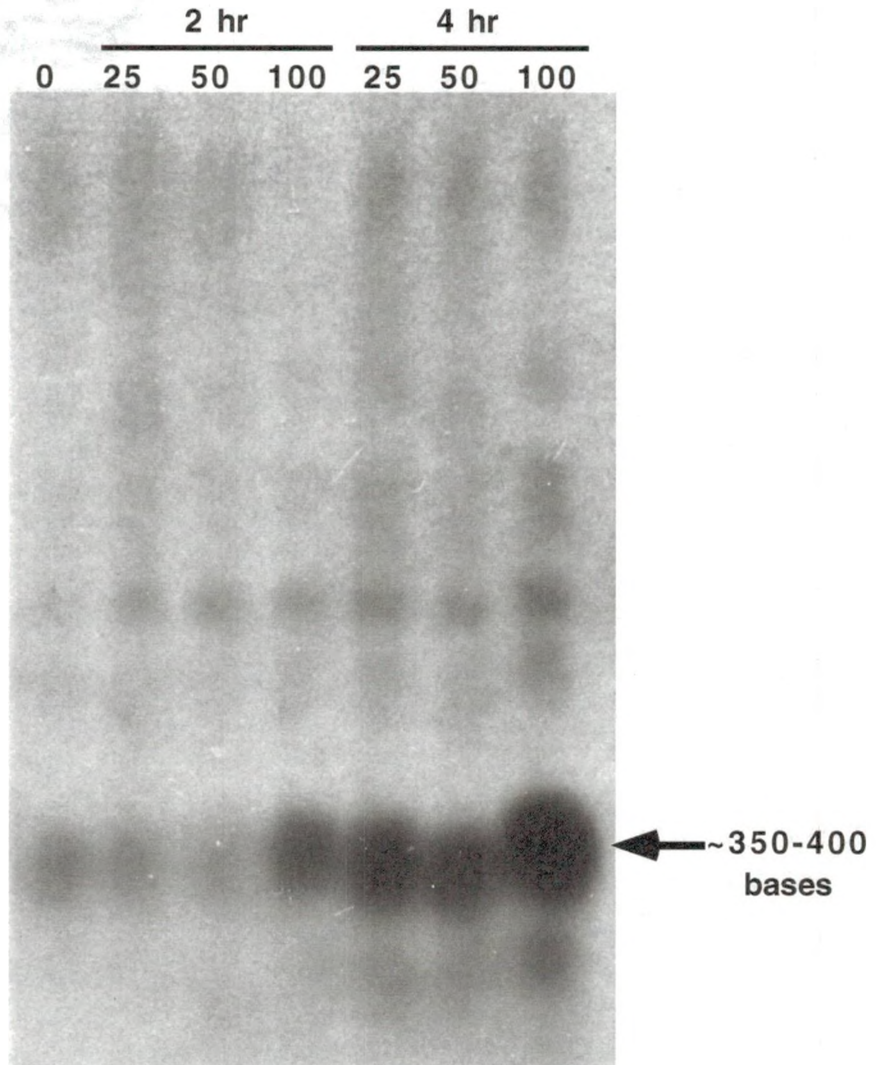


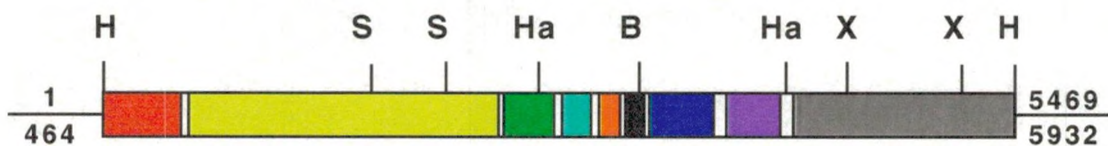
Figure 12. RNA induction of *C. acidophila* strain 122 by cadmium. Total RNA was isolated, quantitated, size fractionated (10 μg/lane) at 20V in a 1.2% agarose formaldehyde gel for 18 hrs, and transferred to a nylon membrane. The resulting blot was probed with ^{32}P labeled plasmid pJB1. Lane numbers indicate $Cd(NO_3)_2$ concentration (μM) and exposure time (hour) prior to RNA isolation. Lane 1 (labeled '0') is the total RNA isolated from unexposed *C. acidophila* control.

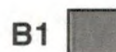
5.5 kb insert. Figure 13 shows a restriction map of the 5 kb region. In addition, since this insert was ultimately sequenced, coding regions are illustrated by color. For continuity, regions are described according to their map locations from the F3 clone. All regions of interest, including restriction endonuclease sites, can be viewed in the mtDNA map and the complete mtDNA sequence (Figures 3 and 4, respectively). However, the fragment sizes are listed according to size in the legend to Figure 13.

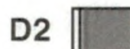
The insert is 5469 bases in size, mapping between and including nucleotide positions 464 – 5932 of the F3 mtDNA sequence, and contains a number of putative coding regions. This region includes 323 bp of the *nad2* 3′ coding region, full copies of the *cob*, *rnl d*, rns a, *rnl a*, *tRNA met 1* and 2, *rnl e*, *rnl c* genes, and 1270 bp of the 5' *cox 1* coding region. The first round of analysis included probes prepared from fragments A1, A2, B1, B2 and C (Figure 13). Northern analysis revealed that the 2482 base *Sca*I/*Xba*I fragment C encoded the strongly induced transcript. As seen in Figure 13, fragment C contains 525 bases of the 3' end of *cob*, complete coding regions for *rnl d*, *rns a*, *rnl a*, both tRNAs, *rnl e*, *rns c*, and 279 bases of the *cox 1* gene. Fragment C was then cleaved by the blunt end endonuclease *Hae*III. Three fragments obtained from the digestion (D1, D2 and D3) were labeled and used as probes against northern blots. Fragment D3, which contains 40 bases of the 3′ end of *rnl d*, complete copies of *rnl a* and *e*, *rns a* and *c*, and both tRNAs was shown to contain the region of interest. Finally, the fragment D3 was cut by *Bst*N1 once, producing fragments E1 and E2 and effectively bisecting the RNA coding regions.

When D3 was hybridized to total RNA from cadmium challenged *C. acidophila* cells, a number of signals were evident as in Figure 14 A. The top (larger sized) band

Figure 13. Restriction map and probe construction.
Diagram of restriction mapping for the pJb2.5 insert. Each fragment was labeled with ^{32}P and used to localize the origin of the 350-400 base RNA transcript produced in response to cadmium. These probes were: A1 = *Hind*III/*Sca*I fragment (1739 bp), A2 = *Sca*I/*Sca*I fragment (257 bp), B1 = *Xba*I/*Hind*III fragment (91 bp), B2 = *Xba*I/*Xba*I fragment (900 bp), C = *Sca*I/*Xba*I fragment (2482 bp), D1 = *Sca*I/*Hae*III fragment (693 bp), D2 = *Hae*III/*Xba*I fragment (336 bp), D3 = *Hae*III/*Hae*III fragment (1453 bp), E1 = *Hae*III/*Bst*NI (573 bp), E2 = *Bst*NI/*Hae*III (880 bp). The restriction enzyme abbreviations are: H = *Hind*III. S = *Sca*I, Ha = *Hae*III, B = *Bst*N1, X = *Xba*I. Blunt end *Hae*III fragment (D3*) was cloned into the blunt end *Sma*I site of pGEM-3Zf® (p1500). Numbers flanking the top map are; Top numbers = actual length of the insert. Bottom numbers = Map positions of *C. acidophila* mtDNA as seen in Figure 3.







= *nad 2*

= *cob*

= *rnl d*

= *rns a*

= *rnl a*

= *tRNA met 1 & 2*

= *rnl e*

= *rns c*

= *cox 1*

represents the strongly induced transcript of interest. When an identical lane of RNA was probed with fragment E2 (Figure 14 A) the lower bands disappeared, indicating that the transcript of interest originated from either the *rnl e* coding region or the *rns c*. Since the putative coding sequence for both these regions are so close in size (*rnl e* = 396, *rns c* 348) it was difficult to separate them on agarose gels. To discover which of the coding regions was responsible for producing the signal, oligomers from previous sequencing efforts were end labeled with γ-^{32}P and used to probe dot blots containing total cellular RNA of cadmium challenged *C. acidophila* cells.

