National Library of Canada
Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services Branch
Direction des acquisitions et des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

The Ribosomal RNA Gene Region of *Acanthamoeba castellanii* Mitochondrial DNA: Organization, Mode of Expression and Evolution

by

Kim Marie Lonergan

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

at

Dalhousie University
Halifax, Nova Scotia
October, 1993

© Copyright by Kim Marie Lonergan, 1993

National Library of Canada
Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services Branch
Direction des acquisitions et des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-93700-9

Canada

Name Kim Lonergan

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Biology (Molecular) — SUBJECT TERM

0307 — SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences

Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences

Applied Mechanics 0346
Computer Science 0984
Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

For my parents, Colin and Phyllis Lonergan

Table of Contents

	Page
Table of Contents	v
List of Figures	xi
List of Tables	xv
Abstract	xvi
List of Abbreviations	xvii
Acknowledgements	xviii
I. Introduction	1
A. Mitochondrial Genomes	2
(i). Animal mitochondria	3
(ii). Plant mitochondria	6
(iii). Chlorophyte mitochondria	8
(iv). Fungal mitochondria	9
(v). Ciliate protozoan mitochondria	11
(vi). Kinetoplastid protozoan mitochondria	12
(vii). Summary	13
B. Ribosomal RNAs	14
(i). Animal mitochondrial rRNAs	15
(ii). Plant Mitochondrial rRNAs	16
(iii). Chlorophyte mitochondrial rRNAs	16
(iv). Fungal mitochondrial rRNAs	17
(v). Ciliate mitochondrial rRNAs	18
(vi). Kinetoplastid mitochondrial rRNAs	18
C. Organelle Genome Evolution	
(i). Autogenous origin versus serial endosymbiosis	18
(ii). Molecular data supporting endosymbiosis	
a. Molecular phylogeny based upon rRNA sequence data	20
1. Mitochondria	21
2. Plastids	21
b. Protein sequence data supporting endosymbiosis	

1. Mitochondria 21
2. Chloroplasts 22
c. Other data supporting endosymbiosis
1. Mitochondria 22
2. Chloroplasts 23
(iii). The process of endosymbiosis 23
(iv). Summary 25

D. Mitochondrial Genome Evolution
(i). Monophyletic versus polyphyletic origin 25
(ii). Means to an answer 27

E. Introns
(i). Five different classes of introns 27
(ii). Group I introns 28
(iii). Group II introns 32
(iv). Non-ribosomal group I and II introns 34
(v). Ribosomal RNA introns 35
a. SSU rDNA introns 35
b. LSU rRNA introns 36
(vi) Intron ORFs 36
a. Group I introns 37
b. Group II introns 38
(vii). Intron evolution
a. Distribution of group I introns 38
b. Intron mobility 40

F. RNA Editing 42
(i). Editing of mRNA 42
a. Co-transcriptional editing 42
b. Post-transcriptional editing
1. Insertional editing 42
2. Substitutional editing 43
(ii). Editing of structural RNAs 46
(iii). Mechanisms 47

(iv). Specificity determinants 49

G. Transfer RNAs
(i). Structural characteristics
a. Secondary structure 51
b. Tertiary structure 53
(ii). Post-transcriptional processing
a. Endonucleolytic cleavage events 53
b. Modified nucleosides 54
(iii). Acceptor identity elements 56
(iv). Interaction with aminoacyl tRNA synthetases 57

H. Summary of Research Presented 58

II. Materials and Methods

A. Materials 59

B. Methods
(i). Culture conditions for *Acanthamoeba castellanii* 59
(ii). Preparation of nucleic acids from *Acanthamoeba castellanii* 60
a. Method 1 (DNA only) 60
b. Method 2 (DNA and RNA) 61
c. Nuclease treatment of nucleic acids 62
d. Purification of nucleic acids by density-gradient centrifugation 62
(iii). Cloning and sequencing of *A. castellanii* mitochondrial DNA
a. Plasmid cloning 63
1. Competent *E. coli* cells 64
2. Transformation of *E. coli* with plasmids 64
3. Alkaline plasmid mini-preps 65
4. Plasmid clones obtained 66
5. Preparation of plasmid dsDNA template for sequencing 66
b. M13 subcloning 67

1. Transformation of *E. coli* with M13 67
2. M13 subclones obtained 68
3. Nested deletion series 68
4. Preparation of single-stranded template DNA 71
c. DNA sequencing 71
(iv). Direct analysis of ribosomal RNAs
a. Purification by gel electrophoresis 72
b. Enzymatic sequencing of 5'-end-labeled RNA 72
c. Terminal nucleotide analysis 73
d. Chemical sequencing of 3'-end-labeled RNA 73
(v). Primer extension analysis and cDNA sequencing 74
a. Ribosomal RNAs 75
b.Transfer RNAs 75
c. Messenger RNA encoding cytochrome oxidase, subunits 1 and 2 76
(vi). Terminal deoxynucleotidyl transferase treatment for elimination of ambiguities in tRNA sequence analysis 76
(vii). Southern hybridization analysis of tRNA genes 77
(viii). Northern hybridization analysis of tRNAs 78
(ix). Potential secondary structures 79
(x). Analysis of open reading frames 79
(xi). Phylogenetic analysis of mitochondrial rRNA genes 79

III. Results 81

A. Gene Organization and Sequence 81

B. Ribosomal RNA Genes of *Acanthamoeba castellanii* Mitochondria 88
(i). Potential secondary structure of the SSU rRNA 99
(ii). Potential secondary structure of the LSU rRNA 110
(iii). Primary sequence comparisons of rRNA genes 117
a. Pairwise comparisons 117
b. Phylogenetic treeing analysis 119

C. Introns of the LSU rRNA Gene 131

(i). Secondary structures 131
(ii). Open reading frames 136
(iii). Comparison of AcLSU.m3 and CrLSU.c 137
(iv). AcLSU.m1 and AcLSU.m2 comparisons 137

D. Transfer RNA Editing 141

E. Cytochrome Oxidase 166

IV. Discussion

A. Evolutionary Affiliation of the Mitochondrial Genome of *Acanthamoeba castellanii* 180
(i). Ribosomal RNA secondary structure analysis 182
(ii). Primary sequence comparisons 182
(iii). Ribosomal RNA phylogenetic analysis 183
(iv). Phylogenetic treeing problems 184
a. *A. castellanii* 184
b. *P. wickerhamii* 185
c. *C. reinhardtii* 186
(v). Conclusions
a. Phylogenetic affiliation of *A. castellanii* mitochondria 187
b. Chlorophyte mitochondrial phylogeny 188

B. Evolution of Introns in *A. castellanii* Mitochondrial *rnl* 189

C. Transfer RNA Editing
(i). Characteristics 197
(ii). Mechanism 199
(iii). Specificity 199

D. Cytochrome Oxidase 200

E. Future Directions
(i). Ribosomal RNAs of *A. castellanii* mitochondria
a. A monophyletic origin of plant mitochondria 203

b. Problems with chlorophytes 204
c. *Chlamydomonas reinhardtii* 204
(ii). Further investigation of group I introns of *A. castellanii* mitochondria 205
(iii). Transfer RNA editing 206
(iv). Cytochrome oxidase 209

V. References 211

List of Figures

		Page
Figure 1.	Generalized structure of group I intron.	29
Figure 2.	Generalized structure of group II intron.	33
Figure 3.	Cloverleaf secondary structure (A) and tertiary structure (B) of tRNA.	52
Figure 4.	Electron micrograph of a whole *A. castellanii* cell.	82
Figure 5.	Growth curve of a suspension culture of *A. castellanii.*	84
Figure 6.	Fractionation of total cellular DNA from *A. castellanii* by CsCl density gradient centrifugation.	85
Figure 7.	Arrangement of LSU and SSU rRNA genes (*rnl* and *rns*, respectively) and flanking regions in *A. castellanii* mtDNA.	87
Figure 8.	DNA sequence of the LSU and SSU rRNA genes and flanking regions of *A. castellanii* mtDNA.	89
Figure 9.	Gel profile of M13 deletion series subclones used for dideoxy sequencing.	95
Figure 10.	Gel profile of high-molecular-weight RNA from *A. castellanii* mitochondria.	97
Figure 11.	Diagram of the potential secondary structure of mitochondrial SSU rRNA from *A. castellanii*, as deduced from the DNA sequence.	100
Figure 12.	Confirmation of 5'-terminal heterogeneity of rRNAs from *A. castellanii* mitochondria.	104
Figure 13.	Determination of the 5' termini of rRNAs from *A. castellanii* mitochondria.	106
Figure 14.	Identification of 3'-terminal nucleotides of SSU and LSU rRNAs from *A. castellanii* mitochondria.	108

Figure 15. Diagram of the potential secondary structure of mitochondrial LSU rRNA from *A. castellanii*, as deduced from the DNA sequence. 111

Figure 16. Confirmation of intron sequences within the LSU rDNA. 115

Figure 17. Mitochondrial phylogenetic tree based on SSU rRNA sequence data. 122

Figure 18. Distance (neighbor-joining) mitochondrial phylogenetic tree based on LSU rRNA sequence data. 124

Figure 19. Maximum likelihood mitochondrial phylogenetic tree based on LSU rRNA sequence data. 126

Figure 20. Maximum parsimony phylogenetic tree based on LSU rRNA sequence data. 128

Figure 21. Diagrams showing the potential secondary structures of *A. castellanii* mitochondrial *rnl* introns, as derived from the DNA sequence. 132

Figure 22. Amino acid sequence alignment of intron ORFs encoded by *A. castellanii* mitochondrial AcLSU.m3 and *C. reinhardtii* chloroplast CrLSU. 138

Figure 23. Comparison of the primary nucleotide sequence within the core region of the *rnl* introns of *A. castellanii* mitochondria (AcLSU.m3) and *C. reinhardtii* chloroplast (CrLSU). 139

Figure 24. Amino acid sequence alignment of AcLSU.m1 and AcLSU.m2 intron ORFs with those from: (A) *C. eugametos* chloroplast *rnl* (CeLSU.5); and (B and C) *Podospora anserina* mitochondrial *nad*3 (PaND3). 142

Figure 25. Organization of tRNA gene cluster 1 (A) and tRNA gene cluster 2 (B) in *A. castellanii* mtDNA. 144

Figure 26. Potential secondary structure of *A. castellanii* mitochondrial proline tRNA as derived form the gene sequence. 146

Figure 27. Potential secondary structure of *A. castellanii* aspartate tRNA as derived from the gene sequence. 147

Figure 28. Potential secondary structures of cluster-2 tRNAs encoded by *A. castellanii* mtDNA, as derived from the gene sequence. 148

Figure 29. Potential secondary structures of *A. castellanii* mitochondrial cluster-1 tRNAs as derived from the gene sequence. 150

Figure 30. Autoradiograms showing the results of northern hybridization analysis of cluster-2 tRNA gene products using gene-specific oligonucleotide probes. 152

Figure 31. Comparison of tRNA gene sequence and tRNA mature transcript sequence. 154

Figure 32. Potential secondary structure of *A. castellanii* mitochondrial aspartate tRNA showing deduced editing sites. 157

Figure 33. Potential secondary structures of *A. castellanii* mitochondrial cluster 2 tRNAs as derived from the gene sequence, showing changes induced by editing. 158

Figure 34. Southern hybridization analysis of the $tRNA^{Ala}$ gene of cluster 2 using a gene-specific oligonucleotide probe. 160

Figure 35. Southern hybridization analysis of Met1, Asp and Met2 tRNA genes of *A. castellanii* mtDNA cluster 2 using gene-specific oligonucleotide probes. 163

Figure 36. Summary of confirmed (uppercase) and predicted (lowercase) editing within the acceptor stems of tRNAs of gene cluster 1 (A) and cluster 2 (B) in *A. castellanii* mitochondria. 165

Figure 37. Comparison of $tRNA^{Met1}$ and $tRNA^{Ala}$ gene sequences and the corresponding tRNA precursor sequences. 167

Figure 38. Amino acid sequence alignment of *A. castellanii* mitochondrial *cox*1, as inferred from the gene sequence. 169

Figure 39. Amino acid sequence alignment of *A. castellanii* mitochondrial *cox*2, as inferred from the gene sequence. 172

Figure 40. Codon usage for *A. castellanii* mitochondrial *cox*1 (A) and *cox*2 (B) as determined from the gene sequences. 174

Figure 41. Autoradiogram of cDNA sequence flanking the 5'-terminal coding region of the *cox*1 transcript from *A. castellanii* mitochondria. 175

Figure 42. Autoradiogram of cDNA sequence flanking the 5'-terminal coding region of the *cox*2 transcript from *A. castellanii* mitochondria. 177

Figure 43. Codon usage in CrLSU.c intron ORF (A), and *C. reinhardtii* chloroplast protein-coding genes for *pet*A (B) and *rps*12 (C). 194

Figure 44. Hydropathy (Kyte Doolittle) plots of *A. castellanii* Cox2 (A), and *Triticum aestivum* (wheat) Cox2 (B). 202

List of Tables

		Page
Table 1.	Lengths (nt) of variable regions in SSU and LSU rRNAs.	103
Table 2.	Degree of identity (%) in pairwise comparisons within universal regions of mitochondrial and eubacterial SSU rRNA sequences.	118
Table 3.	Degree of identity (%) in pairwise comparisons within conserved regions of mitochondrial and eubacterial LSU rRNA sequences.	120

Abstract

For this thesis, the organization, expression and evolutionary affiliations of the mitochondrial genome in the non-photosynthetic amoeboid protozoan, *Acanthamoeba castellanii*, have been investigated. A specific question that was posed was, "Is the mitochondrial genome of *A. castellanii* more closely affiliated with that of land plants or with that of the chlorophycean alga, *Chlamydomonas reinhardtii*?". (The latter share a recent common nuclear ancestor but differ strikingly in mitochondrial genome sequence and organization.) To answer this question, the DNA sequence of the *A. castellanii* mitochondrial rRNA genes and flanking regions has been determined. The LSU rRNA gene is flanked by two clusters of tRNA genes, cluster 1 (upstream) and cluster 2 (downstream), whereas the SSU rRNA gene is flanked by tRNA gene cluster 2 (upstream) and the cytochrome oxidase subunit I gene (downstream), with very little intergenic space and with all of these genes in the same transcriptional orientation. The tRNAs of cluster 2, and likely those of cluster 1, are altered by a novel type of RNA editing such that a conserved tRNA secondary structure is restored. The *Acanthamoeba* rRNA genes display primary sequence and secondary structure elements strikingly similar to those of their eubacterial homologs. In this regard, they resemble much more closely the mitochondrial rRNA genes of land plants than those of *C. reinhardtii*. Phylogenetic analysis suggests a monophyletic origin of mitochondria from within the α-Proteobacteria. There are no intronic sequences within the SSU rRNA gene, but the LSU rRNA gene harbors three group I introns, all located within the 3' portion of the RNA. One of these introns is found at precisely the same nucleotide position as the group I intron of the LSU rRNA gene of *C. reinhardtii* chloroplast. These two introns appear to be structurally homologous, raising questions about their evolutionary origin. A cytochrome oxidase subunit 2 gene is located immediately downstream of the cytochrome oxidase subunit 1 gene and it is likely that these two genes are co-transcribed and, perhaps, co-translated.

List of Abbreviations

A, adenine
aa, amino acid
bp, base pair
C, cytosine
° C, degree Celsius
dNTP, 2'-deoxynucleoside-5'-triphosphate
ddNTP, 2', 3'-dideoxynucleoside-5'-triphosphate
DNA, deoxyribonucleic acid
ds, double-stranded
EDTA, ethylenediaminetetraacetate
g, gram
g, gravity
G, guanine
h, hour
I, inosine
kbp, kilobase pair
kDa, kiloDalton
μl, microliter
μM, micromolar
μg, microgram
mg, milligram
ml, milliliter
mM, millimolar
m, minute
mt, mitochondrial
M, molar
n, nuclear
ng, nanogram
nm, nanometer
nM, nanomolar
nt, nucleotide
rpm, revolutions per minute
N, any nucleotide
R, purine
RNA, ribonucleic acid
r, ribosomal
s, sec
ss, single-stranded
SDS, sodium dodecyl sulfate
T, thymine
Tris, tris(hydroxymethyl)-aminomethane
UV, ultraviolet
U, uracil
vol., volume
Y, pyrimidine

Acknowledgements

I take this opportunity to formally thank those people who have helped me, in one way or another, to meet the challenge of completing the research described in this thesis and in obtaining this degree.

I thank my supervisor, Dr. Mike Gray for allowing me the opportunity to do research in his laboratory. I am particularly appreciative of the freedom given me to work according to my own schedule and of the patience and supervision that Dr. Gray has displayed during this time period.

I would like to thank the external examiner, Dr. D. Wolstenholme and the other members of the examining committee, Dr. R. Lee, Dr. C. Helleiner and Dr. P. Liu for their time spent in reviewing this thesis.

I am grateful to other past and present members of the Gray lab, with special reference to Murray Schnare and David Spencer (two long-term friends), for making my time spent here more enjoyable than it otherwise might have been. Murray is the person who convinced me to enter a Ph. D. program in Biochemistry (although he says that he does not remember doing so). David Spencer's tendency to 'drop whatever he is doing' to help other people has certainly benefited me. This was especially apparent during the period of phylogenetic treeing analysis which was conducted during the latter phase of this program. Pam, Paul, Pat, Mike, Kathy, Spencer and Tatsuya have all helped at various times throughout the duration of the program.

It is my pleasure to extend a very special " thankyou " to Wan Lam for being my constant companion despite the fact that we have been physically apart for the past two years. Wan not only has provided me with frequent technical advice, but has continuously given me the inspiration and motivation that is so important when a challenge of this magnitude is undertaken.

I thank all members of my immediate family for support and words of encouragement which helped me much more than what they could realize.

I acknowledge that the final versions of the ribosomal RNA potential secondary structures were kindly provided by R. R. Gutell. The presented versions of the intron potential secondary structures were critiqued by F. Michel.

Financial support from the Walter C. Sumner scholarship is gratefully acknowledged.

I. Introduction

Mitochondria and plastids are membrane-bounded organelles found within the cytoplasm of eukaryotic cells. Most eukaryotic cells possess mitochondria which provide the cell with aerobic respiration through oxidative phosphorylation, resulting in the generation of ATP. In addition to mitochondria, plant and algal cells have plastids, which provide the cell with photosynthetic ability. Within these organelles are genomes that have the form of relatively small molecules of typically circular double-stranded DNA, from which proteins characteristic of the distinct and specialized functions performed by mitochondria and plastids are produced. These organelles are critically dependent upon the nuclear genome of the cell for the replication and expression of their own genomes and thus for executing their specialized functions. Theories pertaining to the origin and evolution of mitochondria and plastids represent a frequently debated area of evolutionary molecular biology. The research presented in this thesis addresses this evolutionary issue.

I begin this Introduction with an overview of mitochondrial genome structure and coding potential from representative eukaryotic phyla, followed by a description of mitochondrial ribosomal RNA genes. Next, I describe opposing theories proposed to explain the origin of plastids and mitochondria within eukaryotic cells, followed by a description of molecular data that lend support to one of these two theories. This is followed by an analysis of some of the pecularities that persist in our understanding of mitochondrial genome evolution, culminating in the rationale for this particular research project. In the latter sections of the Introduction I discuss specific aspects of mitochondrial gene expression that have particular relevance to other areas of the research project presented in this thesis; these include a description of the evolution of introns within organellar genomes and of RNA editing. This is followed by a general

description of transfer RNAs that is pertinent to the RNA editing data presented in this thesis.

A. Mitochondrial Genomes

A typical eukaryotic cell contains hundreds of mitochondria, with many genome equivalents per mitochondrion. As mentioned above, mitochondria provide cellular ATP via oxidative phosphorylation. This process is catalyzed by multisubunit protein complexes I to IV of the electron transport chain together with complex V, all of which are integrated into the inner mitochondrial membrane along with the electron carriers ubiquinone and cytochrome *c*. In human mitochondria, complex I (NADH dehydrogenase) consists of more than 40 protein subunits; complex II (succinate-ubiquinone reductase), 4 subunits; complex III (ubiquinol cytochrome *c* oxidoreductase), 11 subunits; complex IV (cytochrome oxidase), 13 different subunits; and complex V (ATP synthase), 12 polypeptides (Wallace, 1986).

Functioning mitochondria are crucial for the survival of eukaryotic cells in both unicellular and multicellular organisms. The essential importance of mitochondrial function for multicellular organisms is made evident by the characterization of severe degenerative disease syndromes shown to be the result of human mitochondrial genetic aberrations (Wallace, 1993).

Mitochondrial genomes range in size from 14 kbp (some animals) to as much as 2400 kbp (some land plants). Mitochondrial DNAs (mtDNAs) are usually circular but in some organisms are linear molecules. The DNA of mitochondrial genomes has no associated histone proteins such as are found in the chromosomes of nuclei.

Many mitochondrial genomes have been completely sequenced and many more are in the process of being sequenced. All mitochondrial genomes encode a variable number of protein subunits of the electron transport chain as well as small subunit (SSU) and large subunit (LSU) ribosomal RNAs (rRNAs). The complete

characterization of a number of phylogenetically diverse mitochondrial genomes should greatly increase our awareness and understanding of mitochondrial genome evolution.

(i). Animal mitochondria

(For a recent review of animal mitochondrial genomes see Woistenholme, 1992.) The mitochondrial genome of multicellular animals (metazoa) ranges in size from 14 kbp to 42 kbp and is (almost always) in the form of double-stranded, covalently closed DNA. Many of these genomes from various phyla have been completely sequenced (Bibb et al., 1981; Clary and Wolstenholme, 1985; Okimoto et al., 1991; Okimoto et al., 1992) and are notable for an almost complete absence of non-coding DNA, which makes them the most economical of mitochondrial genomes. The coding potential of animal mtDNAs is highly conserved. The 16,569-kbp human mitochondrial genome encodes 13 polypeptides (all involved in oxidative phosphorylation), LSU and SSU rRNAs, and 22 transfer RNAs (tRNAs). Protein-coding genes include those specifying three subunits (I, II and III) of cytochrome oxidase in complex IV (*cox*1, *cox*2, and *cox*3), apocytochrome *b* in complex III (*cob*), subunits 6 and 8 of the ATPase complex V (*atp*6, *atp*8), and 7 subunits of the NADH dehydrogenase in complex I (*nad*1, *nad*2, *nad*3, *nad*4, *nad*5, *nad*6, *nad*4L) (Wallace, 1993). The only known animal mitochondrial genomes that do not encode all of these 13 polypeptides are those of nematode worms and *Mytilus edulis*, both of which lack *atp*8 (Hoffmann et al., 1992). Some novel proteins encoded by animal mitochondrial DNA in specific phyla include a potential mis-match repair enzyme in the Cnidaria and an unidentified open reading frame (ORF) in one of the nematodes (Wolstenholme, 1992).

Gene order is identical within the mammalian, *Xenopus* and fish mitochondrial genomes (Tzeng et al., 1992), but otherwise differs among phyla and sometimes within phyla (nematodes) (Wolstenholme, 1992). There are three exceptions to the circular molecules of animal mtDNA and these occur in the hydra (phylum Cnidaria), where

two 8 kbp linear molecules constitute the mitochondrial genome (Warrior and Gall, 1985), and in Scyphozoa and Cubozoa (Bridge et al., 1992).

The only non-coding region found within mammalian and amphibian mitochondrial genomes contains the two promoters directing transcription from both strands [designated heavy (H) and light (L) according to base composition] (reviewed by Clayton, 1992). The origin of H-strand replication is contained within the so-called 'D-loop region' of the non-coding portion of the mtDNA. Size differences among animal mitochondrial genomes can usually be attributed to variation in the length of the D-loop region. For mammalian mtDNAs, genes from each strand are co-transcribed and mature transcripts are processed from precursor transcripts by endonucleolytic cleavages. Processing signals might consist of tRNA sequences in some animal mitochondria, but obviously cannot in others. In some cases of overlapping genes (*atp*6, *atp*8; *nad*4, *nad*4L), bicistronic transcripts have been detected (Ojala et al., 1981; Fearnley and Walker, 1986).

In some cases, A residues are added to the 3'-termini of processed transcripts so that translational termination codons are created (Ojala et al., 1981). With the exception of a recently reported example of tRNA editing (Janke and Pääbo, 1993), this is the only example in animal mitochondria of post-transcriptional processing that resembles RNA editing observed in some other mitochondrial systems (see F. RNA editing).

With one exception the 22 mitochondrially-encoded tRNAs of all animal mitochondria are sufficient to decode the mRNAs specifying the 13 mitochondrially encoded proteins; this is accomplished by unusual wobbling at the first position of the anticodon such that one tRNA species is able to recognize a four-codon family (see G. tRNAs). Thus it would not be necessary to import tRNAs from the cytoplasm into mitochondria. The vast majority of animal mitochondrial tRNAs have unusual primary sequences and secondary structures within the dihydrouridine (DHU) and the T-

pseudouridine-C (TΨC) loops (Wolstenholme, 1992). This is especially pronounced in the nematodes where the DHU arm and/or the TΨC arm are either completely missing or severely truncated, indicating a minimal functional tRNA structure (Okimoto and Wolstenholme, 1990).

Protein-coding transcripts in animal mitochondria deviate from the universal genetic code in a number of respects (Wolstenholme, 1992). Initiation appears to be specified by a variety of codons specific to particular animal phyla; the only two that appear to use AUG exclusively for initiation are Amphibia and Cnidaria. AUA and AUU as well as AUG seem to direct initiation in most animal mitochondria, and AUC, GUG, UUG and GUU are apparently used as initiation codons in certain animal phyla. It is suspected that AGA and AGG are recognized as termination codons in vertebrate mitochondria, but are recognized as serine codons in all invertebrate mitochondria except Cnidaria (where these codons specify arginine). In the Echinodermata and the Platyhelminthes, AAA codes for Asn instead of Lys. UGA is recognized as tryptophan rather than as a termination codon in all animal mitochondria.

The mitochondrial genome of the Cnidaria (sea anemone, hydra), a very primitive animal phylum, is unusual among animal mitochondrial genomes in a number of respects (Wolstenholme, 1992). Only two tRNAs are encoded by the mitochondrial genome of the sea anemone, *Metridium senile*, thus necessitating import of cytosolic tRNAs to support mitochondrial protein synthesis. The mtDNA-encoded two tRNAs have standard primary sequence and potential secondary stucture, in contrast to the majority of other animal mitochondrial tRNAs. This is also the only known animal mitochondrial genome to have introns, of which there are two (Beagley et al., 1992). Cnidarian mitochondria follow the universal genetic code except that, as in other animal mitochondria, UGA specifies tryptophan instead of termination. As mentioned previously, the mitochondrial genome of the hydra is one of the few linear animal mitochondrial genomes known.

Animal mitochondrial genomes have an accelerated rate of primary sequence evolution thought to reflect the absence of a functioning mitochondrial DNA repair system; animal nuclear genes evolve at a much slower rate (Brown et al., 1982; Vawter and Brown, 1986; Monnerot et al., 1990). However, there is evidence to suggest that a mis-match repair process might be functional in the Cnidaria (Beagley et al., 1992).

(ii). Plant mitochondria

Photosynthetic multicellular organisms comprise the non-vascular plants or bryophytes and the vascular plants, including angiosperms (flowering plants). The plant mitochondrial genome varies greatly in size (187 to 2400 kbp) and undoubtedly represents the largest of the mitochondrial genomes. The only plant mitochondrial genome that has been completely sequenced is that of the bryophyte, *Marchantia polymorpha*, a liverwort: at 187 kbp, it represents the smallest characterized plant mitochondrial genome (Oda et al., 1992). Although complete sequence information is missing for mitochondrial genomes from any of the flowering plants, the available data indicate that there are major differences between the liverwort and angiosperm mitochondrial genomes, as outlined below.

The topology of the higher plant mitochondrial genome is distinctive: this genome consists of a series of subgenomic circular molecules presumably generated from a hypothetical master circle by homologous recombination thought to be initiated by a series of direct and inverted repeats within the mtDNA (Lonsdale et al., 1984; Palmer and Shields, 1984). The complexity of the topology depends upon the number of repeated sequences contained within the master circle. The propensity for recombination leads to considerable variation in genome arrangement among various plant species. Recombination appears not to occur within the liverwort mitochondrial genome although some repeat elements are found within this genome (Oda et al., 1992).

Much of the "extra" DNA of land plant mitochondrial genomes is thought to be non-coding, intergenic spacer DNA, and thus it is suspected that the coding capacity of plant mitochondrial genomes is not much greater than that of animal mitochondrial genomes. However, sequencing of the liverwort mtDNA indicates that as many as 94 potential genes are encoded by this genome, consisting of 30 identified proteins and a number of unidentified open reading frames (ORFs). In addition to those proteins known to be encoded by the animal mitochondrial genome (with the exception of *atp*8), the liverwort mtDNA encodes *atp*A, *atp*9 and many small subunit and large subunit ribosomal proteins (*rps* and *rpl*) (Oda et al., 1992). It is likely that both bryophyte and metaphyte mitochondrial genomes encode reverse transcriptase-like proteins (Schuster and Brennicke, 1987).

Plant mitochondrial genomes obey the universal genetic code. In plant mitochondria UGA codes for termination and not tryptophan as is the case in most other mitochondrial genomes.

Both group I and group II introns are found in liverwort mtDNA, but only group II introns (cis- and trans-splicing) have thus far been found in metaphyte mtDNA (see E. Introns).

Plant mitochondrial DNA has a relatively slow rate of divergence at the primary sequence level (Wolfe et al., 1987; Palmer and Herbon, 1988). Primary sequence comparisons between liverwort and angiosperm mitochondrial genomes have not yet been published.

Gene organization in *Marchantia* mitochondria is condensed, with genes transcribed from both strands of the DNA, whereas in angiosperm mitochondria the genes are typically dispersed throughout the mitochondrial genome with little conservation of gene order [one notable exception is the conserved linkage of 18S and 5S rRNA genes (Huh and Gray, 1982).]. The ribosomal protein genes of the *Marchantia* mitochondrial genome are mostly found within two clusters, one

consisting of 12 genes arrayed in the same order as the *Escherichia coli* ribosomal protein genes of the S10, *spc* and α operons.

The mitochondrial tRNA genes of *Marchantia* tend to be located within a few major clusters while in angiosperm mitochondria they are more dispersed throughout the genome. It is known that at least a few of the angiosperm mitochondrial tRNAs are nucleus-encoded and imported into the mitochondria (Maréchal-Drouard et al., 1988). The liverwort mitochondrial genome encodes 29 tRNAs and it is suspected that at most, two additional tRNAs are imported into mitochondria from the cytoplasm (Oda et al., 1992).

The wheat mitochondrial genome contains tRNA genes whose sequences indicate an origin in the chloroplast, and it appears that chloroplast-like genes produce functional mitochondrial tRNAs (Joyce and Gray, 1989). This contribution of chloroplast DNA to the plant mitochondrial genome can be quite extensive (Stern and Palmer, 1984). There is no indication of chloroplast-like sequences in the mitochondrial genome of *Marchantia*, indicating that transfer of genes from the chloroplast to the plant mitochondrion likely occurred after the separation of bryophytes and angiosperms.

Plant mitochondrial mRNAs are extensively edited by C-to-U and sometimes U-to-C conversion (see F. RNA editing). Again, the mitochondrial genome of the liverwort is distinct from that of higher plants in that it seems unlikely that liverwort mitochondrial mRNAs are edited (Oda et al., 1992).

For more details on plant mitochondrial genomes see reviews by Levings and Brown, 1989; Gray et al., 1992; Hanson and Folkerts, 1992.

(iii). Chlorophyte mitochondria

The only molecular data available for the mtDNAs of photosynthetic, unicellular eukaryotes comes from within the chlorophycean algal lineage, which displays substantial variation regarding mitochondrial genome organization and expression.

The mitochondrial genome of the chlorophyte *Chlamydomonas reinhardtii*, a 15.8 kbp linear DNA, has been completely sequenced (Boer and Gray, 1988). All genes are tightly linked and some standard mitochondrial protein-coding genes, including *cox*2 and *cox*3 as well as all those encoding subunits of the ATPase complex, are missing. This genome encodes *cox*1 and *cob* genes, several subunits of NADH dehydrogenase (*nad*1, 2, 4, 5, 6) and a reverse transcriptase-like protein (*rtl*). It appears that the majority of *C. reinhardtii* mitochondrial tRNAs are nucleus-encoded and imported into the organelle, as only 3 tRNAs are encoded by the mtDNA. There are no introns within this genome, whose most distinctive feature is the fragmented arrangement of rRNA genes (see B. Ribosomal RNAs). There appears to be variation in mtDNA organization within the genus *Chlamydomonas* (Denovan-Wright and Lee, 1992).

The only other reports of chlorophyte mitochondrial genomes analyzed by gene sequencing are those from *Prototheca wickerhamii* (Wolff and Kück, 1990; Wolff et al., 1993), and *Scenedesmus obliquus* (Kück et al., 1990). Only the sequence of *cox*1, and partial sequences of LSU rRNA and SSU rRNA genes have been published for *Prototheca*, whereas a partial LSU rRNA gene sequence is known for *Scenedesmus*.

(iv). Fungal mitochondria

The fungi are classified as zygomycetes, basidomycetes and filamentous and non-filamentous ascomycetes. Mitochondrial genome sizes vary within the extreme range of 17.3 kbp (for the ascomycetous fission yeast, *Schizosaccharomyces pombe*) to 176 kbp (for the basidiomycete, *Agaricus bitorquis*), but more typical sizes among the various phyla are 30 to 80 kbp of circular double-stranded DNA. Fungal mitochondrial genomes have been investigated in the most detail among ascomycetes, with the complete genome sequence of *Podospora anserina* being available (Cummings et al., 1990). Ascomycete mitochondrial genomes are noted for their abundance of both

group I and group II intron sequences within protein-coding and structural genes. The presence or absence of introns accounts for the extreme size variation often observed among homologous genes from different fungal genera and within genera. For a recent review of fungal mitochondrial genomes, see Clark-Walker (1992).

Despite the significant difference in size between the human mitochondrial (16.5 kbp) and *Podospora anserina* mitochondrial (87 to 100.3 kbp) genomes, the coding capacity of each mtDNA is very similar. Each genome encodes the same set of 13 polypeptides, two rRNA genes and 22 (human) or 25 (fungal) tRNA genes. In addition, the mitochondrial genome of *Podospora* contains a few unidentified open reading frames. Such overall genome size differences are due mainly to an increase in the intergenic regions within fungal mtDNA relative to animal mtDNA and also to the larger size of individual genes in the fungal mitochondrial genome, due mainly to the presence of multiple introns within certain genes of some fungal mtDNAs. In fact there is significant size variation among mitochondrial genomes even within the filamentous ascomycetes; once again, this is due mainly to the presence of optional introns. The mitochondrial genome of *Podospora anserina* contains 36 introns in total which encompasses greater than 50% of the mitochondrial genome. Although the much smaller mitochondrial genome of *Aspergillus nidulans* (33-40 kbp) has a similar number of protein-coding genes to that found in *Podospora* mitochondrial DNA, it has only 5 introns dispersed among three genes. Many of the mitochondrial introns in the filamentous ascomycetes are optional in that their presence varies both inter- and intraspecifically (see E. Introns).

About 90% of the mitochondrial genome of the budding yeast, *Saccharomyces cerevisae*, has been sequenced, and over 50% of the genome is in the form of A+T-rich intergenic regions, with G+C clusters sequestered within these regions (DeZamaroczy and Bernardi, 1986). It is believed that the G+C clusters are mobile

elements (Weiller et al., 1989) and are involved in genome rearrangements through excision and recombination events (Clark-Walker, 1989).

Although the yeast mitochondrial genome is roughly the same size as that of the filamentous ascomycetes, it encodes only 8 polypeptides of known identity; one of these, specified by *var*1, is thought to be a ribosome-associated protein. Completely missing in this genome are any subunits of the NADH dehydrogenase complex, but a number of open reading frames of unknown genetic significance are present. Introns are found within the *cob* gene and the cytochrome oxidase genes and were among the first characterized organellar introns. Twenty-four tRNAs are encoded in the *S. cerevisiae* mitochondrial genome.

Deviations from the universal genetic code are observed within some fungal mitochondria, where UGA codes for tryptophan rather than termination in ascomycetous budding yeast and in filamentous ascomycetes. In some ascomycetous budding yeasts AUA codes for methionine instead of isoleucine and CUN codes for threonine instead of leucine (Fox, 1987).

In contrast to what is observed in mammals, there is evidence to suggest that evolution of fungal nucleus-encoded sequences is faster than that of mitochondrion-encoded sequences (Clark-Walker, 1991).

(v). Ciliate protozoan mitochondria

Most information regarding mitochondrial genomes of the protists (besides the trypanosomes and chlorophytes) is derived from studies of the ciliate protozoans, *Tetrahymena* and *Paramecium*. Both *Tetrahymena* and *Paramecium* have linear mitochondrial genomes ~50 and 41 kbp, respectively), in contrast to the circular genomes of most other mitochondria.

The mitochondrial genome of *Paramecium aurelia* has been completely sequenced (Pritchard et al., 1990) and encodes 14 identified proteins and 17 unidentified ORFs in addition to rRNA genes (see B. Ribosomal RNAs). The

identified proteins include *cox*1, *cox*2 (but not *cox*3), *atp*9 (but not *atp*6 or *atp*8), *cob*, *nad*1, *nad*2, *nad*3, *nad*4, *nad*5 (but not *nad*6 or *nad*4L), ribosomal proteins *rps*12, *rps*14, *rpl*12 and *rpl*14, and additional *nad* genes. All of the derived protein sequences of these genes are very divergent compared to their homologs from other mitochondria (reviewed by Cummings, 1992).

Only 3 tRNAs are encoded within the *Paramecium* mitochondrial genome, thus implying that many tRNAs are imported into the mitochondria from the cytosol. It also appears that the majority of mitochondrial tRNAs of *Tetrahymena* are nucleus-encoded and imported into the mitochondria (Suyama, 1986). No introns have been identified in the *Paramecium* mitochondrial genome and not much non-coding DNA is present in this genome, as indicated by transcript studies.

Similar to animal mitochondria, unusual initiation codons seem to be employed in ciliate mitochondria, including GUG, AUU, UUG and AUC, although this has not been directly demonstrated. As in most other mitochondria, UGA codes for tryptophan.

Tetrahymena has a duplicated inversion at both ends of the mitochondrial genome; this contains the LSU rRNA gene and repeated sequences (Goldbach et al., 1978). The G+C content is very low in the *Tetrahymena* mitochondrial genome (~25%), whereas the *Paramecium* mitochondrial genome has a more standard G+C content of 41%.

(vii). Kinetoplastid protozoan mitochondria

In addition to other genera, the kinetoplastid protozoans include *Trypanosoma*, *Crithidia* and *Leishmania*. Considering the early phylogenetic position of the nuclear genome of these organisms (Sogin et al., 1986), the kinetoplastid mitochondrion presumably represents an early acquisition of mitochondria within the eukaryotic lineage. Kinetoplastid mtDNA comprises 40-50 maxicircles, each ranging from 19 to 39 kbp in size in different species, plus 5,000-10,000 minicircles, with a species size

range of 0.9 to 2.5 kbp, that are catenated together within the single mitochondrion of these organisms (see Stuart and Feagin, 1992 for review). All of the mitochondrial proteins and rRNAs are encoded in the maxicircle DNA, whose gene content and order are highly conserved among kinetoplastid species. It is unlikely that the minicircle DNA encodes proteins. One of the peculiarities of kinetoplastid mitochondria is the extensive editing of mRNA transcripts (see F. RNA Editing). The guide RNAs (gRNAs) that direct the editing are encoded by both the maxicircle and minicircle DNAs (Maslov and Simpson, 1992).