Figure 14 B shows an autoradiograph of the dot blot hybridization experiments. Two samples are shown for each experiment. The bottom dot of each lane contains the plasmid p1500, which includes fragment D3 from the above experiments as a positive control (see Figure 3). The top dot of each blot contains 10 µg/dot of total cellular RNA isolated from *C. acidophila* cells challenged with 100 µM $Cd(NO_3)_2$. The cellular RNA was then hybridized to primers complementary to the coding strand. Figure 14 B shows the results of these hybridization efforts. Blots one and two hybridized to oligomers targeting the *rnl e* gene and showed a strong signal. PFL5 of lane 3 is a 15 nucleotide oligomer that overlaps the putative *rnl e* coding region by 3 bases, with the remaining 12 nucleotides overlapping an intergenic region. This hybridization produced no signal, which was used to modify our sequence coding determinations. Blots four and five were probed with oligomers targeted against the *rns c* transcript. The blots revealed that transcript production from the *rns c* coding region are many fold lower than that of the *rnl e* coding region. As a control for the amount of RNA loaded into the blot, all blots were stripped of probe and rehybridized with γ-^{32}P PFL7, which had been shown

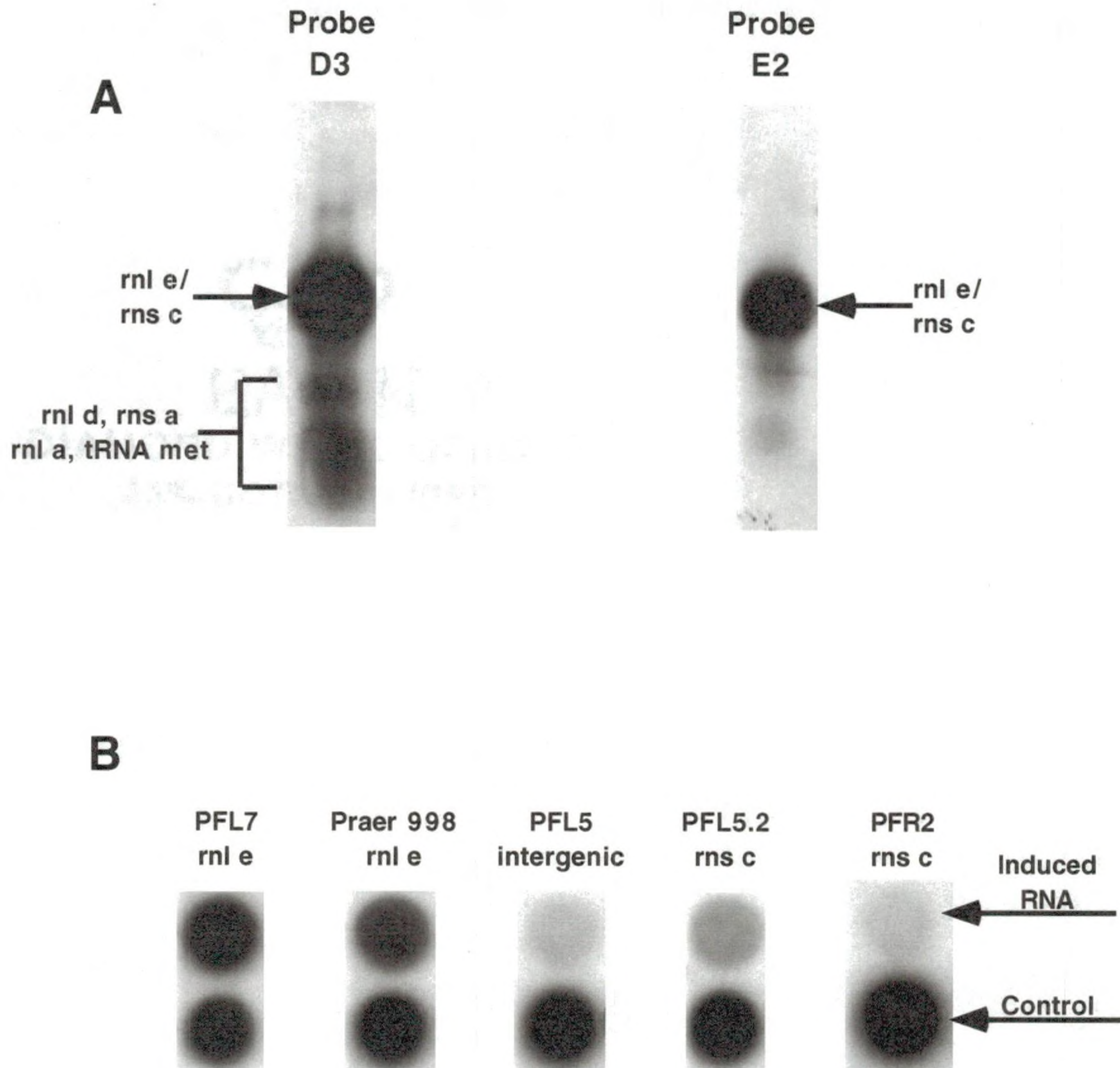


Figure 14. Localization of $Cd(NO_3)_2$ induced transcript. (A) Nylon strips of cadmium-induced RNA (10 µg/lane). Strip one was probed with ^{32}P p1500 (fragment D3 - Figure 13). Strip two was probed with ^{32}P *Bst*N1/*Hae*III fragment (fragment E2 - Figure 13). (B) Dot blot of cadmium induced RNA (10 µg/dot) probed with ^{32}P labeled oligomers. Top dot = cadmium-induced RNA, bottom dot = control p1500 DNA. Lanes are labeled with: Top - oligomer names; Bottom - the targeted rRNA transcripts as determined by DNA sequencing. PFL5 corresponds to putative intergenic region between *rnl e* and *rns c*.

previously to produce a strong radioactive signal (Lane 1, Figure 14 B). All dot blots showed similar levels of radioactivity by autoradiography, indicating that similar amounts of intact RNA were loaded into each dot. Taken together, these results indicate that the rRNA large subunit fragment *rnl e* is the transcript produced in response to cadmium challenge.

DISCUSSION

In this study two aspects of *Chlamydomonas acidophila* were investigated. The first involved *C. acidophila's* mitochondrial genome. Previous experiments have suggested that this molecule is circular and roughly 20 kb in size (Spanier, unpublished results). An attempt to clone the molecule resulted in the isolation of a 15 kb fragment. This fragment was sequenced and encoded genetic elements were shown to be essentially colinear with the *C. eugametos* mitochondrial genome. However, a number of distinct differences appear in the mitochondrial genomes of both species. The differences include a decreased amount of intergenic space in *C. acidophila* relative to *C. eugametos* and intronic variation between the two species. The second subject of study focused on the relatively high cadmium tolerance of *C. acidophila*. Previous investigations had shown that a segment of DNA amplifies its copy number in response to cadmium (Spanier, unpublished results). This DNA has been shown to be the *C. acidophila* mitochondrial genome (this work). Part of this genome was cloned and found to encode rRNA species that increases its copy number (primarily *rnl e*) in response to cadmium.

Mitochondrial Genome

Sequence analysis reveals that the 15 kb mtDNA fragment contains only a partial set of genetically encoded mitochondrial genes. Figure 15 directly compares the partial sequence of the *C. acidophila* mitochondrial genome with the equivalent region of the

Figure 15. Direct comparison of the *C. acidophila* and *C. eugametos* mitochondrial genomes.
Regions of significant difference between the two genomes are highlighted in **bold** type.
Intron names and open reading frames are listed in the Results section.
[a] The *C. eugametos nad 2* gene is 1491 nucleotides in length.
[b] The *C. eugametos nad 5* gene is 3897 nucleotides in length (including introns).

Figure 15

Region	*C. acidophila*		*C. eugametos*	
	nt position	Length	nt position	Length
nad2	3-788	786	6969-7754[a]	786
Intergenic	**789-811**	**23**	**7755-7866**	**112**
cob				
Exon 1	812-1240	429	7867-8259	393
Intron	**1241-2270**	**1030**	**8260-9500**	**1241**
Exon 2	2271-2984	714	9501-10250	750
Intergenic	2982-3003	22	10251-10304	54
rnl d	3004-3192	189	10305-10479	175
Intergenic	3193-3212	20	10480-10519	40
rns a	3213-3320	108	10520-10630	111
Intergenic	3321-3334	14	10631-10661	31
rnl a	3335-3500	166	10662-10831	170
Intergenic	**3501-3596**	**96**	**10832-11297**	**466**
$tRNA^{met1}$	3597-3669	73	11298-11370	73
Intergenic	**3670-3683**	**14**	**11371-12270**	**900**
$tRNA^{met2}$	3684-3756	73	12271-12343	73
Intergenic	3757-3762	6	-	-
rnl e				
Exon 1	3763-4150	388	12344-12595	252
Intron	-	-	**12596-13132**	**537**
Exon 2	-	-	13133-13360	228
Intergenic	4151-4230	80	13361-13425	65
rns c	4230-4652	423	13426-13933	508
Intergenic	4653-4662	10	13934-13968	35
cox 1				
Exon 1	4663-4938	276	13969-15063	1095
Intron 1	**4939-6173**	**1235**	**15064-16066**	**1003**
Exon 2	6174-6752	579	16067-16501	435
Intron 2	**6753-7893**	**1141**		
Exon 3	7894-8571	678		
Intergenic	8572-8607	36	16502-16518	17
$tRNA^{trp}$	8608-8681	74	16519-16591	73
Intergenic	**8682-8819**	**138**	**16592-16598**	**7**
$tRNA^{gln}$	8820-8890	71	16599-16669	71
Intergenic	**8891-8902**	**12**	**16670-16740**	**71**
nad 1				
Exon 1	8903-9796	894	16741-17340	600
Intron			**17341-17655**	**315**
Exon 2			17656-17943	288
Intergenic	9797-10712	916	17944-19000	1057
nad 5				
Exon 1	10713-11150	438	19001-19438	438
Intron 1	11151-12210	1060	19439-20596	1158
Exon 2	12211-12492	282	20597-20878	282
Intron 2	12493-13749	1257	20879-21886	1008
Exon 3	13750-13956	207	21887-22093[b]	207
Intron 3	**13957-15007**	**1051(?)**		
Total	3-15007	15005	6969-22093	15125

C. eugametos mitochondrial genome. While the remainder of the *C. acidophila* mitochondrial genome has yet to be sequenced, the fragment that has been sequenced bears a striking resemblance to the mitochondrial genome of *C. eugametos* (Denovan-Wright *et al.*, 1998). The two genomes are essentially colinear and share the sequence order of: *nad 2*, *cob*, *rnl e*, *rns a*, *rnl a*, *tRNA*^{met1} and *tRNA*^{met2}, *rnl e*, *rnl c*, *cox 1*, *tRNA*^{trp}, *tRNA*^{gln}, *nad 1* and *nad 5*.

C. eugametos contains additional genes than those contained in Figure 15. Those genes include the rRNA genes *rns b*, *rnl b*, *rnl c*, *rnl f*, and the electron transport proteins *nad 4* and *nad 6*. These genes, including the missing coding regions for *nad 2* and *nad 5*, account for about 8 kb of DNA missing from the F3 clone. This places *C. acidophila's* mitochondrial genome close in size to *C. eugametos* at about 23 kb. Considering the colinearity of these two species and low sequence divergence in protein coding regions, *C. acidophila* is predicted to contain these gene sequences.

The reason why only 15 kb of sequence, rather than the entire mitochondrial genome, of *C. acidophila* was cloned is unclear. Previous investigations suggested that the mitochondrial genome of *C. acidophila* is circular (Spanier, unpublished results). Circularity was proposed from three independent lines of evidence: (1) restriction analysis, (2) partial digest with *Bam*HI that reportedly cut the mtDNA only once, and (3) the isolation of the mtDNA by alkaline lysis. In preparation for the cloning procedure, mitochondrial DNA was isolated by alkaline lysis then cut with *Bam*HI. *C. acidophila's* entire mtDNA should have been ligated within the *Bam*HI site of LambdaGem-11®. Further, all isolated clones were truncated to some extent and each lost one of their *Bam*HI sites during the cloning procedure. One possibility for the loss of sequence may

be attributed to the amplification procedure. Phage particles were amplified in the *E. coli* strain LE392. Strain LE392 has been reported to create anomalous recombination with eukaryotic DNA. The cloning procedure may have best been served by using *E. coli* strain KW251 which has mutations in a number of recombination pathways. Another possibility is that the mitochondria of *C. acidophila* is populated with mtDNAs in different forms, including linear DNA. This possibility has been addressed in a number of different reviews (Bendich, 1993; Nosek *et al.*, 1998), and the actual form (or forms) of *C. acidophila's* mitochondrial genome may need to be revisited.

Apart from the *rtl* (reverse transcriptase like) gene of *C. reinhardtii*, an identical set of mitochondrially encoded genes are encoded by *C. eugametos* (Denovan-Wright et al., 1998), C. *moewusii* (Lee *et al.*, 1991), *C. reinhardtii* (Gray & Boer, 1988), *C. smithii*, and *Chlorogonium elongatum* (Kroymann & Zetsche, 1998). While the remaining coding regions of the electron transport chain (*nad 4* and *nad 6*) and ribosomal RNA (*rns b*, *rnl b*, *rnl c*, *rnl f*) have yet to be sequenced from the *C. acidophila* mitochondrial genome, those genes are expected to be present. This suggests that the common ancestor of the above named species already possessed a highly reduced coding capacity compared to other green algae, e.g., *Prototheca wickerhamii* (Wolff *et al.*, 1994).

Phylogenetic studies with nuclear and chloroplast rRNA have indicated that *C. eugametos*, *C. moewusii*, and *C. elongatum* are more closely related to each other than to *C. reinhardtii* and *C. smithii*, and also indicated that *Chlamydomonas* is not monophyletic (Buchheim *et al.*, 1996). This data is corroborated by the recent sequencing of both *C. eugametos* and *C. elongatum's* mitochondrial genomes (Denovan-Wright et al., 1998; Kroymann & Zetsche, 1998). The data gained by this investigation indicates that the

closest relative to *C. acidophila's* mitochondrial genome is *C. eugametos* followed by *C. elongatum*. Indeed a number of similar characteristics are evident in these three species. All three have been suggested to be circular. *C. acidophila* and *C. eugametos* are colinear and while *C. elongatum* does not share this colinearity, the breakpoints for the rRNA coding regions and many of the protein coding genes are identical (Kroymann & Zetsche, 1998). For example, *C. elongatum* encodes *rnl d*, *rns a*, *rnl a*, *tRNAmet*, *rnl e* and *rns c* in a module similar to *C. eugametos* and *C. acidophila*. By contrast, *C. reinhardtii* contains a linear ~15 kb mitochondrial genome which, though fragmented, shows no relationship to the gene arrangement in *C. acidophila*.

Another characteristic feature shared among *C. acidophila*, *C. eugametos*, and *C. elongatum* is that all three encode their genes head-to-tail on a single strand. In contrast, *C. reinhardtii* transcribes *nad5*, *nad4*, and *cob* from one strand of DNA while the other genes are transcribed from the other (Michaelis *et al.*, 1990). Other protists such as *Acanthamoeba castellanii*, *Dictyotelium discoideum*, *Monsiga brevicollis*, and *Pedinomonas minor* share this single stranded coding (Gray *et al.*, 1998). It has been suggested that this represents a special property of these mitochondrial genomes, such as a single promoter from which all genes are transcribed (Denovan-Wright *et al.*, 1998; Kroymann & Zetsche, 1998). However, the differential expression of the *rnl e* and *rns c* genes under cadmium challenge in Figure 14 weakens this argument. The results of the cadmium induction experiments suggest that *C. acidophila* may contain a number of transcriptional units. Unfortunately, no promoter element has been identified in either *C. eugametos* or *C. elongatum*, and no promoter regions were defined in this study.

Intergenic Regions

Intergenic spacers in *C. acidophila* can be small (e.g. 23 base pair spacer between *nad 2 / cob*) or as large as 916 base pairs (*nad 1 / nad 5* spacer), and together with introns account for over half of the nucleotides in the *C. acidophila* 15 kb segment of the mitochondrial genome (8161 total bp = 6774 bp intronic + 1387 bp intergenic spacer). Database searches at both NCBI and EBI with both BLAST and FASTA search methods revealed no similarities with other intergenic regions. One of the most prominent features of the *C. eugametos* intergenic regions is the presence of two copies of large direct repeats in the two largest intergenic regions of the mitochondrial genome. However, counterparts in the *C. acidophila* mitochondrial genome could not be found. Indeed, these two direct repeats have no equivalent in any *Chlamydomonas* species studied so far. Furthermore, none of the repetitive structures reported in *C. eugametos* or *C. elongatum* show any equivalents in *C. acidophila.*

C. acidophila does have two direct repeats (Figure 10 B1) but they show no relationship to *C. eugametos*. Also, the two repeats are separated by only 83 bp and reside in the same intergenic region. In contrast, the linear mtDNA of *C. reinhardtii* possesses terminal inverted repeats of about 530 base pairs as well as an internal 86 base pair repeat of the two outermost sequences. It has been suggested that the internal repeat plays a role in the replication of this linear molecule (Vahrenholz *et al.*, 1993).

C. acidophila's largest intergenic region is characterized by an elevated G + C content with respect to the rest of the F3 fragment, especially towards the coding region of *nad 5*. In Figure 10 A2 three spikes of elevated G + C content can be seen. The first of these spikes contains the short tandem repeat sequence (STR) $(ACGCGGAGCTG)_5$,

which occurs only in this part of the *C. acidophila* mitochondrial genome. Database searches returned similarities for this sequence, however, none carried the tandem repeated nature of this sequence and no counterpart could be found in any *Chlamydomonas* species, including *Chlorogonium elongatum*. While the etiology of this sequence is unclear, its repetitive nature may be attributed to slip-strand mispairing which occurs in combination with inadequate DNA mismatch repair pathways (Strand *et al.*, 1993). Repetitive DNA has been reported to contain a 'peculiar' tertiary structure (Coggins & O'Prey, 1989). This structure allows mismatches between neighboring repeats, and depending on the strand orientation, repeats can be inserted or deleted during DNA duplication (Chiurazzi *et al.*, 1994; Henderson & Petes, 1992; van Belkum *et al.*, 1998).