The maxicircles consist of a protein-coding region comprising 15-17 kbp and a variable non-coding segment. Variations in maxicircle length among species is due largely to heterogeneity in the length and number of repeats contained within the variable region, as the genes and their order are conserved among species. Genes are tightly packed with overlapping termini in many cases and are transcribed from both strands of the DNA. Ten characteristic mitochondrial proteins and probably 8 additional unidentified proteins are encoded in maxicircle DNA; these include 5 subunits of the NADH dehydrogenase complex as well as *cox*1, *cox*2, *cox*3, *cob*, *atp*6, *rps*12 and possibly another subunit of NADH dehydrogenase. It is possible that edited versions of certain transcripts might be translated to produce altered versions at specific stages of the life cycle of the organism (Stuart and Feagin, 1992).

No tRNA genes have been found within the kinetoplastid DNA and thus the tRNAs needed for mitochondrial function are likely to be encoded in the nuclear genome and imported into the mitochondrion (Simpson et al., 1989; Hancock and Hadjuk, 1990).

(vi). Summary

Mitochondrial genomes encode LSU and SSU rRNAs, transfer RNAs (tRNAs) and a small set of proteins. Although the same basic set of respiratory chain proteins is encoded in all mitochondrial genomes, the only protein-coding genes that are invariably

present are *cox*1 and *cob*. Whereas the coding potential of mitochondrial genomes in diverse eukaryotes is similar, the organization and mode of expression varies greatly, challenging our understanding of the evolution of mitochondrial genomes.

B. Ribosomal RNAs

Because rRNA gene sequences are used most frequently for phylogenetic analysis (see C. Organelle Genome Evolution), the mitochondrial rRNAs are described here separately for each of the major eukaryotic phyla.

The LSU and SSU rRNAs, in combination with many ribosomal proteins, constitute the large and small subunits, respectively, of eubacterial, eukaryotic, archaebacterial and organellar ribosomes. Recent experimental data provide evidence supporting a catalytic role of the rRNAs in the translation of messenger RNA (mRNA) to proteins (Noller et al, 1992; Piccirilli et al., 1992; Pace 1992).

Potential secondary structures, derived from comparative analysis of rRNA gene sequence data from diverse sources and using *E. coli* 23S and 16S rRNAs as models (Gutell et al., 1992a), have been compiled for both LSU (Gutell et al., 1993) and SSU (Gutell et al., 1985; Gutell, 1993) rRNAs. The secondary structures are generated by the formation of short- and long-range, intramolecular, complementary base pairings that bring distant regions of the molecule together to form a characteristic core stucture. Those regions of the SSU rRNA that are highly conserved in secondary structure and primary sequence, even among distantly related organisms, are referred to as the universal regions (U1 to U8) (Gray et al., 1984). Other portions of the molecule that vary significantly in primary sequence and secondary structure, even among closely related orgaisms, are termed variable regions. Those regions that encompass neither the universal core nor the variable regions are referred to as semi-conserved and are found interspersed among the universal and variable regions. In the LSU rRNA, long-range intramolecular base pairings define 6 discrete structural domains (designated

I through VI). There is a general conservation of overall rRNA secondary structure among the eukaryotes, the eubacteria and the archaebacteria (Noller, 1984), presumably reflecting their descent from a common ancestor, the progenote (Woese and Fox, 1977; Darnell and Doolittle, 1986).

E. coli SSU rRNA is 1542 nt in length while the LSU rRNA is 2904 nt long; both contain a modest number of modified nucleosides (Cunningham et al., 1990), the functions of which are not known. Functional sites have been identified within both rRNAs and generally are confined to the more conserved regions of the secondary structures (Dahlberg, 1989; Raué et al., 1990; Noller, 1991, 1993). Some tertiary level interactions have been determined from comparative sequence analysis (Gutell et al., 1992a, Gutell, 1993).

(i). Animal mitochondrial rRNAs

The mitochondrial rRNA genes of all classes of vertebrates and *Drosophila* are physically linked in the order 5'-SSU rRNA, $tRNA^{Val}$, LSU rRNA-3', whereas in other metazoan mitochondrial genomes these genes are usually separated by protein-coding genes. The rRNA genes of animal mitochondria are relatively small: SSU, 955 nt and LSU, 1582 nt in the case of mouse mitochondria; they are designated 12S (SSU rDNA) and 16S (LSU rDNA). The rRNAs from the nematode, *Caenorhabditis elegans,* are exceptionally small at only 697 nt and 953 nt for the SSU and LSU rRNAs, respectively (Wolstenholme, 1992).

Potential secondary structure elements constituting the universal regions, U1 to U8, are present in the animal mitochondrial SSU rRNAs (Gutell et al., 1985), indicating that the relatively small size of these rRNAs is due to deletions within the variable regions of the rRNA genes. The LSU rRNAs from animal mitochondria deviate significantly from the *E. coli* homologs in secondary structure especially within the 5'-half of the molecule, within which the nematode sequences are limited to a few conserved regions (Gutell et al., 1990, 1992b, 1993).

(ii). Plant mitochondrial rRNAs

Both the liverwort and seed plant mitochondrial genomes encode a 5S rRNA in addition to the LSU and SSU rRNAs (Huh and Gray, 1982; Oda et al., 1992). The only other mitochondrial genome known to encode a 5S rRNA is that of the chlorophyte, *Prototheca wickerhamii* (Burger et al., 1992). Chlorophytes are thought to be ancestral to land plants in eukaryotic phylogeny (Sogin et al., 1986; Gunderson et al., 1987), and so it is intriguing that only the chlorophyte/metaphyte mitochondrial lineage appears to encode a 5S rRNA sequence. The 5S rRNA gene is found immediately downstream of the SSU rRNA gene in wheat mitochondria, but a $tRNA^{Met}$ gene is interposed between the two in the *Marchantia* mitochondrial genome; these genes are likely co-transcribed (Gray and Spencer, 1983; Oda et al., 1992). The SSU and LSU rRNA genes are physically linked in the liverwort mitochondrial genome but are far apart in the wheat mitochondrial genome.

Among mitochondrial SSU rRNA genes, those of land plants, along with those from some fungi, are among the largest known, (1955 nt for the mature 18S rRNA of wheat mitochondria) (Spencer et al., 1984). Relative to *E. coli* SSU (16S) rRNA, most of the extra sequence of plant mitochondrial SSU rRNA is found within three of the variable regions of the secondary structure. Despite its large size, the wheat mitochondrial SSU rRNA gene is closest among mitochondrial genes to the homologous eubacterial gene, both in primary sequence and potential secondary structure (Spencer et al., 1984).

A number of modified nucleosides, all involving base methylations, have been detected in the 3'-terminal hairpin structure of the wheat mitochondrial SSU rRNA (Schnare and Gray, 1982; Schnare, 1984). These modifications, one of which has not been found in other mitochondrial SSU rRNAs, are similar to those known for eubacterial SSU rRNA (Schnare, 1984).

(iii). Chlorophyte mitochondrial rRNAs

The rRNA genes of *Chlamydomonas reinhardtii* are unique among mitochondria in that they are encoded as sub-genic modules scattered throughout 6 kbp of the genome, and interspersed with protein-coding and tRNA genes (Boer and Gray, 1988). Four modules comprise the SSU rRNA and at least 8 modules the LSU rRNA. Despite the fragmented nature of the *C. reinhardtii* rRNAs, intermolecular base pairings result in secondary structures that generally conform to those of *E. coli.* The most noticeable deviation of the *C. reinhardtii* LSU secondary structure is the apparent absence of the 5.8S-like region of the molecule located at the 5'-terminal region of *E. coli* LSU rRNA.

In contrast to this gene organization, the mitochondrial rRNA genes of the chlorophytes *Prototheca wickerhamii* and *Scenedesmus obliquus* are continuous (Wolff and Kück, 1990; Kück et al., 1990) and the LSU rRNA gene of *P. wickerhamii* contains two introns in the 3'-half of the molecule (Wolff et al., 1993). The mitochondrial genome of *P. wickerhamii* also encodes a 5S rRNA, as mentioned above.

(iv). Fungal mitochondrial rRNAs

SSU and LSU rRNAs have been sequenced from a number of fungi. Potential secondary structure diagrams can be found in Lang et al., 1987; Huttenhofer et al., 1988; Cummings et al., 1989b; Gutell et al., 1985; Gutell et al., 1990, 1992b and 1993. Relative to that of *E. coli*, the SSU rDNA from *S. cerevisiae* contains 3 A+T-rich clusters but at 1651 nt, the corresponding rRNA is not much larger in overall size than *E. coli* 16S rRNA (Huttenhofer et al., 1988). The SSU rRNA gene from *Podospora anserina* mitochondria is much larger, at 1980 nt, mainly due to extra sequences at the termini and within certain variable regions (Cummings et al., 1989b).

The LSU rRNA gene from a number of ascomycete fungi contains a group I intron (which is a mobile intron in *Saccharomyces cerevisiae*) at an identical insertion site within domain V (see E. Introns). The intron from *Neurospora* is believed to encode ribosomal protein S5 (Burke and RajBhandary, 1982).

(v). Ciliate mitochondrial rRNAs

Both the SSU and LSU rRNA genes of *Tetrahymena* and *Paramecium* are split into two coding modules, designated α and β (Schnare et al., 1986; Heinonen et al., 1987; Seilhamer et al., 1984); α represents the 5'-terminal portion of the molecule and β represents the remainder of the molecule. The spacer separating the two coding modules of the *Tetrahymena* SSU rRNA is an A+T-rich segment of DNA located within a variable region of the rRNA structure. The α and β coding modules are separated by a tRNA gene and are rearranged relative to the normal transcriptional orientation. Intermolecular base pairings between the two coding modules result in a convential secondary structure for both rRNAs. The LSU rRNA gene is present in two copies in *Tetrahymena* within terminal inverted repeats of the mtDNA. A modified nucleoside pair, invariably present in the 3'-terminal hairpin structure of SSU rRNAs, is not found in the *Tetrahymena* species (Schnare et al., 1986).

(vi). Kinetoplastid mitochondrial rRNAs

At ~1150 nt for the LSU rRNA and ~600 nt for the SSU rRNA, the rRNAs of kinetoplastid mitochondria, along with those from the nematodes, are among the smallest known (Eperon et al., 1983; Sloof et al., 1985; de la Cruz et al., 1985). They contain posttranscriptionally added poly (U) tails at the 3'-termini (Adler, 1991). Potential secondary structures for the LSU rRNA have been derived from the gene sequences and are distinguished by a notable absence of secondary structure, especially limited within the 5'-half of the molecule (Gutell et al., 1992b, 1990). What is able to be modeled likely represents a minimal structural core necessary for functional ribosomes. It is not suspected that these kinetoplastid rRNAs are edited.

C. Organelle Genome Evolution

(i). Autogenous origin versus serial endosymbiosis

Two theories have been advanced to explain the origin of plastids and mitochondria within eukaryotic cells. One theory proposes that these organelles arose from within the cell by compartmentalization of nuclear genes, i.e., an autogenous origin (Raff and Mahler, 1972). This process presumably would have involved a sequestration of a small number of functionally related genes to a special compartment of the cell, perhaps similar to the nucleolar organization of ribosomal RNA (rRNA) genes of eukaryotic cells. The resulting organelles would have been completely dependent upon the nuclear genome for expression, with some autonomy perhaps gained continuously with time.

The opposing theory of endosymbiosis assumes a xenogenous origin of organelles. The endosymbiotic theory of eukaryotic cellular evolution proposes that self-replicating organelles such as mitochondria and chloroplasts were once free-living organisms (eubacteria) that entered into a symbiotic relationship with a primitive eukaryotic cell, thereby providing the host cell with aerobic respiration and/or photosynthetic ability (Margulis, 1981; Taylor, 1987; Gray and Doolittle, 1982; Gray, 1989, 1992; Cavalier-Smith, 1987). The endosymbiosis would have been followed by a massive loss of genetic material from the endosymbionts in order to account for the much smaller genomes of organelles relative to eubacterial genomes. This would be accompanied by an ever-increasing functional dependence of the organelle upon the nuclear genome of the host cell.

As both types of organelle have their own genomes, these alternative theories of origin and evolution can be subjected to molecular analysis. Comparisons of nuclear, organellar and eubacterial DNA sequence data should be able to reveal whether organellar DNA is more similar to nuclear DNA than it is to eubacterial DNA (which would support an autogenous origin in which nuclear and organellar genomes shared a common ancestor more recently with each other than either did with other genomes), or to eubacterial DNA (in support of a xenogenous origin,with eubacterial and organellar genomes sharing the more recent common ancestor).

(ii). Molecular data supporting endosymbiosis

a. Molecular phylogeny based upon rRNA sequence data. Molecular phylogeny is the study of the origin and evolution of taxonomic units (genes, genomes, organisms) at the molecular level. This frequently involves the analysis of sequence data from homologous genes of the taxonomic units under investigation. With the exception of viral genomes, rRNA genes are present in all types of genomes (nuclear, eubacterial, archaebacterial and organellar) and thus are suitable for these types of studies. The universal regions of SSU rRNAs, U1 to U8 (and the semi-conserved regions to varying extents), can be used for primary sequence analysis in phylogenetic studies because nucleotide positions within these regions can be unambiguously aligned, even among distantly related organisms. This allows for the comparison of phylogenetically equivalent nucleotide positions among homologous rRNA genes of diverse organisms. Generally, interpretation of this type of analysis assumes that the evolution of the gene used for the comparisons reflects that of the genome and therefore that of the organelle or cell in its entirety. (However, this is not always the case as explained below.)

SSU rRNA gene sequence data have most frequently been used for phylogenetic studies of organelles, and the data are typically presented in the form of phylogenetic trees. Phylogenetic trees display the evolutionary relationships among organisms in terms of ascent and descent based upon sequence comparisons. Various methods are used for the construction of phylogenetic trees (Li and Graur, 1991). Maximum parsimony methods assume that the most parsimonious tree, i.e., the phylogenetic tree that requires the minimal number of nucleotide substitutions among all possible trees to explain the topology (branching patterns), is the correct one (Stewart, 1993; Fitch, 1977). For distance matrix methods, evolutionary distances are computed based upon the number of nucleotide substitutions that appear to have occurred between pairs of taxonomic units, and trees are constructed taking into account the different rates

of evolution among taxonomic units (Saitou and Nei, 1987). A third method of tree construction is maximum likelihood (Felsenstein, 1988).

1. Mitochondria. Early molecular data suggesting that mitochondria were derived from eubacterial endoysmbionts came from studies of the SSU rRNA gene of wheat mitochondria (Bonen et al., 1977) and from the complete SSU rRNA gene sequences of fungal and animal mitochondria; in the latter cases, the mitochondrial sequences appeared to be more similar to their *E. coli* homologs than to the nuclear ones (Küntzel and Köchel, 1981). Compelling evidence for a eubacterial origin of mitochondria was derived from an analysis of the complete nucleotide sequence of the SSU rRNA gene of wheat mitochondria, where sequence similarity within the universal regions was much greater to the eubacterial homolog (88% identity) than to a nuclear homolog (61% identity) (Spencer et al., 1984). Further phylogenetic analysis based upon mitochondrial and eubacterial SSU rRNA gene sequence data indicates that mitochondria originated from within the α-subdivision of the purple photosynthetic bacteria (Proteobacteria) (Yang et al., 1985; Gray et al., 1989).

Mitochondrial 5S rRNA sequence data also suggest that plant mitochondria originated from within the α-Proteobacteria, although inferences based upon 5S sequence data are statistically less significant than those based upon the much larger SSU rRNA gene sequence (Delihas and Fox, 1987).

2. Plastids. Phylogenetic analysis of SSU rRNA gene sequence data indicates that all types of plastids originated from within the cyanobacteria (Bonen and Doolittle, 1976; Giovannoni et al., 1988; Turner et al., 1989; Douglas and Turner, 1991; Urbach et al., 1992) and that plastid origins are sometimes attributable to secondary endosymbiosis (Douglas et al., 1991).

b. Protein sequence data supporting endosymbiosis

1. Mitochondria. There are only two protein-coding genes that are conserved among all known mitochondrial genomes (see A. Mitochondrial Genomes) and

consequently protein sequence data have not contributed significantly to the analyses of mitochondrial genome evolution. The mitochondrial cytochrome *c* resembles the cytochrome *c* from the α-Proteobacteria in structure and in sequence information (Dickerson, 1980; Dayhoff and Schwartz, 1981). One limitation of these data is that the gene for mitochondrial cytochrome *c* is nucleus-encoded (presumably as a result of gene transer from the mtDNA). Genes for the respiratory proteins are either mitochondrially encoded or nucleus-encoded but not both, thus making phylogenetic analysis among the three genomes (eubacterial, nuclear and mitochondrial) difficult. Gene sequence data for subunit 9 of ATP synthase (*atp*9) indicate a eubacterial origin of mitochondria from α-Proteobacteria (Dayhoff and Schwartz, 1981; Recipon et al., 1992).

2. Chloroplasts. Cytochrome *c* gene sequence data indicate that chloroplasts are cyanobacterial descendants (Schwartz and Dayhoff, 1981). Other protein data that support an origin of chloroplasts from cyanobacteria include those from ATP synthase, subunit 6 (*atp*6) (Cozens et al., 1986), *tuf*A (encoding translational elongation factor Tu) (Baldauf and Palmer, 1990), and RNA polymerase C1 subunit (*rpo*C1) (Palenik and Haselkorn, 1992.)

(iv). Other data supporting endosymbiosis

1. Mitochondria. There are a number of biochemical similarities between mitochondria and α-Proteobacteria such as *Paracoccus* and *Rhodopseudomonas*; these similarities support an origin of mitochondria from within this group of eubacteria. Many features of the electron transport chain appear to be shared preferentially between mitochondria and *Paracoccus /Rhodopseudomonas,* to the exclusion of other eubacteria (Whatley, 1981; John, 1987). The cellular membrane of *Paracoccus* and *Rhodopseudomonas* has invaginations resembling mitochondrial cristae (Taylor, 1987).

However, due to the extreme diversity in mitochondrial genome organization and modes of expression that has become apparent with accumulating molecular data (see A. Mitochondrial Genomes), most evidence supporting an endosymbiotic origin of mitochondria has come from rRNA gene sequence studies, as described above.

2. Chlorpolasts. The presence of phycobilins is a unique common trait shared between the cyanobacteria and the rhodophytic algae, supporting an origin of rhodophyte plastids directly from cyanobacteria (Gray, 1992). The complete sequencing of several chloroplast genomes has revealed a similar arrangement of genes within specific operons (Cozens and Walker, 1987; Steinmuller et al., 1989; Nixon et al., 1989; Young et al, 1981; reviewed by Gray, 1992). Finally, an intron is found in exactly the same place in a $tRNA^{Leu}$ gene from both cyanobacteria and chloroplasts (Kuhsel et al., 1990). There are similarities in transcription and translation between cyanobacteria and chloroplasts (Douglas and Turner, 1991; Hanley-Bowdoin and Chua, 1987; Gruissem, 1989).

(iii). The process of endosymbiosis

Organellar genomes are notably smaller than their ancestral eubacterial genomes. For example, the size of the largest chloroplast genome (400 kbp) is much smaller than the size of the smallest cyanobacterial genome (3,100 kbp), whereas the smallest animal mitochondrial genome is only 14 kbp in size. Obviously such small-sized genomes would not have had a coding capacity sufficient to sustain an ancestral free-living organism as the endosymbiotic theory predicts.

However, this genome size discrepancy can be accommodated within the endosymbiotic theory. The majority of proteins needed for organelle function are nucleus-encoded and are imported into the respective organelle, where they function in combination with organellar gene products in executing the specialized functions of the organelles. It is believed that these nucleus-encoded organellar proteins represent those that were intitially encoded by the endosymbiont and that were part of a massive

transfer of genetic information to the host cell nuclear genome. Indeed, many of these nucleus-encoded proteins have been recognized from phylogenetic analysis as chloroplast- or mitochondrion-specific (Recipon et al., 1992). Because all mitochondrial and all chloroplast genomes generally encode a similar set of proteins (see A. Mitochondrial Genomes), the bulk of this transfer would likely have occurred relatively early, i.e., soon after the endosymbiotic event took place.

Recently, certain homologous genes have been found to be encoded in the chloroplast genome of some organisms but in the nuclear genome of closely related organisms, suggesting that the transfer of genetic information from the chloroplast to the host cell nuclear genome is an ongoing process. Such transfer involves the eventual loss of the functional gene from the organellar genome (Baldauf and Palmer, 1990; Baldauf et al., 1990; Gantt et al., 1991). The *tuf*A gene has been transferred from the chloroplast genome to the nuclear genome likely within the algal lineage ancestral to the land plant lineage (Baldauf and Palmer, 1990; Baldauf et al., 1990). The large subunit ribosomal protein gene, *rpl*22, has been transferred from the chloroplast genome to the nuclear genome within the land plant lineage (Gantt et al., 1991). The Rubisco SSU gene is found in the plastid genome of the rhodophyte and chromophyte algae but in the nuclear genome of chlorophytes and metaphytes (Palmer, 1985).

There is also evidence for recent transfer of genetic information from the mitochondrial to the nuclear genome in land plants. The cytochrome oxidase subunit II (*cox*2) gene, which is normally encoded in mtDNA, was transferred to the nuclear genome (from which it is now expressed) sometime after the separation of monocotyledonous and dicotyledonous plants (Nugent and Palmer, 1991; Covello and Gray, 1992). Subsequent to transfer, *cox*2 was lost from the mitochondrial genome in two species of legumes but retained in other legumes (Nugent and Palmer, 1991) but with loss of coding potential (Covello and Gray, 1992).

These examples of recent gene transfer are not restricted to the chlorophyte/metaphyte lineage, as a similar situation has been described within the fungal lineage. Here, *atp*9 is encoded within the mitochondrial genome of *Saccharomyces* whereas in *Neurospora* and *Podospora*, the gene is expressed from a nucleus-encoded copy (van den Boogaart, et al., 1982; Brown et al., 1985).

(iv). Summary

From rRNA gene sequence comparisons, it has become evident that the organelles of eukaryotic cells originated as eubacterial endosymbionts. Mitochondria and chloroplasts have been shown to be derived from two distinct eubacterial lineages (mitochondria from within the α-Proteobacteria and plastids from within the Cyanobacteria) and it is obvious that the nuclear genome is not derived from either of these organisms or their recent ancestors (Pace et al., 1986). These observations provide evidence for the xenogenous origin of these organelles (Gray and Doolittle, 1982; Gray, 1989, 1992).

Recent studies indicate that transfer of organelle genes to the nuclear genome of the host cell is an ongoing process, allowing one to imagine a once-extensive organellar genome being progressively condensed to the small genomes observed today.

D. Mitochondrial Genome Evolution

(i). Monophyletic versus polyphyletic origin

The most primitive eukaryotic cells that contain conventional mitochondria, as judged from their early position on the eukaryotic phylogenetic tree (Sogin et al., 1986), are the flagellated, euglenoid protozoans. Therefore it is assumed that an initial endosymbiotic event whereby mitochondria were acquired by the eukaryotic lineage would have occurred sometime before the divergence that led to *Euglena* but after the divergence that led to the microsporidia and diplomonads, which are primitive eukaryotic cells lacking mitochondria (Sogin et al., 1989).

One of the many unanswered questions concerning acquisition of mitochondria and their subsequent evolution is whether all present-day mitochondria are descendents of a common eubacterial ancestor as a result of a single endosymbiotic event, or whether mitochondria are derived from different eubacteria, presumably still from within the α-Proteobacteria, as a result of multiple endosymbiotic events that occurred throughout eukaryotic cellular evolution. That is, are mitochondria monophyletic or polyphyletic? Considering the similarities in the coding potential of various mitochondrial genomes, it can be argued that mitochondria arose only once from a single endosymbiont; however, considering the variation in mitochondrial genome organization, it could be proposed that this indicates multiple origins.

If mitochondria are monophyletic, then phylogenetic trees based upon nuclear gene sequences and phylogenetic trees based upon mitochondrial gene sequences should be congruent, that is, the tree topology or branching order should be the same. Stated another way, if mitochondria are monophyletic, then phylogeny of the mitochondrial genome should reflect that of the host cell nuclear genome.

The mitochondrial SSU rRNA phylogenetic tree suggests that the mitochondrial genomes of land plants have a stronger affiliation with their eubacterial ancestors than do the mitochondrial genomes of all other eukaryotic organisms analyzed so far (Gray et al., 1989). If mitochondria are monophyletic, and considering the topology of this tree, it would then be expected that the plant nuclear lineage would be ancestral to the nuclear lineages of all other eukaryotes. However, this is not observed. Eukaryotic phylogenetic trees based upon nuclear SSU rRNA gene sequences clearly indicate that plants, animals, fungi and the ciliate protozoans diverged from one another at roughly the same time during eukaryotic cellular evolution (Sogin et al, 1986; Gunderson et al., 1987). Thus the plant nuclear lineage is not ancestral to other eukaryotic lineages. In fact, the chlorophyte lineage, as represented in phylogenetic trees by *Chlamydomonas reinhardtii*, is the one most closely related to the plant lineage in these nuclear trees. In

contrast to this affiliation of nuclear genomes, the mitochondrial genome of *C. reinhardtii* is very distantly related to land plant mitochondrial genomes (Gray et al., 1989).

One explanation of this dichotomy in branching orders between plant nuclear and plant mitochondrial genomes is that the mitochondrial genome of land plants is of a different evolutionary origin, reflecting an independent, relatively recent endosymbiotic event. That is, these observations may suggest that mitochondria are polyphyletic (Gray et al., 1989).

However, there is much variation in molecular data obtained from chlorophyte mitochondrial genomes (see A and B of this Introduction). Sequencing of the SSU rRNA from the chlorophyte, *Prototheca wickerhamii*, suggests that the relatively strong eubacterial affiliation previously found uniquely within the metaphyte lineage, extends outside of this lineage to include at least some of the chlorophytes (Wolff and Kück, 1990).

(ii). Means to an answer

In an attempt to resolve some of these pecularities of mitochondrial phylogeny, it is important to obtain molecular data from within the green algae and also from those non-photosynthetic organisms that are phylogenetically close to the chlorophyte-metaphyte clade. *Acanthamoeba castellanii* is a non-photosynthetic amoeboid protozoan that occupies a position close to green algae and land plants in eukaryotic phylogenetic trees (Sogin et al., 1986; Gunderson et al., 1987). For this thesis, I have determined the DNA sequence of the ribosomal RNA genes and flanking regions of *A. castellanii* mtDNA, with the aim of gaining insight into the evolutionary affiliation(s) of the mitochondrial genome of this non-photosynthetic protist.

E. Introns

(i). Five different classes of introns

Introns are DNA sequences that interrupt genes and are excised from a precursor transcript with the concomitant ligation of non-contiguous coding segments (exons) by a process described as RNA splicing. Nuclear mRNA introns (reviewed by Guthrie, 1991; Ruby and Abelson, 1991), nuclear tRNA introns (Greer and Abelson, 1984), and group I and group II introns are distinguishable by structural features and splicing mechanisms (Cech, 1990). Group III introns are believed to be variations of group II introns (Christopher and Hallick, 1989) while "twintrons" are introns within introns (Copertino and Hallick, 1991; Copertino et al., 1991; Copertino et al., 1992). Of the above, only group I and group II introns are found within organelle genomes, and hence only these will be described here.

Group I and group II introns have highly conserved but distinct secondary structure cores consisting of helices and loops generated by intramolecular base pairings (Michel and Dujon, 1983). The splicing mechanisms of group I and II introns have been studied in detail and are similar in that they involve a two-step transesterification reaction; however, they are also distinct in a number of respects. Group I and II introns may or may not contain open reading frames (ORFs) which may or may not be fused in-frame with the upstream exon. Group I introns have been found within both mRNA and structural genes of organellar genomes, rRNA genes of nuclear origin, and tRNA genes of bacterial origin. Group II introns have thus far been found only within mRNA and structural genes of organellar genomes.

(ii). Group I introns

A highly conserved secondary structure core of group I introns has been derived from comparative sequence analysis. (See Burke et al., 1987 for a consensus of structural conventions and nomenclature for group I introns.) The intron core region is generated by intramolecular complementary base pairings and consists of a series of helices, P1 to P9, and their associated loops, L1to L9 (numbered 5' to 3' along the RNA; Waring and Davies, 1984; Cech, 1988) (see Fig. 1). All group I introns have conserved

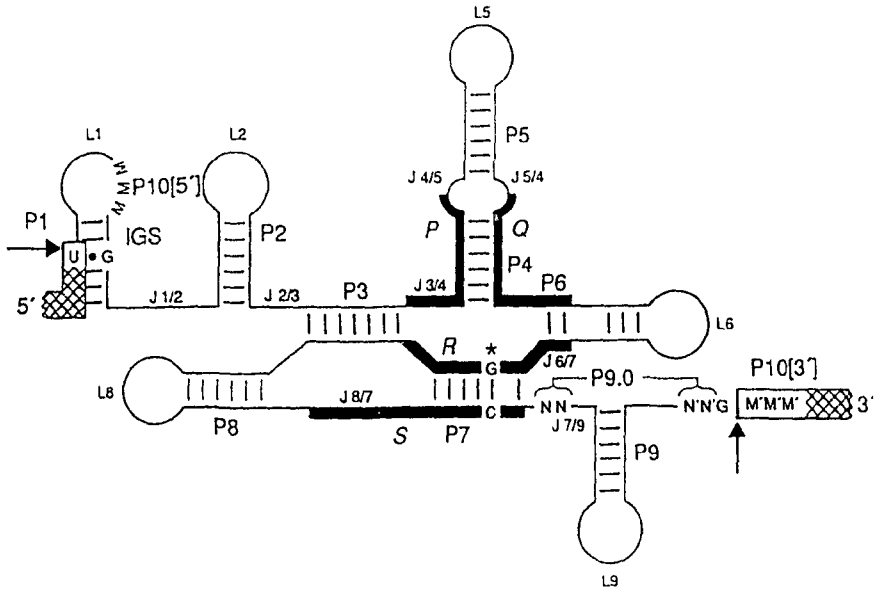


Figure 1. Generalized structure of group I intron. Arrows indicate splice sites. Conserved sequence elements P, Q, R, and S are indicated by the bold lines. Asterisk refers to the guanosine-binding site; IGS refers to the internal guide sequence; hatched areas indicate exon sequences; NN and N'N' refer to nucleotides involved in a P9.0 pairing; MMM and M'M'M' refer to nucleotides involved in a P10 pairing. Adapted from Saldanha et al., 1993.

primary sequence elements designated P, Q, R and S (Waring and Davies, 1984; Cech, 1988), which are found within the core region; each of these elements has base-paired and unpaired segments. Most sequence variation between closely related introns occurs preferentially at the extremities of the hairpin loops of the secondary structure. There is some conservation of splice-site sequences: U immediately preceeds the 5' splice site and is paired with a conserved G in the intron sequence of helix P1, and a G immediately preceeds the 3' splice site of all group I introns. Tertiary interactions have been elucidated for group I introns (Michel and Westhof, 1990).

Group I introns can be further classified into subgroups A, B, C and D on the basis of the presence (subgroup A) or absence (subgroups B, C and D) of helix P7.1 of the secondary structure. Based on sequence elements P, Q, R and S and finer details of the secondary structure, group I introns can be further categorized within the subgroups A, B, C and D, resulting in a total of eleven divisions of group I introns (Michel and Westhof, 1990).

A number of group I introns encode ORFs localized to peripheral regions of the core structure. These ORFs frequently encode components involved in intron maturation and/or propagation as described in more detail below. The size of group I introns ranges from 205 nt for the eubacterial tRNA group I intron lacking an ORF (Reinhold-Hurek and Shub, 1992) to greater than 2 kbp for the rRNA intron from fungal *(Neurospora)* mitochondria (Burke and RajBhandary, 1982).

Splicing of group I introns (reviewed by Cech, 1990) is dependent upon the highly conserved secondary structure of the RNA. Splicing of all group I introns proceeds by two transesterification reactions involving nucleophilic attack by the 3' hydroxyl of a free guanosine at the 5'- splice site as the first step of splicing (with the concomitant addition of the guanosine to the 5' end of the intron) and nucleophilic attack by the free 3' hydroxyl of the 3'-terminal nucleotide of the 5' exon at the 3'-splice site as the second step of splicing. This second step of splicing results in ligation

of the two exons and release of the linear intron. The free guanosine of the first step of splicing is held in a catalytically favorable position by the guanosine-binding site located within P7 of the intron core secondary structure. The 3'-terminal G of the intron interacts with the guanosine-binding site during the second step of splicing.

Two regions of potential base pairing of exon sequences with intron sequences are thought to direct splice site selection and ligation of the exons. The 3' terminal 4 to 7 nt of the 5' exon interacts with the 5'-terminal region of the intron (referred to as the internal guide sequence, IGS) to form the P1 pairing in all group I introns. For some introns the IGS potentially engages in an alternative pairing which includes the 5'-terminal region of the 3' exon to form the P10 structure. The potential P10 pairing is believed to have a role in the selection of the 3' splice site and likely directs exon ligation by mediating juxtaposition of the exons (Davies et al., 1982). The 3'-splice site binding sequence (SSBS) is believed to also have a role in selection of the 3' splice site (Michel et al., 1990).

In vitro experiments indicate that the only information required for splicing of some group I introns is contained within the intron itself (Kruger et al., 1982; van der Horst and Tabak, 1985). These introns are referred to as self-splicing and the RNA as catalytic RNA. However, the splicing of these introns occurs at a much reduced efficiency in the absence of proteins, and these same introns also require proteins to splice *in vivo*. It is believed, therefore, that the associated protein components function to stabilize the RNA secondary structure into a catalytically active conformation to optimize splicing. Some group I introns absolutely require protein components for *in vitro* splicing; frequently these components are encoded by the intron ORF (maturases; Lazowska et al., 1980, 1989) or are the products of nuclear genes (Garriga and Lambowitz, 1986; Gampel and Cech, 1991; Kittle et al., 1991). As the mechanism of intron splicing is identical regardless of whether the splicing occurs dependently or

independently of proteins, it is believed that RNA represents the catalytic component of all group I introns.

(iii). Group II introns

Group II introns have a highly conserved secondary structure comprised by domains I through VI with some degree of primary sequence conservation (Michel and Dujon, 1983; Michel et al., 1989). Domain I consists of a complex helix generated by short- and long-range base pairings, while the helices of domains II through VI are generated by short-range base pairings (see Fig. 2). Domains V and VI are the most highly conserved among group II introns. Group II introns can be further classified into subgroups A or B (Michel et al., 1989). Group II introns frequently begin with the semi-conserved sequence GUGCG and terminate with the dinucleotide RY or CY (Kück et al., 1990). Splicing of group II introns differs from that of group I introns in that 5'-splice site cleavage is initiated by nucleophilic attack of the 2'-hydroxyl of an internal, unpaired A residue located within helix VI (7 to 8 nt upstream of the 3' end of the intron), referred to as the branch point A (analogous to the branch point A of nuclear mRNA introns). This first step in splicing results in the release of the 5' exon and formation of the intron-3' exon lariat intermediate. The second transesterification reaction is mediated by nucleophilic attack of the 3'-hydroxyl of the 5' exon on the 3' splice site. This results in concomitant release of the intron in lariat form and ligation of the two exons to complete splicing. Some group II introns have the ability to self-splice *in vitro* (VanDerVeen et al., 1986; Peebles et al, 1986; Schmelzer and Schweyen, 1986; Winkler and Kück, 1991).

Some group II introns in the chloroplasts and mitochondria of plants and the chloroplasts of green algae participate in trans-splicing (Kohchi et al., 1988; Wissinger et al., 1991; Conklin et al., 1991; Chapdelaine and Bonen, 1991; Goldschmidt-Clermont et al., 199[illegible]). In trans-splicing events the exons are often separated by great distances [greater than 10 and 35 kbp for certain *nad*1 exons in *Oenothera*

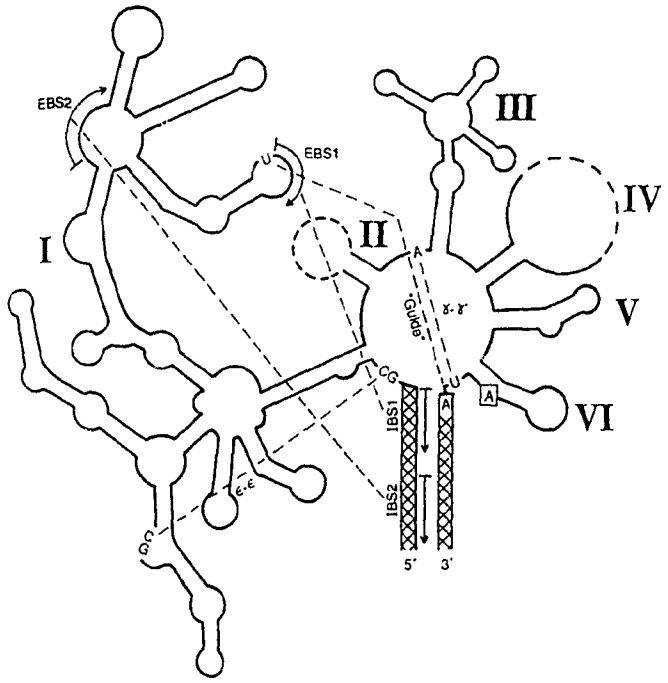


Figure 2. Generalized structure of group II intron. Intron domains of secondary structure are indicated by I to VI; EBS refers to exon binding site; IBS refers to intron binding site; hatched regions indicate exon sequences; the A enclosed in the box in domain VI is the branch site (bulged) A conserved in group II introns. Adapted from Saldanha et al., 1993.

mitochondria (Wissinger et al., 1991)] and the conserved secondary structure of discontinuous group II introns is thought to be maintained by intermolecular base pairings within the helical regions of the intron.

(iv). Non-ribosomal group I and II introns

Non-ribosomal group I and group II introns are found only in organellar genomes. Group I and group II introns have been found most frequently in mRNA genes of fungal mitochondria (Michel et al., 1982; Michel and Westhof, 1990). These include *cox*1 (Cummings et al., 1989c; Lambert et al., 1983), *cob* (Cummings et al., 1989d; Collins et al., 1988; Lazowska et al, 1980), *atp*6 (Cummings and Domenico, 1988) and genes for various subunits of the NADH dehydrogenase complex (Cummings and Domenico, 1988). In contrast to the abundance of introns in fungal mitochondrial genomes, there is only one report of introns in animal mtDNA, that being group I introns found within *cox*1 and *nad*5 of the sea anemone, *Metridium senile* (Beagley et al., 1992).

No group I introns have been detected in angiosperm mitochondrial genomes, but group II introns have been found within the *cox*2 gene of some plant species (Michel and Dujon, 1983; Pruitt and Handon, 1989), within *rps*3 (Hunt and Newton, 1991), and within genes encoding various subunits of the NADH dehydrogenase complex of plant mitochondria. As described above, trans-splicing as well as regular cis-splicing group II introns are found in these latter genes (Wissinger et al., 1991; Conklin et al., 1991; Chapdelaine and Bonen, 1991).

In contrast to the situation in angiosperms, both group I and II introns are found in the mitochondrial genome of the liverwort, *Marchantia polymorpha* (Oda et al., 1992). Although a completely sequenced mitochondrial genome from an angiosperm is not yet available, sequences of homologous genes such as *cox*1 (which contains group I introns in liverwort but not in angiosperms) can be directly compared. Introns have been detected within the mitochondrial genomes of certain genera of the chlorophytes,

such as *Prototheca wickerhamii* (Wolff et al., 1993), but appear to be completely absent from others, such as *Chlamydomonas reinhardtii* (Boer and Gray, 1988). Many group II introns are found within the chloroplast genes of *Euglena gracilis* (Christopher and Hallick, 1989). Trans-splicing by group II introns occurs in plant (Kohchi et al., 1988) and green algal (Goldschmidt-Clermont et al., 1991) chloroplast transcripts. A group I intron is found in the *rpl2* gene of the maize chloroplast (Hoch et al., 1991).

Introns have also been found in organelle tRNA genes. Group I introns have been identified within the chloroplast $tRNA^{Leu}$ genes of various plant species (Cech, 1988; Steinmetz et al., 1982) and group II introns within the chloroplast $tRNA^{Ala}$ and $tRNA^{Ile}$ genes (Michel and Dujon, 1983). A group I intron occurs at the identical site in the $tRNA^{Leu}$ gene of cyanobacteria and plant chloroplasts (Kuhsel et al., 1990). Group I tRNA gene introns have also been detected among members of the α-subdivision of the Proteobacteria (Reinhold-Hurek and Shub, 1992). All tRNA introns are inserted within the anticodon loop. A $tRNA^{Ser}$ intron is reported to occur within the liverwort mitochondrial genome but details have not been published (Oda et al., 1992).