While no homologue for this sequence could be found, STRs are becoming increasingly recognized in DNA sequences (Bork *et al.*, 1998; Epplen *et al.*, 1998; van Belkum *et al.*, 1998). Repeated sequences sharing a long evolutionary history to a genome have even been shown to adopt important functions (Britten, 1997). Such functions have been postulated in *C. eugametos* (Kroymann & Zetsche, 1998). However, the lack of conservation of any repetitive element between *C. eugametos* and the closely related *C. acidophila* seems to preclude this. This observation is corroborated by Kroymann and Zetsche who suggest that these elements are simply mobile 'selfish DNA' within the mitochondrial genome (Kroymann & Zetsche, 1998). However it may be premature to invoke a 'junk' or 'selfish' DNA function for these repeated DNAs. It must be offered that the sequence analysis for the F3 fragment was rather limited, and may not have been sensitive enough to the similarities between the species. One possibility may

be the retention of information via a higher order structure than primary sequence. The precedence for this is easily observed in introns. Related introns may show very little sequence similarity, but the information for function is retained in the secondary and tertiary structures.

Introns

All intronic regions of *C. acidophila* were shown to belong to the group I class of introns and account for nearly half of the DNA sequence in the F3 clone (6774 base pairs). These introns showed little relationship with each other, indicating that lateral transfer within the *C. acidophila* mitochondrial genome was unlikely. However they did show sequence and structural homology to group I introns of other species. The lack of published information on the intron subtypes of the closely related *C. eugametos* makes it difficult to infer the movement of these introns. The only other closely related species shown to possess introns are *C. smithii* and *C. elongatum.* Investigation of their mitochondrial introns suggests a common ancestry for some of the introns.

Ancestry of introns (and their open reading frames) can result from either vertical transmission or lateral transfer between taxa. A high degree of similarity of both introns and open reading frames should be expected when lateral transfer occurs. For example, a transfer between an alga and fungus should show a high degree of similarity if the transfer does not date too far back in the past. In contrast, little similarity should be observable in two cases: (1) when introns in different taxa are vertically inherited from a common ancestor, and the respective taxa evolve independently of one another for a long period of time, and (2) when lateral transfer dates far back in the past. Ohta et al. (Ohta

et al., 1993) drew the conclusion that vertical transmission was the simplest explanation for the presence of cognate introns in liverwort and fungal mitochondrial genomes.

Of *C. acidophila*, *C. eugametos*, and *C. elongatum* the most similar introns in insertion site and open reading frames are *nad5i1* and *nad5i2*. This low level of divergence implies that these introns are a recent addition to the *Chlamydomonas* lineage. Further, if vertical inheritance is postulated, then the insertion event of these introns predates the split of these three species.

Of the remaining group I introns, the *C. acidophila cobi1* and *coxi2* share sequence similarity and an identical insertion site with *C. elongatum's cobi2* and *cox1i1*. It is interesting to note that *C. eugametos* contains no introns at these insertion sites. Since *C. acidophila* is clearly a closer relative to *C. eugametos* than to *C. elongatum* it is tempting to speculate that these two introns originated in this clad prior to the *C. eugametos/C. acidophila* split from *C. elongatum*. This would imply that *C. eugametos* did possess these introns, but lost them some time after the split with *C. acidophila*.

The remaining two introns in *C. acidophila*, *cox1i1* and *nad5i3*, seem to have no counterpart in any other *Chlamydomonas* species. The sequences in the NCBI database showing the most similarity to *cox1i1* were from *Emericella nidulans* and *Podospora anserina*. However, organisms with an identical insertion site were not found. It is interesting to note that *E. nidulans* and *P. anserina* insert their introns (*nox2* and *cox1i3* respectively) six base pairs downstream from the *C. acidophila* insertion site.

An even more puzzling intron in *C. acidophila* is *nad5i3*. Not only does it contain a highly degenerate GIY YIG homing endonuclease motif, but no other organism with an identical insertion site could be found. In addition, the insertion site of *nad5i3* is atypical.

The most conserved nucleotides displayed by group I introns include an exonic 'U' preceding the 5' splice site, and an intronic 'G' residue preceding the 3' splice site. Preceding the 5' splice site in *nad5i3* the 'U' residue is substituted by a 'G'. Substitutions such as these have been reported for other introns. An intron in the *cox 1* gene of *Aspergillus* for instance shows 'C' as the replacing residue at the 5' splice site. Investigations revealed that the intron did not edit the 'C' residue to achieve a 'U' at the 5' splice site (Hur *et al.*, 1997; Hur & Waring, 1995). Other studies show that some group I introns are capable of self splicing *in vitro* even when an experimentally introduced 'A' was substituted for the 'G' at the 3' splice site (Beagley *et al.*, 1996; Golden & Cech, 1996; Michel *et al.*, 1989). It remains to be determined if the atypical 5' splice site of *nad5i3* in *C. acidophila* undergoes RNA editing or if self splicing activity is affected.

Cadmium Tolerance

Chlamydomonas acidophila was shown by previous investigators to be more tolerant of high cadmium concentrations than *C. reinhardtii* (Spanier and Braunner, unpublished results). The present study explored the molecular mechanisms that respond to cadmium toxicity. One way to analyze cadmium response mechanisms is to characterize the induction of specific RNA transcripts when cells are exposed to cadmium. Previous experiments had shown that *C. acidophila* increases the copy number of its mitochondrial genome in response to cadmium (Spanier, unpublished results). To ascertain if any transcripts were induced from this DNA, cloned mtDNA fragments were used to probe the northern blots of cadmium challenged *C. acidophila* cells.

A number of small transcripts were shown to be induced in response to cadmium. Those transcripts originated from the regions encoding ribosomal RNA (*rnl c*, *rnl a*, *rnl e*, *rns a* and *rns c*). Further, the transcript showing the greatest induction seemed to originate from the *rnl e* region. It must be mentioned that this induction did not manifest during every cadmium exposure, indicating that some other variable may need to be accounted for (data not shown).

Further analysis revealed that *rnl e* was induced in response to cadmium, though it may not be the most highly induced transcript (i.e., induced to the highest copy number). Initial transcript inspection was carried out by probing with DNA labeled by random primer labeling. This procedure incorporates ^{32}P labeled dCTP into the probing molecule. Transcripts, depending on length, could hybridize a disproportionate amount of ^{32}P dCTP. For example, *rnl e* is proposed to be 388 nucleotides long and (in an idealized situation) could hybridize 81 ^{32}P dCTP residues. In a similar fashion *rnl a* could hybridize 21 ^{32}P dCTP residues. Theoretically, *rnl e* could hybridize ~4 fold more radioactive dCTPs than *rnl a*, creating the illusion that *rnl e* is present in a higher copy number than *rnl a*. In fact, the total number of potentially hybridized ^{32}P dCTP for the ribosomal RNA transcripts *rnl d*, *rnl a* and *rns a* is 78 which is still less than the 81 ^{32}P dCTP hybridized by *rnl e*. In order to remove this bias, hybridization procedures could be repeated by constructing a single end labeled radioactive probe. In this way one transcript would be labeled with only one radioactive moiety. Therefore, transcript copy number would be better represented in the autoradiograph. End primer labeling was eventually employed but only in discerning *rnl e* from *rns a*.

The results of the induction experiments beg two questions. (1) What is the role of mitochondrial rRNA transcript production in response to cadmium exposure; and (2) Why doesn't *rns c* show an induction pattern similar to the other rRNA transcripts? One of the caveats of this study resulted from not obtaining control data on mitochondrial transcript production during dark cycles. Cells were grown under continuous light both during maintenance and experimental procedures. Previous studies have indicated that plant mitochondria are transcriptionally inactive during light phases. Mitochondrial genes have been shown to express their transcipts mainly in the dark when the energy demands of the cell cannot be met by photosynthesis (Salganik *et al.*, 1991). Data on this phenomena in *C. acidophila* would have been helpful.

Previous investigations have indicated that cadmium localizes mainly in the chloroplast (Nagel *et al.*, 1996) and that this is a major site of cadmium toxicity in *C. reinhardtii* (Voigt *et al.*, 1998). One possible scenario in *C. acidophila* is that the loss of energy generation by the inhibition of photosynthetic activity is compensated for by the increased activity of the mitochondria. During Northern analysis, when total RNA was probed with pJB1, only ribosomal RNA was shown to be highly induced. Cloned onto the plasmid pJB1 are, in addition to the genes for rRNA, the coding regions for *cob*, *nad 2* and *cox 1*. Since these genes for the electron transport chain failed to show high induction rates similar to the rRNA, a translational control mechanism might be postulated. This data is in conflict with other research that has demonstrated that dark grown plant cells showed increased transcript production for the electron transport proteins *cox 1* and *2*, *cob* and *ATPase* (Salganik et al., 1991). Clearly more research on this phenomena is needed.

One of the more confounding results is the high level of expression of some of the rRNA transcripts, while the transcript production of *rns c* was almost non-existent. One would expect that all of the rRNA transcript segments should be produced at equal ratios. Equal ratios of *C. acidophila's* fragmented rRNA should be necessary to construct a complete and functional ribosome. Interpretation of this phenomena may indicate that this response mechanism is a side effect of cadmium. In order to more clearly define this cadmium response, data on the normal function of the mitochondria during both light and dark phases are necessary.

APPENDICES

APPENDIX A
Media

M media (minimal)

Component	Stock	*Stock wt	Volume
trace metal mix	10x	see below	1.0 ml
Na citrate$\cdot 2H_20$	10.0%	50.0 gr	5.0 mls
$FeCl_3 \cdot 6H_2O$	1.0%	5.0 gr	1.0 mls
$CaCl_2 \cdot 2H_20$	5.3%	26.5 gr	1.0 mls
$MgSO_4 \cdot 7H_20$	10.0%	50.0 gr	3.0 mls
NH_4NO_3	10.0%	50.0 gr	3.0 mls
KH_2PO_4	10.0%	50.0 gr	1.0 mls
K_2HPO_4	10.0%	50.0 gr	1.0 mls

*stocks are w/v; per 500 mls

Trace minerals solution

Component	grams/liter stock solutions
H_3BO_3	0.100
$ZnSO_4 \cdot 7H_20$	0.100
$CoCl_2 \cdot 6H_2O$	0.020
$Na_2MoO_4 \cdot 2H_20$	0.020
$CuSO_4$	0.004

APPENDIX B
Oligomers

Oligomer name	Length (nt)	Sequence 5´-3´
LLA20015[b]	23	CGC AAC TCG TGA AAG GTA GGC GG
LRA97[b]	29	CCA GAG GTT CAT TAC TGA ACA CTC GTC CG
PACML1[b]	19	GGA CAA TCA GCA GGA AAC C
PACMR1[b]	18	CGC TCG TAA TGT TGG TTG
PAT1[b]	15	GAG TTC CAG ACC AAC
PAT2[b]	18	CTC GAA GTC CTT CTC CAG
PFL3[a,b,c]	15	CGA GTC GAT ATA GAG
PFL5[a,b,c]	15	CAT GTA AAG GTG CAC
PFL5.1[a,b,c]	20	AGC TAG TCA GCC ATG CAA TG
PFL5.2[a,b,c]	22	CAG CTA GTC AGC CAT GCA ATG C
PFL7[a,b,c]	15	ATT TTG CCG AGT TCC
PFR3[a,b,c]	15	CCG AAA CTT GTC TGA
PFR4[a,b,c]	15	TCA TGG GTC AGT CAC
PFR6[a,b,c]	15	TAC CAC AAA CCA ACG
PFR1140[b]	21	GCT GCA CAA GTC CGC ATG GTC
PL5AC[b]	15	GCT GCG ACT GTT TAC
PRAER756[b,c]	17	CTG TTT ACC AAA AAC AC
PRAER998[b,c]	15	AGG GTC TCT TCG TCC
PS3AC[b]	15	GCT ACA ATG GAT GGT
SP055-1 (SP6)[a,b]	19	GAT TTA GGT GAC ACT ATA G
SP070-1 (T7)[a,b]	17	AAT ACG ACT CAC TAT AG

a = Oligomer used in sequencing
b = Oligomer used in PCR
c = Oligomer used as probe

APPENDIX C
Amino Acid Single Letter Code

Alanine	A	Leucine	L
Arginine	R	Lysine	K
Asparagine	N	Methionine	M
Aspartic acid	D	Phenylalanine	F
Cysteine	C	Proline	P
Glutamic acid	E	Serine	S
Glutamine	Q	Threonine	T
Glycine	G	Tryptophan	W
Histidine	H	Tyrosine	Y
Isoleucine	I	Valine	V

APPENDIX D
List of Abbreviations

ATP	=	Adenosine triphosphate
bp	=	base pairs
Cd	=	cadmium
cob	=	apocytochrome b
cox	=	cytochrome oxidase
ddNTPs	=	dideoxy nucleotide triphosphates
dNTPs	=	deoxy nucleotide triphosphates
DNA	=	Deoxyribonucleic acid
g	=	gravity
HSP	=	heat shock protein
kb	=	kilobases
LSU (or *rnl*)	=	large subunit ribosomal RNA
μM	=	micromolar
M	=	Molar
μl	=	microliter
ml	=	milliliter
mM	=	millimolar
mtDNA	=	mitochondrial DNA

List of abbreviations (continued)

N	=	normality
nad	=	NADH dehydrogenase
ng	=	nanogram
nt	=	nucleotides
OD	=	optical density
orf	=	open reading frame
PCR	=	polymerase chain reaction
RNA	=	ribonucleic acid
rpm	=	revolutions per minute
rtl	=	reverse-transcriptase like
S	=	Svedberg unit
SSU (or *rns*)	=	small subunit ribosomal RNA
STR	=	short terminal repeat
tRNA	=	transfer ribonucleic acid
ura	=	uracil

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**(3), 403-10.

Andersson, S. G. & Kurland, C. G. (1998). Reductive evolution of resident genomes. *Trends Microbiol* **6**(7), 263-8.

Andersson, S. G., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T., Alsmark, U. C., Podowski, R. M., Naslund, A. K., Eriksson, A. S., Winkler, H. H. & Kurland, C. G. (1998). The genome sequence of Rickettsia prowazekii and the origin of mitochondria. *Nature* **396**(6707), 133-40.

Andersson, S. G., Zomorodipour, A., Winkler, H. H. & Kurland, C. G. (1995). Unusual organization of the rRNA genes in Rickettsia prowazekii. *J Bacteriol* **177**(14), 4171-5.

Banroques, J., Perea, J. & Jacq, C. (1987). Efficient splicing of two yeast mitochondrial introns controlled by a nuclear-encoded maturase. *Embo J* **6**(4), 1085-91.

Bauman, J. W., Liu, J. & Klaassen, C. D. (1993). Production of metallothionein and heat-shock proteins in response to metals. *Fundam Appl Toxicol* **21**(1), 15-22.

Baumann, P., Baumann, L., Lai, C. Y., Rouhbakhsh, D., Moran, N. A. & Clark, M. A. (1995). Genetics, physiology, and evolutionary relationships of the genus Buchnera: intracellular symbionts of aphids. *Annu Rev Microbiol* **49**, 55-94.

Beagley, C. T., Okada, N. A. & Wolstenholme, D. R. (1996). Two mitochondrial group I introns in a metazoan, the sea anemone Metridium senile: one intron contains genes for subunits 1 and 3 of NADH dehydrogenase. *Proc Natl Acad Sci U S A* **93**(11), 5619-23.

Belfort, M. & Perlman, P. S. (1995). Mechanisms of intron mobility. *J Biol Chem* **270**(51), 30237-40.

Beyersmann, D. & Hechtenberg, S. (1997). Cadmium, gene regulation, and cellular signalling in mammalian cells. *Toxicol Appl Pharmacol* **144**(2), 247-61.

Bingham, F. T. (1979). Bioavailability of Cd to Food crops in relation to heavy metal content of sludge-amended soil. *Environ Health Perspect* **28**, 39-43.

Boer, P. H. & Gray, M. W. (1988a). Genes encoding a subunit of respiratory NADH dehydrogenase (ND1) and a reverse transcriptase-like protein (RTL) are linked to ribosomal RNA gene pieces in Chlamydomonas reinhardtii mitochondrial DNA. *Embo J* **7**(11), 3501-8.

Boer, P. H. & Gray, M. W. (1988b). Scrambled ribosomal RNA gene pieces in Chlamydomonas reinhardtii mitochondrial DNA. *Cell* **55**(3), 399-411.

Boer, P. H. & Gray, M. W. (1988c). Transfer RNA genes and the genetic code in Chlamydomonas reinhardtii mitochondria. *Curr Genet* **14**(6), 583-90.

Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M. & Yuan, Y. (1998). Predicting function: from genes to genomes and back. *J Mol Biol* **283**(4), 707-25.

Borner, G. V., Morl, M., Janke, A. & Paabo, S. (1996). RNA editing changes the identity of a mitochondrial tRNA in marsupials. *Embo J* **15**(21), 5949-57.

Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V. & Schuster, W. (1993). The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. *FEBS Lett* **325**(1-2), 140-5.

Bridge, D., Cunningham, C. W., Schierwater, B., DeSalle, R. & Buss, L. W. (1992). Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. *Proc Natl Acad Sci U S A* **89**(18), 8750-3.

Britten, R. J. (1997). Mobile elements inserted in the distant past have taken on important functions. *Gene* **205**(1-2), 177-82.

Buchheim, M. A., Lemieux, C., Otis, C., Gutell, R. R., Chapman, R. L. & Turmel, M. (1996). Phylogeny of the Chlamydomonadales (Chlorophyceae): a comparison of ribosomal RNA gene sequences from the nucleus and the chloroplast. *Mol Phylogenet Evol* **5**(2), 391-402.

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**(18), 9651-6.

Burger, G., Lang, B. F., Reith, M. & Gray, M. W. (1996). Genes encoding the same three subunits of respiratory complex II are present in the mitochondrial DNA of two phylogenetically distant eukaryotes. *Proc Natl Acad Sci U S A* **93**(6), 2328-32.

Burke, J. M., Belfort, M., Cech, T. R., Davies, R. W., Schweyen, R. J., Shub, D. A., Szostak, J. W. & Tabak, H. F. (1987). Structural conventions for group I introns. *Nucleic Acids Res* **15**(18), 7217-21.