(v). Ribosomal RNA introns

In addition to being found in various organellar mRNA genes, group I and group II introns are located in highly conserved regions of the LSU and SSU rRNA genes of many organisms.

a. SSU rDNA introns. There are only a few reported examples of introns within SSU rRNA genes and these include group I introns only. Group I introns are found in highly conserved regions of SSU rRNA genes from green algal chloroplasts (Durocher et al., 1989), the mitochondria of the liverwort, *Marchantia polymorpha* (Oda et al., 1992), and fungal (Sogin and Edman, 1989) and algal (Huss and Sogin, 1990; Davila-Aponte et al., 1991; Wilcox et al., 1992) nucleus.

b. LSU rDNA introns. Group I introns have been found in highly conserved regions of the LSU rRNA genes from fungal mitochondria (Burke and RajBhandary, 1982; Netzker et al., 1982; Cumming et al., 1989a; Jacquier and Dujon, 1983), green algal chloroplasts (Turmel et al., 1993; Turmel et al., 1991; Rochaix et al., 1985) and mitochondria (Kück et al., 1990), and protozoan nucleus (Wild and Gall, 1979; Muscarella and Vogt, 1989). The only group II introns found within the LSU rRNA gene are from the mitochondria of the green alga, *Scenedesmus obliquus* (Kück et al., 1990) and the mitochondria of the liverwort, *Marchantia polymorpha* (Oda et al., 1992). Introns are found in archaebacterial LSU rRNA genes at locations close to the positions of group I or group II introns in the same genes of eukaryotic nuclei and organelles (Garrett et al., 1991).

To date, considering known introns in chloroplast, mitochondrial and nuclear LSU rRNA genes within a range of eukaryotes and from archaebacteria, these constitute 22 unique insertion sites, consisting of 16 group I (Turmel et al., 1993), 4 group II (Kück et al., 1990; Oda et al, 1992; Garrett et al., 1991) and 3 archaebacterial (Garrett et al., 1991) introns. All of the above introns are found within highly conserved regions of either domains II, IV or V of the LSU rRNA secondary structure.

(vi). Intron ORFs

Group I and group II introns frequently encode ORFs that produce functional proteins. The ORFs of rRNA introns are localized to peripheral regions of the intron core structure. Although the intron ORFs of protein-coding genes are usually in phase with the 5' exon reading frame, the bulk of the intron ORF sequence is also localized to peripheral regions of the intron core stucture (Michel et al., 1982; Lazowska et al., 1980). It is believed that such intron ORFs are translated as precursor, hybrid proteins containing upstream exon sequences in addition to intron ORF sequences, and are subsequently processed by proteolytic cleavage (Weiss-Brummer et al., 1982). In

agreement with this, it has been demonstrated that biological activity (see below) resides within the intron ORF portion of the protein (Delahodde et al., 1989).

There are two major functions of the intron ORF-encoded proteins. Maturases are necessary for the splicing of the intron, probably by way of enhancing stabilization of the higher-order structure of the intron RNA during catalysis (Lazowska et al., 1989, 1980; Carignani et al., 1983; Banroques et al., 1986; Delahodde et al., 1989). The essential nature of the maturase is revealed by the loss of splicing, and thus absence of a mature transcript and loss of cell viability, following mutagenesis and inactivation of the intron ORF (Lazowska et al., 1980; Delahodde et al., 1989). Site-specific, double-stranded endonucleases are encoded by some intron ORFs and probably function to direct the propagation of the intron via gene conversion (Jacquier and Dujon, 1985; Dujon, 1980; Muscarella and Vogt, 1989; Quirk et al., 1989; Wenzlau et al., 1989; Colleaux et al., 1986; Delahodde et al., 1989; Gauthier et al., 1991; Durrenberger and Rochaix, 1991; see below). [Endonucleases encoded by protein intervening sequences also exert a direct role in the propagation of these sequences into intron-free genes (Shub and Goodrich-Blair, 1992).]

There is one example of a group I intron ORF that encodes ribosomal protein S5 in fungal mitochondria (Burke and RajBhandary, 1982).

a. Group I introns. Most group I intron-encoded proteins have a net positive charge due to elevated lysine contents (Collins et al., 1988), although the overall amino acid sequence among intron ORFs is generally not highly conserved. The basic character of intron-encoded proteins is apparent regardless of any proposed function. The size of group I intron ORFs from protein-coding genes is typically ~250 amino acid residues in length after predicted protein processing to eliminate the upstream exon sequences, but varies from less than 150 amino acid residues for a nuclear rRNA intron (Muscarella and Vogt, 1989) to over 400 amino acid residues for some fungal mitochondrial rRNA introns (Burke and RajBhandary, 1982).

The most obvious sequence conservation found among intron ORFs resides in the semi-conserved and related dodecapeptides, P1 and P2, which are found in the majority of mitochondrial group I intron ORFs, where they are separated by ~100 amino acid residues (Cummings et al., 1989; Waring et al., 1982). These sequence elements are also found in site-specific endonucleases from fungi (Nakagawa et al., 1991) or other (Shub and Goodrich-Blair, 1992) organisms. For many of these intron ORFs containing P1 and P2, the protein products have been shown to have either maturase or endonuclease activities. That maturases and endonucleases are closely related is indicated by the observation that of two fungal mitochondrial intron ORFs that share extensive amino acid sequence similarity beyond P1 and P2, one functions as a maturase while the other functions as a site-specific endonuclease (Delahodde et al., 1989; Wenzlau et al., 1989). The endonuclease can be activated to display maturase activity by a single base substitution (Dujardin et al., 1982). These observations suggest that there is an evolutionary relationship between endonucleases and maturases (Shub and Goodrich-Blair, 1992; Goguel et al., 1992). The endonucleases encoded by bacteriophage T4 intron ORFs lack both of the conserved dodecapeptides (Quirk at al., 1989a).

b. Group II introns. ORFs of group II introns are frequently located within the loop of domain IV. These ORFs encode reverse transcriptase-like proteins that function as maturases to facilitate splicing of the intron and that might also function in the propagation of the intron (Michel and Lang, 1985; Wahleithner et al., 1990). The maturases of group I and II introns typically splice only the intron that encodes them. The group II maturases are not related to the group I maturases.

(vii). Intron evolution

a. Distribution of group I introns. As described above, group I introns are widely distributed, being present within nuclear, mitochondrial, chloroplast and eubacterial lineages of diverse organisms. However, this distribution is very irregular in that the

presence or absence of introns within a particular lineage often does not parallel the phylogeny of that lineage. Irregularities observed include the following situations:

1). There are many examples where homologous genes of closely related organisms differ by the presence or absence of an intron. Only some of these examples are cited here. The mitochondrial *cox*1 gene of *Podospora anserina* harbors 14 group I introns while the homologous gene from the closely related ascomycete, *Neurospora crassa*, has only 0 to 4 introns depending upon the strain (Cummings et al., 1989c). Six group I introns are found in *cox*1 of the liverwort, a primitive plant (Ohta et al., 1993), while the homologous gene from angiosperms lacks introns completely (Bonen et al, 1987; Hiesel et al., 1987). *Chlamydomonas* species differ in the presence or absence of introns within the chloroplast LSU rDNA in a manner that does not reflect the gene phylogeny (Turmel et al., 1993). Different species of *Tetrahymena* differ in the presence or absence of the nuclear LSU rDNA intron, again in a manner that does not parallel the organismal phylogeny (Sogin et al., 1986). Even different strains of the same species may differ in the presence or absence of a particular intron (optional introns) as has been shown for the nuclear LSU rDNA intron of *Physarum polycephalum* (Muscarella & Vogt, 1989). Different species of *Saccharomyces* (Jacquier and Dujon, 1983) or different strains of *Saccharomyces cerevisiae* (Jacquier and Dujon, 1985) differ in the presence or absence of the mitochondrial LSU rDNA intron.

The following examples of irregular distribution are based upon the relative relatedness (similarity) of introns to determine homology and can be somewhat subjective. Also, for these comparisons, it is important to consider the possibility that intronic ORFs can been transmitted independently of the intron core structure, as has been demonstrated in *Neurospora* mitochondria (Mota and Collins, 1988) and also seems to have occurred in *Podospora* mitochondria (Cummings and Domenico, 1988).

2). There are examples where closely related introns are found at different sites within the same gene, or within entirely different genes. Two closely related (based upon primary sequence similarity within the intron core region) mitochondrial introns of *Podospora anserina* are found within different structural domains of the LSU rRNA gene, one insertion site of which is unique to *Podospora* (Cummings et al., 1989a). The group I introns of the *nad*4 and the *atp*6 genes of *Podospora anserina* mitochondria share extensive similarity within the core region, suggesting a common ancestry (Cummings and Domenico, 1988).

3). Sometimes closely related introns are found at the same site of homologous genes in distantly related organisms. In the distantly related fungi, *S. pombe* and *A. nidulans*, introns that are inserted at the same site in *cox*1 are more similar than the fungal phylogeny would predict (Lang, 1984). Similar introns are found at the same positions within mitochondrial *cox*1 from the chlorophyte *Prototheca wickerhamii* and from certain fungi (Wolff et al., 1993), and from the liverwort *Marchantia polymorpha* and certain fungi (Ohta et al., 1993). In these latter two examples, it is reasoned that these intron similarities are the result of vertical transmission of the introns from a common ancestral organism.

These observations of irregular distribution of group I introns suggest that the present population of introns has been influenced by factors beyond simple inheritance. Considering only vertical transmission, it is necessary to invoke multiple events of intron loss with retention of specific introns in select lineages in order to explain some of the observations cited above. A more likely explanation would involve intron gain within select lineages by horizontal transmission in combination with vertical transmission and intron loss. With the discovery of mobile introns, this seems to be a likely scenario to explain at least some of the above situations.

b. Intron mobility. Intron mobility refers to the gain or loss of an intron (reviewed by Lambowitz, 1989). Intron gain can occur at the DNA level mediated by site-specific

endonuclease activity encoded by intronic ORFs (reviewed by Dujon, 1989). Many mobile introns, such as the mitochondrial LSU rDNA intron (omega intron) of *S. cerevisiae* and the nuclear LSU rDNA intron of *P. polycephalum*, encode site-specific endonuclease activity that mediates the duplicative transfer of the intron to homologous genes of intron-minus strains via gene conversion (Jacquier and Dujon, 1985; Colleaux et al., 1986; Muscarella and Vogt, 1989). This process results in a co-conversion of DNA flanking sequences over a few hundred base pairs both upstream and downstream of the intron.

Intron gain could potentially occur at the RNA level by reverse splicing of an excised intron into compatible RNA, followed by reverse transcription to produce a cDNA that is subsequently integrated into the genome by homologous recombination (Woodson and Cech, 1988). This process might be relevant for those introns that do not encode endonucleases, such as the *Tetrahymena* nuclear LSU rRNA gene intron. It has been suggested that the reverse transcriptase-like products of group II intron ORFs might function to promote mobility of the introns by this mechanism (Lambowitz, 1989). It is possible that this mechanism operates in the transposition of introns to different sites of non-homologous genes. Intron loss could conceivably occur by homologous recombination of a cDNA copy of a spliced transcript into the genome, thereby replacing the intron-containing gene (Dujon, 1989).

It has been suggested that acquisition of the endonuclease-encoding ORFs of group I introns has occurred relatively recently (Belfort, 1990) and that although such an acquired mobility could account for the irregular distribution of homologous introns via horizontal transfer within the eukaryotic lineages, it would not account for their origin. According to the introns-ancient theory (Darnell and Doolittle, 1986), introns would have existed in the ancestral prokaryotic and progenote assemblages (Belfort, 1991). The cyanobacterial $tRNA^{Leu}$ introns are not likely to represent recent acquisitions as they appear to have been maintained in the homologous plant chloroplast

gene as judged by identical insertion sites and a significant level of primary sequence conservation between the eubacterial and chloroplast introns (Kuhsel et al., 1990; Cech, 1988).

F. RNA Editing

RNA editing results when the transcript sequence differs from that specified by the gene sequence. This deviation can be the result of nucleotide insertion, nucleotide deletion or nucleotide change. Most known editing systems function within organelles of eukaryotic cells and most often alter mRNAs. In the latter cases, editing often results in an alteration of the coding potential of the edited transcript, in the process producing a functional from an otherwise non-functional transcript (reviewed by Cattaneo, 1991).

(i). Editing of mRNA

Editing of mRNAs can occur co-transcriptionally but most often takes place post-transcriptionally.

a. Co-transcriptional editing. The insertion of one or more guanosine residues at a defined site of the P (phosphoprotein) gene transcript of paramyxovirus results in the production of multiple proteins with identical amino-terminal but different carboxy-terminal portions, due to reading frame shifts and overlapping ORFs induced by the editing (Pelet et al., 1991). Insertion occurs 3' to a specific stretch of purines and is believed to be mediated by polymerase stuttering, i.e., pausing of the polymerase followed by 5' slippage of the 3' end of the mRNA. Editing thus occurs co-transcriptionally in these cases.

b. Post-transcriptional editing.

1. Insertional editing. Insertional editing also occurs post-transcriptionally. Insertion/deletion of single or multiple uridine residues is seen at single or multiple sites throughout trypanosomal (kinetoplastid protozoan) mitochondrial mRNAs (reviewed

by Stuart, 1991a and b). This editing generates translatable transcripts from at least seven distinct mRNAs in each of *Trypanosoma*, *Leishmania* and *Crithidia*, through more or less extensive modification of otherwise non-functional mRNAs (Stuart, 1991a; Stuart, 1991b; Benne et al., 1986; Feagin et al., 1988; Feagin et al., 1987; Shaw et al., 1988). This editing creates ORFs, corrects reading frame shifts and sometimes creates initiation and termination codons. Transcripts from homologous genes are edited to different extents in different organisms. This editing is developmentally regulated in *Trypanosoma brucei* where certain transcripts are edited at particular stages of the life cycle and not at other stages, while other transcripts are constitutively edited. The mechanistic basis of this developmental regulation is not known.

Insertional editing occurs in other mitochondrial systems besides that of the trypanosomes. Single cytidine insertion occurs at 54 sites within the α-subunit of the mitochondrial ATP synthase transcript of the slime mold, *Physarum polycephalum*. This editing creates an ORF and 54 new translation codons (comprising ~10% of the total protein) (Mahendran et al., 1991). The cytidine insertions occur uniformly throughout the transcript except near the 5' and 3' termini of the transcript, which appear to be devoid of editing sites. The mechanism and specificity determinants of this editing process are not known.

Addition of A residues (polyadenylation) in verterbrate mitochondria results in the generation of a termination codon (UAA) for many mRNAs (Ojala et al., 1981).

2. Substitutional editing. Another type of post-transcriptional editing is substitutional editing, where one type of residue is replaced by another. RNA editing occurs in most angiosperm mitochondrial protein-coding transcripts analyzed (reviewed by Gray and Covello, 1993; Wissinger et al., 1992) and much less frequently in non-translated regions of plant mitochondrial transcripts (Schuster et al., 1991). Editing in plant mitochondria involves pyrimidine transitions, mainly C to U but sometimes U to

C; it often occurs extensively throughout the transcript. As editing in plant mitochondria almost always takes place at the first or second position of a codon, it almost invariably alters the amino acid sequence of the edited protein, to the extent that up to 10% of the total amino acid sequence may be altered as a result of the editing process (Gray and Covello, 1993). This usually improves the evolutionary conservation of the protein (Gray and Covello, 1993; Wissinger, et al., 1992; Covello and Gray, 1989; Covello and Gray, 1990; Schuster, et al., 1991) which implies that the editing is biologically relevant. In this regard, the editing-induced replacement of an arginine by a cysteine at position 228 in *cox*2 of wheat mitochondria and the replacement of a threonine by a methionine at position 235 in the same protein in most mitochondria, are likely to represent changes essential for the function of the protein, as these particular residues have been implicated in the binding of copper in the active site of the enzyme (Covello and Gray, 1990). Cys-288 is universally conserved in all other plants besides wheat and in all non-plants, whereas Met-235 is universally conserved in non-plant species. Editing in plant mitochondria can also create initiation (*nad*1) (Chapdelaine and Bonen, 1991) or termination (*atp*9) (Begu et al., 1990) codons in wheat mitochondria.

Some editing events are silent, i.e., the amino acid identity is not altered. Editing at these sites tends to be less frequent than at non-silent sites and is often manifested as partial editing, where only a proportion of the transcripts are edited (Schuster et al., 1991).

It seems unlikely that RNA editing occurs in the mitochondria of the liverwort *Marchantia polymorpha* (a bryophyte). Complete sequencing of this genome (Oda et al., 1992) indicates that unedited transcripts would be adequate to produce standard, evolutionarily conserved mitochondrial proteins.

There are only a few examples of RNA editing in chloroplast transcripts, and these involve C to U conversion (reviewed by Gray and Covello, 1993) as in plant

mitochondria. These editings create initiation codons (Hoch et al., 1991; Kudla et al., 1992) where ACG is converted to AUG, or alter internal amino acid residues (Maier et al., 1992). In the latter example, the amino acid changes result in the replacement of non-conserved amino acids with phylogenetically conserved ones.

Editing also occurs within group II introns of plant mitochondrial mRNAs. One of the two cis-splicing group II introns of the *nad*1 transcript of *Oenothera* mitochondria is edited within the ORF region by multiple C to U substitutions. Many of these editing events within the intron ORF (*mat-r*), which likely encodes a reverse transcriptase-like protein, result in a change of amino acid residue but do not enhance the similarity to other intronic-encoded group II maturases; therefore, the significance of this editing is not clear. However, it should be noted that less than half of the *mat-r* transcript has been examined for editing. One of the two trans-splicing group II introns of the same transcript is edited within domain VI by two C to U changes, one of which restores standard base pairing within this helix (Wissinger et al., 1991, 1992). However, this appears to be partial editing, as only one out of ten transcripts analyzed was altered at this position; thus, the relevance of editing at this position for splicing is questionable. The other editing event within domain VI occurs 4 nt upstream and is immediately adjacent to the bulged A of group II introns (branch point A) known to be important for catalysis (see E. Introns). Editing at this position results in a secondary structure that is similar to other group II introns and was found in all transcripts analyzed, indicating its possible relevance to splicing (Wissinger et al., 1991). Also, a trans-splicing group II intron of the *nad*2 transcript is edited twice (Binder et al., 1992) and the cis-splicing group II intron of the *nad*5 pre-mRNA is also edited (Knoop, et al., 1991). However, the significance of these particular editings is not obvious.

Substitutional RNA editing also has been shown to occur in nucleus-encoded transcripts specifying mammalian apolipoprotein B (apoB) (Chen et al., 1987; Powell et al., 1987; Tennyson et al., 1989) and a mouse glutamate receptor subunit B

(Sommer et al., 1991). In the apoB mRNA, editing by C-to-U conversion within a CAA codon creates a translational termination codon internal to the reading frame of the transcript (i.e., CAA specifying glutamine is altered to UAA specifying termination). This results in the generation of a high-molecular-weight form ($apoB_H$) from the unedited transcript and a low-molecular-weight form ($apoB_L$) of apoB from the edited transcript. This edited site is located at the junction of two functional domains where $apoB_H$ contains a domain for lipoprotein assembly and a second domain for LDL receptor binding; $apoB_L$ contains only the domain for lipoprotein assembly (Hodges and Scott, 1992). This editing is tissue-specific for the intestine in humans and for the liver and intestine in rodents and is developmentally regulated. A second editing site, of unknown significance, has been found to occur within the 3'-untranslated region of the human apoB transcript (Hodges and Scott, 1992). Editing at this site appears to be only partial.

In another example of substitutional editing of nucleus-encoded transcripts, a single A residue is changed to G in the mRNAs for 3 out of 6 glutamate receptor subunits of glutamate-gated ion channels in mouse and rat brain (Sommer et al., 1991). The subunit transcripts are edited with different efficiencies, and this editing results in the replacement of a glutamine residue (CAG codon) by an arginine residue (CGG codon) in the encoded protein. This significantly affects the properties of the encoded protein such that the properties of ion flow are altered.

(ii). Editing of structural RNAs

There have been only a few reports of editing of structural RNAs. The mitochondrial LSU rRNA from the plant, *Oenothera berteriana* (evening primrose), appears to be edited at two positions, both of which are within single-stranded regions of the secondary structure and each of which involves pyrimidine transitions. However, as this is partial editing where most of the transcripts are not altered, it probably is not essential for ribosome function (Schuster et al, 1991).

There are two reports of editing of mammalian nucleus-encoded transfer RNAs. Two distinct isoacceptor tRNAs for selenocysteine, both recognizing UGA codons, are found in bovine liver (Diamond et al., 1990). These tRNAs differ from each other by several pyrimidine transitions. Only one tRNA selenocysteine gene could be found in the nuclear genome. One of the tRNA sequences differs at 2 positions from the gene sequence whereas the other differs at 3 positions. Failure to find more than one gene implies that RNA editing is responsible for the existence of these two distinct tRNAs. All but one of the differences are localized to single-stranded regions of the tRNA so that the functional significance of this editing is not clear.

In the second example of tRNA editing (Beier et al., 1992), the sequence of a rat liver cytoplasmic $tRNA^{Asp}$ differs from its gene sequence by two pyrimidine transitions within the anticodon loop. The U next to the first position of the anticodon, which is edited to C in the transcript, is highly conserved in tRNAs and thus the changes induced by this editing are away from a conserved pattern. Once again, the functional significance of this editing is not clear.

A third, more recent example of tRNA editing has been reported to occur in the mitochondria of a marsupial (the opossum, *Didelphis virginiana*) (Janke and Pääbo, 1993). In this case, a C-to-U change at the second position of the anticodon is expected to change the codon recognition from glycine (GCC) to aspartate (GUC). This appears to be another instance of partial editing.

Having described some specific examples of RNA editing, I would now like to summarize where things stand regarding our understanding of the mechanisms and specificities of the various editing processes.

(iii). Mechanisms

All types of post-transcriptional editing could occur by at least one of three potential mechanisms, as outlined by Wissinger et al. (1992); these are deamination, base substitution, or nucleotide replacement. All documented types of substitutional editing

involve either pyrimidine transitions or purine transitions and can be explained by some form of base conversion. Base conversion, as mediated by deamination of a cytidine to a uridine, could account for the C-to-U editing that occurs in plant mitochondria and mammalian apoB mRNA. An *in vitro* editing system has been established for apoB mRNA and studies indicate that a 125 kDa protein component (or components) is responsible for the editing activity, with no indication of involvment of an RNA species (Driscoll and Casanova, 1990; Greeve et al., 1991; reviewed by Hodges and Scott, 1992). This, in combination with the finding that the edited cytidine nucleotide is not removed during the editing process (Bostrom et al., 1990; Hodges et al., 1991), implies that the editing is likely to be mediated by a site-specific cytidine deaminase. Base replacement (transglycosylation, see below) is another possible mechanism for this editing. Similarily, through the action of a site-specific adenosine deaminase, adenosine bases could be converted to inosine, which perhaps could be chemically modified further to generate guanine for those editing events involving purine transitions such as editing of mRNAs encoding receptor subunits of glutamate-gated ion channels in mouse brain (Sommer et al., 1991).

Base substitution or replacement involves the removal of a base by an RNA glycosylase, which then can be exchanged for another base (transglycosylation). The polynucleotide chain is not broken in this process. In fact, the formation of queuosine at the first (Wobble) position of the anticodon in certain mammalian and eubacterial tRNAs occurs as the result of base replacement (Okada and Nishimura, 1979; Okada, et al., 1979).

Nucleotide replacement involves excision of a nucleotide from the sugar-phosphate backbone followed by replacement by another nucleotide; alternatively, multiple nucleotides can be inserted or deleted at the position of cleavage of the polynucleotide chain. This mechanism, necessitating breakage of the polynucleotide backbone, occurs in kinetoplastid mitochondrial editing through the intervention of

guide RNAs. Guide RNAs (gRNAs) are small transcripts ~60 nt in length encoded by kinetoplastid DNA, with 3'-poly (U) tails being added post-transcriptionally. They bind to complementary regions of the precursor mRNA (Cech, 1991; Blum et al., 1991). The primary sequence of the gRNAs appears to supply the information for editing, which is initiated by mismatches in the gRNA:mRNA hybrid. The precursor mRNA is cleaved, U residues are inserted or deleted, and then the edited mRNA is re-ligated. It is possible that the process of uridine addition/deletion in trypanosomes requires multiple enzymatic activities consisting of an endonuclease, an exonuclease or uridylyl terminal transferase, and RNA ligase (Blum et al., 1990; Simpson, 1990). However, there is no experimental evidence implicating these enzymatic activities in the editing process. Another model has been proposed by Cech (1991) and is supported by data from Blum et al. (1991). In this model, two transesterification reactions, perhaps catalyzed by the gRNA itself and/or the precursor mRNA, are thought to mediate the cleavages and ligations necessary for uridine deletion/addition in a manner similar to that observed in the case of RNA splicing/reverse splicing.

There is some evidence to suggest that nucleotide replacement might actually operate for plant mitochondrial editing also (Gualberto et al., 1991). *In vitro* assays for editing in plant mitochondria (Araya et al., 1992) may prove useful in the delineation of editing mechanisms.

(iv). Specificity determinants

Not a lot is known about the specificity of editing. Determinants of editing specificity could potentially involve the primary sequence and/or secondary structure of the targeted site.

The gRNAs that direct uridine insertion/deletion of trypanosome mitochondrial mRNAs are complementary not only to a region of the edited transcript but also to a region of the precursor transcript immediately 3' of the site to be edited, called the 3'-anchor sequence. Before editing, the gRNA binds to the pre-mRNA within the 3'-

anchor sequence (Blum et al., 1990). Editing in trypanosome mitochondria proceeds with a 3'-to-5' polarity, reflecting the fact that the downstream edited transcript provides the 3' anchor sequence for the immediately adjacent upstream (5') editing event (Masslov and Simpson, 1992; Sturm et al., 1992). Therefore the 3'-anchor sequence, which is a product of the previous editing event, determines the editing specificity for all but the initial editing at the 3' extremity of the editing domain in which multiple gRNAs are involved. In other words, prior editing at an adjacent downstream site determines the editing of the immediately adjacent upstream site by allowing proper anchor formation. Further studies implicating the gRNAs as determinants of editing specificity demonstrate that misguiding leads to misediting in *Leishmania* mitochondrial mRNAs (Sturm et al, 1992). Obviously, in this case primary nucleotide sequence in both the precursor mRNA and the gRNA is important in determining specificity.

There is recent evidence to suggest that the specificity for editing of apoB mRNA resides within an 11-nt stretch +5 nt to +15 nt downstream of the editing site of the mRNA but not immediately surrounding the editing site (Hodges and Scott, 1992). This sequence motif is highly conserved among mammalian apoB mRNAs, suggesting that the primary nucleotide sequence is crucial for the editing specificity.

Specificity of editing of the mRNAs for mouse glutamate receptor subunits is not likely to involve primary sequence determinants immediately surrounding the editing site. One of the subunits that is produced from edited mRNA, subunit B, belongs to a class of co-expressed receptors whose other members are produced from mRNAs that are almost identical in nucleotide sequence around the editing site (the transmembrane region II); yet the other three members of this class are not edited at all (Sommer et al., 1991). However, it is possible that nucleotide sequences more distal to the editing site, and perhaps more divergent among receptor subunit mRNAs, might exert a role in defining the specificity of editing.

The co-transcriptional editing that leads to G insertion is believed to be mediated by polymerase stuttering in paramyxovirus. This editing is thought to be directed by the primary sequence immediately preceding the editing site, which is believed to cause the polymerase to stutter for some reason (Pelet et al., 1991).

G. Transfer RNAs

(i). Structural characteristics

a. Secondary structure. As deduced from comparative sequence analysis, standard tRNA molecules, which include those encoded by prokaryotic, eukaryotic nuclear, chloroplast and mitochondrial genomes (with the exception of most metazoan mitochondrial tRNAs), assume a highly conserved cloverleaf secondary structure generated through the interaction of complementary intramolecular base pairings (Holley et al., 1965; Levitt, 1969). These interactions result in the generation of the 7 bp acceptor stem, the dihydrouridine (DHU) arm, the anticodon stem and loop, the variable loop and the TΨC (ribothymidine, pseudouridine, cytidine) arm (see Fig. 3). The fourth nucleotide from the 3'-terminus of the tRNA is referred to as the discriminator nucleotide (Crothers et al., 1972) and is unpaired. All tRNAs are numbered from 1 to 73 beginning at the 5' end and proceeding to the 3' end according to a standardized scheme (Sprinzl et al., 1989). The lengths of individual tRNA species can vary reflecting variation in the number of nucleotides within the D loop and variable arm.

There is some conservation of primary sequence among tRNA molecules. All tRNAs have CCA as the 3'-terminal nucleotides; these are either gene-encoded (eubacterial tRNAs) or added post-transcriptionally (eukaryotic cytoplasmic tRNAs). Those sequence positions that are conserved among all known tRNAs are referred to as invariant and occur mainly within the loop regions of the secondary structure.

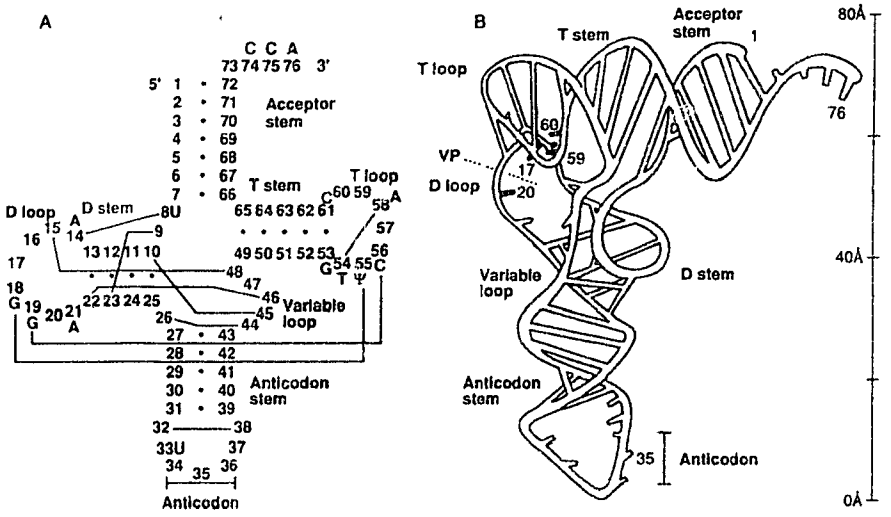


Figure 3. Cloverleaf secondary structure (A) and tertiary structure (B) of tRNA. In A, nucleotides are replaced by numbers to indicate the standard 76 positions, and absolutely conserved nucleotides are identified. Adapted from McClain, 1993.

Positions within the tRNA that are either always a pyrimidine or always a purine are referred to as semi-invariant.

There are examples of unconventional tRNAs encoded by animal mitochondria where the DHU arm or the TΨC arm and the variable loop are truncated to varying degrees and where the tRNAs appear to be functional as such (Okimoto and Wolstenholme, 1990). These observations suggest that only one of these two arms is essential, presumably to allow folding of the tRNA into an L-shaped structure, and help to delimit a minimal size for tRNA function.

b. Tertiary structure. The three-dimensional structures of a number of tRNAs have been determined from X-ray crystallographic studies (Kim et al., 1974; Kim, 1978a; Westhof et al., 1985; Rould et al., 1989; Basavappa and Sigler, 1991; Woo et al., 1980). The tertiary structures assume a characteristic L shape. Base pairings observed in the secondary structure are maintained in the tertiary structure. The acceptor stem and the TΨC stem together form one of the arms of the L while the anticodon stem and the DHU stem form the other arm. The CCA is positioned at one end, the anticodon at the other end and the DHU loop and the TΨC loop form the corner of the L (Kim, 1978b) (see Fig. 3). Hydrogen bonds mainly involving the invariant and semi-invariant nucleotides within the DHU and TΨC loops are involved in maintaining the tertiary structure.

(ii). Post-transcriptional processing

a. Endonucleolytic cleavage events. Post-transcriptional processing of mitochondrial tRNAs resembles that for nucleus-encoded tRNAs in that single endonucleolytic cleavages of precursor RNAs generate the mature 5' and 3' termini (Hanic-Joyce and Gray, 1990; Chen and Martin, 1988; Manam and Van Tuyle, 1987). Such processing differs from eubacterial tRNA processing, where the 3' endonucleolytic cleavage occurs at a distance followed by exonucleolytic trimming to obtain the mature 3'-terminus (Deutscher, 1984).

Endonucleolytic removal of the 5'-leader sequence from precursor tRNAs in eubacteria, eukaryotes and probably organelles is catalyzed by RNase P. Experiments indicate that the main recognition element for RNase P is the fully base-paired acceptor stem of the tRNA (McClain et al., 1987; Carrara et al., 1989). Other experiments identify a conserved C-G base pair of the TΨC stem as a determinant for 3'-end processing of some fungal mitochondrial tRNAs (Zennaro et al., 1989). Addition of CCA to the 3' end of the processed tRNA is catalyzed by nucleotidyltransferase activity in plant mitochondrial extracts (Hanic-Joyce and Gray, 1990).

b. Modified nucleosides. Transfer RNAs from all three kingdoms are noted for an abundance of modified and hypermodified nucleosides (reviewed by Bjork et al., 1987). All modified nucleosides are derivatives of the four standard nucleosides (guanosine, adenosine, uridine, cytidine), and modification frequently involves methylation of the 2' hydroxyl group of the ribose and the addition of thio, carboxyl, isopentenyl, hydroxyl, threonyl, lysyl or amino groups to the base (Nishimura, 1979). These modifications occur post-transcriptionally and in all but a few cases (see below) are alterations of pre-existing nucleosides within the tRNA. In eubacterial tRNAs, as many as 21 modified nucleosides are found mainly, but not exclusively, within the loop regions of the secondary structure, and in eukaryotic tRNAs as many as 46 modified nucleosides are found dispersed over the stems and loops of the molecule (Bjork et al., 1987).

Modified nucleosides are frequently found at the Wobble (first) position of the anticodon where they influence codon choice (see Bjork et al., 1987). Inosine and the hypermodified nucleoside queuosine are found at this position of the anticodon in certain tRNAs, where their presence bestows an expanded codon recognition capability on the modified tRNA. Formation of queuosine occurs post-trancriptionally by base replacement (Okada, et al., 1979; Okada and Nishimura, 1979). Inosine can base pair with cytidine, uridine or adenosine. Queuosine, which replaces G at the Wobble

position of unmodified histidine, tyrosine, asparagine and aspartate isoacceptors, base pairs non-preferentially with uridine and cytidine (Meier et al., 1985). On the other hand, some thio derivatives of uridine show preferential pairing with A over G due to conformation of the modified nucleoside (Yokoyama et al., 1985). When they occur in the Wobble position of a tRNA, they restrict rather than expand codon recognition properties.

Another example in which modification at the Wobble position of the tRNA anticodon alters the aminoacylation specificity of the synthetase occurs in *E. coli* $tRNA^{Ile}$. The modified base found at the Wobble position of the $tRNA^{Ile}$ anticodon CAU is lysidine (L), a derivative of the gene-encoded cytidine generated by the attachment of a lysine residue to the 2-position of the cytidine ring structure. Before modification of the C to L, the tRNA is charged *in vitro* with methionine, as expected (in *E. coli*, the anticodon is the sole recognition signal for the methionyl tRNA synthetase); following modification, the tRNA is charged with isoleucine instead of methionine. This results in the incorporation of isoleucine instead of methionine into protein in response to AUA codons (Muramatsu et al., 1988).

In animal mitochondrial systems, an unmodified U in the Wobble position of the anticodon can presumably pair with any one of the four codons at the third position and thus read all four codons of a four-codon family (Anderson et al., 1981; Lagerkvist, 1981). This allows protein synthesis to occur in the presence of only 22 tRNAs in animal mitochondria. Conversely, for two-codon families, the U at the Wobble position of the anticodon is frequently modified, and this modification is believed to restrict codon recognition to G or A at the third position (Randerath et al., 1981a, 1981b; Dubin et al., 1986).

The role of modified nucleosides outside of the anticodon is less certain, but some studies indicate that they are important for the discrimination between cognate and noncognate tRNA/synthetase recognition, thereby increasing the fidelity of protein

synthesis (Perret et al., 1990). There are indications that modification can result in conformational changes in the tRNA resulting in a less ordered tertiary structure (Sampson and Uhlenbeck, 1988; Beresten et al., 1992). Modification of fungal and plant initiator $tRNA^{Met}$ appears to mediate discrimination between elongation and initiation events during translation by preventing elongation (Kiesewetter et al., 1990; Basavappa and Sigler, 1991).

(iii). Acceptor identity elements

During aminoacylation distinct aminoacyl-tRNA synthetases catalyze the covalent attachment of a specific amino acid to the 3' terminus of a given tRNA. All tRNAs have a similar secondary and tertiary structure and it is essential that the synthetase recognize the cognate tRNAs (comprised of isoaccepting species) while at the same time discriminating against noncognate tRNAs to maintain the fidelity of protein synthesis. Acceptor identification is achieved by an interplay of positive elements of tRNA structure for the cognate synthetase (referred to as recognition elements) and negative structural elements that function in the discrimination among synthetases to prevent mischarging. Determinants of amino acid acceptor specificity, i.e., identity elements, have been defined for a number of tRNAs from *E. coli* (Shimizu et al., 1992) and from yeast (Pütz et al., 1991; Sampson et al., 1989) by investigating the effects of specific base changes on charging and mischarging by cognate and noncognate synthetases, respectively (reviewed by Schulman, 1991). Positions and total number of recognition elements vary among tRNAs but these determinants reside predominately within the anticodon and/or within the first three base pairs of the acceptor stem plus the discriminator nucleotide.

Among the few tRNAs where changes in the anticodon bases have little or no effect on recognition by the synthetase in either *E. coli* or yeast are $tRNA^{Ala}$, $tRNA^{Ser}$ and $tRNA^{Leu}$. The first three nucleotide pairs of the acceptor stem, the unpaired discriminator nucleotide at position 73 at the 3' end of the molecule and also one

nucleotide within the DHU loop (G20) of the secondary structure have been shown to contribute to *E. coli* $tRNA^{Ala}$ acceptor identification, with the G3:U70 pair of the acceptor stem making the major contribution (McClain et al., 1991; Hou and Schimmel, 1988). G3:U70 is uniquely and, with the exception of some animal mitochondrial tRNAs (D. R. Wolstenholme, personal communication), invariably conserved among alanine-specifying tRNAs. A 24-nt RNA microhelix containing the $tRNA^{Ala}$ acceptor stem and the 5'-half of the DHU stem as a loop can be charged with alanine (Hou et al., 1989). This represents a minimal tRNA structure that can act as a substrate for aminoacylation.

(iv). Interaction with aminoacyl tRNA-synthetases

The amino acid is attached to the 2' or 3' hydroxyl group of the terminal A of the 3'-CCA sequence by specific aminoacyl-tRNA synthetases during charging. X-ray crystallographic data of an enzyme-substrate complex can provide information regarding the structural specificity of the precise interactions. The crystal structure of the complex consisting of *E. coli* $tRNA^{Gln}$, the glutaminyl-tRNA synthetase and ATP has been derived from an electron density map (Rould et al., 1989). The synthetase makes contact with the cognate tRNA along the entire inside of the L of the tertiary structure, and specific interactions have been identified. In particular, the three anticodon nucleotides interact with amino acids within the carboxy-terminal portion of the synthetase, being sequestered in protein pockets (Rould et al., 1991). The acceptor end of the tRNA, along with the ATP and glutamine, bind by way of sequence-specific interactions within a structure referred to as the dinucleotide fold, located within the active site of the enzyme. Here, the first three base pairs of the acceptor stem, the discriminator nucleotide and the 3'-terminal CCA make direct contact through hydrogen bonding with amino acid side chains of specific residues of the enzyme in a manner that likely contributes significantly to $tRNA^{Gln}$ identification. All of the sequence-specific interactions occur via the exocyclic (position 2) amino group of guanine nucleotides that

are available for hydrogen-bonding. Likewise, in *E. coli* $tRNA^{Ala}$, an important role as an identity element has been ascribed to the position 2 amino group of the guanine of G3:U70 of the acceptor stem (Musier-Forsyth et al., 1991).