Carignani, G., Groudinsky, O., Frezza, D., Schiavon, E., Bergantino, E. & Slonimski, P. P. (1983). An mRNA maturase is encoded by the first intron of the mitochondrial gene for the subunit I of cytochrome oxidase in S. cerevisiae. *Cell* **35**(3 Pt 2), 733-42.

Castresana, J., Lubben, M., Saraste, M. & Higgins, D. G. (1994). Evolution of cytochrome oxidase, an enzyme older than atmospheric oxygen. *Embo J* **13**(11), 2516-25.

Cavalier-Smith, T. (1992). The number of symbiotic origins of organelles. *Biosystems* **28**(1-3), 91-106.

Cech, T. R., Zaug, A. J. & Grabowski, P. J. (1981). In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27**(3 Pt 2), 487-96.

Cedergren, R., Gray, M. W., Abel, Y. & Sankoff, D. (1988). The evolutionary relationships among known life forms. *J Mol Evol* **28**(1-2), 98-112.

Chiurazzi, P., Kozak, L. & Neri, G. (1994). Unstable triplets and their mutational mechanism: size reduction of the CGG repeat vs. germline mosaicism in the fragile X syndrome. *Am J Med Genet* **51**(4), 517-21.

Cho, Y., Qiu, Y. L., Kuhlman, P. & Palmer, J. D. (1998). Explosive invasion of plant mitochondria by a group I intron [see comments]. *Proc Natl Acad Sci U S A* **95**(24), 14244-9.

Clark-Walker, G. D. (1992). Evolution of mitochondrial genomes in fungi. *Int Rev Cytol* **141**, 89-127.

Claros, M. G., Perea, J., Shu, Y., Samatey, F. A., Popot, J. L. & Jacq, C. (1995). Limitations to in vivo import of hydrophobic proteins into yeast mitochondria. The case of a cytoplasmically synthesized apocytochrome b. *Eur J Biochem* **228**(3), 762-71.

Clyman, J. & Belfort, M. (1992). Trans and cis requirements for intron mobility in a prokaryotic system. *Genes Dev* **6**(7), 1269-79.

Coggins, L. W. & O'Prey, M. (1989). DNA tertiary structures formed in vitro by misaligned hybridization of multiple tandem repeat sequences. *Nucleic Acids Res* **17**(18), 7417-26.

Collard, J. M. & Matagne, R. F. (1990). Isolation and genetic analysis of Chlamydomonas reinhardtii strains resistant to cadmium. *Appl Environ Microbiol* **56**(7), 2051-5.

Colleaux, L., d'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Galibert, F. & Dujon, B. (1986). Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into E. coli as a specific double strand endonuclease. *Cell* **44**(4), 521-33.

Colleaux, L., Michel-Wolwertz, M. R., Matagne, R. F. & Dujon, B. (1990). The apocytochrome b gene of Chlamydomonas smithii contains a mobile intron related to both Saccharomyces and Neurospora introns. *Mol Gen Genet* **223**(2), 288-96.

Covello, P. S. & Gray, M. W. (1992). Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome c oxidase (cox2) in soybean: evidence for RNA-mediated gene transfer. *Embo J* **11**(11), 3815-20.

Crawford, B. D., Enger, M. D., Griffith, B. B., Griffith, J. K., Hanners, J. L., Longmire, J. L., Munk, A. C., Stallings, R. L., Tesmer, J. G., Walters, R. A. & et al. (1985). Coordinate amplification of metallothionein I and II genes in cadmium- resistant Chinese hamster cells: implications for mechanisms regulating metallothionein gene expression. *Mol Cell Biol* **5**(2), 320-9.

Cummings, D. J., McNally, K. L., Domenico, J. M. & Matsuura, E. T. (1990). The complete DNA sequence of the mitochondrial genome of Podospora anserina. *Curr Genet* **17**(5), 375-402.

Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A. & Scazzocchio, C. (1982). Making ends meet: a model for RNA splicing in fungal mitochondria. *Nature* **300**(5894), 719-24.

De Knecht, J., Dillen, V., Koevoets, P. L. M., Schat, H., Verkleij, J. A. C. & Ernst, W. H. O. (1994). Phytochelatins in cadmium-sensitive and cadmium-tolerant *Silene vulgaris*. *Plant Physiol.* **104**, 225-261.

Denovan-Wright, E. M. & Lee, R. W. (1994). Comparative structure and genomic organization of the discontinuous mitochondrial ribosomal RNA genes of Chlamydomonas eugametos and Chlamydomonas reinhardtii. *J Mol Biol* **241**(2), 298-311.

Denovan-Wright, E. M. & Lee, R. W. (1995). Evidence that the fragmented ribosomal RNAs of Chlamydomonas mitochondria are associated with ribosomes. *FEBS Lett* **370**(3), 222-6.

Denovan-Wright, E. M., Nedelcu, A. M. & Lee, R. W. (1998). Complete sequence of the mitochondrial DNA of Chlamydomonas eugametos. *Plant Mol Biol* **36**(2), 285-95.
Dujon, B., Belfort, M., Butow, R. A., Jacq, C., Lemieux, C., Perlman, P. S. & Vogt, V. M. (1989). Mobile introns: definition of terms and recommended nomenclature. *Gene* **82**(1), 115-8.

Epplen, J. T., Maueler, W. & Santos, E. J. (1998). On GATAGATA and other "junk" in the barren stretch of genomic desert. *Cytogenet Cell Genet* **80**(1-4), 75-82.

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* **1143**(3), 243-71.

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**(4), 879-87.

Fearnley, I. M. & Walker, J. E. (1992). Conservation of sequences of subunits of mitochondrial complex I and their relationships with other proteins. *Biochim Biophys Acta* **1140**(2), 105-34.

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics* **78**(2), 737-56.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M. & et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd [see comments]. *Science* **269**(5223), 496-512.

Gadd, G. M. (1992). Metals and microorganisms: a problem of definition. *FEMS Microbiol Lett* **79**(1-3), 197-203.

Garcia-Horsman, J. A., Barquera, B., Rumbley, J., Ma, J. & Gennis, R. B. (1994). The superfamily of heme-copper respiratory oxidases. *J Bacteriol* **176**(18), 5587-600.

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 U S A* **93**(25), 14614-7.

Gillham, N. W. (1994). *Organelle genes and genomes*, Oxford University Press, Inc.,, New York.

Golden, B. L. & Cech, T. R. (1996). Conformational switches involved in orchestrating the successive steps of group I RNA splicing. *Biochemistry* **35**(12), 3754-63.

Grant, C. A., Buckley, L. D. & Selles, F. (1998). Cadmium accumulation in crops. *Can. J. Plant Sci.* **78**, 1-17.

Gray, M. W. (1988). Organelle origins and ribosomal RNA. *Biochem Cell Biol* **66**(5), 325-48.

Gray, M. W. (1992). The endosymbiont hypothesis revisited. *Int Rev Cytol* **141**, 233-357.

Gray, M. W. (1993). Origin and evolution of organelle genomes. *Curr Opin Genet Dev* **3**(6), 884-90.

Gray, M. W. (1998). Mass migration of a group I intron: promiscuity on a grand scale [comment]. *Proc Natl Acad Sci U S A* **95**(24), 14003-5.

Gray, M. W. & Boer, P. H. (1988). Organization and expression of algal (Chlamydomonas reinhardtii) mitochondrial DNA. *Philos Trans R Soc Lond B Biol Sci* **319**(1193), 135-47.

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**(4), 865-78.

Gray, M. W. & Schnare, M. N. (1990). *The ribosome. Structure, function and evolution.*, American Society for Microbiology, Washington, D. C.

Grohmann, L., Brennicke, A. & Schuster, W. (1992). The mitochondrial gene encoding ribosomal protein S12 has been translocated to the nuclear genome in Oenothera. *Nucleic Acids Res* **20**(21), 5641-6.

Gutell, R. R. (1994). Collection of small subunit (16S- and 16S-like) ribosomal RNA structures: 1994. *Nucleic Acids Res* **22**(17), 3502-7.

Gutell, R. R., Gray, M. W. & Schnare, M. N. (1993). A compilation of large subunit (23S and 23S-like) ribosomal RNA structures: 1993. *Nucleic Acids Res* **21**(13), 3055-74.

Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., Spielmann, A. & Stutz, E. (1993). Complete sequence of Euglena gracilis chloroplast DNA. *Nucleic Acids Res* **21**(15), 3537-44.

Harmens, H., Den Hartog, P. R., Bookum, T. & Verkleij, J. A. C. (1993). Increase zinc tolerance in Silene vulgaris (Moench) Garcke is not due to increased production of phytochelatins. *Plant Physiol.* **103**, 1305-1309.

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**(6), 2879-87.

Henderson, S. T. & Petes, T. D. (1992). Instability of simple sequence DNA in Saccharomyces cerevisiae. *Mol Cell Biol* **12**(6), 2749-57.

Horner, D. S., Hirt, R. P., Kilvington, S., Lloyd, D. & Embley, T. M. (1996). Molecular data suggest an early acquisition of the mitochondrion endosymbiont. *Proc R Soc Lond B Biol Sci* **263**(1373), 1053-9.

Howden, R., Goldsbrough, P. B., Andersen, C. R. & Cobbett, C. S. (1995). Cadmium-sensitive, cad1 mutants of Arabidopsis thaliana are phytochelatin deficient. *Plant Physiol* **107**(4), 1059-66.

Howe, G. & Merchant, S. (1992). Heavy metal-activated synthesis of peptides in *Chlamydomonas reinhardtii*. *Plant Physiol.* **98**, 127-136.

Hur, M., Geese, W. J. & Waring, R. B. (1997). Self-splicing activity of the mitochondrial group-I introns from Aspergillus nidulans and related introns from other species. *Curr Genet* **32**(6), 399-407.

Hur, M. & Waring, R. B. (1995). Two group I introns with a C.G basepair at the 5' splice-site instead of the very highly conserved U.G basepair: is selection post- translational? *Nucleic Acids Res* **23**(21), 4466-70.

Jaeger, J. A., Turner, D. H. & Zuker, M. (1989). Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci U S A* **86**(20), 7706-10.

Jaeger, J. A., Turner, D. H. & Zuker, M. (1990). Predicting optimal and suboptimal secondary structure for RNA. *Methods Enzymol* **183**, 281-306.

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem Sci* **23**(10), 403-5.

Johansen, S., Embley, T. M. & Willassen, N. P. (1993). A family of nuclear homing endonucleases. *Nucleic Acids Res* **21**(18), 4405.

Jones, K. C., Jackson, A. & Johnston, A. E. (1992). Evidence for an increase in the Cd content of herbage since 1860's. *Environ. Sci. Technol.* **26**, 834-836.

Kabata-Pendias, A. & Pendias, H. (1992). Trace elements in soils and plants., pp. Chapt. 5. CRC Press, Boca Raton, FL.

Keller, L. R. (1995). Enrichment of mRNA encoding flagellar proteins. *Methods Cell Biol* **47**, 551-7.

Kesseler, A. & Brand, M. D. (1994). Effects of cadmium on the control and internal regulation of oxidative phosphorylation in potato tuber mitochondria. *Eur J Biochem* **225**(3), 907-22.

Kneer, R. & Zenk, M. H. (1992). Phytochelatins protect plant enzymes from heavy metal poisoning. *Phytochemistry* **31**, 2663-2667.

Kroymann, J. & Zetsche, K. (1998). The mitochondrial genome of Chlorogonium elongatum inferred from the complete sequence. *J Mol Evol* **47**(4), 431-40.

Kuboi, T., Noguchi, A. & Yazaki, J. (1986). Family-dependent cadmium accumulation characteristics in higher plants. *Plant Soil* **92**, 405-415.

Kuck, U., Godehardt, I. & Schmidt, U. (1990). A self-splicing group II intron in the mitochondrial large subunit rRNA (LSUrRNA) gene of the eukaryotic alga Scenedesmus obliquus. *Nucleic Acids Res* **18**(9), 2691-7.

Lambowitz, A. M. & Belfort, M. (1993). Introns as mobile genetic elements. *Annu Rev Biochem* **62**, 587-622.

Lambowitz, A. M. & Perlman, P. S. (1990). Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends Biochem Sci* **15**(11), 440-4.

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 [see comments]. *Nature* **387**(6632), 493-7.

Lazowska, J., Jacq, C. & Slonimski, P. P. (1980). Sequence of introns and flanking exons in wild-type and box3 mutants of cytochrome b reveals an interlaced splicing protein coded by an intron. *Cell* **22**(2 Pt 2), 333-48.

Lee, J. G. & Morel, F. M. (1995). Replacement of zinc by cadmium in marine phytoplankton. *Marine Ecology Progress Series* **127**(1-3), 305-309.

Lee, R. W., Dumas, C., Lemieux, C. & Turmel, M. (1991). Cloning and characterization of the Chlamydomonas moewusii mitochondrial genome. *Mol Gen Genet* **231**(1), 53-8.

Lisacek, F., Diaz, Y. & Michel, F. (1994). Automatic identification of group I intron cores in genomic DNA sequences. *J Mol Biol* **235**(4), 1206-17.

Lykke-Andersen, J., Aagaard, C., Semionenkov, M. & Garrett, R. A. (1997). Archaeal introns: splicing, intercellular mobility and evolution. *Trends Biochem Sci* **22**(9), 326-31.

Ma, D. P., King, Y. T., Kim, Y., Luckett, W. S., Jr., Boyle, J. A. & Chang, Y. F. (1992). Amplification and characterization of an inverted repeat from the Chlamydomonas reinhardtii mitochondrial genome. *Gene* **119**(2), 253-7.

Macnair, M. R. (1993). Tansley Review No. 49. The genetics of metal tolerance in vascular plants. *New Phytol.* **124**, 541-559.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular cloning. A laboratory manual.*, Cold Spring Harbor: Cold Spring Harbor Laboratory.

Margulis, L. (1975). Symbiotic theory of the origin of eukaryotic organelles; criteria for proof. *Symp Soc Exp Biol* **29**, 21-38.

Margulis, L. (1976). Genetic and evolutionary consequences of symbiosis. *Exp Parasitol* **39**(2), 277-349.

Masters, B. A., Kelly, E. J., Quaife, C. J., Brinster, R. L. & Palmiter, R. D. (1994). Targeted disruption of metallothionein I and II genes increases sensitivity to cadmium. *Proc Natl Acad Sci U S A* **91**(2), 584-8.

McClain, W. H. (1993). Rules that govern tRNA identity in protein synthesis. *J Mol Biol* **234**(2), 257-80.

McHugh, J. P. & Spanier, J. G. (1994). Isolation of cadmium sensitive mutants in Chlamydomonas reinhardtii by transformation/insertional mutagenesis. *FEMS Micro. Letters* **124**, 239-244.

Michaelis, G., Vahrenholz, C. & Pratje, E. (1990). Mitochondrial DNA of Chlamydomonas reinhardtii: the gene for apocytochrome b and the complete functional map of the 15.8 kb DNA. *Mol Gen Genet* **223**(2), 211-6.

Michalska, A. E. & Choo, K. H. (1993). Targeting and germ-line transmission of a null mutation at the metallothionein I and II loci in mouse. *Proc Natl Acad Sci U S A* **90**(17), 8088-92.

Michel, F. & Ferat, J. L. (1995). Structure and activities of group II introns. *Annu Rev Biochem* **64**, 435-61.

Michel, F., Hanna, M., Green, R., Bartel, D. P. & Szostak, J. W. (1989). The guanosine binding site of the Tetrahymena ribozyme [see comments]. *Nature* **342**(6248), 391-5.

Michel, F., Jacquier, A. & Dujon, B. (1982). Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* **64**(10), 867-81.

Michel, F. & Westhof, E. (1990). Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* **216**(3), 585-610.

Mohr, G. & Lambowitz, A. M. (1991). Integration of a group I intron into a ribosomal RNA sequence promoted by a tyrosyl-tRNA synthetase. *Nature* **354**(6349), 164-7.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mut. Res.* **1**, 2-9.

Nagel, K., Adelmeir, U. & Voigt, J. (1996). Subcellular distribution of cadmium in the unicellular green alga *Chlamydomonas reinhardtii*. *J. Plant Physiol.* **149**, 86-90.

Nagel, K. & Voigt, J. (1989). In vitro evolution and preliminary characterization of a cadmium-resistant population of *Chlamydomonas reinhardtii*. *Applied and Environ. Micro.* **55**(2), 526-528.

Nedelcu, A. M. (1997). Fragmented and scrambled mitochondrial ribosomal RNA coding regions among green algae: a model for their origin and evolution. *Mol Biol Evol* **14**(5), 506-17.

Nedelcu, A. M. & Lee, R. W. (1998). A degenerate group II intron in the intronless mitochondrial genome of Chlamydomonas reinhardtii: evolutionary implications [letter]. *Mol Biol Evol* **15**(7), 918-22.

Nosek, J., Tomaska, L., Fukuhara, H., Suyama, Y., Kovac, L. (1998). Linear mitochondrial genomes: 30 years down the line. *Trend Genet* 14(5), 184-188.

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

Ohta, E., Oda, K., Yamato, K., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Ohyama, K. & Michel, F. (1993). Group I introns in the liverwort mitochondrial genome: the gene coding for subunit 1 of cytochrome oxidase shares five intron positions with its fungal counterparts. *Nucleic Acids Res* **21**(5), 1297-305.

Palmer, J. D. (1997). Organelle genomes: going, going, gone! *Science* **275**(5301), 790-1.

Palmer, J. D. & Herbon, L. A. (1988). Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J Mol Evol* **28**(1-2), 87-97.

Paquin, B. & Lang, B. F. (1996). The mitochondrial DNA of Allomyces macrogynus: the complete genomic sequence from an ancestral fungus. *J Mol Biol* **255**(5), 688-701.

Pearson, W. R. (1994a). Using the FASTA program to search protein and DNA sequence databases. *Methods Mol Biol* **24**, 307-31.

Pearson, W. R. (1994b). Using the FASTA program to search protein and DNA sequence databases. *Methods Mol Biol* **25**, 365-89.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* **85**(8), 2444-8.