From an analysis of these precise interactions and from a more recent analysis (Hayase et al., 1992), potential identity elements of $tRNA^{Gln}$ are now known to include the anticodon nucleotides and two additional nucleotides within the anticodon loop, the discriminator nucleotide, the first three base pairs of the acceptor stem and a nucleotide located within the D stem. There seems to be no question but that the acceptor stem is crucial to tRNA recognition by the cognate synthetase and probably represents an ancient form of recognition between RNA and protein.

H. Summary of Research Presented

For this thesis, the sequence of the ribosomal RNA genes and their flanking regions from the mtDNA of *Acanthamoeba castellanii* has been determined for the purpose of gaining insight into the evolutionary affiliation of the mitochondrial genome of this non-photosynthetic protist. The mitochondrial genome of *Acanthamoeba castellanii* is a circular DNA molecule, ~40 kbp in size, with a calculated G+C content of 32.9% (Mery-Drugeon et al., 1981). A physical map of this genome has been constructed by Bohnert and von Gabain (cited in Byers, 1986).

The LSU rRNA gene of *A. castellanii* mitochondria harbors three group I introns, all containing ORFs. Characteristics of these sequences in terms of intron propagation and evolution constitute another area of research presented in this thesis.

In the course of sequencing the flanking regions of the rRNA genes of *A. castellanii*, a unique type of RNA editing was discovered and is described here and elsewhere (Lonergan and Gray, 1993a).

Finally, the gene sequences for *cox*1 and *cox*2 from *Acanthamoeba castellanii* mitochondria are presented with accompanying transcript analyses.

II. Materials and Methods

A. Materials

Generally, lab chemicals were from BDH, Anachemia, Boehringer Mannheim Biochemicals (BMB) and Sigma Chemical Company. Proteose peptone and yeast extract were obtained from Difco Laboratories. Avian myeloblastosis virus (AMV) reverse transcriptase, polynucleotide kinase, nucleoside deoxy- and dideoxytriphosphates (dNTP and ddNTP) and dextran sulfate were obtained from Pharmacia. Restriction endonucleases were from Pharmacia, New England Biolabs and Bethesda Research Laboratories (BRL). T4 DNA ligase was purchased from BRL, Amersham and BMB. DNase I was obtained from BMB, and the Klenow fragment of *E. coli* DNA polymerase I was purchased from BRL. Terminal deoxynucleotidyl transferase was obtained from Promega while SDS was from BRL. [γ-^{32}P]ATP and [α-^{32}P]dATP were products of ICN Biomedicals. All oligonucleotides used as sequencing primers and hybridization probes were obtained from the Regional DNA Synthesis Laboratory, University of Calgary, Canada.

B. Methods

(i). Culture conditions for Acanthamoeba castellanii

Acanthamoeba castellanii, strain Neff (American Type Culture Collection 30010), was maintained as suspension cultures in growth medium at 30°C[1]. It was necessary to subculture cells into fresh growth medium once every 7 days. Adequate aeration of cultures was maintained by ensuring that the depth of the growth medium never exceeded 1 inch (100 ml culture/500 ml flask) and by the gentle swirling of cultures in a gyrotory water bath shaker (New Brunswick Scientific). Optimal growth medium as

[1] When cells were not used for experiments for long periods of time, they could be maintained as tube cultures in 2 ml of growth medium at room temperature.

defined by Neff et al. (1964) contained 1.5% glucose, 0.75% yeast extract, 0.75% proteose peptone, 2 mM KH_2PO_4, 1 mM $MgSO_4$, 0.05 mM $CaCl_2$, 0.1 mM ferric citrate, 1.0 mg/liter thiamine, 0.2 mg/liter biotin and 1 μg/liter B12, and was adjusted to pH 7.0 with 10 M NaOH before addition of glucose. Glucose and medium (without glucose) were sterilized by autoclaving (121°C) for 15 m before the two were mixed.

Culture growth was monitored by optical density (OD) readings at 550 nm and a standard growth curve was derived from these readings. For nucleic acid extraction, cells were harvested from the early- to mid-log phase of growth as determined by OD_{550} readings and the general appearance of cells under the light microscope.

Electron microscopy of whole cell and mitochondrial preparations was periodically performed courtesy of G. Faulkner and M.-A. Travers of the Department of Microbiology, Electron Microscopic Services, Dalhousie University.

(ii). Preparation of nucleic acids from Acanthamoeba castellanii

Nuclear and mitochondrial nucleic acids were prepared according to one of two methods, depending on whether mitochondrial RNA was required. All procedures were performed at 4°C.

a. Method 1 (DNA only). Mitochondrial DNA (mtDNA) was isolated by cesium chloride (CsCl) density-gradient centrifugation of total cellular DNA. One liter of cells was centrifuged at 2,250*g* for 20 m and the cell pellet was washed once with 20 ml of 150 mM Tris, 25 mM EDTA (pH 8.5). Cells were resuspended in 10 ml of the same buffer and lysed by addition of SDS to 1%, followed by NaOAc to 0.3 M and an equal volume of phenol mix [500 g of phenol, 70 ml of *m*-cresol, 55 ml of H_2O and 0.5 g of 8-hydroxyquinoline, equilibrated against 10 mM Tris-HCl (pH 7.6)]. Nucleic acids were precipitated by addition of 2 vol. 95% ethanol to the aqueous phase. DNA and small RNAs were separated from high-molecular-weight, salt-insoluble RNAs (mostly rRNA) by incubation in 1 M NaCl at 4°C for 16 h. DNA was precipitated from the salt-soluble fraction by the addition of polyethylene glycol (PEG) 6000 (Sigma), to a

final concentration of 10% with incubation on ice for 60 m. After multiple phenol extractions and ethanol precipitations, mtDNA was separated from nuclear DNA (nDNA) and residual RNA by CsCl density-gradient centrifugation in the presence of dye [74 μg/ml; Hoechst 33258 (Sigma)] (see below).

b. Method 2 (DNA and RNA). Mitochondrial DNA and RNA were extracted from isolated mitochondria while nDNA was extracted from a crude nuclear fraction and the resulting nucleic acids were further purified. Nuclei and mitochondria were isolated by differential centrifugation of a whole-cell homogenate by modification of a procedure provided by E.A. Dawidowicz (Tufts University, Boston). A 100-ml log-phase suspension culture was centrifuged at 900*g* for 2 m, and the cells were washed twice with 0.5 vol. 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4, 137 mM NaCl and 2.7 mM KCl [phosphate-buffered saline (PBS)]. Cells were resuspended in 5 ml of homogenization buffer [10 mM Tris-HCl (pH 7.6), 1 mM $ZnCl_2$ and 0.25 M sucrose] and disrupted in a Dounce tissue grinder (35 strokes using pestle A and 5 strokes using pestle B). The homogenate was centrifuged at 900*g* for 10 m to pellet nuclei and intact cells. Nucleic acids were extracted by resuspending the crude nuclear pellet in TE [10 mM Tris-HCl (pH 7.6), 1 mM EDTA-Na_2], with addition of SDS to 1%, NaOAc to 0.3 M and an equal volume of phenol mix, followed by ethanol precipitation from the resulting aqueous phase. Nuclear DNA was further purified by two rounds of CsCl density-gradient centrifugation following RNase A treatment (see below). Mitochondria were recovered by centrifugation of the post-nuclear supernatant at 9,000*g* for 20 m, and were washed twice with 5 ml of 10 mM Tris-HCl, 2 mM EDTA-Na_2, and 0.35 M sucrose (pH 7.7). Nucleic acids were extracted by resuspending the final mitochondrial pellet in TE with addition of SDS to 1%, NaOAc to 0.3 M and an equal volume of phenol mix, followed by ethanol precipitation from the resulting aqueous phase. The mitochondrial nucleic acids were further treated with RNase A for some experiments in which only DNA was required, or with DNase I for some experiments

in which only RNA was required. For those experiments involving direct RNA analysis, high-molecular-weight RNA (mostly rRNA) was separated from small RNAs and DNA by salt fractionation (incubation in 1 M NaCl at 4°C for 16 h).

c. Nuclease treatment of nucleic acids. For DNase treatment, mitochondrial nucleic acids at a ratio of 1 μg to at least 1 unit of DNase I (35 units/μl; RNase-free) were incubated in 50 mM Tris-HCl (pH 7.6) and 10 mM $MgCl_2$ in a total of 100 μl at 37°C for 20 m. Following incubation, samples were diluted to 200 μl with H_2O and extracted with phenol. RNA was recovered by multiple ethanol precipitations, dissolved in H_2O and quantifed by measuring ultraviolet absorbance (OD_{260}). Samples were routinely checked by agarose gel electrophoresis to confirm successful DNase I treatment.

RNase was freed of any contaminating DNase activity by incubation in a boiling water bath for 10 m. For RNase treatment, nuclear nucleic acids (>1000 μg) were incubated with 10 μl RNase A (10 μg/μl) and 50 mM NaCl in TE in a total of 400 μl at 37°C for 30 m. Mitochondrial nucleic acids (<10 μg) were incubated with 3 μl RNase A (10 μg/μl) in TE in 200 μl total volume at room temperature for 20 m. Samples were extracted with phenol and precipitated with ethanol, DNA was dissolved in TE, and the concentration was determined by measuring OD_{260}.

d. Purification of nucleic acids by density-gradient centrifugation. To isolate nDNA minimally contaminated with mtDNA, the crude nuclear fraction, prepared by method 2 above, was treated with RNase A and subjected to two consecutive rounds of density-gradient centrifugation. Eleven g of CsCl were dissolved in 10 ml TE containing the DNA to be fractionated; 50 μl of a 20 mg/ml stock solution of dye [Hoechst 33258 (Sigma)] was added to achieve a final concentration of 74 μg/ml and the sample was mixed by repeated inversion. Samples were centrifuged at 40,000 rpm for 44 h at 20°C in a Ti50 fixed angle rotor in a Beckman L8-70M ultracentrifuge. A long-wave UV lamp (Blak Ray® UVL-21) was used to detect the DNA bands, which

were recovered (in ~2 ml per gradient), applied to a second gradient and centrifuged as above. DNA recovered from the second gradient was mixed with an equal volume of water-saturated butanol. The butanol phase (with dye) was discarded and these extractions were repeated until all traces of dye were eliminated (~4 times in total). After the final butanol extraction, the solution was diluted two-fold with TE and DNA was precipitated by addition of 2.5 vol. 95% ethanol, extracted with phenol, and re-precipitated from ethanol 2 times. The final DNA pellet was dissolved in 50 μl TE and the concentration was determined by OD_{260} measurements.

(iii). Cloning and sequencing of A. castellanii mitochondrial DNA

a. Plasmid cloning. Plasmid vectors used for the cloning of mtDNA were pAT 153 (Twigg and Sherratt, 1980) and PIBI 31. pAT 153 contains ampicillin- and tetracycline-resistance genes and hence growth of bacterial cells on ampilcillin-containing agar plates signifies uptake of plasmid (transformation); lack of growth of the same clones on tetracycline-containing agar plates signifies presence of inserted DNA within the tetracycline gene (i.e., transformation by a recombinant plasmid). PIBI 31 contains the ampicillin-resistance gene and the β-galactosidase gene and hence growth of bacterial cells on ampicillin-containing agar plates signifies transformation with PIBI 31, while white colonies grown in the presence of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) and IPTG (isopropyl-thio-galactoside) signify transformation with recombinant plasmids. PIBI 31 also contains sequences complementary to the universal and reverse sequencing primers of M13 vectors.

Mitochondrial DNA used for cloning and sequencing was prepared by methods 1 and 2 described above. The DNA was digested with appropriate restriction endonucleases (as described below) and ligated into suitably digested plasmid vectors under standard conditions as recommended by the manufacturers. Restriction digests were extracted with phenol and precipitated with ethanol before ligations were set up. Typically, ligations were done in 20 μl reaction volumes containing a 5-fold excess of

the insert DNA (usually ~100 ng) relative to plasmid DNA (usually ~20 ng) in 1/10 vol. 10x ligation buffer [500 mM Tris-HCl (pH 7.6), 100 mM $MgCl_2$, 10 mM DTT, 10 mM ATP, 50% PEG 8000] and 1 μl of T4 DNA ligase (1 unit/μl). Ligations were incubated for at least 4 h at room temperature and sometimes overnight at 4°C, whichever was most convenient. Ligation reactions were heated at 65°C for 5 m immediately after incubation.

1. Competent *E. coli* cells. Recombinant plasmids were transformed into competent *E. coli* cells (strain JM101 or DH5αF'). Two procedures were used for making competent cells, with procedure 2 ultimately being adopted as standard. Procedure (1). Fifty ml of 2x YT medium (1.6 % tryptone, 1.0 % yeast extract, and 1.0 % NaCl) in a 250 ml flask were inoculated with 0.5 ml of a fresh overnight tube culture of JM101 or DH5αF' *E. coli* cells and placed at 37°C for 2 h with shaking. The culture was placed on ice for 10 m and then spun at 2,500 rpm for 7 m. The cell pellet was resuspended in 1/2 vol. 50 mM $CaCl_2$ and left on ice for 40 m. The cells were pelleted and resuspended in 1/20 vol. (i.e., 2.5 ml) 50 mM $CaCl_2$ and 100 - 200 μl aliquots were dispersed to 1.5 ml tubes and kept on ice until use, 8 to 12 h later. Procedure (2). DH5αF' cells were grown to early log phase (OD_{600} = 0.30 to 0.40 nm) in 250 ml of 2x YT medium at 37°C with shaking. The cells were pelleted (4,000 rpm for 15 m) and gently resuspended in 1/10 vol. ice-cold TSB [LB broth (pH 6.1) containing 5% DMSO (dimethyl sulfoxide), 10% PEG 3350, and 20 mM $MgCl_2$] (Chung and Miller, 1988; Chung et al., 1989). Aliquots (200 μl) were dispersed to 1.5 ml tubes and quick-frozen in an ethanol/dry ice bath. Cells were stored at -70°C, with competency being maintained under these conditions for a relativley long period of time (~6 months). Competent cells were not refrozen after being thawed.

2. Transformation of *E. coli* with plasmids. If previously frozen, competent cells were thawed on ice, an aliquot of the ligation reaction (typically 1/5 vol. containing ~4 ng vector DNA) was gently mixed with competent *E. coli* cells (typically

200 μl). If competent cells were prepared according to procedure (1) above, then transformation mixtures were either incubated on ice for 30 to 60 m followed by a heat shock at 37°C for 2 m (JM101, DH5αF') or on ice for 15 m with no heat shock (DH5αF'). If competent cells were prepared according to procedure (2) above, then transformation mixtures were incubated on ice for 5 to 15 m with no heat shock. For plating, 10 μl, 50 μl or 100 μl amounts of transformation mixtures were spread directly onto YT agar plates (0.8% tryptone, 0.5% yeast extract, 0.5% NaCl and 1.5% agar) containing 50 μg/ml ampicillin and incubated at 37°C for ~16 h. When PIBI 31 was the transforming plasmid, 40 μg/ml of X-Gal [2 ml of a 2% solution in *N*, *N*-dimethylformamide (DMF) per liter of YT agar] was also included in the YT agar plates. [Sometimes X-Gal (50 μl of a 2% solution in DMF) and IPTG (20 μl of a 100 mM solution) were spread directly onto the agar plates, which were then allowed to dry before transformed cells were spread.] Colonies that were positive for transformation by recombinant plasmids were picked to 2 ml of 2x YT medium containing 50 μg/ml of ampicillin. The 2 ml cultures were incubated in a roller drum for ~16 h at 37°C.

3. Alkali plasmid mini-preps. Alkali plasmid preps from the tube cultures were carried out by modification of a procedure described by Maniatis et al. (1982), with clones being identified by restriction profiles on agarose gels. Briefly, 1 ml of the 2 ml culture was transferred to an eppendorf tube and centrifuged at 13,000*g* for 1 to 2 m to sediment bacterial cells, and the pellet was then re-centrifuged to eliminate residual supernatant. Cell pellets were resuspended in 100 μl of TEG [25 mM Tris-HCl (pH 8.0), 50 mM EDTA-Na_2, 1% glucose], vortexed and incubated on ice for 5 m. After addition of 200 μl 0.2 M NaOH and 1% SDS, the resulting lysate was vortexed gently and left at room temperature for 2 m. Then 150 μl 7.5 M NH_4OAc were added and the mixture was vortexed gently and incubated on ice for 5 m. The sample was centrifuged at 13,000*g* for 5 m and DNA was precipitated from the supernatant with 2 vol. ethanol. The final DNA pellet was dissolved in 30 ml of TE. Typically, 6 μg of DNA were

obtained by this procedure, an amount sufficient for any required manipulations, such as double-stranded (ds) plasmid DNA sequencing or subcloning into M13 vectors. When a larger-scale plasmid prep was required, it was performed in an essentially identical manner to that described for the isolation of M13 RF DNA [see (iv)., b., 3. Nested deletion series].

4. Plasmid clones obtained. A 6.7 kbp *Bam*HI fragment of mtDNA (prepared by method 1 above) was cloned into the plasmid vector pAT 153. This clone (pKL-1) contained the entire SSU rRNA gene, the 3' half of the LSU rRNA gene, the 5 tRNA genes of cluster 2 (between the LSU and SSU rRNA genes) and the majority of the *cox*1 gene. The 5' half of the LSU rRNA gene and the 5 tRNA genes of cluster 1 (immediately upstream of the LSU rRNA gene) were obtained in an overlaping 7.7 kbp *Mlu*I fragment of mtDNA (prepared by method 2 above) cloned into the plasmid PIBI 31 (the resulting clone was designated pKL-2). The 3'-terminal region of the *cox*1 gene and the 5' terminus of the *cox*2 gene were contained in an adjacent 350 bp *Bam*HI fragment of mtDNA and the majority of the *cox*2 gene was present in an overlapping 5.31 kbp *Sal*I fragment; both were cloned into plasmid vector PIBI 31, giving the recombinant plasmids pKL-3 and pKL-4, respectively. Where applicable, clones to span restriction site junctions were also prepared (*Sal*I 4.8 kbp, pKL-5).

5. Preparation of plasmid dsDNA template for sequencing. The entire insert of pKL-3 and the terminal regions of the inserts of pKL-4 and pKL-5 were sequenced by dsDNA sequencing. Prior to sequencing, ~2 μg of plasmid dsDNA (10 μl) were incubated with 5 μl RNase A (10 μg/μl), 2 μl of 5 M NaCl (50 mM) and 183 μl TE (pH 7.6), in a total volume of 200 μl at 37°C for 60 m. Samples were then extracted with phenol and precipitated with ethanol. The DNA pellet was dissolved in H_2O and denatured in 0.2 M NaOH, 0.2 mM EDTA-Na_2 in a total volume of 100 μl at 37°C for 30 m (Zhang et al., 1988). DNA was precipitated by addition of 10 μl 3 M NaOAc (pH 4.7) and 300 μl 95% ethanol, incubated at -70°C for 30 m, and then

washed with 70% ethanol. The DNA pellet was dissolved in 14 μl H_2O and 7 μl were used for sequencing reactions exactly as for single-stranded (ss) template DNA sequencing (see c. DNA sequencing).

b. M13 subcloning. All plasmid clones except pKL-3 and pKL-5 were subcloned into M13 vectors (Messing, 1983) of the um20, um21 series (International Biotechnologies Inc.) for sequencing. M13 RF DNA [replicative form; for preparation of M13 RF DNA see (iv)., b., 3. Nested deletion series] and recombinant plasmid DNAs were digested with the appropriate restriction endonucleases to obtain the desired subclones [see (iv)., b., 2. M13 subclones obtained] and ligated under standard conditions.

1. Transformation of *E. coli* with M13. Transformation of competent *E. coli* cells with M13 RF DNA was performed exactly as for transformation with plasmid DNA [see (iv)., a., 2.]. For plating, 10 μl, 50 μl or 100 μl aliquots of the transformation mix were combined with 3 ml of top agar (0.8 g tryptone, 0.5 g yeast extract, 0.5 g NaCl, and 0.8 g agar per 100 ml, previously maintained at 50°C), 50 μl of 2% X-Gal (in DMF), 10 μl of 100 mM IPTG (in H_2O) and 300 μl of *E. coli* plating culture (a late log-phase culture) and poured onto YT agar plates. Plates were inverted and incubated at 37°C for 8 to 12 h. White plaques and a blue plaque (see below) were picked to 2 ml of 2x YT medium containing 10 μl of an early log-phase suspension of *E. coli* cells. The 2 ml cultures were incubated in a roller drum for 8 to 10 h at 37°C. The cultures were spun at 2,000 rpm for 10 m to pellet the bacterial cells and were stored at 4°C until use. The supernatant, which contains the ss phage DNA, was analyzed by agarose gel electrophoresis to identify subclones [see (iv)., b., 3. Nested deletion series].

C-tests were performed to monitor the subcloning of complementay strands. Five μl of the two test supernatants were mixed together with 2.5 μl of 5x SDS dye [25% glycerol, 10 mM Tris-HCl (pH 8.0), 5 mM EDTA-Na_2, 0.02% bromophenol blue (BPB), 0.02% xylene cyanol (XC) and 2% SDS] and incubated at 65°C for 1 h.

Samples were loaded directly onto a 0.7% agarose gel (24 cm by 14 cm by 0.4 cm) in TEAS [50 mM Tris-OAc, 2.8 mM EDTA, 20 mM NaOAc, and 20 mM NaCl (pH 8.0)] and electrophoresed at 125 V for ~1 h. Gels were stained with ethidium bromide, destained and photographed. Appropriate supernatants were used to make either template ssDNA for sequencing or to prepare RF M13 dsDNA for deletion series, as required.

2. M13 subclones obtained. M13 subclones used for sequencing of mtDNA included: 1). from pKL-1: *Bam*HI-*Mlu*I, 3.2 kbp subclone; *Mlu*I-*Mlu*I, 1.5 kbp subclone; *Mlu*I-*Bam*HI, 2.2 kbp subclone, 2). from pKL-2: *Xba*I-*Xba*I, 3.2 kbp subclone, and 3). from pKL-4: *Sal*I-*Eco*R1, 1.5 kbp subclone. The complementary strands for all of the above subclones were also obtained.

3. Nested deletion series. To sequence regions of DNA >500 bp in length, multiple series of deletion clones were constructed from recombinant M13 RF DNA and the resulting ss phage DNA was used for sequencing.

Preparation of RF DNA was equivalent to a large-scale version of the alkali plasmid mini-prep described above. In a 500 ml flask, 50 ml of 2x YT medium was inoculated with 200 μl of an overnight tube culture of *E. coli* DH5αF' cells and incubated at 37°C for 1 h. Phage stock (200 μl from the 2 ml supernatant of infected *E. coli* cells) was added and incubation continued at 37°C for an additional 8 to 12 h with constant shaking. Cultures were centrifuged at 10,000 rpm for 5 m and the supernatant either discarded or used to isolate ss phage DNA. The cells were washed with 25 ml of TES [50 mM Tris-HCl (pH 8.0), 10 mM EDTA-Na_2, 50 mM NaCl]. The washed cells were resuspended in 1 ml of TEG and kept on ice for 5 m. Two ml of 0.2 N NaOH, 1% SDS was added with vigorous shaking and vortexing and the mixture left on ice for 10 m; 1.5 ml 7.5 M NH_4OAc (pH 7.5) [or sometimes 5 M KOAc (pH 4.8)] was then added, and the lysate was mixed by shaking and the mixture left on ice for 5 m. Samples were centrifuged at 10,000 rpm for 10 m and the nucleic acids were

precipitated from the supernatants by addition of 2 vol. 95% ethanol followed by incubation on ice for 20 m and centrifugation at 10,000 rpm for 15 m. [To salt out any remaining large RNAs, the pellet was dissolved in 2 ml of TE, 1 ml of 7.5 M NH_4OAc (pH 7.5) was added, and the solution was mixed by vortexing and left on ice for 30 m, after which it was centrifuged at 10,000 rpm for 10 m.] DNA was isolated from the supernatant by ethanol precipitation, dissolved in TE, extracted with phenol, and precipitated again with ethanol. The final RF DNA pellet was dissolved in TE.

M13 RF DNA could be purified away from any contaminating *E. coli* chromosomal DNA by agarose gel electrophoresis. RF DNA (~40 μg) was applied to a low-melting point (LMP) 0.7% Seaplaque agarose gel (24 cm by 14 cm by 0.4 cm) in TEA [0.1 M Tris-OAc, 0.1 M EDTA (pH 8.0)], and electrophoresed at 70 V for ~12 h. The gel was stained with ethidium bromide in the dark, and the RF DNA was visualized using a short-wave UV lamp (Mineralight® UV6-2) and excised from the gel. A 1 ml volume of TE was added to the gel slice in a Corex tube and the agarose was melted by placing the tube at 65°C for 5 to 10 m. The resulting solution was made up to 6 ml with TE, mixed by vortexing and heated again at 65°C for ~2 m. An equal volume of phenol was added with repeated vortexing and the mixture was centrifuged at 10,000 rpm for 15 m. The phenol phase was re-extracted with TE and the resulting aqueous phases were combined. Multiple phenol extractions and ethanol precipitations followed. The purified DNA was dissolved in 30 μl of TE.

Treatment of the M13 RF DNA with exonuclease III and mung bean nuclease was used to generate an overlapping series of deletion clones for sequencing (Henikoff, 1984). RF DNA (5 μg) was digested with appropriate restriction endonucleases to generate a 5' overhang or blunt end next to the insert DNA and a 3' overhang next to the vector DNA within the polylinker region. Digests were extracted with phenol and precipitated with ethanol and the resulting DNA was then dissolved in 20 μl H_2O. Four separate samples of digested DNA were treated with exonuclease III (exoIII) in

reaction volumes of 20 μl containing 5 μl DNA (~1 μg), 2 μl 10x reaction buffer [500 mM Tris-HCl (pH 8.0), 100 mM $MgCl_2$)], 11 μl H_2O and 2 μl exoIII at 20 units/μl. All four samples were placed at 37°C at the same time and aliquots (2 μl) were removed sequentially from the reaction mix every 30 s for 5 m per reaction tube, encompassing a total incubation series of 5, 10, 15 and 20 m. The 2 μl aliquots were added to tubes containing 6 μl of 5x mung bean nuclease buffer [250 mM NaOAc (pH 5.0), 250 mM NaCl, 0.5 mM ZnOAc, 0.005% Triton-X 100, 25% glycerol] on ice. After all exoIII reaction samples had been combined (20 μl per tube), 2 μl H_2O and 2 μl mung bean nuclease (10 units/μl) were added and the mixtures incubated at 37°C for 10 m to degrade the single-stranded 5' protrusions of the RF DNA. The reaction was stopped by addition of 1.5 μl 1 M Tris followed by heating at 70°C for 5 m. Aliquots (5 μl) were removed for agarose gel electrophoresis to monitor the extent of digestion by exoIII. The remaining samples (4 x 26.5 μl) were added to tubes containing 2.5 μl 100 mM $MgCl_2$, 5 μl each of the 4 dNTPs (1 mM), 11 μl H_2O and 1 μl Klenow DNA polymerase (6 units/ μl) and were incubated at 37°C for 20 m for blunt-end formation. Samples were then extracted with phenol and precipitated with ethanol; the DNA was dissolved in 50 μl H_2O and 1/10 vol. (5 μl) was used for ligation in a 100 μl volume containing 2 μl of DNA ligase (1 unit/μl) and 10 μl 10x ligation buffer. Ligation mixtures were typically incubated at room temperature for at least 4 h or, if more convenient, were left overnight initially in a 15°C H_2O bath that gradually warmed to room temperature. One-fifth of the ligation mix (20 μl) was used for transformation of competent *E. coli* cells and individual plaques were picked as described previously.

The size range of M13 inserts was determined by agarose gel electrophoresis. Phage supernatant (40 μl) was mixed with 10 μl of 5x SDS dye. Samples were heated at 65°C for 10 m before loading 10 μl per well on a 1% agarose gel (20 cm by 30 cm by 0.5 cm) in TBE-2 [89 mM Tris, 89 mM borate, 1.8 mM EDTA, (pH 8.1)] and electrophoresing at 150 V for ~20 h. Gels were stained with ethidium bromide and

photographed. Size markers included phage DNA from a blue plaque (i.e., M13 without insert DNA) and phage DNA from the original M13 clone before exoIII digestion (i.e., M13 containing full-size insert DNA).

4. Preparation of single-stranded template DNA. Single-stranded (ss) phage DNA was isolated from the supernatant of *E. coli* cultures infected by single M13 plaques (Heidecker et al., 1980). One ml of supernatant was centrifuged for 2 m at 13,000*g* to pellet the cells. Phage were precipitated from the supernatant by addition of 200 μl 20 % PEG 6000, 2.5 M NaCl, with incubation at room temperature for 15 m and on ice for 30 m. The phage were pelleted by centrifugation at 13,000*g* for 10 m. The supernatant was discarded and the pellet re-centrifuged for 1 m to ensure that all of the PEG was removed. The phage were dissolved in 150 μl TE, an equal volume of phenol mix was added, and the mixture was vortexed for 20 s and left at room temperature for 15 m. The aqueous phase was extracted with phenol two more times and then 1/10 vol. 3 M NaOAc and 2 vol. 95% ethanol were added. Samples were left at -20°C for at least 2 h before centrifugation at 13,000*g* for 15 m. The pellet was washed with 200 μl of 80 % ethanol and allowed to dry in air. The phage ssDNA was dissolved in 20 μl TE and stored at -20°C until use (7 μl were used as sequencing template).

c. DNA sequencing. Both strands of the DNA were sequenced using the dideoxy-chain termination method (Sanger, 1977) of Sequenase™, version 2.0 (U.S. Biochemicals, Cleveland, Ohio). Any sequencing ambiguities were resolved by substituting dITP for dGTP in the sequencing reactions, as recommended by the manufacturer. For some experiments as indicated in the figure legends, oligonucleotides specific for LSU rRNA, SSU rRNA or tRNA were 5'-end labeled with [γ-^{32}P]ATP and polynucleotide kinase and used as sequencing primers (5 ng; 0.5 to 1 pmol), replacing the M13 universal primer. In these experiments, the labeling step of the sequencing reactions using [α-^{32}P]dATP was omitted. Also, the M13 reverse

primer (0.5 to 1 pmol) was used in addition to the universal primer for plasmid dsDNA sequencing. Sequencing reaction products were resolved in 6% polyacrylamide, 7 M urea gels (50 cm by 33 cm by 0.025 cm) containing TBE [50 mM Tris, 50 mM boric acid, 1 mM EDTA (pH 8.1)] by electrophoresis for 3 h and for 8 h at 2000 V. The gels were often dried onto 3MM paper (Whatman) using a BioRad slab gel dryer before exposure to XK-1 X-ray film (Kodak), usually overnight. Primary sequence data were analyzed using the MicroGenie™ (Beckman) (Queen and Korn, 1984), DNA Strider™ 1.1 (Marck, 1989) and FASTA (Pearson and Lipman, 1988) programs.

(iv). Direct analysis of ribosomal RNAs

a. Purification by gel electrophoresis. The salt-insoluble fraction of mitochondrial nucleic acids (mostly rRNA) was either 5'-end labeled (b. and c. below) or 3'-end labeled (d. below) with ^{32}P. Radiolabeled RNA (~10 μg) was dissolved in NMF/urea loading buffer [10 mM Tris-HCl (pH 7.6), 1 mM EDTA-Na_2, 5 M urea, 60% *N*-methylformamide, and 0.1% BPB and XC dyes], heated at 65°C for 5 to 10 m and applied to a 2.5% polyacrylamide, 7 M urea gel (20 cm by 20 cm by 0.15 cm) in TBE. The gel was prerun at 200 V for 1 to 2 h and samples were electrophoresed for 5 h at 350 V. The rRNAs were detected by autoradiography and extracted from the gel according to Rubin (1973). (If the rRNAs were unlabeled, they could be visualized by UV shadowing.) Briefly, gel pieces were homogenized in 2 ml of a solution containing 0.1 M Tris-HCl (pH 9.1), 0.5 M NaCl and 10 mM EDTA-Na_2 plus 2 ml of phenol mix, and incubated with vigorous shaking at 4°C for 16 h. Carrier *E. coli* tRNA (5 to 8 μg) was added to the aqueous phase. After multiple phenol extractions, the RNA was isolated by ethanol precipitation.

b. Enzymatic sequencing of 5'-end-labeled RNA. The mitochondrial salt-insoluble nucleic acid fraction was dephosphorylated by bacterial alkaline phosphatase (BAP), labeled at the 5' end with [γ-^{32}P]ATP and polynucleotide kinase under standard conditions, and purified after gel electrophoresis, as described above. The gel-purified,

5'-end-labeled RNA was partially digested as described by Donis-Keller et al. (1977). Briefly, for RNases T1 (cleaves after G) and Phy M (cleaves U > A), 1 μl amounts of the appropriately diluted ribonucleases were added to 5 μl of the RNA in a solution containing 20 mM Na-citrate (pH 5.0), 7 M urea, 1 mM EDTA-Na_2, 0.05% XC and 0.75 μg of carrier tRNA; for RNase U2 (cleaves after A) the same solution was used except with 20 mM Na-citrate at pH 3.5. Incubation was at 50°C for 30 m. The RNA was also subjected to partial or complete alkaline hydrolysis by incubation in 0.15 M NH_4OH at 90°C for 1 to 2.5 m or for 15 m, respectively, followed by a series of lyophilizations. Samples were loaded directly onto 20% polyacrylamide, 7 M urea gels (32 cm by 40 cm by 0.05 cm) in TBE, prerun at 2000 V for 2 h and electrophoresed at 2000 V for 3 h.

(c). Terminal nucleotide analysis. This was performed according to MacKay et al. (1980). Aliquots of the gel-purified, 5'-^{32}P-end-labeled LSU and SSU rRNAs were hydrolyzed with snake venom phosphodiesterase in a 20 μl reaction volume containing 0.125 M ammonium formate (pH 9.2) at 37°C for 16 h. The hydrolysate was applied to cellulose thin layer chromatography (tlc) plates [pre-run in saturated $(NH_4)_2SO_4$] containing fluorescent indicator (Eastman), along with marker nucleotides (pG, pA, pU and pC). Chromatography in the first dimension was with 95% ethanol:H_2O, 4:1 and in the second dimension with saturated $(NH_4)_2SO_4$:2-propanol, 40:1. Migration of marker nucleotides was visualized with a short-wave UV lamp (Mineralight® UV6-2) and then the tlc plate was exposed to X-ray film.

d. Chemical sequencing of 3'-end-labeled RNA. Ribosomal RNA from the salt-insoluble fraction of mitochondrial nucleic acids was labeled at the 3' end with [5'-^{32}P]pCp (Walseth and Johnson, 1979) by RNA ligase (Peattie, 1979), and purified by gel electrophoresis, as described above. The gel-purified, end-labeled RNA was sequenced using a combination of partial digestion with RNase T1 (to determine G positions) and partial chemical degradation (to determine A, U and C), as described by

Peattie (1979) with only minor modifications. Briefly, base-specific modifications were generated, followed by strand scission induced by treatment with aniline acetate. Carrier *E. coli* tRNA was added at 1 μg/μl of RNA. For the A-specific modification, 1 μl of diethylpyrocarbonate was added to 5 μl of RNA in 200 μl 50 mM NaOAc (pH 4.5) on ice, followed by incubation at 90°C for 3.5 m. The reaction products were precipitated by the addition of 50 μl 1.0 M β-mercaptoethanol, 1.5 M NaOAc and 750 μl 95% ethanol. For the U-specific reaction, RNA was lyophilized and then redissolved in 10 μl of 50% hydrazine, incubated on ice for 8 m and the products of the reaction precipitated with 200 μl 0.3 M NaOAc and 750 μl 95% ethanol. For the C-specific modification, the lyophilized RNA was dissolved in 10 μl 3 M NaCl/hydrazine, incubated on ice for 20 m, and the products precipitated with 80% ethanol. Reacted RNAs in each of the samples were precipitated with 0.3 M NaOAc and 3 vol. ethanol. The RNA pellets were each dissolved in 6.4 μl H_2O and added to 13.6 μl of the aniline acetate [aniline: glacial acetic acid: H_2O, (1:3:3.5)]. Samples were lyophilized multiple times and then dissolved in NMF/urea loading buffer. Sequencing reaction products were resolved in pre-run (2000 V for 3 h) 20% polyacrylamide, 7 M urea gels (32 cm by 40 cm by 0.05 cm) in TBE at 2000 V for 3 h.

(v). Primer extension analysis and cDNA sequencing

All primer extension analysis and cDNA sequencing reactions were performed according to the method of Geliebter (1987) with only minor modifications as indicated, where applicable. Gene-specific oligonucleotides (22-mers) contained 5'-hydroxyls that were labeled with [γ-^{32}P]ATP (7000 Ci/mmol) using polynucleotide kinase under standard conditions (Geliebter, 1987), except that spermidine was used at a final concentration of 1 mM. Mitochondrial RNA was treated with DNase I and 2 - 10 μg were annealed to 5 ng of the radiolabeled oligonucleotide, which served as a primer for extension by avian myeloblastosis virus (AMV) reverse transcriptase in the presence of dNTPs. For cDNA sequencing, ddNTPs were included in the reaction mix

at the concentrations indicated (Geliebter, 1987) unless otherwise specified, and actinomycin D was omitted unless otherwise specified. Stop solution was 95% formamide, 20 mM EDTA, 0.05% BPB and 0.05% XC (as for DNA sequencing samples).

a. Ribosomal RNAs. One to 2 μg of mtRNA was used for the annealing. Oligonucleotides specific for internal regions of either the SSU rRNA or the LSU rRNA were as follows:

(1). SSU rRNA-specific for determination of 5' termini by primer extension analysis:

5'-GTGTTAAAATATGTTACCAGCG-3';

(2). LSU rRNA-specific for determination of 5' termini by primer extension analysis:

5'-CTTTTATGCTATAAGATATTCG-3';

(3). LSU rRNA-specific for cDNA sequence analysis across intron boundaries,

a. within LSU rRNA domain IV:

5'-TTTCACTAGGTTAATACTGGAG-3';

b. within LSU rRNA domain V:

5'-ATCCTCTCGTACAAAAGAAACA-3'.

Primer extension and cDNA sequencing products were electrophoresed in 10% polyacrylamide, 7 M urea gels (50 cm by 33 cm by 0.025 cm) containing TBE at 2000 V for ~3 h. Gels were exposed to Kodak XK-1 X-ray film for ~24 h.

b. Transfer RNAs. Ten μg of mtRNA were annealed to radiolabeled oligonucleotides complementary to an internal region (from the 5' end of the anticodon loop to the 5' end of the TψC loop) of each of the five tRNAs of cluster 2. The sequences of the gene-specific oligonucleotides used for tRNA cDNA sequencing were as follows:

$tRNA^{Met1}$: 5'-ACCAATAACCTTTCGTTATGAG-3';

$tRNA^{Ala}$: 5'-ACTGATATCTAATACGTGCAAG-3';

$tRNA^{Pro}$: 5'-ACCTATAGTCTTCTGTTCCCAA-3';

$tRNA^{Asp}$: 5'-ACCTGCATCTTTTAACTCGACA-3';

$tRNA^{Met2}$: 5'-ACCTACATCGTCAGGGCATGAG-3'.

Sequencing reaction products and primer extension products were electrophoresed in 10% polyacrylamide, 7 M urea gels (50 cm by 33 cm by 0.025 cm) in TBE for ~3 h at 2000 V. Gels were exposed to Kodak XAR-5 X-ray film for 1 to 5 d with or without an intensifying screen.