Popot, J. L. & de Vitry, C. (1990). On the microassembly of integral membrane proteins. *Annu Rev Biophys Biophys Chem* **19**, 369-403.

Robinson, N. J., Tommey, A. M., Kuske, C. & Jackson, P. J. (1993). Plant metallothioneins. *Biochem J* **295**(Pt 1), 1-10.

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 U S A* **93**(25), 14618-22.

Roman, J. & Woodson, S. A. (1995). Reverse splicing of the Tetrahymena IVS: evidence for multiple reaction sites in the 23S rRNA. *Rna* **1**(5), 478-90.

Rusconi, C. P. & Cech, T. R. (1996). The anticodon is the signal sequence for mitochondrial import of glutamine tRNA in Tetrahymena. *Genes Dev* **10**(22), 2870-80.
Ryan, R., Grant, D., Chiang, K. S. & Swift, H. (1978). Isolation and characterization of mitochondrial DNA from Chlamydomonas reinhardtii. *Proc Natl Acad Sci U S A* **75**(7), 3268-72.

Sager, R. & Granick, S. (1954). Nutritional control of sexuality in *Chlamydomonas reinhardtii. J. Gen. Physiol.* **37**, 729-742.

Salganik, R. I., Dudareva, N. A. & Kiseleva, E. V. (1991). Structural organization and transcription of plant mitochondrial and chloroplast genomes. *Electron Microsc Rev* **4**(2), 221-47.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A laboratory Manual*. Second edit, Cold Spring Harbor Laboratory Press.

Saraste, M. (1999). Oxidative Phosphorylation at the fin de sicle. *Science* **283**(5407), 1488-1493.

Saraste, M. & Castresana, J. (1994). Cytochrome oxidase evolved by tinkering with denitrification enzymes. *FEBS Lett* **341**(1), 1-4.

Schmelzer, C., Schmidt, C., May, K. & Schweyen, R. J. (1983). Determination of functional domains in intron bI1 of yeast mitochondrial RNA by studies of mitochondrial mutations and a nuclear suppressor. *Embo J* **2**(11), 2047-52.

Seilhamer, J. J., Gutell, R. R. & Cummings, D. J. (1984). Paramecium mitochondrial genes. II. Large subunit rRNA gene sequence and microevolution. *J Biol Chem* **259**(8), 5173-81.

Shub, D. A., Goodrich-Blair, H. & Eddy, S. R. (1994). Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. *Trends Biochem Sci* **19**(10), 402-4.

Silver, S. (1996). Bacterial resistances to toxic metal ions--a review. *Gene* **179**(1), 9-19.

Silver, S. & Keach, D. (1982). Energy-dependent arsenate efflux: the mechanism of plasmid-mediated resistance. *Proc Natl Acad Sci U S A* **79**(20), 6114-8.

Silver, S. & Misra, T. K. (1988). Plasmid-mediated heavy metal resistances. *Annu Rev Microbiol* **42**, 717-43.

Silver, S. & Walderhaug, M. (1992). Gene regulation of plasmid- and chromosome-determined inorganic ion transport in bacteria. *Microbiol Rev* **56**(1), 195-228.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**(3), 503-17.

Spanier, J. G., Graham, J. E. & Jarvik, J. W. (1992). Isolation and characterization of three new field strains of Chlamydomonas reinhardtii (Chlorophyta). *J. Phycology* **28**, 822-828.

Sprinzl, M., Steegborn, C., Hubel, F. & Steinberg, S. (1996). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **24**(1), 68-72.

Srivastava, A. K. & Schlessinger, D. (1990). Mechanism and regulation of bacterial ribosomal RNA processing. *Annu Rev Microbiol* **44**, 105-29.

Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair [see comments] [published erratum appears in Nature 1994 Apr 7;368(6471);569]. *Nature* **365**(6443), 274-6.

Straton, G. W. & Corke, C. T. (1979). The effect of cadmium ion on growth, photosynthesis, and nitrogenase activity of *Anabaena inaequalis*. *Chemosphere* **5**, 277-282.

Summers, A. O. (1986). Organization, expression, and evolution of genes for mercury resistance. *Annu Rev Microbiol* **40**, 607-34.

Summers, A. O. & Silver, S. (1972). Mercury resistance in a plasmid-bearing strain of Escherichia coli. *J Bacteriol* **112**(3), 1228-36.

Suyama, Y. (1986). Two dimensional polyacrylamide gel electrophoresis analysis of Tetrahymena mitochondrial tRNA. *Curr Genet* **10**(5), 411-20.

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. (1983). The double-strand-break repair model for recombination. *Cell* **33**(1), 25-35.

Tang, D. K., Qiao, S. Y. & Wu, M. (1995). Insertion mutagenesis of Chlamydomonas reinhardtii by electroporation and heterologous DNA. *Biochem Mol Biol Int* **36**(5), 1025-35.

Thompson, A. J. & Herrin, D. L. (1994). A chloroplast group I intron undergoes the first step of reverse splicing into host cytoplasmic 5.8 S rRNA. Implications for intron-mediated RNA recombination, intron transposition and 5.8 S rRNA structure. *J Mol Biol* **236**(2), 455-68.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**(24), 4876-82.

Thorsness, P. E. & Fox, T. D. (1990). Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae. *Nature* **346**(6282), 376-9.
Thorsness, P. E. & Weber, E. R. (1996). Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. *Int Rev Cytol* **165**, 207-34.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. & Yoshikawa, S. (1995). Structures of metal sites of oxidized bovine heart cytochrome c oxidase at 2.8 A [see comments]. *Science* **269**(5227), 1069-74.

Turmel, M., Mercier, J. P. & Cote, M. J. (1993). Group I introns interrupt the chloroplast psaB and psbC and the mitochondrial rrnL gene in Chlamydomonas. *Nucleic Acids Res* **21**(22), 5242-50.

Twiss, M. R. (1990). Copper tolerance of *Chlamydomonas acidophila* (Cholophyceae) isolated from acidic, copper-contaminated soils. *J. Phycol.* **26**, 655-659.

Vahrenholz, C., Riemen, G., Pratje, E., Dujon, B. & Michaelis, G. (1993a). Mitochondrial DNA of *Chlamydomonas reinhardtii*: the structure of the ends of the

linear 15.8-kb genome suggests mechanisms for DNA replication. *Curr Genet* **24**, 241-247.

Vahrenholz, C., Riemen, G., Pratje, E., Dujon, B. & Michaelis, G. (1993b). Mitochondrial DNA of Chlamydomonas reinhardtii: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication. *Curr Genet* **24**(3), 241-7.

van Belkum, A., Scherer, S., van Alphen, L. & Verbrugh, H. (1998). Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* **62**(2), 275-93.

Voigt, J., Nagel, K. & Wrann, D. (1998). A cadmium-tolerant *Chlamydomonas* mutant strain impaired in photosystem II activity. *J. Plant Physiol.* **153**, 566-573.

Waalkes, M. P., Coogan, T. P. & Barter, R. A. (1992). Toxicological principles of metal carcinogenesis with special emphasis on cadmium. *Crit Rev Toxicol* **22**(3-4), 175-201.

Webb, M., Ed. (1979). The chemistry, biochemistry and biology of cadmium. Vol. 2. Basics in Enivronmental Health. Amsterdam: Elsevier/North Holland Biomedical Press.

Weiss-Brummer, B., Rodel, G., Schweyen, R. J. & Kaudewitz, F. (1982). Expression of the split gene cob in yeast: evidence for a precursor of a "maturase" protein translated from intron 4 and preceding exons. *Cell* **29**(2), 527-36.

Wolfe, K. H., Li, W. H. & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A* **84**(24), 9054-8.

Wolff, G., Burger, G., Lang, B. F. & Kuck, U. (1993). Mitochondrial genes in the colourless alga Prototheca wickerhamii resemble plant genes in their exons but fungal genes in their introns. *Nucleic Acids Res* **21**(3), 719-26.

Wolff, G., Plante, I., Lang, B. F., Kuck, U. & Burger, G. (1994). Complete sequence of the mitochondrial DNA of the chlorophyte alga Prototheca wickerhamii. Gene content and genome organization. *J Mol Biol* **237**(1), 75-86.

Wu, M. & Tinoco, I., Jr. (1998). RNA folding causes secondary structure rearrangement. *Proc Natl Acad Sci U S A* **95**(20), 11555-60.

Zenk, M. H. (1996). Heavy metal detoxification in higher plants--a review. *Gene* **179**(1), 21-30.

Zhou, J. M. & Goldsbrough, P. B. (1994). Functional homologs of fungal metallothionein genes from Arabidopsis. *Plant Cell* **6**, 875-884.

Zimmer, M., Luckemann, G., Lang, B. F. & Wolf, K. (1984). The mitochondrial genome of the fission yeast Schizosaccharomyces pombe. 3. Gene mapping in strain EF1 (CBS 356) and analysis of hybrids between the strains EF1 and ade7-50h. *Mol Gen Genet* **196**(3), 473-81.

Zuker, M. (1989). On finding all suboptimal foldings of an RNA molecule. *Science* **244**(4900), 48-52.