To obtain the precursor-$tRNA^{Met1}$ sequence, the $tRNA^{Ala}$-specific oligonucleotide was used as a primer in reverse transcriptase sequencing; likewise, to obtain the precursor-$tRNA^{Ala}$ sequence, the $tRNA^{Asp}$-specific oligonucleotide was used as a primer. Sequences (cDNA) were determined as described above, except that the concentration of ddNTPs used for the RNA sequencing was 1/5 of that described (Geliebter, 1987). Reaction products were analyzed by electrophoresis for 5 to 6.5 h at 2000 V in 6% polyacrylamide, 7 M urea gels (50 cm by 33 cm by 0.025 cm) in TBE. The gels were dried onto 3MM paper (Whatman) using a BioRad slab gel dryer and exposed to X-ray film (XAR-5, Kodak) for at least 14 d in the presence of an intensifying screen.

c. Messenger RNA encoding cytochrome oxidase subunits 1 and 2. Ten μg of mtRNA were annealed to radiolabeled oligonucleotides specific for either the *cox*1 or *cox*2 mRNA. The sequences of the oligonucleotide primers were as follows:

*cox*1-specific: 5'-CGTACCGATATCTTTGTGATTA-3';

*cox*2-specific: 5'-CATAAGCAATTGGCGTTGTAGG-3'.

For these reactions, actinomycin D was present and ddNTPs were used at 1/5 or 1/10 the recommended concentrations (Geliebter, 1987). Sequencing products were resolved in 8% polyacrylamide, 7 M urea gels (50 cm by 33 cm by 0.025 cm) containing TBE, with electrophoresis at 2000 V for 2.5 to 5 h.

(vi). Terminal deoxynucleotidyl transferase treatment for elimination of ambiguities in tRNA sequence analysis

Terminal deoxynucleotidyl transferase, by virtue of its ability to extend ("tail") 3' hydroxyl termini of DNA, was used to eliminate sequencing ambiguities occasionally resulting from internal pausing of reverse transcriptase (DeBorde et al., 1986). For $tRNA^{Met1}$ and $tRNA^{Met2}$, the products of standard RNA sequencing reactions were extracted with phenol, precipitated three times with ethanol (to ensure removal of ddNTPs), dissolved in water, and heated at 80°C for 2 m. Samples were then cooled on ice and incubated with 40 units of terminal deoxynucleotidyl transferase (20 units/μl; Promega) in the presence of 1 mM dATP, 100 mM cacodylic acid (pH 6.8), 1 mM $CoCl_2$, 0.1 mM dithiothreitol, and bovine serum albumin (0.1 mg/ml) in a 20 μl reaction volume at 37°C for 60 m. Samples were precipitated with ethanol and loaded onto the sequencing gel and resolved as usual.

(vii). Southern hybridization analysis of tRNA genes

A. castellanii nDNA and mtDNA were prepared by method 2 and treated with RNase A. The DNAs were digested with restriction endonucleases, and 2.5 μg of nDNA and 2.5 ng of mtDNA were applied per lane of a 0.7% agarose gel (13 cm by 9 cm by 0.3 cm) in 40 mM Tris-OAc, 1 mM EDTA-Na_2 (pH 8.1); restriction fragments were resolved by electrophoresis for 2 h at 150 V. The gel was stained for 30 m with ethidium bromide, destained for 5 m in H_2O, and photographed. To facilitate transfer of large DNA fragments, the gel was incubated in 0.25 M HCl for 10 m at room temperature. The gel was rinsed in H_2O and the DNA was denatured by immersing the gel in 0.5 N NaOH, 1.5 M NaCl for 45 m at room temperature with gentle agitation. The gel was rinsed with H_2O and neutralized by incubating in 2 M NaCl, 1 M Tris-HCl (pH 7.0) for 45 m at room temperature with gentle agitation. DNA was transferred overnight in 10x SSC to GeneScreen Plus nylon membrane (DuPont Canada, Lachine, Quebec). Following transfer, the DNA was denatured by immersing the blot in 0.4 M NaOH for 60 s and was then neutralized in 0.2 M Tris-HCl (pH 7.5), 2x SSC

(standard saline citrate: 1x is 0.15 M NaCl and 0.015 M sodium citrate) for 15 m at room temperature. The blot was allowed to air dry before using.

The blot was prehybridized for 2 h at 65°C in 30 ml of 5x SSC, 20 mM sodium phosphate (pH 7.0), 10x Denhardt's solution (Maniatis et al., 1982), 7% SDS, and 10% dextran sulfate with occasional mixing. The oligonucleotide probe (100 ng) was 5'-end labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase under standard conditions and, after heating at 75°C for 5 m, was added to the prehybridization solution. Hybridization was for ~16 h with occasional mixing at a temperature 6° to 7°C lower than the calculated dissociation temperature (T_d) of the oligonucleotide probe (Zeff and Geliebter, 1987). The blot was washed for a total of 60 m in 2x SSC and 1% SDS and for 15 m in 0.5x SSC and 0.1% SDS at a wash temperature 7° to 12°C below the T_d. Washed blots were exposed to Kodak XAR-5 X-ray film for 12 h to 5 d in the presence of an intensifying screen. For reprobing experiments, the blot was stripped of bound oligonucleotide probe by washing in 1x SSC at 85°C for 10 min. Stripping was confirmed by re-exposure to X-ray film.

The sequence of the oligonucleotide probe specific for the *A. castellanii* nuclear myosin IB gene is 5'-GTCGCCCGTTTGCGCATCGTAA-3'.

(viii). Northern hybridization analysis of tRNAs

Five to 10 μg of mtRNA, previously treated with DNase I, were dissolved in 3 μl NMF/urea loading buffer, heated at 65°C for 10 m, applied to each lane of a 10% polyacrylamide, 7 M urea gel (20 cm by 20 cm by 0.15 cm) in TBE, and electrophoresed for ~4 h at 350 V (or until the xylene cyanol reached the gel bottom, about 17 cm). The gel was stained with ethidium bromide in TBE and photographed. The RNA was transferred to nylon membrane (Biotrans, ICN Biomedicals) using a BioRad electrophoretic transfer-cell (Trans-Blot^{R}SD) according to the manufacturer's specifications. The blot was cut into strips for individual hybridizations. Prehybridization and hybridizations were as described for the Southern blots. The

washing temperature was 6 - 10°C below the T_d of the oligonucleotide probe. Washed blots were exposed to Kodak XAR-5 X-ray film with an intensifying screen for 5 to 14 d.

(ix). Potential secondary structures

Potential secondary structures for ribosomal and transfer RNAs were based upon comparative primary sequence and secondary structure analysis of homologous molecules.

(x). Analysis of open reading frames

Open reading frames (ORFs) were detected and analyzed by DNA Strider™ 1.1 (Marck, 1989). Alignment of *cox*1, *cox*2 and intron ORF sequences and determination of comparative amino acid similarities was based on the results of FASTA searches (Pearson and Lipman, 1988) and Clustal V multiple sequence alignments (Higgins et al., 1992), as indicated in figure legends.

Nucleic acid pairwise comparisons were based upon alignments that take into consideration conservation of secondary structural features within intron core regions and amino acid alignments of the intron ORFs.

(xi). Phylogenetic analysis of mitochondrial rRNA genes

DNA sequence alignments, construction of which is facilitated by consideration of conserved secondary structural features and primary sequence of rRNAs, were used for pairwise comparisons and for phylogenetic treeing analysis. Phylogenetic trees derived from mitochondrial LSU rRNA gene sequences were obtained using maximum likelihood, maximum parsimony and distance methods of tree construction. Maximum likelihood was performed using the DNAML program, version 3.5c of Felsenstein (1993). Maximum parsimony tree construction was performed using the heuristic search method (general) for bootstrap analysis (100 replications, 50% majority rule) of PAUP, version 3.0r (Swofford, 1990). Distance trees were generated using Seqboot (100 replications), DNADIST (ML option D, option M), the distance matrix program of

Neighbor, and Consense and Draw of the PHYLIP program, version 3.5c (Felsenstein, 1993).

III. Results

Figure 4 is an electron micrograph of a whole *A. castellanii* cell. Many mitochondria are visible and there is no indication of intracellular bacteria.

Suspension cultures of *A. castellanii* appeared healthiest, as judged by growth rate and microscopic appearance, between 50 and 60 h after inoculation of culture medium, i.e., within the early- to mid-log phase of growth (Fig. 5). Cells were harvested during this time for nucleic acid extraction in order to minimize the degradation of mitochondrial nucleic acids that occurs predominantly in stationary phase cultures of *Acanthamoeba* prior to the encystment process (Neff et al., 1984). Consistent yields of ~100 μg of mtRNA and ~10 μg of mtDNA were obtained, by method 2, from 100 ml suspension cultures.

Nuclear and mitochondrial DNAs formed distinct, well resolved bands in CsCl density-gradients in the presence of Hoechst dye (Fig. 6).

A. Gene Organization and Sequence

A restriction endonuclease map was previously constructed by Von Gabain and Bohnert for the 40 kbp circular mitochondrial genome of *A. castellanii* (cited in Byers, 1986), and heterologous Southern hybridization data indicated that the SSU rDNA was localized to a 6.7 kbp *Bam*HI fragment, with the LSU rDNA located upstream of the SSU rDNA. This information was utilized for the cloning of these genes in the present study.

A genetic map of the region of mtDNA sequenced in this study (~ 25% of the genome) is illustrated in Fig. 7. Genes are tightly clustered with little intergenic space. As predicted by the preliminary results of Von Gabain and Bohnert, the LSU rRNA gene (*rnl*) is found upstream of the SSU rRNA gene (*rns*) in *A. castellanii* mtDNA, with the two genes separated by a stretch of 634 nt that contains tRNA gene cluster 2.

Figure 4. Electron micrograph of a whole *A. castellanii* cell. Cells were harvested from the early log-phase of growth. After washing twice in PBS, cells were fixed in glutaraldehyde (diluted 1:10 in 0.1 M sodium cacodylate) and processed for electron microscopy. Magnification is 5300x.

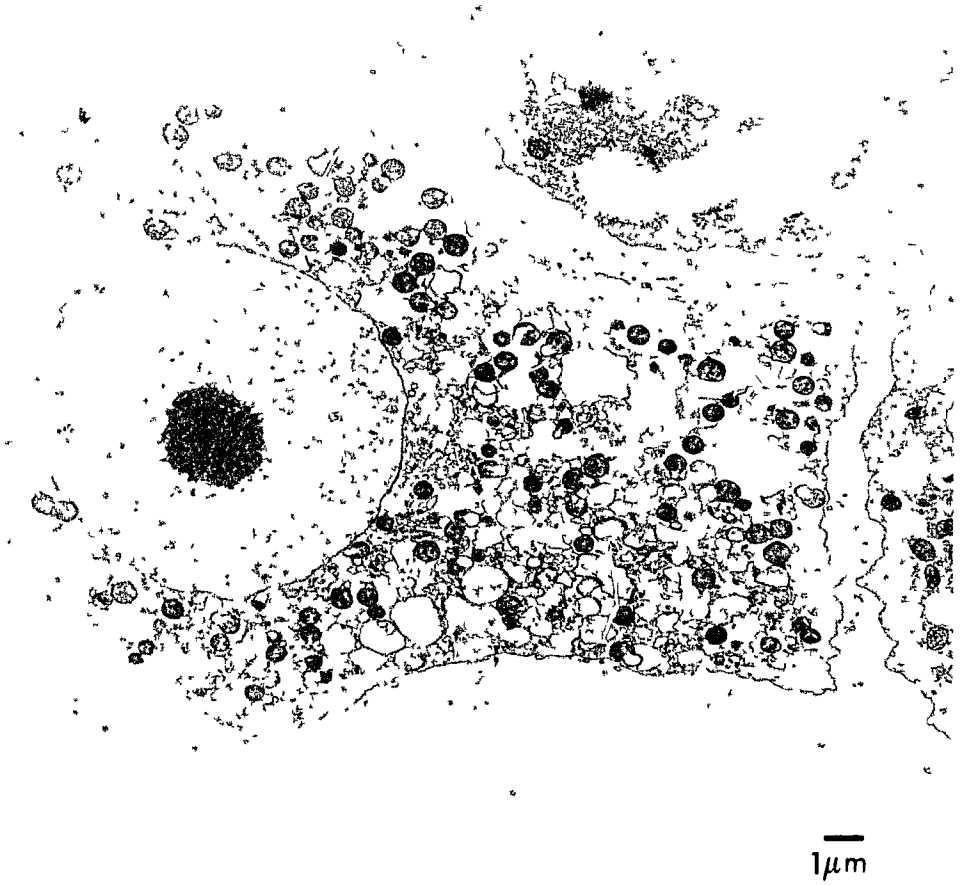


Figure 4

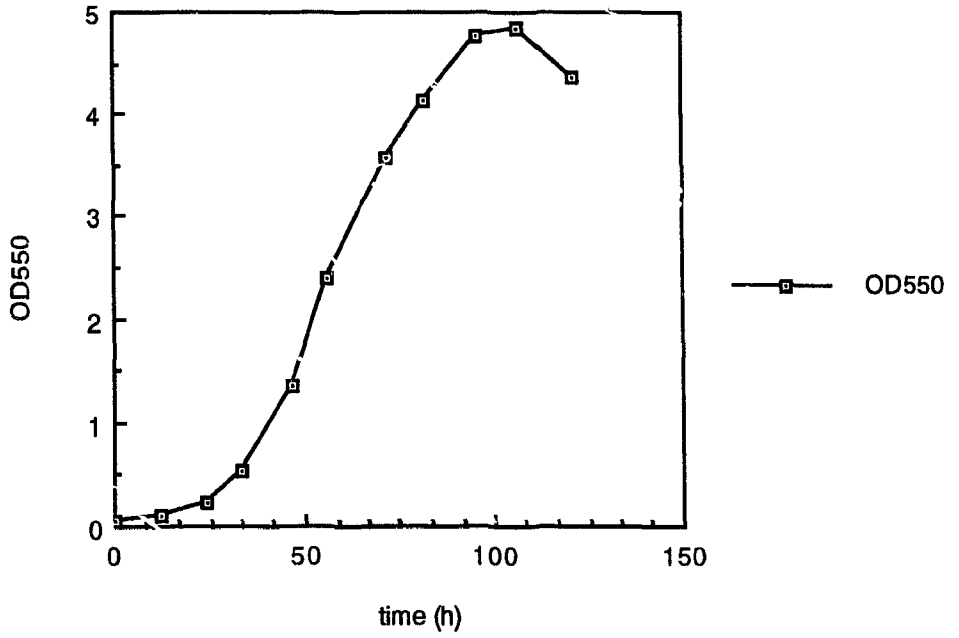


Figure 5. Growth curve of a suspension culture of *A. castellanii.* Growth medium (100 ml) was inoculated with 2 ml of a 10 ml log-phase suspension culture of *A. castellanii* and incubated at 30° C with gentle agitation. Culture growth was monitored by absorbance readings at 550 nm at approximately 12 h intervals.

Figure 6. Fractionation of total cellular DNA from *A. castellanii* by CsCl density gradient centrifugation. DNA was prepared by method 1.

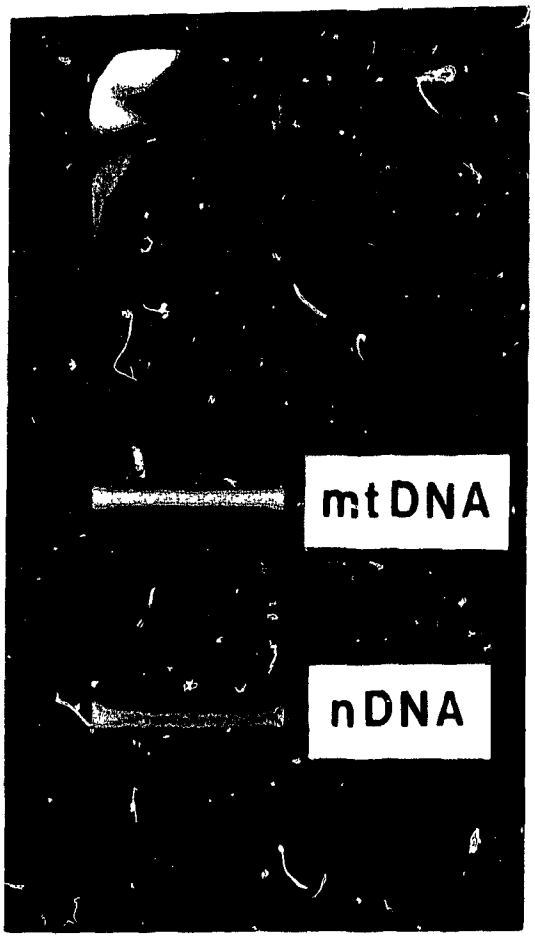


Figure 6

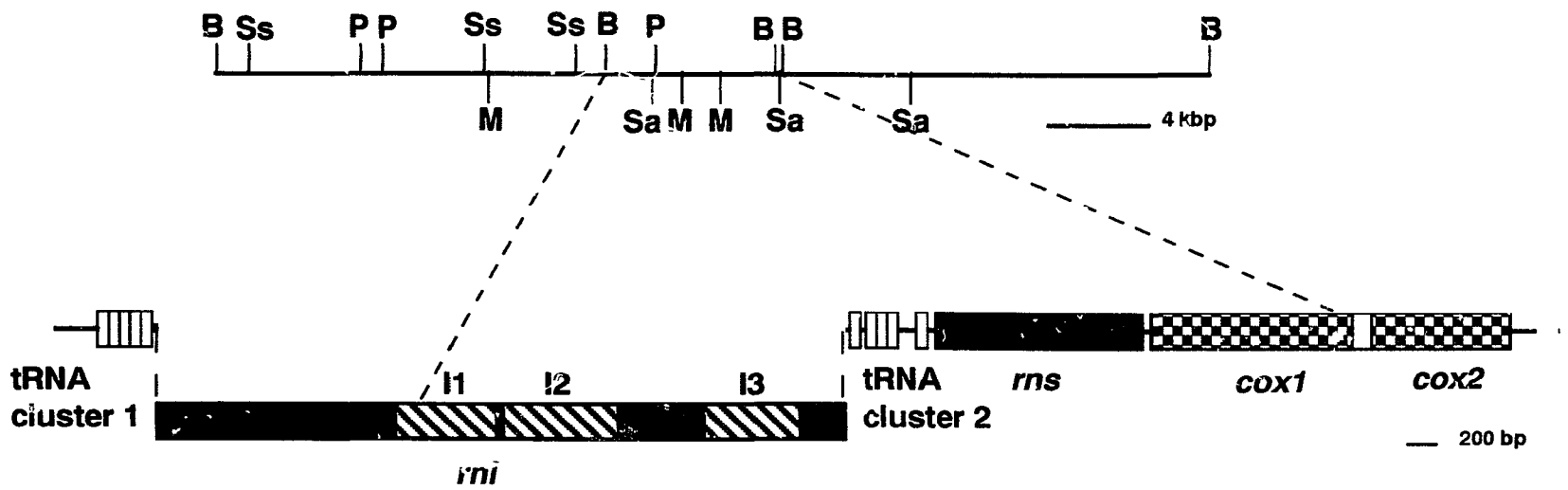


Figure 7. Arrangement of LSU and SSU rRNA genes (*rnl* and *rns*, respectively) and flanking regions in *A. castellanii* mtDNA. The upper part of the figure is a restriction endonuclease map of the mtDNA linearized at a *Bam*HI site. The bottom part of the diagram is a genetic map of the region of mtDNA that has been sequenced for this research project. All genes are in the same transcriptional orientation (5' to 3', left to right). The hatched boxes within the LSU rRNA coding region refer to the locations of intron sequences (I1 to I3).
Other abbreviations: *cox*1, cytochrome oxidase, subunit 1; *cox*2, cytochrome oxidase, subunit 2; B, *Bam*HI; Ss, *Sst*I; P, *Pst*I; M, *Mlu*I; Sa, *Sal*I.

A second cluster of five tRNA genes (cluster 1) is found upstream of *rnl*. The coding region of the cytochrome oxidase subunit 1 gene (*cox*1) begins 54 nt downstream of the 3'-terminus of *rns* while the *cox*2 gene is found immediately downstream of *cox*1. The longest stretch of apparently non-coding DNA within the ~11 kbp of mtDNA sequenced is 136 bp between the $tRNA^{Asp}$ and $tRNA^{Met2}$ sequences of tRNA gene cluster 2. The upper part of Fig. 7 lists some additional restriction endonuclease recognition sites not mapped by Von Gabain and Bohnert (as presented by Byers, 1986).

The complete DNA sequence for this region of the *A. castellanii* mtDNA, i.e., from the beginning of tRNA cluster 1 to the end of the *cox*2 coding region, is presented in Fig. 8. Fig. 9 shows the results of a typical M13 deletion series of overlapping clones from which ss template DNA was made and used for dideoxy sequencing.

Specific features of this region of the mitochondrial genome and its associated transcripts that are presented in this thesis include a phylogenetic analysis of the rRNA genes. The LSU rRNA gene contains three group I introns, which are described in detail below. Transcripts of the tRNA genes show evidence of a novel form of RNA editing, which is also described in detail.

B. Ribosomal RNA Genes of *Acanthamoeba castellanii* Mitochondria

Analysis of RNA extracted from isolated mitochondria (method 2) indicates that the rRNAs are present as high-molecular-weight species (Fig. 10). From a combination of DNA sequence analysis, 3'-terminal chemical sequencing of the RNA, and primer extension analysis to map 5'-termini, the LSU and SSU rRNA genes were found to be 5077 - 5087 nt and 1535 - 1541 nt in length, respectively. Variations in length reflect 5'-end heterogeneity (see below). The relatively large size of the LSU rRNA gene is

Figure 8. DNA sequence of the LSU and SSU rRNA genes and flanking regions of *A. castellanii* mtDNA (the portion of the genome identified in Figure 7). The RNA-like strand of the sequenced region is shown, with numbering from 5' to 3'. A dot is placed directly above the 5'-terminal nucleotide of the coding region of the gene or gene cluster identified to the right (tRNA-1, tRNA cluster 1; tRNA-2, tRNA cluster 2; I, intron; ORF, open reading frame). All potential protein-coding regions are indicated by the single-letter amino acid code and include ORF sequences of the LSU rDNA introns and coding regions of the *cox*1 and *cox*2 genes.

Sequence coordinates (nucleotides):

44 to 426, tRNA cluster 1 (see Figure 25A);

463 to 5549, *rnl* (intron sequences in lowercase letters);

5583 to 6118, tRNA cluster 2 (see Figure 25B);

6184 to 7724, *rns*;

7779 to 9335, *cox*1 (note that the 3'-end of *cox*1 is uncertain and is tentatively assigned on the basis of comparisons with other *cox*1 sequences);

9399 to 10,397, *cox*2.

```
        |   10      |   20      |   30      |   40      |   50      |   60
   1 TCTAGACGCC GTTTGTTCTT ACTTTTTAAA AAGGTTACAA AGCTGGGATA TAGCCAAAGG 60   tRNA-1
  61 GCAAGGCGGT GGGTTTTGAA CCCATGATTA TAGGTTCGAG TCCTATTATC CCAATTTGTA 120
 121 TAGCTCAATT GGTAGAGCAA TGGTCTCTTA AACCAGAGGT TTTGGGTTCG AGTCCCAATA 180
 181 CACCCATTTT AAACCCTTCG TCTAATGGTA GGACGTTAAC CTTTCACGTT AAGAATATGA 240
 241 GTTCGATTCT CATAGGGTTC AATAAAACAC CTATAACTCA GTTGGTTAGA GTGCACGCTT 300
 301 GATAAGCGTG AAGTCGATAG TTCGAGTCTA TCTAGGTTCA CATAAATGGA TATGCTGAAA 360
 361 TGGTAGACAG GCTGATTTTA GAAATCAGTG GGTAATTCCC TATGGGTTCG AGTCCCATTA 420
 421 TCCCCAAATC GTATTTAACT TAAGCGAGCA CAAAGTAAAA AATTGATTAT AAATACAAAT 480 rnl
 481 AAGTAATTAT ATAAAAATTA AAGGGTAAAT AACGAATATC TTATAGCATA AAAGTTTCTG 540
 541 GCGCCTAATC GAGCGAAATA TACAATGAGT CGAAAAATTA CTTAGAATTG TAAATTCCAG 600
 601 CAGAGAAATC TAACATTTAA ATATTGTTCT TACCTAATTA AGTGAAACAT CTCAATAGTT 660
 661 AGAGGAAAGG AAAATAATTA AAGATTTTGT TAGTAATGGT GAGTGAAAAC AAAATAGCCA 720
 721 AAATTTGAAT TTAAGGAAAT AGAAGTCTTA TATGATTTAA GCGCCATAGA AGGTGAAAGC 780
 781 CCTTTATAAT TTCTTATTTT TAATATATAT AATAATGAGT AAAATAAAAT CTGTGAAATT 840
 841 TTATTGGAAA AAGGCAAACC CATTTGCTAA GGCTAAATAC ACTTTTATGA CTAATAGTGA 900
 901 ACAAGTACTG TGAAGGAAAG ATGAAAAGAC AAAGTTAAAA GATCTTGAAA TTGTTTACCT 960
 961 TAAGGATCTA GAGCGCAGCG AATGTGTTAT AGGGTGCCTT TTGCATAATG AGCGGGCGAT 1020
1021 TTCTAAGGAA CTAGCAAACT ATGGTCGTTA GATGAAGGTG AAGTGAAAGC GAGCTTTAAT 1080
1081 TAAGCGTTAA GTTAGTTTTT TGAGACTCGA TACTAAGCGA TCTATTTACT GCCAGATTGA 1140
1141 AAGTACTTTA ACACGTACTG AAGGATCGAA CCCATGTATG TTGCATAATA CCGGGATGAG 1200
1201 TGGTAGCTAG GGGTGAAAGG CTAATCAAGC TTAGTGTTAG CTAGTACTCT ACGAAATCTA 1260
1261 TTTAGGTAGA GCGTTCTGTT CACATTATTG TTGGTAGAGC TCTAGGTAAG CAATGGGAAT 1320
1321 TAATACTTCT ACTGAGTTTA CTTAAACTGC GAATAGCAAT AAATTTTTTT ACAGAGCAGG 1380
1381 CAGATTGCGG GCGCGAAGGT TCGTAATCGA AAGGGAAACA GCCCAGAATA TCAGCTAAAG 1440
1441 TCCATAAAAA TAATTAAGTG TATGAAATAA TCTAAAGTGT TATAACAATT AGGAGGTTGG 1500
1501 CTTAGAAGCA GCCATCCTTT AAAGAAAGCG TAACAGCTCA CTAATTTTAT AAAAGCATTA 1560
1561 TAGGGTTTTA TGATATAACG GACTAAAATT ATTTACTGAA GCTATATGTA AATCGGTAGT 1620
1621 AGAGTATCTT ATATTTTATA GATAAATTTG TGAGAATTTA TTAATATATT AAGAGAAAGT 1680
1681 ATGCCTGTAT GAGTAGCAAA GAATAGTGAG AAACTATTCC ACCGTATATA TAAGGTTTCC 1740
1741 TATTTATAAA TTTATTTTAA TAGGGTTAAT CGGCCCCTAA GTTTTATAAT GACAATTATA 1800
1801 GGCGATGGGT TTTGGGTTAA AATTCTCAAA TTTTTCAGTA AGTGACAAAG GGGGCAGGTT 1860
1861 ATTTAATTTT AATAAAATTA ATTAAGTAAT CAGGACCTTT CCGGAAAAAG CTTTGAAAAT 1920
1921 AAAAAACCGT ACTTAAATCG GCTCCGATAT TTTGATAGAG TATATTAAGG TGATAGGATA 1980
1981 ATAATATTGA AGGAACTCGG CAAAATAGCT CCGTAACTTC GGGAGAAGGA GTACCTTTTG 2040
2041 TAAGCAATTA CTTATAGGTG TAACAGAATA GAGGGTAGCG ACTGTTTACT AAAAACACAG 2100
2101 GACTCCGCTA AGTTGTAAAA CGATGTATAG GGTCTGACAC CTGCCCGGTG CTGAAAGATA 2160
2161 AAAAGGATTG GTAGGAATGC CATGAATATA AATCTCAGTA AACGGCGGCT GTAACCATAA 2220
2221 CGGTCCTAAG GTgaatttgt tgccttctct tttggtaaca aaagagcaaa cattcaaccg 2280 I1
2281 tatgccggaa tatcctcaaa agacttaagg tgcaccaact ttttgttgta gcaaaaatcc 2340
2341 ttagtacacg gggacaatcg gcaggaaacc ttaaaattta ttttgaaagg gatcctcaga 2400
2401 gactatacgt tgaatatcct tatagtagta ataaaagatg aaaatcgata agaattgaat 2460 I1 ORF
                                            M   K   I   D   K   N   W   I
2461 cgtagggttt gttgatggag aaggttgctt ttatatcggt ataaataaat cagttgattc 2520
       V   G   F   V   D   G   E   G   C   F   Y   I   G   I   N   K   S   V   D   S
2521 aaaattaggt tatcaagtac ttcctgagtt tagggtcgtt caacataagc gtgacataaa 2580
       K   L   G   Y   Q   V   L   P   E   F   R   V   V   Q   H   K   R   D   I   K
2581 ggttttatat gctataaaag actttttgg tcacggctcg gttgtttgta acaagtctaa 2640
       V   L   Y   A   I   K   D   F   F   G   H   G   S   V   V   C   N   K   S   N
2641 tggctcagaa atttacgaat accgtgttcg taaattcgag actttacacg acgttatact 2700
       G   S   E   I   Y   E   Y   R   V   R   K   F   E   T   L   H   D   V   I   L
2701 tcctttttt gaatctaacg ggttattaac gagtaagaag tttaattttt tagcatttag 2760
       P   F   F   E   S   N   G   L   L   T   S   K   K   F   N   F   L   A   F   R
2761 agacgttata ttaataatga aaaggcgtga acatttaacc gagtctggct tgtctaaaat 2820
       D   V   I   L   I   M   K   R   R   E   H   L   T   E   S   G   L   S   K   I
2821 aattgatata aaaagtagaa tgaaccgatc ttcaattcat tattaaattt tattttaag 2880
       I   D   I   K   S   R   M   N   R   S   S   I   H   Y
2881 gataaagaaa tagtccgaac ttagcagaaa tgttaagagt gccaatcatt aagctgtttc 2940
2941 tatttataac ataaatatat agcgcgctag attgtgccaa caaattgAGC GAAATTCCTT 3000
3001 GTCGGGTaga attggaactt gcccgattaa cgcagtaatg tgttatagca aatcagctaa 3060 I2
3061 taacggtgaa accctaagtt ttaattttgt tttaaaatat ggcaataccg tgggaacgat 3120
```

Figure 8 continues.

```
         |  10      |  20      |  30      |  40      |  50      |  60
3121 tattctagaa ctacaagatg aataattgac ccgtaacgac tgaatcgaaa gatgagattt 3180
3181 gttttctgaa agacaaataa aacgctgacc cctattgttc aaacttacta tgaagtataa 3240
3241 aaaaatgaga aaaagaatta agaattttta aaatatgaag aatttggatt taaagcaaga 3300 I2 ORF
                                        M  K   N  L  D   L  K  Q  E
3301 aaaagcactg attgatcctt tttggataag cggttttgtt gatggagaag gttgcttttg 3360
      K  A  L   I  D  P   F  W  I  S   G  F  V   D  G  E   G  C  F  C
3361 tattagtttt aatttaaagg aacgtttaac tttaggcata gaagttaggc caagtttttc 3420
      I  S  F   N  L  K   E  R  L  T   L  G  I   E  V  R   P  S  F  S
3421 tatatctcaa actagggata aggaaggttt aaatttgagg tgtttgcagg atttttaaa 3480
      I  S  Q   T  R  D   K  E  G  L   N  L  R   C  L  Q   D  F  L  N
3481 ctttttgat tgtggattta tacgctttc aaaaagagat aatacctgaa agtacgagtg 3540
      F  F  D   C  G  F   I  R  F  S   K  R  D   N  T  W   K  Y  E  C
3541 tcgggattta tccgatatac gctctaaggt gttacctcat tttgaaaaat ttacactaag 3600
      R  D  L   S  D  I   R  S  K  V   L  P  H   F  E  K   F  T  L  R
3601 gactaaaaag cttcgagact ttgagttatt caaagatgtc gttaattctg tcgccagtaa 3660
      T  K  K   L  R  D   F  E  L  F   K  D  V   V  N  S   V  A  S  K
3661 gcaacattta aatgaagttg gtttaaagcg tataatagac atttcttacc aaataaatat 3720
      Q  H  L   N  E  V   G  L  K  R   I  I  D   I  S  Y   Q  I  N  M
3721 gggtaaaaga aaattaacca aggatgagct gttatccaaa atcaatttaa aaaatctgta 3780
      G  K  R   K  L  T   K  D  E  L   L  S  K   I  N  L   K  N  L
3781 attcggattc ctttttaatg tttctaaggt aagttaaggg tgatgatata gtctaaactt 3840
3841 ttttgaaaga aaaagattaa gtttttgGAG TTCCGACCTG CATGAATGGT GTAACGACTT 3900
3901 CCCTACTGTC TCCAGTATTA ACCTAGTGAA ATTAAACTTG CTGTGAAGAT GCAGTATATA 3960
3961 AATGATTGGA CGGAAAGACC CTGTACACCT TTACTATAAT CATAGATTGT TTTTAAAATA 4020
4021 TTGAATATGT AGGATAAACA GCAAGCTTAT TATACTTTAA TGCTAGTTAA TTGTTTAGGC 4080
4081 AATAGTGAAA TACTGTTTTT TTAATTTTTA AAACTAACTA AATTTTAGGA TAGTTTATGA 4140
4141 CTGGTAGTTT GACTGGGGTG GTCGCCTTCT AAATAGTAAC GAAGGCGTGC AAATGGTAAT 4200
4201 AATAATTAAT TAAAAATTAA TTTGAAAATT TAATGGTAAA AGATTGCTTG ACTGAGTAGA 4260
4261 TCGACAGATC GAGTCAGAGA CGAAAGTCGG CCATAGTGAT CCAATAGTGC TGAATGGAAG 4320
4321 GGCTATTGCT TAACGGATAA AAGGTACGTC AGGGATAACA GGCTAATGGG CCCTAAGAGT 4380
4381 CCTTATCGAC GGGCTTGTTT GGTACCTCGA TGTCGACTCA TCACATCCTG GAGCTGCAGG 4440
4441 CGGTTCCAAG GGTTAGGCTG TTCGCCTATT AAAGTGGTAC GTGAGTTGGG TTTAGAACGT 4500
4501 agataacact gcgttggttt gtggtaacat aaaccgtagc attaacttat atcggtgaag 4560 I3
4561 ccctactctt ttttgaactg gaagaggggt aataccgagg gaaaatataa attatttaga 4620
4621 aaagagcaaa atatgcaaaa aaatcaattt aaaaaattaa ataatgaaca attagcctat 4680 I3 ORF
                  M  Q  K   N  Q  F   K  K  L   N  N  E  Q   L  A  Y
4681 ttagctggat ttgttgaggc agatggctgt tttttagttc agatcatccc gggacttcag 4740
      L  A  G   F  V  E  A   D  G  C   F  L  V   Q  I  I  P   G  L  Q
4741 tataggtata aacatactat acgcataagt atagttttct atcaaaaaaa agataagcat 4800
      Y  R  Y   K  H  T  I   R  I  S   I  V  F   Y  Q  K  K   D  K  H
4801 tgatattttc ttcagcttaa aaatttaata ggtttaggct ctattcgttt tagaaatgat 4860
      W  Y  F   L  Q  L  K   N  L  I   G  L  G   S  I  R  F   R  N  D
4861 gggatgcttg aatattctat tactggttta agcttggtaa ataagttttt agagatgcta 4920
      G  M  L   E  Y  S  I   T  G  L   S  L  V   N  K  F  L   E  M  L
4921 tttccatatt taattttaaa aaagaatctt gctgttttaa tctttcgcat tattaagggg 4980
      F  P  Y   L  L  L  K   K  N  L   A  V  L   I  F  R  I   I  K  G
4981 ttaaacgatg tgaaaaacga ggccggcttt ttggagcgttt gtaaattagt agatgaagtc 5040
      L  N  D   V  K  N  E   A  G  F   L  E  V   C  K  L  V   D  E  V
5041 gccgatcata cgtattctaa gaaaagaaag aatacttcct taactgtaaa aaatagttta 5100
      A  D  H   T  Y  S  K   K  R  K   N  T  S   L  T  V  K   N  S  L
5101 ttattacctg tagagactga ggagtagttt ttgttataaa aactattccc gataattcca 5160
      L  L  P   V  E  T  E   E
5161 aacttcctcg gaattataat acgttaactc cataataaat ttttttatgg atgaagatat 5220
5221 agtccatgcc actagtaata gtggattttc atgCGTGAGA CAGTTCGGTT CCTATCTGTC 5280
5281 ATTTTGTAAA AACTAAAAAA ATGTTTCTTT TGTACGAGAG GATTGAGAAA TGTATACCAC 5340
5341 TGGTAAATCG ATTGTACTGT CAAAGGCATT GTCGAGTAGC TACGTATAAA TTTTATAATT 5400
5401 GCTGATAGCG TATAAGCAAG AAAATGATTT TTATGTTTTT TTAAACGGTT GTAGATAATA 5460
5461 ACTTTTTATA GGCGCGAAAT TAAATTTGGT AACGAATGGT TTACGCGTAC TAATTATTAA 5520
5521 ATTTTTTATA TATGATTTTT TTGCAATTTT TAAAAAAAAA TATCGAGGGT AAAAAATCCG 5580
5581 ATAATCCTAT AGTTTAATGG TAAAATATTT CGCTCATAAC GAAAGGTTAT TGGTTCGATT 5640 tRNA-2
```

Figure 8 continues.

```
          |  10       |  20       |  30       |  40       |  50       |  60
5641 CCAGTTAGGC CTATTTTATT AGGGAGATTG TAATTAGCCG CTTGGTTGCA TAGTTTAATG 5700
5701 GTAAAATCAA TACCTTGCAC GTATTAGATA TCAGTTCGAT TCTGATTGCG TCCACTTTGT 5760
5761 GCAGAGTGTA GCATAATAGG TAATGTGTCT GTTTTGGGAA CAGAAGACTA TAGGTTCGAG 5820
5821 TCCTGTCACT CTGAGCGTTT AGTGTAACTT AATTGGTAAA GAGTTAACTT GTCGAGTTAA 5880
5881 AAGATGCAGG TTCGAGTCCT GTCACTCTCA CTACTTAGAT TTTTAAATTA CCATAAATTA 5940
5941 CTTAAACTAG TATTTTTTTG AAAACTGCCT TCTTATATCC TATATATTTC AGCTGTCTTT 6000
6001 TTTTTTTTTA TATAAAGGTT ACTTAAATTT GTTTTTTTAT TGGCTTTTGG AGGTATTGTA 6060
6061 AATGGAAACA TATCAGGCTC ATGCCCTGAC GATGTAGGTT CGAGTCCTAC CCTCCTGATT 6120
6121 TTTGGTAAAG TTATATTTAG TAAATTTCAA AGATATTATA TTATATATAA GTAATTCCAT 6180
6181 TACTTGTATA AACAATCGTT GGGTTTTATT CTGGCTCCGA GTGAACGCTG GTAACATATT 6240 rns
6241 TTAACACATG CAAATCGAAC GAAAAGTAAA TTTACTTTAA GTGGTGACCT GGTGAGTAGT 6300
6301 ATAGGAGAAT TGAAGTGTGG GTTAGGCTAA ATTCCTAAAA TTATTAAAAA TAAAAGATTT 6360
6361 ATATCGCCTA CAATAGAGCT TCTAAAGGAT TAGGTAGTTG TTGTAGTAAT TATTCAACAA 6420
6421 GCCGACGATG CTTAGCTGGT TTGAGAGAAT GATCAGCCAC ATTGGGACTG AAAACGGCCC 6480
6481 AAGTCTTCTA AAAGGGCAGC AGTGGAGAAT ATTTGACAAT GGGCGAAAGC TTGATCGAGT 6540
6541 AATGTTACGT GGGTGATCAG AAGGCTTAAT TGCCGTAAAA CTCGTTTCGT CAAGGATTAT 6600
6601 AATGCAATCG TACTTGAAGG AGAAGCCCTG ACTAACTTCG TGCCAGCAGT CGCGGTAATA 6660
6661 CGGAGAGGGC GAGTGTTATT CGCATTGACT GGGTGTAAAG GGTGCGTAGG TGGTTTGTTA 6720
6721 AGTTTCAAAG TAAATGTTAA ACATAAAGTT TAAAGGCTTT TTAATACTAA CAGACTAGAG 6780
6781 TTTATTAGAA GATAGCTGAA CATTAAGTGT AAAGGTGAAA TTTTTCGATA TTTAAAGGAC 6840
6841 ATGCAATAGG CGAAGGCTGC TATCTATTTT AATACTGACA CTGAGGCACA AAAGCATAGG 6900
6901 GATCAAACAG GATTAGAGAC CCTGGTAGTC TATGCCCTGA CCGATGAATA TTGATTATAC 6960
6961 TGCGAGGTGT AATTAAGCTA ACGCGTAAAT ATTCCACCTG AGTACTACGA CCGCAAGGTT 7020
7021 AAAACTCAAA GAAACTGACG GAAGATTGTA TAAGCGGTGG AGTATGTGGT TTAATTTGAT 7080
7081 ATAACGCAAA GAATCTTACC AACCCTTGAA TACAAGTTAT TTTAGCTGTG ATAATGGAAT 7140
7141 TTATTTTCAT TCTCATAAGC TCATTATAAG GATCACCTGA GGAGGTGTAT CATATTGACT 7200
7201 TGTACAGGTG CTGCATGGTT GTCGGCAGTT CGTGTTGTGA AATGTTTTGT TAATTCATTT 7260
7261 AACGGACGTA ACCTTCAGCT TTATTTGTTT TTGAAAAAAT AGAGCAATAT TAATTTGAAA 7320
7321 ATTAATTGAG AGGAGAAGAA GTCAAATCAT TATGGCCTAT ATAGGTTGGG CTACACACGT 7380
7381 ACTACAATGG TAATTACAAT GTGAAGCGAA AATGCGAGTT TTAGCTAATC ATTAAAAATT 7440
7441 AATCTTAGTT CTGATTACTC TCTGTAAGTC GAGAGTATGA AGTAGGAATC GCTAGTAATC 7500
7501 GCAGAATAGA ATGTTGCGGT GAATATATTC TCAATCTTTG TACACACCGC CCGTCACGCC 7560
7561 ATGGGAATTG GTTTGATTAG AAGTGGTATA GGCTCAGATT CTGATGTCTT TTAATTTTTT 7620
7621 TTAAAGTTAT CGCATCTTGG TACTAATTAC GACAATTAGA TTAAGGACTG GGGTGAAGTC 7680
7681 GTAACAAGGT AGCGGTAGGG GAACCTGCTG CTGGACAAAA TTTTTTTAAA GGATACGAAT 7740
7741 ACTTAGACTT AAATAATATA AAATTACAAA TTTTAACGAT GATAAATACA TTATTGAATA 7800 cox1
                                          M   I   N   R   L   L   N
7801 ATTTGACAAG CTTTTTTACT GATAATCGTT GATTATTTTC TACTAATCAC AAAGATATCG 7860
     N   L   T   S   F   F   T   D   N   R   W   L   F   S   T   N   H   K   D   I
7861 GTACGTTATA TTTAATATTT GGTGGCTTTT CTGGTATTAT AGGTACTATA TTTTCTATGA 7920
     G   T   L   Y   L   I   F   G   G   F   S   G   I   I   G   T   I   F   S   M
7921 TTATAAGGTT AGAATTAGCT GCTCCGGGTT CTCAAATATT GAGTGGTAAT AGCCAACTTT 7980
     I   I   R   L   E   L   A   A   P   G   S   Q   I   L   S   G   N   S   Q   L
7981 ACAACGTTAT TATTACGGCG CATGCCTTTG TTATGATTTT CTTTTTTGTT ATGCCGGTTA 8040
     Y   N   V   I   I   T   A   H   A   F   V   M   I   F   F   F   V   M   P   V
8041 TGATAGGCGG TTTCGGAAAT TGATTCGTTC CTTTAATGAT TGGTGCTCCT GATATGGCTT 8100
     M   I   G   G   F   G   N   W   F   V   P   L   M   I   G   A   P   D   M   A
8101 TTCCAAGGCT TAACAATATA AGTTTTTGAT TACTACCGCC TTCCTTATTT CTATTATTAT 8160
     F   P   R   L   N   N   I   S   F   W   L   L   P   P   S   L   F   L   L   L
8161 GTTCTTCATT AGTAGAATTT GGGGCAGGAA CCGGCTGAAC AGTTTACCCG CCGTTAAGTT 8220
     C   S   S   L   V   E   F   G   A   G   T   G   W   T   V   Y   P   P   L   S
8221 CTATTGTAGC TCACTCCGGC GGTTCCGTCG ATTTAGCAAT ATTTAGTTTA CATCTTGCAG 8280
     S   I   V   A   H   S   G   G   S   V   D   L   A   I   F   S   L   H   L   A
8281 GTATATCTTC TCTTCTAGGC GCTATAAACT TTATTACTAC GATATTTAAT ATGAGAGTAC 8340
     G   I   S   S   L   L   G   A   I   N   F   I   T   T   I   F   N   M   R   V
8341 CTGGTTTATC TATGCATAAA TTACCATTAT TTGTTTGGTC TGTGTTGATT ACAGCCTTTT 8400
     P   G   L   S   M   H   K   L   P   L   F   V   W   S   V   L   I   T   A   F
8401 TGTTATTATT TTCTCTACCT GTATTAGCAG GTGCTATTAC TATGCTTTTA ACTGATAGAA 8460
     L   L   L   F   S   L   P   V   L   A   G   A   I   T   M   L   L   T   D   R
```

Figure 8 continues.

```
          |  10      |  20      |  30      |  40      |  50      |  60
 8461 ATTTTAATAC TAGCTTTTTT GATCCATCAG GCGGAGGCGA TCCTATCCTT TATCAACATT 8520
      N  F  N  T   S  F  F   D  P  S   G  G  G  D   P  I  L   Y  Q  H
 8521 TATTCTGATT TTTTGGTCAT CCTGAAGTTT ATATTTTAAT TTTACCAGCT TTTGGTATAG 8580
      L  F  W  F   F  G  H   P  E  V   Y  I  L  I   L  P  A   F  G  I
 8581 TCAGCCAAAT TATCGGTACT TTTTCCAATA AATCTATTTT TGGTTATATA GGAATGGTTT 8640
      V  S  Q  I   I  G  T   F  S  N   K  S  I  F   G  Y  I   G  M  V
 8641 ACGCAATGTT GCCTATTGCA GTTTTAGGTT TTATTGTATG GGCTCATCAC ATGTATACAG 8700
      Y  A  M  L   P  I  A   V  L  G   F  I  V  W   A  H  H   M  Y  T
 8701 TAGGTTTAGA TGTTGATACT AGAGCTTACT TCACTGCGGC AACTATGATG ATTGCTGTTC 8760
      V  G  L  D   V  D  T   R  A  Y   F  T  A  A   T  M  M   I  A  V
 8761 CAACTGGTAT AAAGATCTTT AGTTGAATTG CTACGTTGTG AGGCGGCCAA ATAGTAAGAA 8820
      P  T  G  I   K  I  F   S  W  I   A  T  L  W   G  G  Q   I  V  R
 8821 AGACTCCATT GCTTTTTGTT ATAGGTTTTT TAATCCTATT TACCCTAGGT GGTTTAACTG 8880
      K  T  P  L   L  F  V   I  G  F   L  I  L  F   T  L  G   G  L  T
 8881 GTATTGTTTT ATCAAACGCT GGTCTTGATA TAATGTTACA TGATACTTAT TACGTCGTAG 8940
      G  I  V  L   S  N  A   G  L  D   I  M  L  H   D  T  Y   Y  V  V
 8941 CGCATTTTCA TTATGTGTTA TCCATGGGTG CTGTTTTTGC TTTTTTTGCT GGTTTTTATT 9000
      A  H  F  H   Y  V  L   S  M  G   A  V  F  A   F  F  A   G  F  Y
 9001 ACTGGTTTTG AAAAATTTCT GGCTATACTT ATAATGAAAT GTATGGTAAT GTTCATTTTT 9060
      Y  W  F  W   K  I  S   G  Y  T   Y  N  E  M   Y  G  N   V  H  F
 9061 GGTTAATGTT TATTGGCGTT AATTTAACAT TTTTCCCGAT GCATTTCGTT GGTTTAGCGG 9120
      W  L  M  F   I  G  V   N  L  T   F  F  P  M   H  F  V   G  L  A
 9121 GTATGCCACG AAGGATCCCC GATTATCCTG ATAATTATTA TTATTGAAAT ATATTATCCT 9180
      G  M  P  R   R  I  P   D  Y  P   D  N  Y  Y   Y  W  N   I  L  S
 9181 CCTTTGGTTC TATAATATCT TCAGTAAGTG TTATAGTATT TTTTTATTTA ATATATTTAG 9240
      S  F  G  S   I  I  S   S  V  S   V  I  V  F   F  Y  L   I  Y  L
 9241 CTTTTAATAA TAATAATACC CCTAAGTTGA TAAAATTAGT TCACTCTATT TTTGCTCCTT 9300
      A  F  N  N   N  N  T   P  K  L   I  K  L  V   H  S  I   F  A  P
 9301 ATATCAACAC GCTATCTAAG AATTTATTAA CTTTTGCTTC TATAAAGTCG ACATCTGACT 9360
      Y  I  N  T   L  S  K   N  L  L   T  F
 9361 CTTCTTTTTT TAAGTTTAGC AAGTTTTTTA TTTTTTTTAT GGTGTCTTTA TCCGTTCTTT 9420 cox2
                                                M   V  S  L   S  V  L
 9421 TTATTTTTTA TGATTCTCTG TTATGCTTAA ACGATCATAC TAATTCTTGA AAAATAGGTT 9480
      F  I  F  Y   D  S  L   L  C  L   N  D  H  T   N  S  W   K  I  G
 9481 TTCAGGATCC TACAACGCCA ATTGCTTATG GTATTATAAA ACTTCATGAT CATATACTTT 9540
      F  Q  D  P   T  T  P   I  A  Y   G  I  I  K   L  H  D   H  I  L
 9541 TTTTTTTAGC CGTTATTTTA TTTGTTGTAG GTTATCTTTT ACTTTCTACC TATAAAAAAT 9600
      F  F  L  A   V  I  L   F  V  V   G  Y  L  L   L  S  T   Y  K  K
 9601 TCTATTATGG TTCGTTAAAT ATTGATCTTC CCGAATCAAA ACGTATTTCT TTATTCGATA 9660
      F  Y  Y  G   S  L  N   N  D  L   P  E  S  K   R  I  S   L  F  D
 9661 CATTAATAAA CACTTATAAG GAAAATTTAT CATTTAATGT TACTAATAGA ACTTATAATA 9720
      T  L  I  N   T  Y  K   E  N  L   S  F  N  V   T  N  R   T  Y  N
 9721 TAAATCATGG TACAACAATA GAGATAATTT GGACTATTTT ACCAGCATTT ATACTTTTAT 9780
      I  N  H  G   T  T  I   E  I  I   W  T  I  L   P  A  F   I  L  L
 9781 TTATTGCTGT TCCATCATTT GCTCTACTTT ATGCTATGGA TGAGATTATT GATCCGGTTT 9840
      F  I  A  V   P  S  F   A  L  L   Y  A  M  D   E  I  I   D  P  V
 9841 TGACTGTTAA GGTAATAGGC CATCAATGGT ATTGAAGTTA CGAGTACTCG GATTATTCTG 9900
      L  T  V  K   V  I  G   H  Q  W   Y  W  S  Y   E  Y  S   D  Y  S
 9901 TAGTTTACTC TAACAGAATG TTAGATTATG ATTCTATCGA TAGGTTTGCT GCTATGGAAA 9960
      V  V  Y  S   N  R  M   L  D  Y   D  S  I  D   R  F  A   A  M  E
 9961 TGATGTACAA GGGCATGGGG TATTTAAAAG ATCGTTCTCT ATTAAGTTAT TTATACATAC 10020
      M  M  Y  K   G  M  G   Y  L  K   D  R  S  L   L  S  Y   L  Y  I
10021 CTATGGTCAT TCCTGAAACA ACTATCAAAT TTGATAGTTA TATGATTCAT GAGGCTGAAT 10080
      P  M  V  I   P  E  T   T  I  K   F  D  S  Y   M  I  H   E  A  E
10081 TAAATCTAGG TGATCTAAGG TTATTAAAAA CTGACATGCC CCTTTTTTTA CCAAAAAATA 10140
      L  N  L  G   D  L  R   L  L  K   T  D  M  P   L  F  L   P  K  N
10141 CCCATATTAG ACTTTTAATT ACCTCTTCTG ATGTTTTACA TAGTTGAGCT GTTCCGTCAT 10200
      T  H  I  R   L  L  I   T  S  S   D  V  L  H   S  W  A   V  P  S
10201 TTGGTGTCAA AGTTGATGCC GTTCCTGGAA GATTAAACCA GACTTCCCTG TATTTAAAAA 10260
      F  G  V  K   V  D  A   V  P  G   R  L  N  Q   T  S  L   Y  L  K
```

Figure 8 continues.

```
          |   10      |   20      |   30      |   40      |   50      |   60
10261 ACACAGGAAC ATTCTATGGC CAATGTAGTG  AGCTTTGTGG CGTTAATCAT GCTTTTATGC 10320
      N  T  G  T   F  Y  G   Q  C  S    E  L  C  G   V  N  H   A  F  M
10321 CTATAGAAGT TTATGTGGTT AATCCTGTTT  ATTTCTATAA TTACGTCTAC ATTTATTTTA 10380
      P  I  E  V   Y  V  V   N  P  V    Y  F  Y  N   Y  V  Y   I  Y  F
10381 AGAATTTTAA TTTAATATAG AAAGAATAAG  AGGCGGAGAA ACATAAGTTT AACTCTGTTT 10440
      K  N  F  N   L  I
10441 TGTTCTTTTG CATATACAAA AGTAAACGGG  TTTCAGGTTT GAAACCTTAG TTTTAGTAAT 10500
10501 GAGTTAACAT TTTTAGAACT ATAGTTAATT  AAATTACCTG AAACCGTTCC G          10551
```

Figure 8 continued.

Figure 9. Gel profile of M13 deletion series subclones used for dideoxy sequencing. This figure shows an M13 sizing gel of exoIII-deleted clones generated from a subclone of pKL-2. The migration positions of an M13 clone without insert DNA [B, (blue plaque)] and an M13 clone with a full-sized insert DNA (F) are indicated. Numbers at the top refer to individual subclones from which ss template DNA could be prepared for sequencing. Sizes of adjacent subclones differ by ~300 nt.

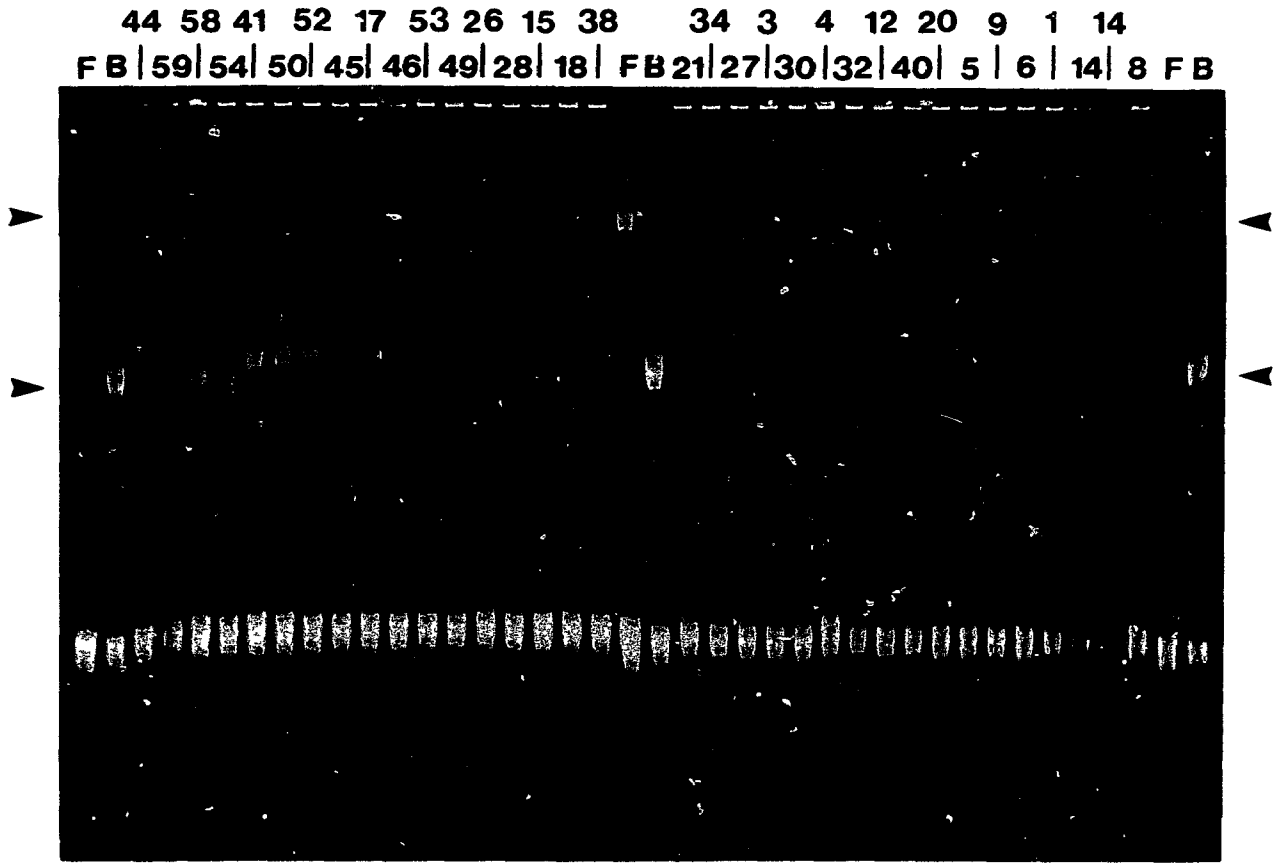

44 58 41 52 17 53 26 15 38 34 3 4 12 20 9 1 14
F B 59 54 50 45 46 49 28 18 F B 21 27 30 32 40 5 6 14 8 F B

Figure 9

Figure 10. Gel profile of high-molecular-weight RNA from *A. castellanii* mitochondria. Mitochondrial nucleic acids were prepared by method 2 and aliquots were resolved by agarose gel electrophoresis and then stained with ethidium bromide. N, nucleic acid from the crude nuclear pellet; C, nucleic acid from the cytoplasmic (high-speed supernatant) fraction; M, nucleic acid from the mitochondrial fraction. The migration positions of selected *Crithidia fasciculata* cytoplasmic rRNA size markers (Collings, 1991) are indicated to the left.

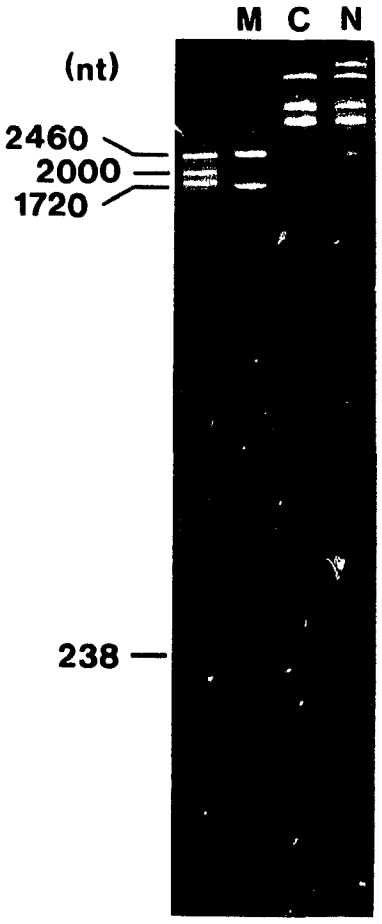


Figure 10

explained by the presence of three group I introns within the 3'-half of this gene (Fig. 8).

(i). Potential secondary structure of the SSU rRNA

Potential secondary structures for the rRNAs have been derived from the gene sequences. The potential secondary structure of *A. castellanii* mitochondrial SSU rRNA is shown in Fig. 11. The structure shown here is modeled after that of *E. coli* SSU (16S) rRNA (Woese, 1984; Gutell et al., 1993), from which it differs only within discrete variable regions (V1 - V9) of the molecule. The length of the longest *A. castellanii* SSU rRNA transcript (1541 nt) is one nucleotide shorter than that of *E. coli.* The *A. castellanii* mitochondrial SSU rRNA secondary structure can be unambiguously aligned over 75% of its length with those of eubacteria and wheat mitochondria.

Major departures from the *E. coli* 16S rRNA secondary structure include:

(1). a shortened version of the structure within V2 (*E. coli* coordinates 143 - 220). As in eubacteria, three helices are found within V2 of *Acanthamoeba* SSU rRNA, whereas the mitochondrial homolog in other eukaryotes has only two V2 helices. As in other mitochondria, two of these helices in *Acanthamoeba* are truncated compared to *E. coli* 16S rRNA and one is truncated relative to the *Agrobacterium tumefaciens* homolog.

(2). 46 extra nucleotides within V3 (*E. coli* coordinates 998 - 1043);

(3). 37 extra nucleotides within V5 (*E. coli* coordinates 1436 - 1458);

(4). 23 fewer nucleotides in V6 (*E. coli* coordinates 447 - 487) (V6 in *A. castellanii* has 2 extra nucleotides and a similar secondary structure compared to V6 in *A. tumefaciens*, and an identical secondary structure compared to wheat mitochondrial V6).

The helix of V7 (*E. coli* coordinates 589 - 650) has been drastically truncated in fungal, animal, ciliate protozoan and *C. reinhardtii* mitochondria but not in

Figure 11. Diagram of the potential secondary structure of mitochondrial SSU rRNA from *A. castellanii*, as deduced from the DNA sequence. The RNA-like strand of the DNA is shown and the secondary structure is modeled after that of *E. coli* SSU (16S) rRNA. Numbering is from 5' to 3'; dashes occur at every 50th nucleotide position, dots at every 10th position. 5'-End heterogeneity is indicated by the straight arrows. Boundaries of variable regions (V1 to V9) are delineated by solid lines. The structure presented for V3 is hypothetical as it is not based upon a comparative analysis.

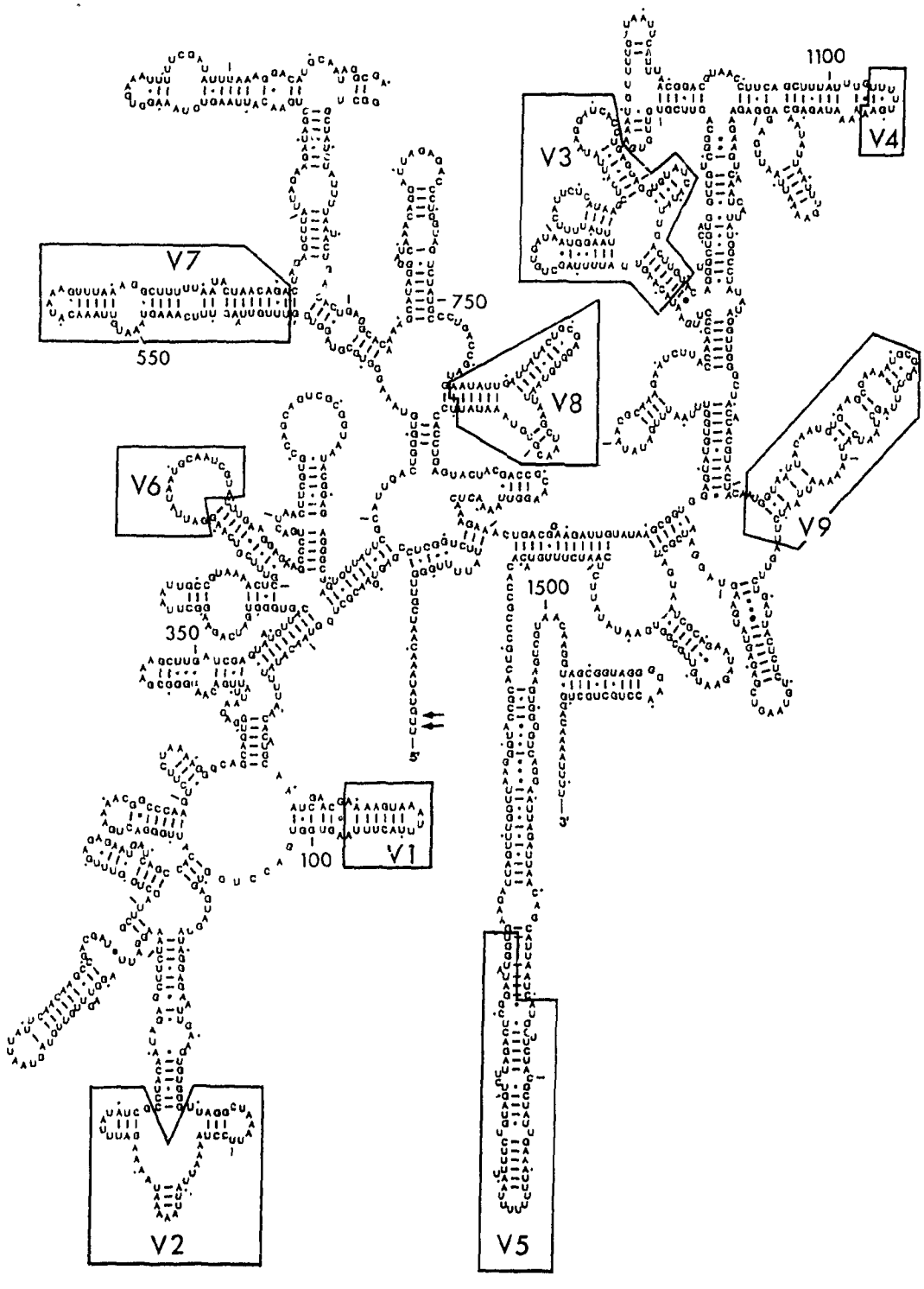


Figure 11

Acanthamoeba mitochondria, where a 1-nt deletion only has occurred relative to *E. coli*, or in higher plants, which show a 2-nt deletion.

In the diagram of Fig. 11, V3 is shown structured: however, it should be noted that this is a hypothetical structure that is not yet supported by comparative sequence analysis or other evidence.

These comparisons of potential secondary structure between *A. castellanii* and *E. coli* SSU rRNAs are summarized in Table 1.

In the case of the two regions of extra nucleotides within *A. castellanii* mitochondrial SSU rRNA relative to *E. coli* 16S rRNA (V3 and V5), it is not suspected that either constitutes an internally transcribed spacer because the ensuing processing events would generate RNA species whose sizes would not correspond with that observed (Fig. 10).

From enzymatic sequencing of 5'-end-labeled, gel-purified SSU rRNA, 5' end heterogeneity was suspected and this was confirmed by end-terminal analysis using tlc (Fig. 12). There are seven consecutive 5'-termini representing three major (5'-U, #1, 5'-U, #2; 5'-G, #3) and four minor transcripts, as identified by primer extension analysis (Fig. 13). The positions of these are indicated in the secondary structure of Fig. 11. Minor amounts of an extended transcript beginning 38 nt upstream from the first major U terminus were also observed (not indicated in Fig. 11 as part of the secondary structure, but noted in Fig. 13), the 5'-extension portion of which is very AU-rich.

The 3'-terminus of the SSU rRNA was determined by direct RNA sequencing of the gel-purified RNA, and was found to be homogeneous (Fig. 14). Altered reactivities in enzymatic and/or chemical sequencing reactions suggested the presence of modified nucleosides within the 3'-terminal region of the *A. castellanii* mitochondrial SSU rRNA. Based on the known occurrence of two adjacent N^6, N^6-dimethyladenosine residues (Nichols and Lane, 1965) at positions 24 and 25 of *E. coli* SSU rRNA

Table 1. Lengths (nt) of variable regions in SSU and LSU rRNAs.

Variable Region	*E. coli*	*A. castellanii* mitochondria
SSU rRNA		
V1	32	18
V2	78	45
V6	41	18
V7	62	61
V8	56	46
V3	46	92
V4	18	8
V9	60	60
V5	22	59
LSU rRNA		
a	56	27
b	193	94
c	15	0

Figure 12. Confirmation of 5'-terminal heterogeneity of rRNAs from *A. castellanii* mitochondria. Autoradiograms display terminal nucleotide analysis of 5'-^{32}P-end-labeled SSU and LSU rRNAs by thin layer chromatography (tlc). Horizontal and vertical arrows identify the first and second dimensions of chromatography, respectively. The origin (of sample application) is denoted by a single dot on each autoradiogram. Nucleotide standards not corresponding to a radioactive nucleotide are outlined; the identification of radioactive nucleotides is also indicated according to the co-migration of unlabelled nucleotide standards.

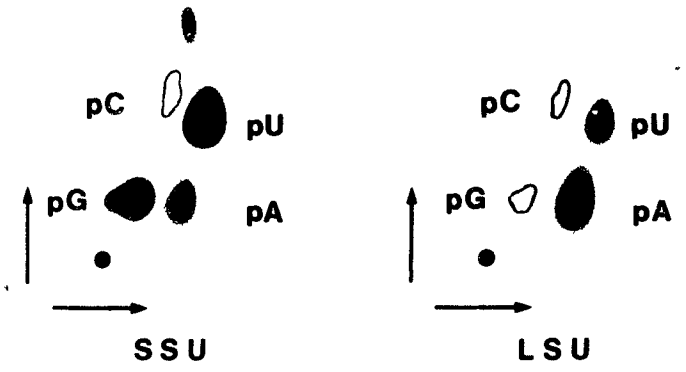


Figure 12

Figure 13. Determination of the 5' termini of rRNAs from *A. castellanii* mitochondria. Oligonucleotides specific for the SSU and the LSU rRNAs were used as primers for extension by reverse transcriptase with mitochondrial RNA as template (lane E), and as primers for dideoxy sequencing of mitochondrial rDNA (lanes G, A, T, C). The sequence indicated to the right of each autoradiogram represents that of an RNA precursor, with the arrowheads indicating the 5' termini of the most abundant RNA species. The arrowheads to the left of each autoradiogram identify a band corresponding to the 5'-terminus of a longer (presumably precursor) transcript.

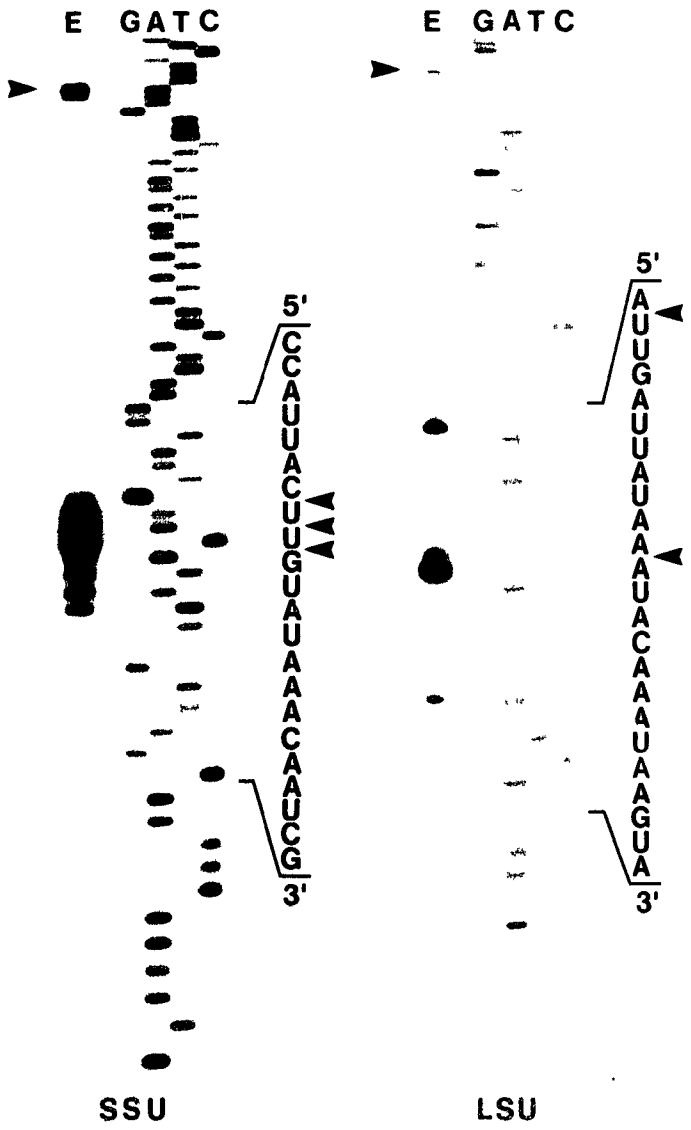

E GATC
E GATC
5'
CCAUUACUUGUAUAAACAAUCG
3'
5'
AUUGAUUAUAAAUACAAAUAAGUA
3'
SSU
LSU

Figure 13

Figure 14. Identification of 3'-terminal nucleotides of SSU and LSU rRNAs from *A. castellanii* mitochondria. (A). Autoradiogram of RNA sequencing gel of 3'-end-labeled SSU rRNA. The nucleotide sequence of the RNA is presented as a secondary structure diagram to the right, and the positions of modified nucleosides, identified by inference only, are indicated by arrowheads. (B). Autoradiogram of RNA sequencing gel of 3'-end-labeled LSU rRNA. The nucleotide sequence of the RNA is indicated along the side of the autoradiogram.

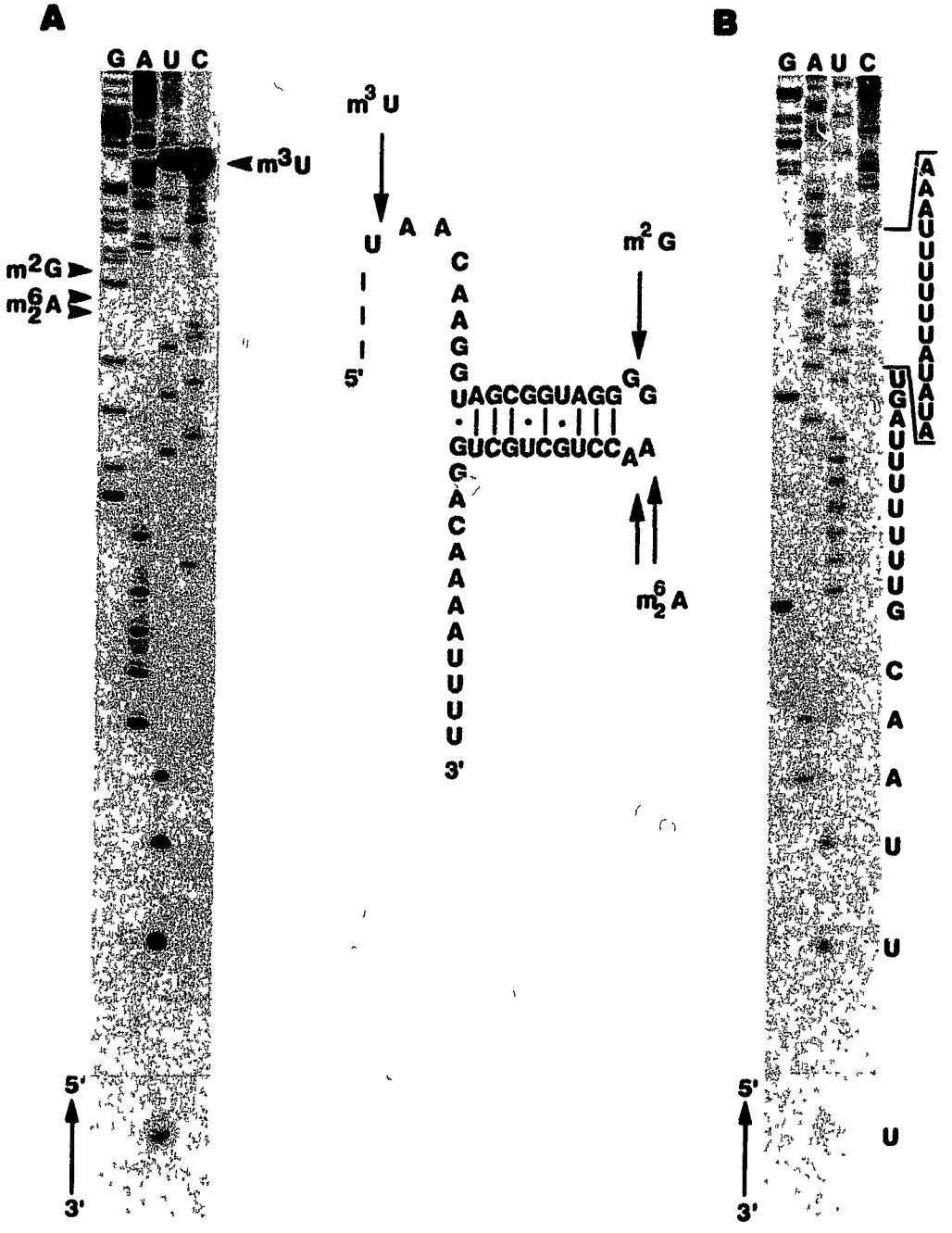

A
G A U C
m3U
m2G
m62A
5'
3'
m3 U
U A A
C
A
A
G
G
5'
UAGCGGUAGG G G
GUCGUCGUCC A A
G
A
C
A
A
A
A
U
U
U
U
3'
m2 G
m62 A
B
G A U C
A
A
A
U
U
U
U
U
U
A
U
A
U
A
U
U
G
U
A
U
U
U
U
U
U
G
C
A
A
U
U
U
5'
3'

Figure 14

(Brosius et al., 1978; Schnare and Gray, 1982), it is inferred that the chemical cleavage-resistant A residues at the analogous postions in *A. castellanii* mitochondrial SSU rRNA (22 and 23 nt from the 3'-end) are likely to be identically methylated. Similarly, the G residue located 25 nt from the 3'-terminus of the SSU rRNA is resistant to RNase T1 (Fig. 14). This reactivity is diagnostic of an N^2-methylguanosine residue at the same position in *E. coli* SSU rRNA (van Charldorp et al., 1981). The G-methylation at this position has thus far been considered to be restricted to eubacterial SSU rRNAs. An anomalous C-specific cleavage occurs at a 3-methyluridine residue located 45 and 42 nt from the 3'-end of *E. coli* and wheat mitochondrial SSU rRNA, respectively (Schnare, 1984). This unusual cleavage is also observed at the U residue at position 43 from the 3'-end of the *A. castellanii* mitochondrial SSU rRNA (Fig. 14). Thus, in terms of post-transcriptional modifications, *A. castellanii* mitochondrial SSU rRNA resembles eubacterial SSU rRNA to a greater extent than observed for the homologous SSU rRNA in other eukaryotes. The locations of these modified nucleosides within the secondary structure of the 3'-terminal region of the *A. castellanii* mitochondrial SSU rRNA are shown in Fig. 14.

(ii). Potential secondary structure of the LSU rRNA

The potential secondary structure of the mitochondrial LSU rRNA from *A. castellanii* is shown in Fig. 15A (5'-half) and Fig. 15B (3'-half). As for the SSU rRNA, the secondary structure shown here is modeled after the homologous 23S rRNA from *E. coli* (Gutell et al., 1993). There are three structural domains (I, II, and III) within the 5'-half of the molecule and three within the 3'-half of the molecule (IV, V and VI), as defined by long-range helical interactions. Generally the potential secondary structure of the *A. castellanii* mitochondrial LSU rRNA is very similar to that from *E. coli.* The few places where the secondary structure has obvious differences include the 5'-terminal region (region "a"; *E. coli* coordinates 121 - 176), where three helices are

Figure 15. Diagram of the potential secondary structure of mitochondrial LSU rRNA from *A. castellanii*, as deduced from the DNA sequence. (A) 5' - half; (B) 3' - half. The RNA-like strand of the DNA is shown and the secondary structure is modeled after that of *E. coli* LSU rRNA. Numbering is from 5' to 3'; dashes are placed at every 50th nucleotide position, dots at every 10th position. 5'-End heterogeneity is indicated by the arrow; structural domains are denoted by roman numerals. The 5'- and 3'-terminal regions of the molecule are thought to be involved in a potential base-pairing interaction, as indicated. The nucleotide positions of intron insertion sites for AcLSU.m1 (I1), AcLSU.m2 (I2) and AcLSU.m3 (I3) are indicated. Variable regions of secondary structure that are described in the text are denoted by 'a, 'b', and 'c'.

A

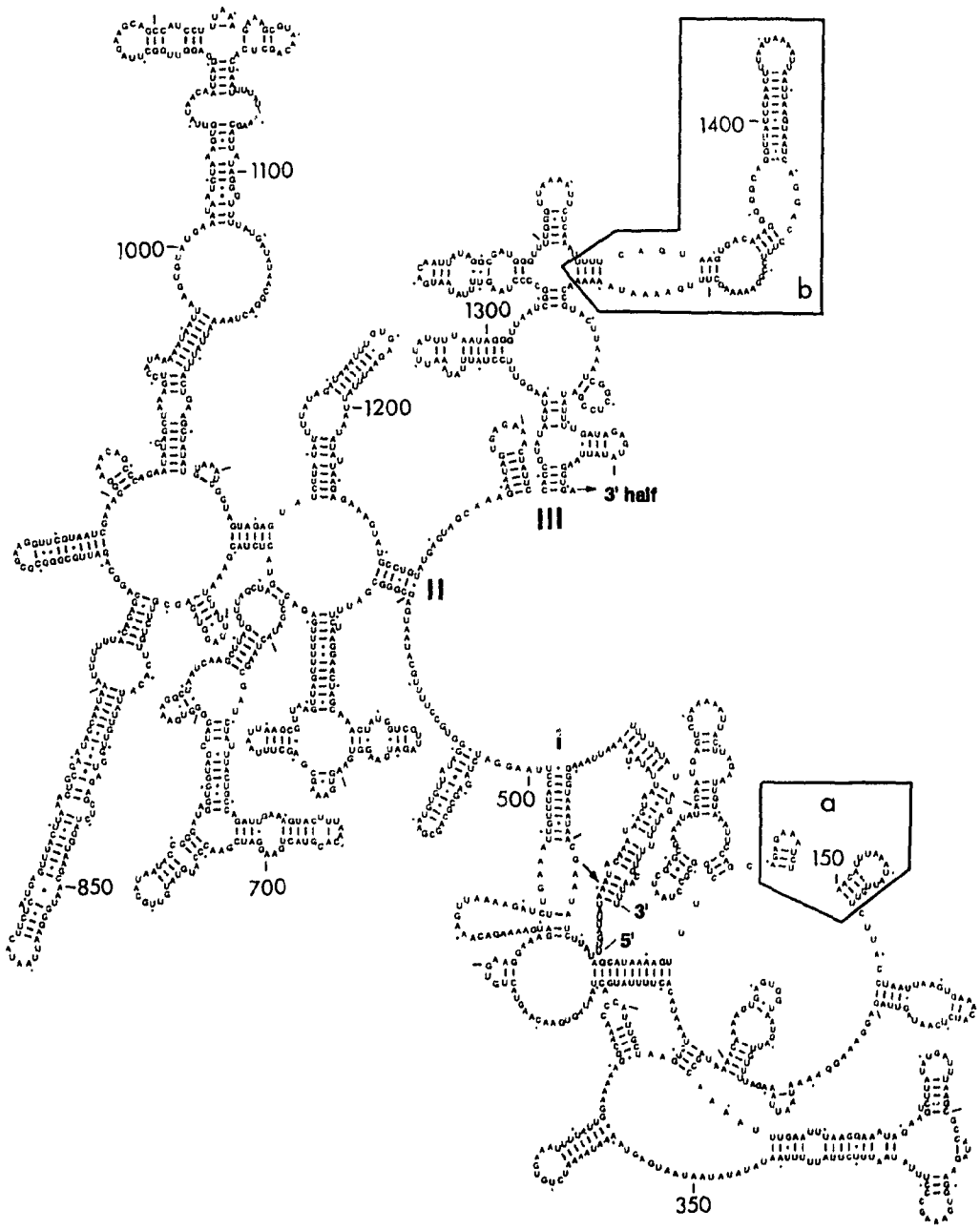

1400
b
1100
1000
1300
1200
3' half
III
II
I
500
a
150
850
700
3'
5'
350

Figure 15

B

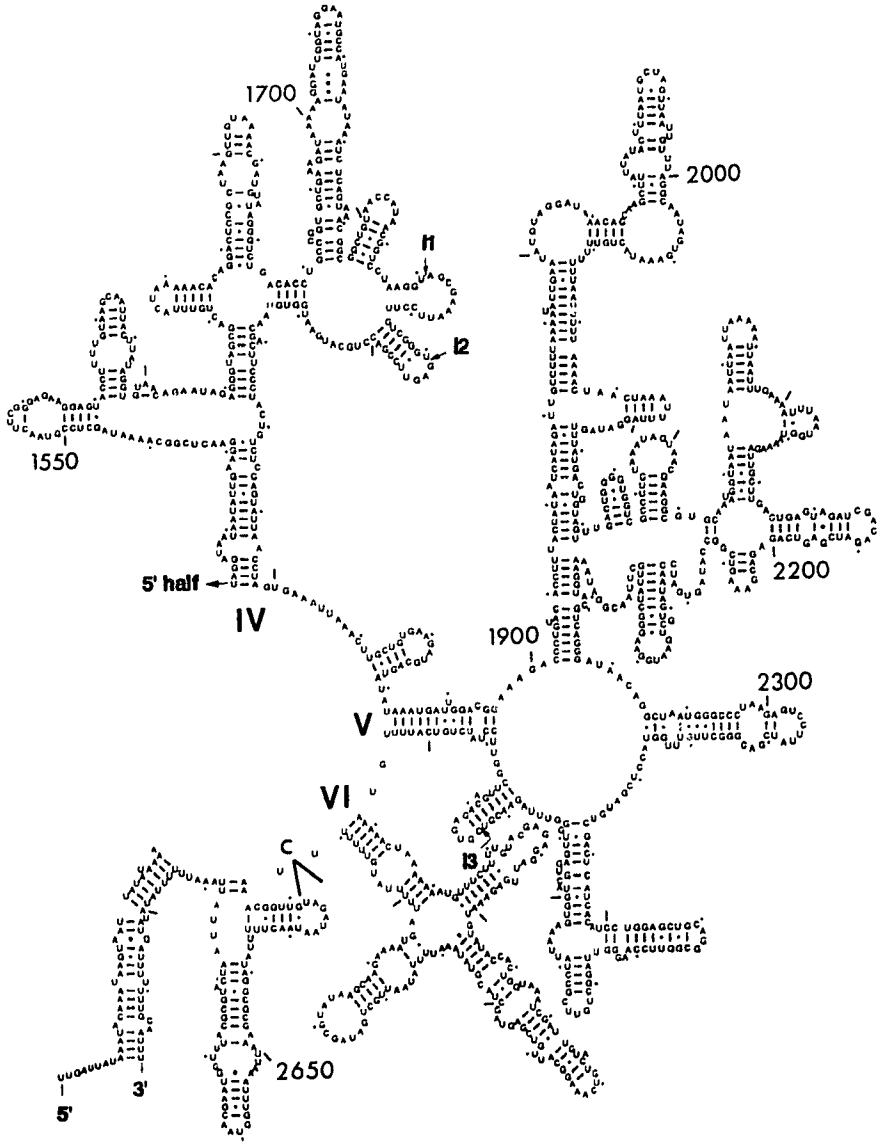


Figure 15

found in *E.coli* and many eubacteria, whereas *Acanthamoeba* and land plant (angiosperm) mitochondria have only two. The mitochondrial LSU rRNA from ciliates and fungi does not have any helices in this region of the molecule. The region within V16 in the 5'-half of the molecule (region "b"; *E. coli* coordinates 1405 - 1597) is shorter in *Acanthamoeba* but longer in angiosperms. The secondary structure of the 3'-half of the LSU rRNA is very similar to that of *E. coli.* One difference is that the helix at *E. coli* coordinates 2791 - 2805 (region "c") is absent in the *Acanthamoeba* mitochondrial sequence, as it is in those from all other mitochondria.

These comparisons of potential secondary structure between *A. castellanii* and *E. coli* LSU rRNAs are summarized in Table 1.

Three group I introns are found within highly conserved regions of domains IV and V at the locations indicated (Fig. 15B). That these nucleotide sequences within the DNA actually represent intron sequences has been verified by cDNA sequencing spanning the intron insertion sites (Fig. 16). The expected size of the LSU rRNA after intron removal is 2709 - 2719 nt, 185 - 195 nt smaller than that of the *E. coli* homolog. From the results of gel electrophoresis (Fig. 10), the size of the LSU rDNA transcript appears to be somewhat smaller than 2709 nt. However, considering that it is difficult to accurately estimate the sizes of very high molecular weight species in this type of gel system and that except for the intron sequences, the *A. castellanii* LSU rDNA contains no extra nucleotides relative to the *E. coli* LSU rRNA secondary structure over its entire length, the available data suggest that the *A. castellanii* mitochondrial LSU rRNA is not fragmented.

The 5'-terminus of the LSU rRNA is heterogeneous according to end-terminal analysis by tlc (Fig. 12). Primer extension analysis identifies one major and two minor 5'-termini (Fig. 13), all of which are indicated in the secondary structure of Fig. 15A. Similar to the finding for the SSU rRNA, an extended transcript is observed beginning 38 nt upstream from the major 5'-terminus of the LSU rRNA (Fig. 13). This position

Figure 16. Confirmation of intron sequences within the LSU rDNA. Autoradiograms are of cDNA sequencing gels generated by reverse transcriptase dideoxy sequencing primed from oligonucleotides specific for (A), the sequence spanning insertion sites for introns AcLSU.m1 (I1 in Fig. 12B) and AcLSU.m2 (I2) , and (B), the sequence spanning the insertion site for intron AcLSU.m3 (I3). Arrows indicate the positions of intron excision. The actual RNA nucleotide sequence is shown to facilitate comparison with Figure 12B.

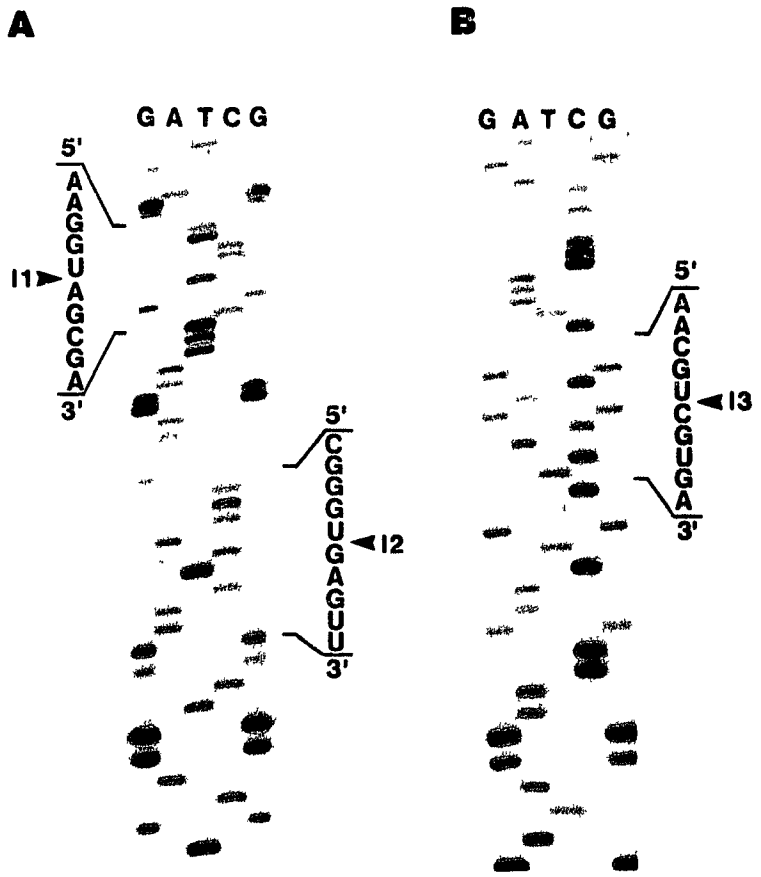

A
B
G A T C G
G A T C G
5'
AAGGUAGCGA
3'
I1
5'
CGGGUGAGUU
3'
I2
5'
AACGUCGUGA
3'
I3

Figure 16

corresponds to the 3'-terminal nt of $tRNA^{Leu}$ of tRNA cluster 1. Despite the fact that the 5'-end of the extended transcript is the same distance from the first major terminus for each rRNA, there is no obvious similarity in primary sequence in this region. The 3'-terminus of the LSU rRNA is homogeneous, as documented in Fig. 14.

(iii). Primary sequence comparisons of rRNA genes

To study the evolutionary affiliation of the *Acanthamoeba* mitochondrial genome, the primary nucleotide sequences of the rRNA genes were compared with the homologous mitochondrial sequences from other organisms and with their eubacterial homologs. The degree of primary sequence similarity was calculated from pairwise nucleotide comparisons of aligned sequences and attempts to define more precise evolutionary affiliations were made from phylogenetic treeing analysis. For SSU rRNA analysis, comparisons were restricted to the universal regions, U1 to U8 (541 positions) (Gray et al., 1984); for LSU rRNA analysis, 1093 phylogenetically conserved positions were compared, the co-ordinates of which are identified in the legend to Table 3.

a. Pairwise comparisons. The results of pairwise comparisons of SSU rRNA gene sequences are presented in Table 2. There are three main points to be noted from this table: 1). The SSU rRNA genes of *Acanthamoeba*, wheat and *Prototheca* show a slightly stronger affiliation with the *Agrobacterium* homolog than with other eubacteria. This is expected as *Agrobacterium* is a member of the α-subdivision of the Proteobacteria and has previously been shown to be a specific eubacterial relative of mitochondria (Yang et al., 1985). 2). Compared to the SSU rRNA sequence of *Agrobacterium*, the mitochondrial SSU rRNAs of *Acanthamoeba*, wheat and *Prototheca* all show >80% sequence identity within the universal regions, whereas the other mitochondrial SSU rRNA sequences analysed here show significantly lower similarity. 3). The bottom half of the table indicates that according to SSU rRNA gene sequence data, the *Acanthamoeba* mitochondrion is much more closely related to

Table 2. Degree of identity (%) in pairwise comparisons within universal regions of mitochondrial and eubacterial SSU rRNA sequences.

	Aca	Tae	Pwi	Cre	Pan	Tpy	Hsa
Atu(α)	82	91	87	66	67	69	59
Pte(β)	80	87	84	67	67	69	58
Eco(γ)	79	87	85	65	67	68	58
Bsu	76	86	81	64	66	66	59
Ani	76	84	81	64	66	67	59
Aca	-	81	76	61	65	68	57
Tae		-	86	65	68	71	57
Pwi			-	66	67	71	57
Cre				-	61	61	50
Pan					-	60	54
Tpy						-	53

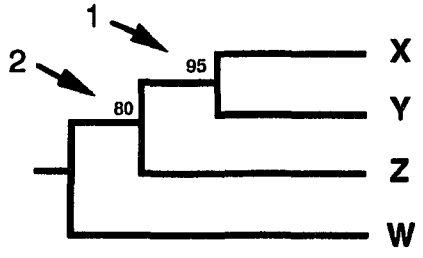
Universal regions are defined with numbering according to the *E. coli* SSU (16S) rRNA sequence and include 541 positions: U1, 11 - 62; U2, 348 - 403; U3, 500 - 557; U3a, 564 - 574; U4, 766 - 817; U5, 880 - 993; U6, 1215 - 1238; U7, 1308 - 1418; U8, 1483 - 1534 (see Gray et al., 1984). α, β, and γ represent the alpha, beta and gamma divisions of the purple non-sulfur bacteria (Proteobacteria), respectively. Abbreviations and references including GenBank accession numbers (Neefs et al., 1991; De Rijk et al., 1992): Atu, *Agrobacterium tumefaciens* (M11223, Yang et al., 1985); Pte, *Pseudomonas testosteroni* (M11224, Yang et al., 1985); Eco, *Escherichia coli* (V00348, Brosius et al., 1981); Bsu, *Bacillus subtilis* (K00637, Green et al., 1985); Ani, *Anacystis nidulans* (X03538, Tomioka and Sugiura, 1983); Aca, *Acanthamoeba castellanii* (this thesis); Tae, *Triticum aestivum* (wheat) (K01229, Spencer et al., 1984); Pwi, *Prototheca wickerhamii* (X15435, Wolff and Kück, 1990); Tpy, *Tetrahymena pyriformis* (M12714, Schnare et al., 1986); Pan, *Podospora ansernia* (X14734, Cummings et al., 1989b); Cre, *Chlamydomonas reinhardtii* (M25119, Boer and Gray, 1988); Hsa, *Homo sapiens* (V00710, Eperon et al., 1980).

wheat and *Prototheca* mitochondria than it is to mitochondria from *C. reinhardtii*, animals, fungi or ciliate protozoa.

The results of pairwise comparisons of LSU rRNA gene sequences are presented in Table 3. These comparisons indicate that: 1). The mitochondria of land plants clearly, and of *Acanthamoeba* to a lesser degree, remain the most closely related of the mitochondria to their eubacterial ancestor, but this affiliation is not as pronounced for the LSU rRNA gene sequence data set as for the SSU rRNA data set (Table 2). 2). In agreement with the SSU rRNA gene sequence analysis, *Acanthamoeba* mitochondria appear to be much more closely related to land plant mitochondria than to *Chlamydomonas* mitochondria. 3). By LSU rRNA analysis, the *Acanthamoeba* mitochondrial sequence is not as distantly related to non-plant mitochondrial sequences as observed for the SSU rRNA sequence analysis.

In summary, as deduced from pairwise primary sequence comparisons of SSU (Table 2) and LSU (Table 3) rRNA sequences, *A. castellanii* mitochondria show the strongest affiliation with plant mitochondria of all mitochondrial representatives, in addition to a strong eubacterial affiliation.

b. Phylogenetic treeing analysis. As an explanation of how to interpret a phylogenetic tree, consider the following hypothetical tree:



In this tree, the more closely related taxonomic units (organism, gene, species), X and Y, share a more recent common ancestor (whose phylogenetic position is indicated at node 1) with one another than either X or Y does with taxonomic unit Z (the

Table 3. Degree of identity (%) in pairwise comparisons within conserved regions of mitochondrial and eubacterial LSU rRNA sequences.

	Aca	Tae	Cre	Tpy	Ncr	Sce
Rca(α)	75	81	71	71	68	64
Pce(β)	73	80	70	70	68	66
Eco(γ)	73	81	70	70	66	65
Ani	73	77	70	69	65	63
Aca	-	76	68	72	63	65
Tae		-	74	72	67	65
Cre			-	66	63	62
Tpy				-	66	66
Ncr					-	68

Sequences were aligned based on a combination of primary sequence and secondary structure characteristics and 1093 positions were used for the analysis (same alignment as for LSU rRNA phylogenetic trees). These positions included the following nucleotide coordinates of *E. coli* LSU (23S) rRNA: 561 - 588, 669 - 703, 729 - 837, 941 - 1004, 1051 - 1108, 1151 - 1163, 1186 - 1197, 1248 - 1261, 1295 - 1312, 1599 - 1645, 1648 - 1706, 1752 - 1803, 1814 - 1837, 1903 - 2076, 2244 - 2287, 2345 - 2394, 2422 - 2625, 2646 - 2695, 2712 - 2741. α, β and γ represent the alpha, beta and gamma divisions of the purple non-sulfur bacteria (Proteobacteria), respectively. Abbreviations and references including GenBank accession numbers (Gutell et al., 1992b): Rca, *Rhodobacter capsulatus* (X06485, Regensburger et al., 1988a); Pce, *Pseudomonas cepacia* (X16368, Hoepfl et al., 1989); Eco, *Escherichia coli* (J01695, Brosius et al., 1980); Ani, *Anacystis nidulans* (X00512, Douglas and Doolittle, 1984); Aca, *Acanthamoeba castellanii* (this thesis); Tae, *Triticum aestivum* (wheat) (M37274, Falconet et al., 1988); Cre, *Chlamydomonas reinhardtii* (M22649, Boer and Gray, 1988); Tpy, *Tetrahymena pyriformis* (M58011, Heinonen et al., 1990); Ncr, *Neurospora crassa* (X55443, Abdel-Malik et al., 1990); Sce, *Saccharomyces cerevisiae* (J01527, Sor and Fukuhara, 1983).

phylogenetic position of the most recent common ancestor of X, Y amd Z is at node 2). All of the phylogenetic trees presented in this thesis are unrooted, i.e., the direction of evolution is not implied. In some phylogenetic trees (as indicated in the figure legends) the horizontal distance of the branch lengths is proportional to the evolutionary distance (i.e., the number of nucleotide substitutions) of the associated lineage. Bootstrap values are indicated at the nodes of bifurcating branches of parsimony and distance trees. Bootstrapping is a statistical method designed by Diaconis and Efron (1983) and adapted to molecular phylogeny by Felsenstein (1985) to evaluate the significance or reliability of tree topologies. Bootstrapping evaluates whether phylogenetic relationships are maintained in the presence of random variation introduced into the data set, i.e., it asks, "How robust are the affiliations?" Multiple series of random nucleotide substitutions are made to the original data set, and phylogenetic trees are generated from each altered data set (usually 100 replicas). A consensus tree is derived from the replica trees and bootstrap values are expressed as percentiles. Thus, in the sample tree shown above, the bootstrap value of 95 indicates that even when random sequence variation is introduced, X and Y are affiliated in 95% of the replica trees. In 80% of the replica trees, all three taxonomic units, X, Y and Z are affiliated. The value of 95% implies that the grouping of X and Y is statistically significant whereas that of X, Y and Z is less significant.

Phylogenetic trees were constructed from SSU and LSU rRNA sequence data. A neighbor-joining distance matrix method was used for the construction of the SSU rRNA phylogenetic tree (Fig. 17). Three different treeing methods, neighbor-joining distance matrix (Fig. 18), maximum likelihood (Fig. 19), and maximum parsimony (Fig. 20), were used for construction of the LSU rRNA phylogenetic trees.

Although there are some inconsistencies among the various trees as noted below, and bootstrap values are generally low except for very closely related taxa, some generalizations can be made from the phylogenetic analysis.

Figure 17. Mitochondrial phylogenetic tree based on SSU rRNA sequence data. This tree was constructed using the neighbor-joining distance matrix method of PHYLIP, version 3.5c (Felsenstein, 1993). Sequences comprising universal regions only (U1 to U8) were included in the analysis. Branch lengths are not proportional to sequence divergence. Bootstrap (or consense) values (% occurrence) are indicated along each bifurcating branch. All sequences that are not mitochondrial (MT) are eubacterial. α, β, and γ represent the alpha, beta, and gamma divisions of the purple, non-sulfur bacteria (Proteobacteria), respectively.

References including GenBank accession numbers not given in Table 1, and which also can be found in Neefs et al., 1991; De Rijk et al., 1992: *Saccharomyces cerevisiae* (V00704, Li et al., 1982); *Paramecium tetraurelia* (K01751, Seilhamer et al., 1984); *Oenothera berteriana* (X61277, Brennicke et al., 1991; Knoop, 1991); *Marchantia polymorpha* (M68929, Oda et al., 1992); *Ehrlichia risticii* (M21290, Weisburg et al., 1989); *Rhodobacter sphaeroides* (X53853, Dryden and Kaplan, 1990); *Pseudomonas aeruginosa* (X06684, Toschka et al., 1988); *Pirellula marina* (X62912, Stackebrandt, 1991); *Thermus thermophilus* (M26923, M26924, Hartman and Erdmann, 1989; Hartman et al., 1989); and *Mycoplasma capricolum* (X00921, X00922, Iwami et al., 1984; Sawada et al., 1984).

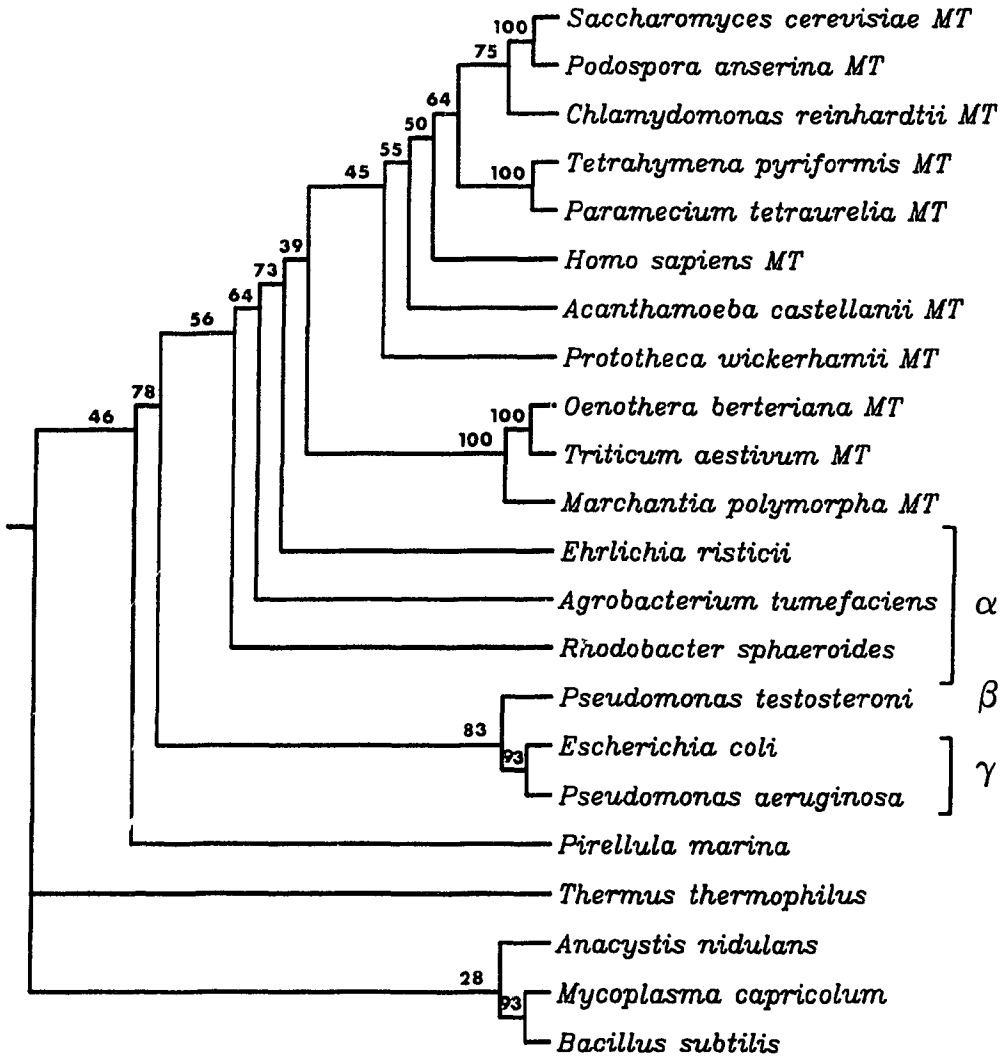

Saccharomyces cerevisiae MT
Podospora anserina MT
Chlamydomonas reinhardtii MT
Tetrahymena pyriformis MT
Paramecium tetraurelia MT
Homo sapiens MT
Acanthamoeba castellanii MT
Prototheca wickerhamii MT
Oenothera berteriana MT
Triticum aestivum MT
Marchantia polymorpha MT
Ehrlichia risticii
Agrobacterium tumefaciens
Rhodobacter sphaeroides
Pseudomonas testosteroni
Escherichia coli
Pseudomonas aeruginosa
Pirellula marina
Thermus thermophilus
Anacystis nidulans
Mycoplasma capricolum
Bacillus subtilis
α
β
γ
100
75
64
50
55
45
100
39
73
64
56
78
46
100
100
83
93
28
93

Figure 17

Figure 18. Distance (neighbor-joining) mitochondrial phylogenetic tree based on LSU rRNA sequence data. Sequence positions included in the analysis are indicated in the legend of Table 2. Bootstrap (consense) values (% occurrence) are indicated along the branch lengths, which are not proportional to sequence divergence. All sequences that are not mitochondrial (MT) or chloroplast (CP) are eubacterial. α, β, and γ represent the alpha, beta, and gamma divisions of the purple, non-sulfur bacteria (Proteobacteria), respectively.

References and GenBank accession numbers not given in Table 2, and which can also in found in Gutell et al., 1992b: *Podospora anserina* (X14735, Cummings et al., 1989); *Schizosaccharomyces pombe* (X06597, Lang et al., 1987); *Paramecium tetraurelia* (K01749, Seilhamer et al., 1984); *Zea mays* (K01868, Dale et al., 1984); *Oenothera berteriana* (X02559, Manna and Brennicke, 1985); *Marchantia polymorpha* (M68929, Oda et al., 1992); *Pseudomonas aeruginosa* (Y00432, Toschka et al., 1987); *Pirellula marina* (X07408, Liesack et al., 1988); *Streptomyces ambofaciens* (M27245, Pernodet et al., 1989); *Micrococcus luteus* (X06484, Regensburger et al., 1988b); *Thermus thermophilus* (X12612, Hoepfl et al., 1988); *Bacillus subtilis* (K00637, M10606, X00007, Green et al., 1985); *Chlamydomonas reinhardtii* CP (X16686, X16687, Lemieux et al., 1989).

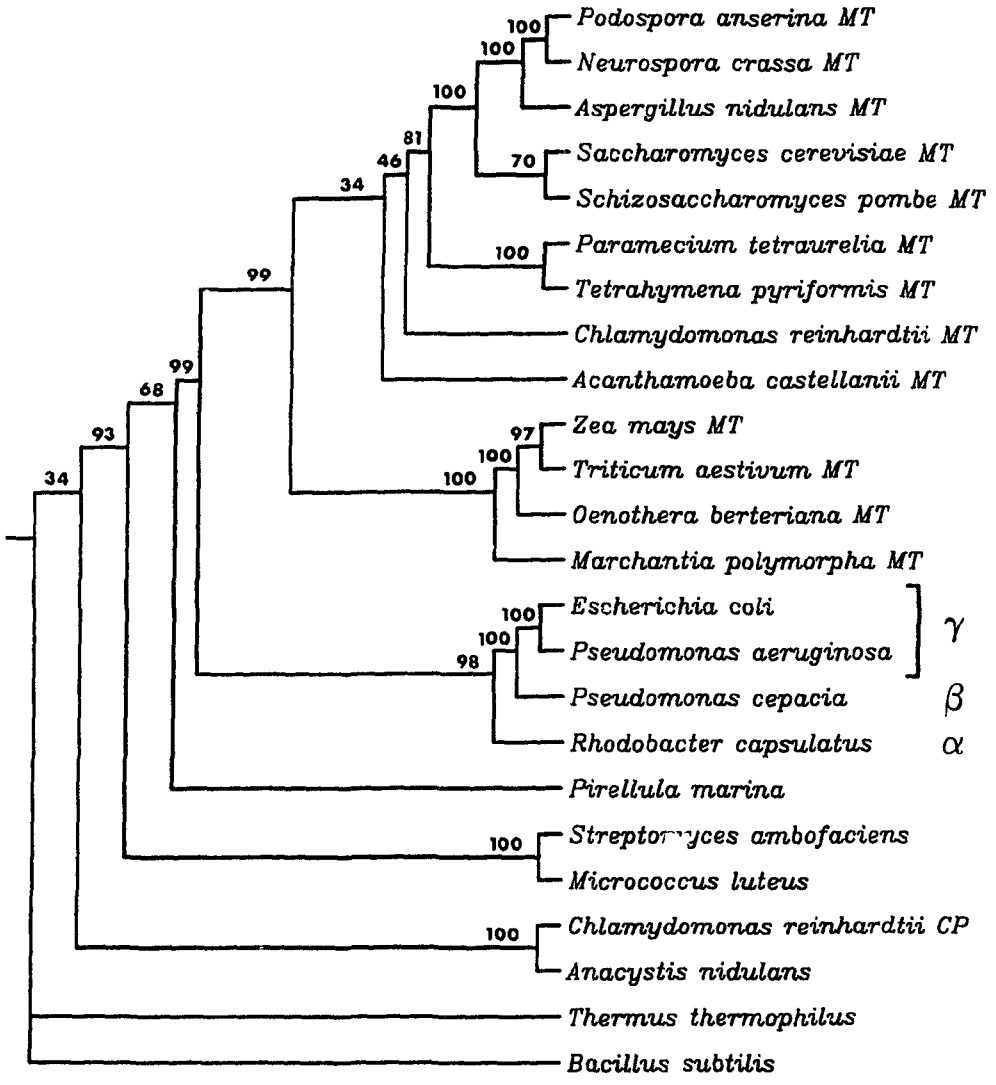

Podospora anserina MT
Neurospora crassa MT
Aspergillus nidulans MT
Saccharomyces cerevisiae MT
Schizosaccharomyces pombe MT
Paramecium tetraurelia MT
Tetrahymena pyriformis MT
Chlamydomonas reinhardtii MT
Acanthamoeba castellanii MT
Zea mays MT
Triticum aestivum MT
Oenothera berteriana MT
Marchantia polymorpha MT
Escherichia coli
Pseudomonas aeruginosa
Pseudomonas cepacia
Rhodobacter capsulatus
Pirellula marina
Streptomyces ambofaciens
Micrococcus luteus
Chlamydomonas reinhardtii CP
Anacystis nidulans
Thermus thermophilus
Bacillus subtilis
γ
β
α
100
100
100
81
46
34
70
100
99
99
68
93
34
97
100
100
100
100
98
100
100

Figure 18

Figure 19. Maximum likelihood mitochondrial phylogenetic tree based on LSU rRNA sequence data. Branch lengths, in horizontal dimension, are proportional to the extent of sequence divergence. Sequence positions included in the analysis are indicated in the legend of Table 2. All sequences that are not mitochondrial (MT) or chloroplast (CP) are eubacterial. α, β, and γ represent the alpha, beta, and gamma divisions of the purple, non-sulfur bacteria (Proteobacteria), respectively.

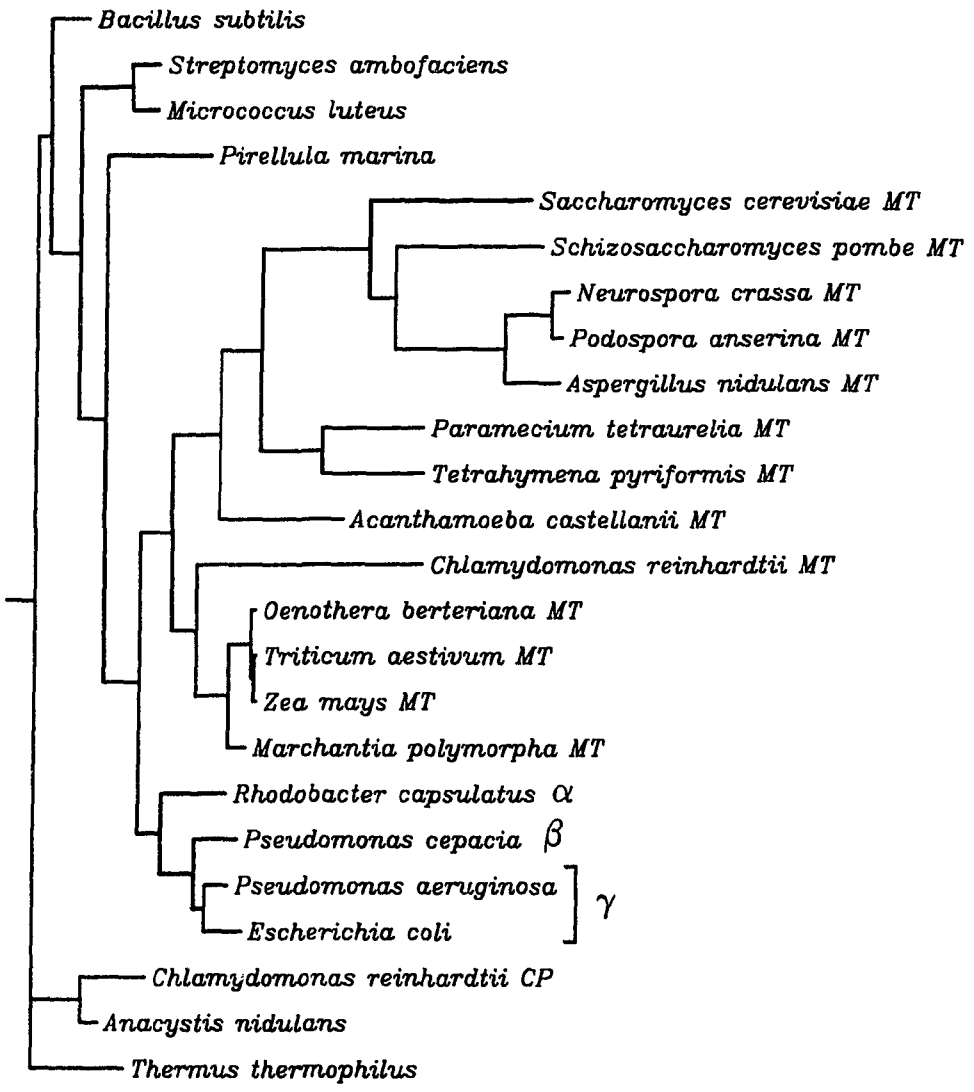

Bacillus subtilis
Streptomyces ambofaciens
Micrococcus luteus
Pirellula marina
Saccharomyces cerevisiae MT
Schizosaccharomyces pombe MT
Neurospora crassa MT
Podospora anserina MT
Aspergillus nidulans MT
Paramecium tetraurelia MT
Tetrahymena pyriformis MT
Acanthamoeba castellanii MT
Chlamydomonas reinhardtii MT
Oenothera berteriana MT
Triticum aestivum MT
Zea mays MT
Marchantia polymorpha MT
Rhodobacter capsulatus α
Pseudomonas cepacia β
Pseudomonas aeruginosa
Escherichia coli
γ
Chlamydomonas reinhardtii CP
Anacystis nidulans
Thermus thermophilus

Figure 19

Figure 20. Maximum parsimony phylogenetic tree based on LSU rRNA sequence data. Sequence positions included in the analysis are indicated in the legend of Table 2. Bootstrap (consense) values (% occurrence) are indicated along the branch lengths, which are proportional, in the horizontal dimension, to sequence divergence. All sequences that are not mitochondrial (MT) or chloroplast (CP) are eubacterial. α, β, and γ represent the alpha, beta, and gamma divisions of the purple, non-sulfur bacteria (Proteobacteria), respectively.

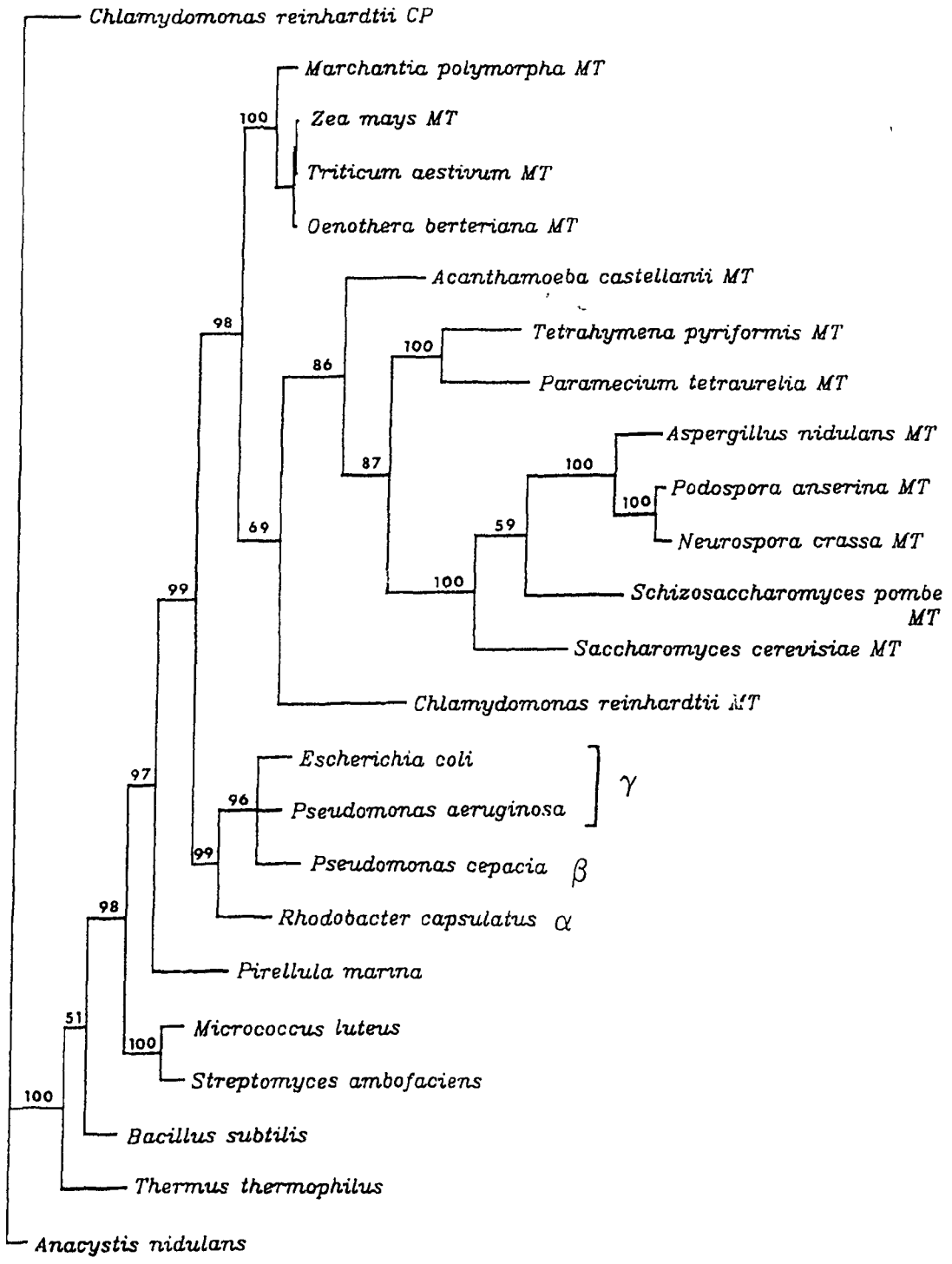

Chlamydomonas reinhardtii CP
Marchantia polymorpha MT
Zea mays MT
Triticum aestivum MT
Oenothera berteriana MT
Acanthamoeba castellanii MT
Tetrahymena pyriformis MT
Paramecium tetraurelia MT
Aspergillus nidulans MT
Podospora anserina MT
Neurospora crassa MT
Schizosaccharomyces pombe MT
Saccharomyces cerevisiae MT
Chlamydomonas reinhardtii MT
Escherichia coli
Pseudomonas aeruginosa
γ
Pseudomonas cepacia β
Rhodobacter capsulatus α
Pirellula marina
Micrococcus luteus
Streptomyces ambofaciens
Bacillus subtilis
Thermus thermophilus
Anacystis nidulans
100
98
86
100
87
100
100
69
59
100
99
97
96
99
98
51
100
100

Figure 20

(1). All mitochondrial sequences consistently tree together.

(2). According to the SSU rRNA tree, all mitochondria originate from within the α subdivision of the Proteobacteria, represented here by *Ehrlichia risticii*, *Agrobacterium tumefaciens* and *Rhodobacter sphaeroides*. The origin of mitochondria from within the α-Proteobacteria is supported but not confirmed by the LSU rRNA trees.

(3). In all of the trees *Acanthamoeba* mitochondrion is ancestral to fungal and ciliate mitochondria.

(4). In none of the trees does *Acanthamoeba* branch with plants.

(5). The relationship between *Chlamydomonas* mitochondria and other mitochondria is not clear, as is evident from the different topologies depending upon the treeing method used and the accompaning low bootstrap values.

(6). Although the bootstrap values for the LSU rRNA distance tree are low (Fig. 18), only in 1% of the trees does *Acanthamoeba* specifically tree with plants (according to the Phylip outfile data), and under no circumstances in the parsimony tree does *Acanthamoeba* specifically tree with plant mitochondria (from PAUP outfile data). Similarily, only 4% of the SSU rRNA distance trees place *Acanthamoeba* mitochondria specifically with plant mitochondria.

In summary, rRNA sequence comparisons indicate the following:

(1). The pronounced primary sequence similarity between mitochondrial and eubacterial SSU rRNAs extends outside of the metaphyte/chlorophyte lineage to include *A. castellanii.*

(2). In terms of organization, primary sequence and secondary structure, the mitochondrial rRNA genes of *A. castellanii* more closely resemble those from land plants than those from the green alga, *C. reinhardtii.*

(3). Phylogenetic analysis places *Acanthamoeba* mitochondria as ancestral to fungal and ciliate mitochondria and not specifically related to plant mitochondria.

C. Introns of the LSU rRNA Gene

Three introns are found within highly conserved regions of domains IV and V of *A. castellanii* mitochondrial *rnl*; these are referred to here as AcLSU.m1; AcLSU.m2 and AcLSU.m3 according to their order of occurrence within the gene (Fig. 15 B). As noted earlier [B. (ii)], the existence of these introns has been confirmed by cDNA sequence analysis of the mature LSU rRNA (Fig. 16). AcLSU.m1 is located at a site equivalent to nucleotide position 1931 of the *E. coli* LSU rRNA, AcLSU.m2 at the equivalent of *E. coli* position 1951, and AcLSU.m3 at *E. coli* position 2593.

(i). Secondary structures

Each of the *A. castellanii* LSU rDNA introns has a potential secondary structure characteristic of a group I intron (Burke et al., 1987) (Fig. 21A, B and C). Conserved secondary structure elements P1 - P10, can be identified for each intron and the conserved primary sequence elements P, Q, R and S (Cech, 1988) can be found within each core region. Based upon these sequence elements, the *A. castellanii* rDNA introns can be further categorized into one of the eleven subgroups of group I introns (Michel and Westhof, 1990). AcLSU.m1 has no extra nucleotides between P7 and P3 and can be assigned to subgroup IB4 (Fig. 21A). AcLSU.m2 and AcLSU.m3 each have helices P7.1 and P7.2 and can be placed within subgroup IA3 (Fig. 21B, C). Distinguishing features of subgroup 1A3 include the following: P3 consists of 5 nt instead of the more commonly occurring 6 - 7 nt; helices P3 and P4 are separated (i.e., J3/4) by 5 nt compared to the more common 3 nt separation; the last 2 nt of sequence element Q (i.e., the first 2 nt of P6) are GG instead of the more commonly occurring GC; the first 3 nt of sequence element R (i.e., J6/7) are GUA instead of the usual UCA for subgroup IA introns.

Group I intron splice sites display some nucleotide conservation: a U immediately precedes the 5' splice site (paired with a G in the intron sequence) and a G

Figure 21. Diagrams showing the potential secondary structures of *A. castellanii* mitochondrial *rnl* introns, as derived from the DNA sequence. The RNA-like strand of the DNA is shown. The secondary structures are drawn according to the convention proposed by Burke et al. (1987). Intron sequence is represented by uppercase letters and exon sequence by lowercase letters. Arrows refer to nucleotide positions that define the splice sites. Helices constituting the structural core of group I introns, P1 - P9, are indicated. Potential P9.0 (indicated by dashed line) and P10 (upper right) pairings are also shown. Conserved sequence elements P, Q, R and S are shown in italics. Solid lines indicate continuity of primary sequence and do not indicate omission of any nucleotides, except as indicated within each open reading frame (ORF) region of the intron secondary structure. Dots signify the initiation and termination codons of the ORFs. Asterisks in (A) indicate the potential for base pairing. (A) AcLSU.m1; (B) AcLSU.m2 ; and (C) AcLSU.m3.

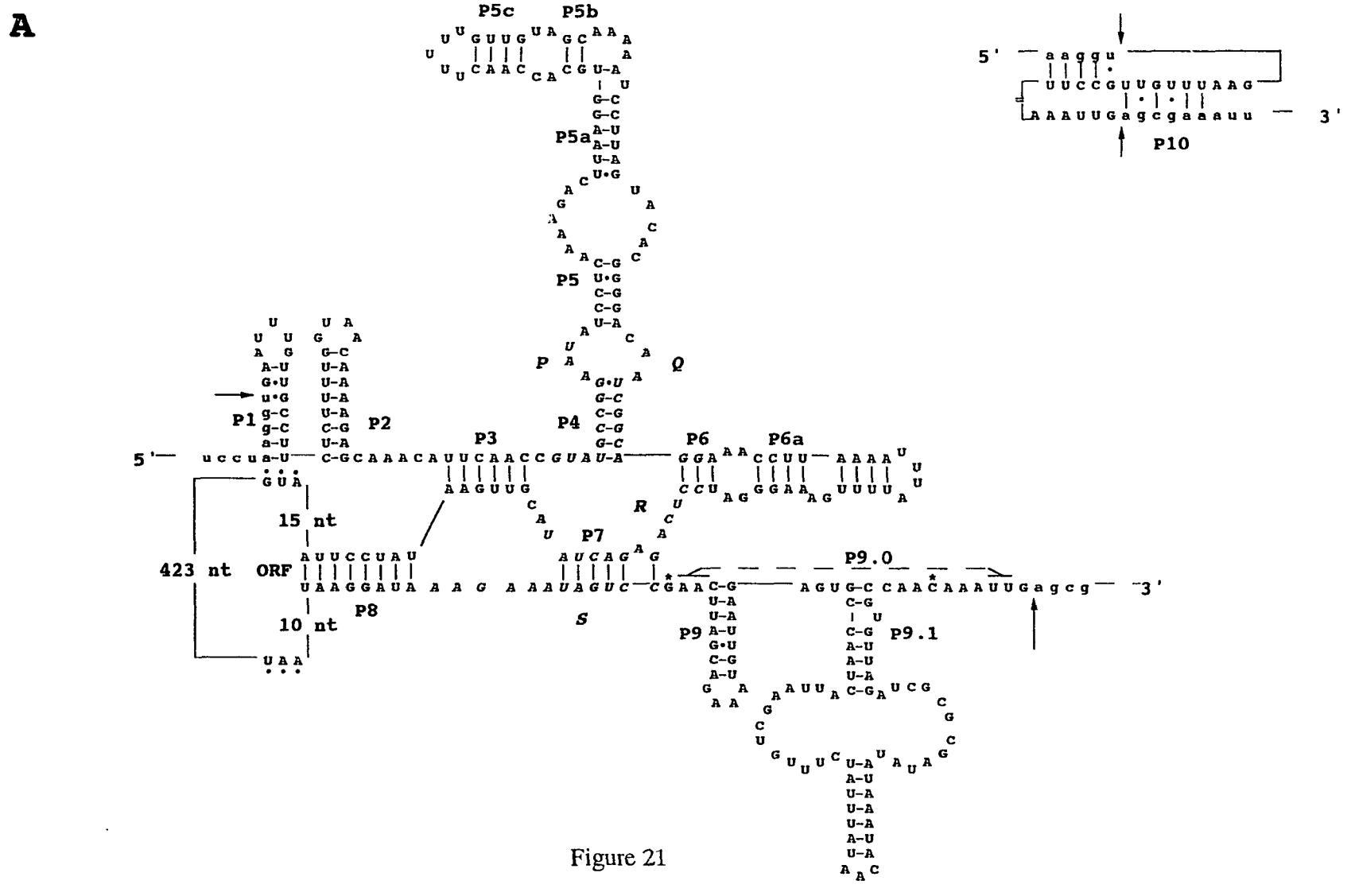

A
P5c
P5b
P5a
P5
P
Q
P1
P2
P3
P4
P6
P6a
R
P7
S
P8
P9
P9.0
P9.1
P10
15 nt
10 nt
423 nt
ORF
5'
3'

Figure 21

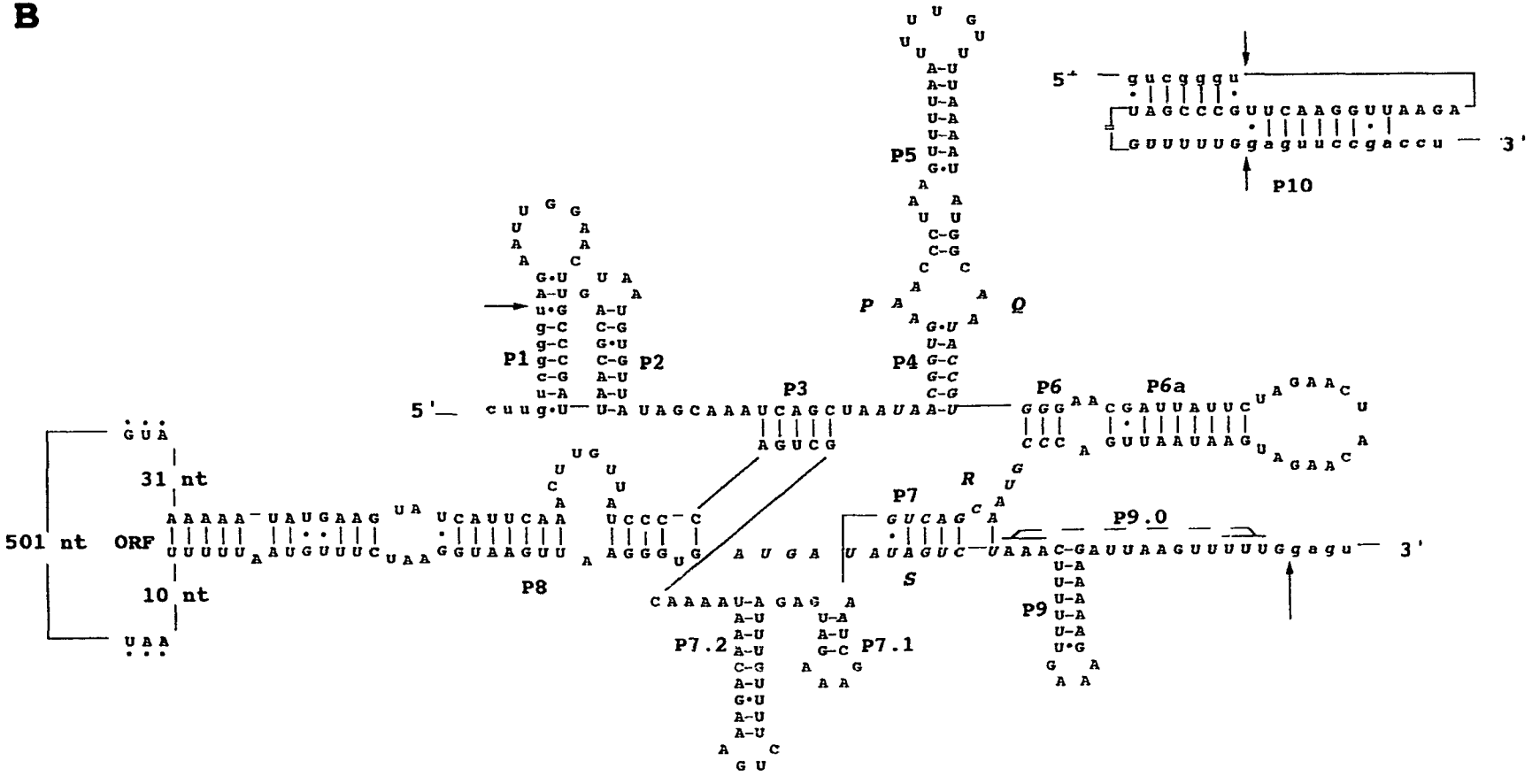

B
501 nt
ORF
31 nt
10 nt
P1
P2
P3
P4
P5
P6
P6a
P7
P7.1
P7.2
P8
P9
P9.0
P10
P
Q
R
S
5'
3'

Figure 21

C

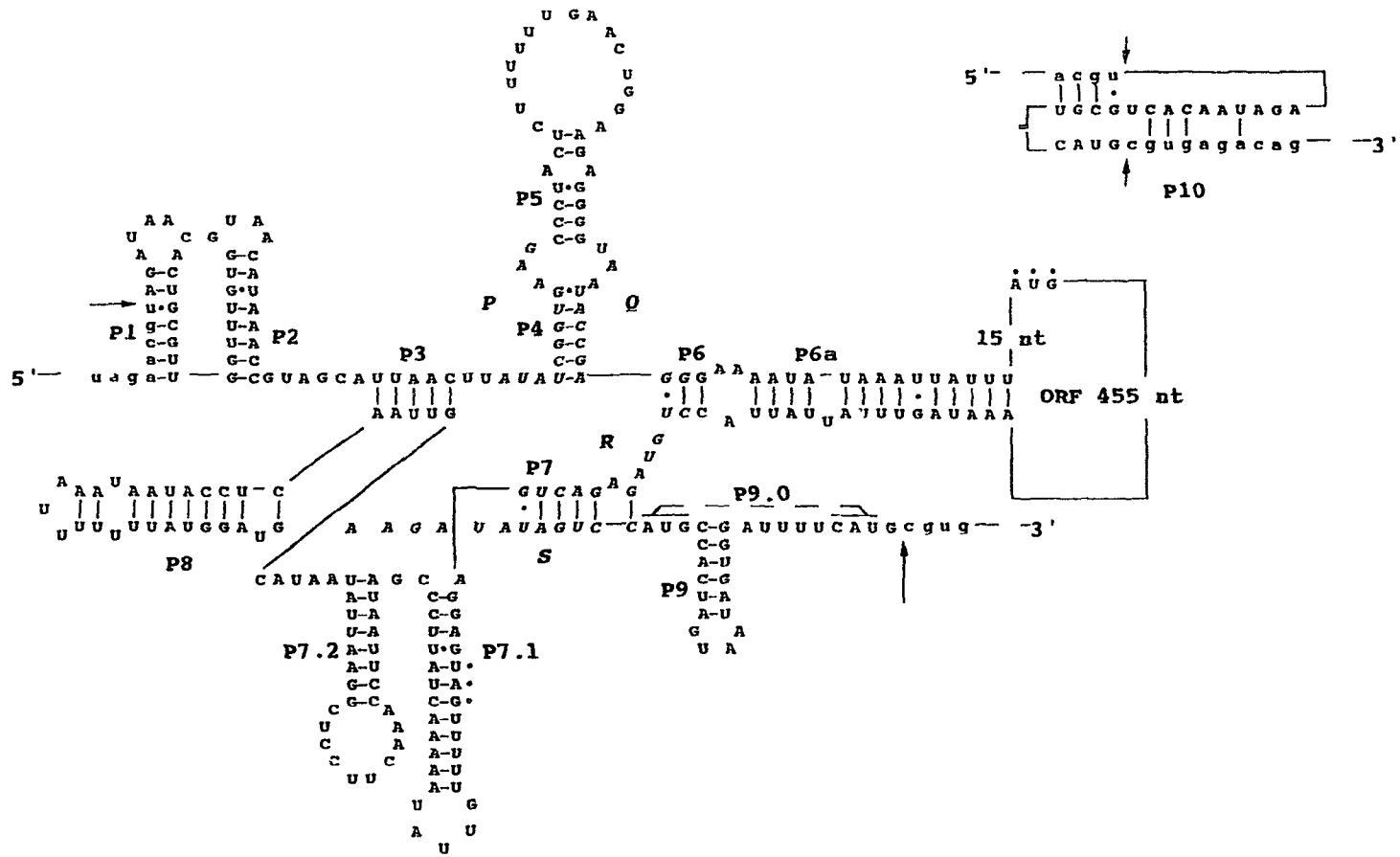

P1
P2
P3
P4
P5
P6
P6a
P7
P7.1
P7.2
P8
P9
P9.0
P10
P
Q
R
S
5'
3'
15 nt
ORF 455 nt

Figure 21

immediately precedes the 3' splice site. An internal guide sequence (IGS) involved in a potential P10 pairing (in addition to P1 pairing) is convincingly present for AcLSU.m1 and AcLSU.m2 (Fig. 21 A and B), but less so for AcLSU.m3 (Fig. 21 C). P10 pairing for AcLSU.m1 and AcLSU.m2 could potentially consist of 6 and 7 consecutive bp between the IGS and the 3' exons, respectively. However, a possible P10 pairing for AcLSU.m3 would be expected to consist of only three consecutive base pairs between the IGS and the 3' exon and only after an unpaired U : c, with the c being the first nucleotide of the 3' exon. There is a possibility for base pairing to occur between two consecutive nucleotides located in the J7/9 region (referred to as the 3'-splice site binding sequence) with two consecutive nucleotides located immediately 5' to the terminal G for all three of the *A. castellanii rnl* introns (P9.0 pairing). For AcLSU.m3 this could consist of a 3 bp interaction in which the CAU immediately 5' to the terminal G pairs with AUG of the J7/9 region (Fig. 21 C).

(ii). Open reading frames

All three of the *A. castellanii* mitochondrial *rnl* introns have ORFs (Fig. 8 and 21A, B, C). As judged from an analysis of *cox*1 and *cox*2 gene sequences, codon usage in *Acanthamoeba* mitochondria follows the universal genetic code except that UGA apparently specifies tryptophan (see E. Cytochrome Oxidase). Termination codons, UAA and UAG, are abundant in the two alternate reading frames for all three introns. As in other group I rDNA introns, the ORFs reported here are localized to regions peripheral to the intron core, except that the ORF of AcLSU.m3 does extend marginally into the core region of this intron (Fig. 21C).

The AcLSU.m1, AcLSU.m2 and AcLSU.m3 intron ORFs have the potential to encode proteins of 142 amino acid (aa) residues (mol. wt. 16.5 kDa), 168 aa (19.7 kDa) and 164 aa (19.1 kDa), respectively. All three of the potential *Acanthamoeba* intron-encoded proteins would be relatively rich in phenylalanine and, similar to other intron-encoded proteins (Collins et al., 1988), would have a net positive charge due to

elevated lysine contents . Other notable features characteristic of intronic ORF amino acid composition (Doolittle, 1986) include elevated isoleucine levels for AcLSU.m1 ORF and elevated leucine levels for AcLSU.m2 ORF and AcLSU.m3 ORF. Only one of the two expected conserved dodecapeptides (P1 or P2) is found in each of the ORFs of *Acanthamoeba* LSU rDNA introns.

(iii). Comparison of AcLSU.m3 and CrLSU.c

The chloroplast *rnl* intron of *C. reinhardtii* (CrLSU.c) (Rochaix et al., 1985) is similar to AcLSU.m3 in a number of respects:

(1). The two introns are found at identical nucleotide insertion sites within domain V of homologous *rnl* genes of the mitochondrial (AcLSU.m3) and the chloroplast (CrLSU.c) genomes.

(2). The intron ORFs of AcLSU.m3 and CrLSU.c differ in size by only one amino acid residue. As with the AcLSU.m3 ORF, that of CrLSU.c has only one of the conserved P1/P2 dodecapeptides. The two ORFs share 39% aa identity over 150 amino acid residues (Fig. 22). According to the normalized alignment score (Doolittle, 1986), these ORFs are within the range of certain homology.

(3). Structural similarities are observed within the core regions of AcLSU.m3 and CrLSU.c introns. Both introns belong to group I subgroup IA3 (Michel and Westhof, 1990) and in each case the ORF they contain is located within L6. Figure 23 indicates the nature and extent to which identical nucleotides are shared within the core regions of the AcLSU.m3 and CrLSU.c introns.

In summary, AcLSU.m3 and CrLSU.c are located at precisely the same site within homologous genes; they share a relatively high degree of primary sequence identity within the core region of the intron; and the potential ORF-encoded proteins that they specify are homologous.

(iv). AcLSU.m1 and AcLSU.m2 comparisons

```
                                 ............
AcLSU.m3   MQKNQFKKLNNFQLAYLAGFVEADGCFLVQIIPGLQYRYKHTIRISIVFY
CrLSU.c        MNTKYNKEFLLYLAGFVDGDGSIIAQIKPNQSYKFKHQLSLAFQVT

AcLSU.m3   QKKDKHWYFLQLKNLIGLGSIRFRNDGMLEYSITGLSLVNKFLEMLPPYL
CrLSU.c    QKTQRRWFLDKLVDEIGVGYVRDR-GSVSDYILSEIKPLHNFLTQLQPFL

AcLSU.m3   ILKKNLAVLIFRIIKGLNDVKNEAG-FLEVCKLVDEVADHTYSKKRKNTS
CrLSU.c    KLKQKQANLVLKIIWRLPSAKESPDKFLEVCTWVDQIAALNDSKTRKTTS

AcLSU.m3   LTVKNSLLLPVETEE
CrLSU.c    ETVRAVLDSLSEKKKSSP
```

Figure 22. Amino acid sequence alignment of intron ORFs encoded by *A. castellanii* mitochondrial AcLSU.m3 and *C. reinhardtii* chloroplast CrLSU.c. Dots above the sequences identify the conserved dodecapeptide motifs of the ORFs. Identical amino acid residues are enclosed within boxes. Sequence alignments were obtained from a GenBank search (FASTA; Pearson and Lipman, 1988).

Figure 23. Comparison of the primary nucleotide sequence within the core region of the *rnl* introns of *A. castellanii* mitochondria (AcLSU.m3) and *C. reinhardtii* chloroplast (CrLSU.c). The structure shown represents the AcLSU.m3 intron and residues shared between AcLSU.m3 and CrLSU.c are circled.

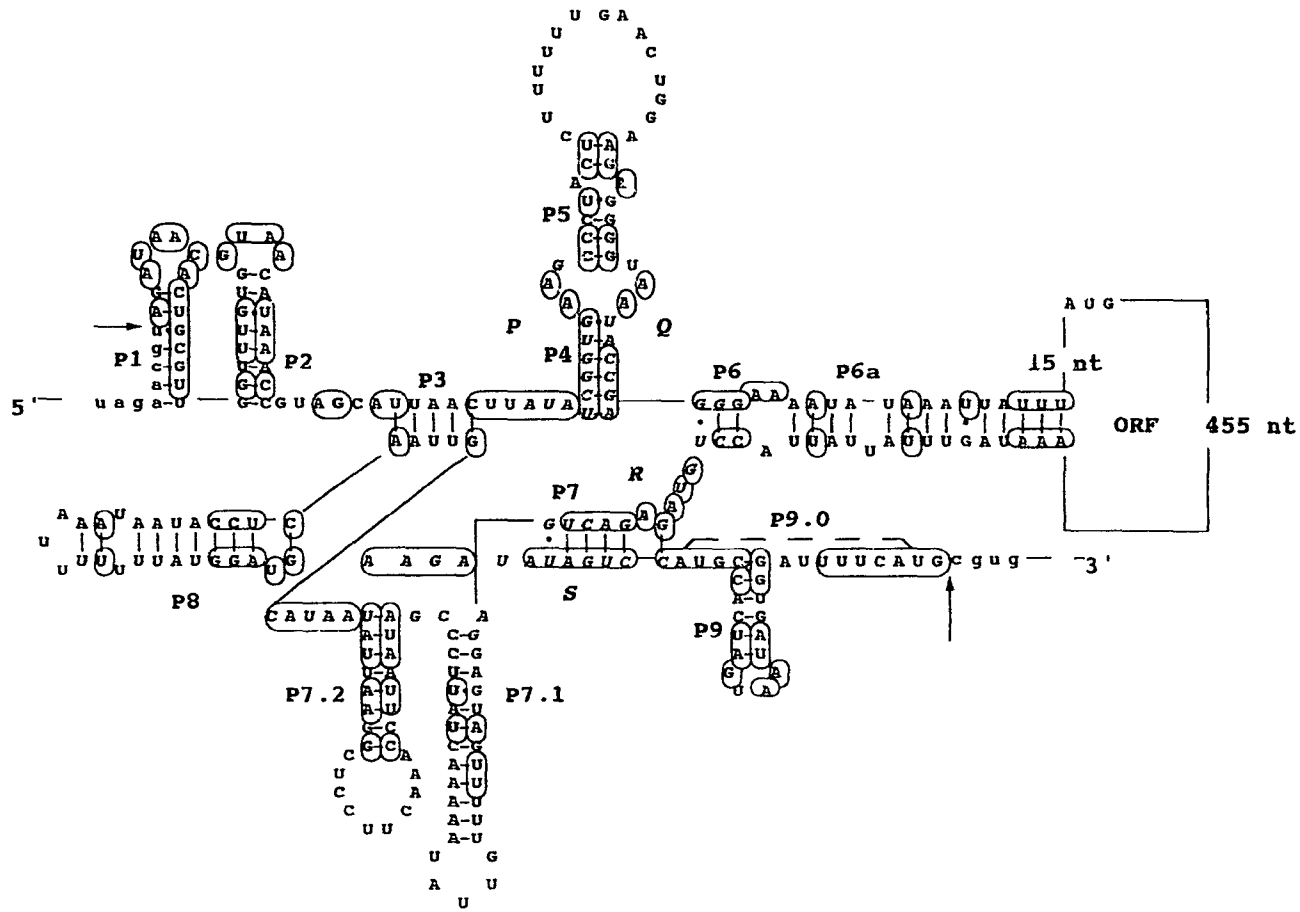

P1
P2
P3
P4
P5
P6
P6a
P7
P7.1
P7.2
P8
P9
P9.0
P
Q
R
S
5'
3'
15 nt
ORF
455 nt
AUG

Figure 23

AcLSU.m1 and AcLSU.m2 intron ORFs can be aligned with intron ORFs from both homologous and non-homologous genes, such as those belonging to the fifth intron of *Chlamydomonas eugametos* chloroplast *rnl* (CeLSU. 5) and to the intron of the *nad*3 gene of *Podospora anserina* mitochondria (PaND3) (Fig. 24).

The intron core regions of AcLSU.m1 and AcLSU.m2 are not closely related to one another, as AcLSU.m1 is a member of subgroup 1B4 whereas AcLSU.m2 belongs to subgroup 1A3. Neither AcLSU.m1 nor AcLSU.m2 display primary sequence or secondary structure similarity within the core region to PaND3, which is a subgroup 1C intron. The ORFs of AcLSU.m1 and AcLSU.m2 are both located within L8 of the intron core region, whereas that portion of the ORF of PaND3 containing the two dodecapeptides and alignable with the AcLSU.m1 and AcLSU.m2 ORFs is found within L9.2. (Note that the ORF of PaND3 is continuous with the upstream exon sequence.) Thus, whereas the intron ORFs of AcLSU.m1 and AcLSU.m2 appear to be related to one another and to the ORF of PaND3, the core regions of the three introns do not share this relationship.

D. Transfer RNA Editing

Two clusters of tRNA genes are found within the region of the *A. castellanii* mtDNA that I have sequenced. Cluster 1 is located immediately upstream and cluster 2, immediately downstream of *rnl* (Fig. 7). Each cluster consists of five closely linked tRNA genes: tRNAs specific for glutamine, lysine, glutamate, isoleucine and leucine comprise cluster 1, while tRNAs specific for methionine (two), alanine, proline and aspartate comprise cluster 2. A genetic map of the regions of *A. castellanii* mtDNA containing these tRNA clusters is shown in Fig. 25.

The potential secondary structures for the tRNAs of clusters 1 and 2 have been modeled from the gene sequences and these are illustrated in the next four figures (Figs. 26 - 29). The $tRNA^{Pro}$ has a standard cloverleaf secondary structure generated

Figure 24. Amino acid sequence alignment of AcLSU.m1 and AcLSU.m2 intron ORFs with those from: (A) *C. eugametos* chloroplast *rnl* (CeLSU.5); and (B and C) *Podospora anserina* mitochondrial *nad*3 (PaND3). Numbers in B and C refer to amino acid positions of the PaND3 intron ORF used for the alignment. The alignment in A was generated using the Clustal V Multiple Sequence Alignment Program whereas the alignments in B and C were generated from a GenBank search (FASTA; Pearson and Lipman, 1988). Dots above the sequences identify the dodecapeptide motifs. Identical amino acid residues shared between aligned sequences are enclosed within boxes.

A.

```
AcLSU.m1 MK-----------------------------------ID-----------
AcLSU.m2 MKNLDLKQE--------------------------KALID-----------
CeLSU.5  MSNFILKPGEKLPQDKLEELKKINDAVKKTKNFSKYLIDLRKLFQIDEVQ
                 ...........
AcLSU.m1 -----KNWIVGFVDGEGCFYIGINKSVDSKLGYQVLPEFRVVQHKR----
AcLSU.m2 -----PFWISGFVDGEGCFCISFNLKERLTLGIEVRPSFSISQTRDKEGL
CeLSU.5  VTSESKLFLAGFLEGEASLNISTKKLATSKFGLVVDPEFNVTQHVN----

AcLSU.m1 DIKVLYAIKDFFGHGSVVCNKSNGSEIYEYRVRKFETLHDVILPFFESNG
AcLSU.m2 NLRCLQDFLNFFDCGFIRFSKRDNTNKYEC--RDLSDIRSKVLPHFEKFT
CeLSU.5  GVKVLYLALEVFKTGRIR-HKSGSNATLVLTIDNRQSLEEKVIPFYEQYV

AcLSU.m1 L---LTSKKFNFLAFRDVILIMKRREHLTESGL-SKIIDI-------K--
AcLSU.m2 L---RIKKLRDFELFKDVVNSVASKQHLNEVGL-KRIIDISYQINMGK--
CeLSU.5  VAFSSPEKVKRVANFKALLELFNNDAHQDLEQLVNKILPIWDQMRKQQGQ

AcLSU.m1 -----SRMKRSST-------HY-
AcLSU.m2 -----RKLTKDELLSKINLKNL-
CeLSU.5  SNEGFPNLEAAQDFARNYKKGIK
```

B.

```
                                 ...........
AcLSU.m1                MKIDKNWIVGFVDGEGCFYIGINKSVDSKLGYQVLPEFR
PaND3    ...KRNFSTLESKLNPSYISGFVDGEGSFMLTIIKDNKYKLGWRVVCRFV
           132

AcLSU.m1 VVQHKRDIKVLYAIKDFFGHGSVVCNKSNGSEIYEYRVRKFETLHDVILP
PaND3    ISLHKKDLSLLNKIKEFFDVGNVFLMTKDSA---QYRVESLKGL-DLIIN

AcLSU.m1 FFESNGLLISKKFNFLAFRDVILIMKRREHLTESGLSKIIDIKSRMNRSS
PaND3    HFDKYPLIIKKQADYKLFKMAHNLIKNKSHLTKEGLLELVAIKAVINNGL

AcLSU.m1 IHY
PaND3    NNDLSIAFPGIN...
                   286
```

C.

```
                                         ...........
AcLSU.m2                MKNLDLKQEKALIDPFWISGFVDGEGCFCISFNLKERLT
PaND3    ...SIAFPGINTILRPDTSLPQILNPFWLSGFVDAEGCFSVVVFKSKTSK
           279

AcLSU.m2 LGIEVRPSFSISQTRDKEGLNLRCLQDFLNFFDCGFIRFSKRDNTNKYEC
PaND3    LGEAVKLSFILTQSNRDEYL----IKSLIEYLGCGNTSLDPRGTID-FKV

AcLSU.m2 RDLSDIRSKVLPHFEKFTLRTKKLRDFELFKDVVNSVASKQHLNEVGLKR
PaND3    TNFSSIKDIIVPFFIKYPLKGNKNLDFTDFCEVVRLMENKSHLTKEGLDQ

AcLSU.m2 IIDISYQINMGKRKLTKDELLSKINLKNL
PaND3    IKKIRNRMNTNRK
                  433
```

Figure 24

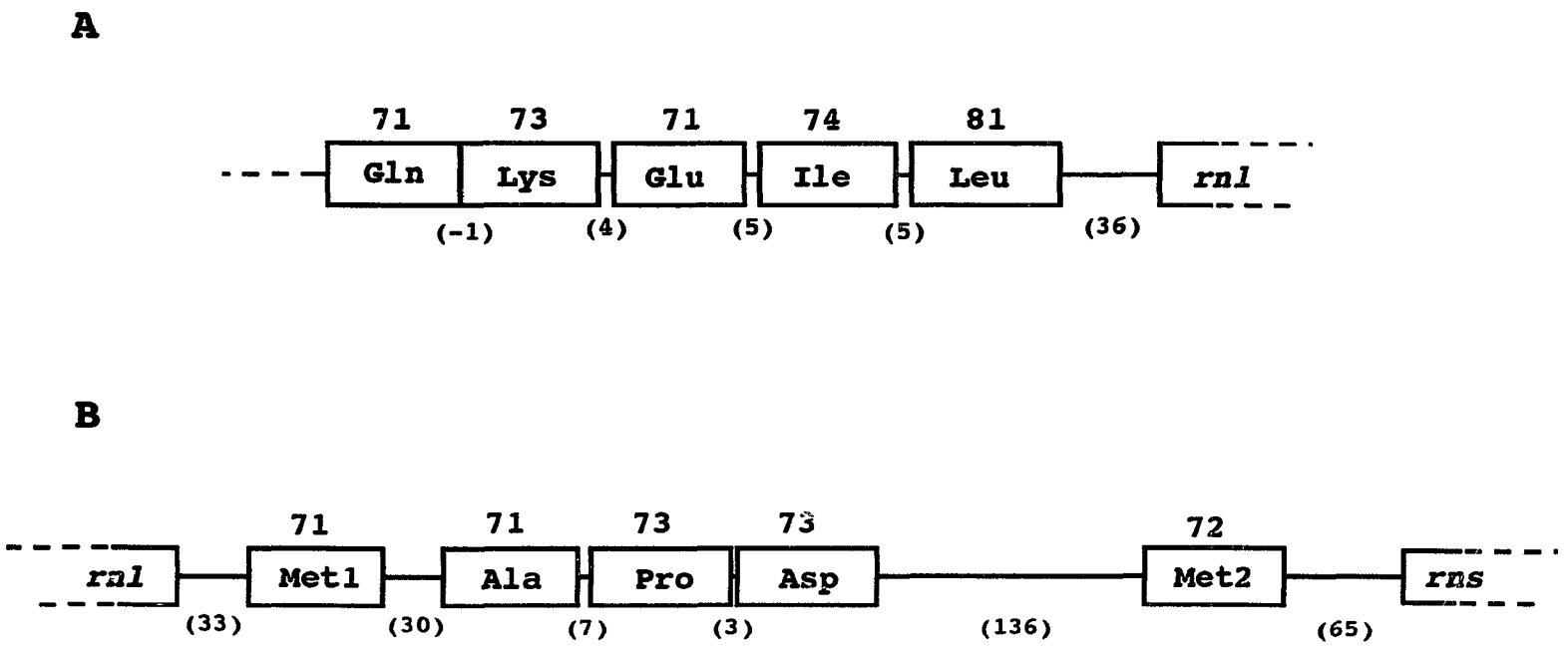


Figure 25. Organization of tRNA gene cluster 1 (A) and tRNA gene cluster 2 (B) in *A. castellanii* mtDNA. Gene identification is based upon anticodon sequence. Numbers above the boxes indicate the gene lengths, while numbers beneath and in parentheses indicate lengths of intergenic regions. All genes are in the same transcriptional orientation (5' to 3', left to right).

by intramolecular base pairings. These interactions result in the formation of the anticodon stem and loop, the DHU stem and loop, the TΨC stem and loop, and the acceptor stem (Fig. 26).

The primary sequence for $tRNA^{Asp}$, as derived from the gene sequence, has the potential to form all but one of the standard helices of tRNA secondary structure. The acceptor stem, which typically consists of a 7-bp helix, contains three mismatched base pairs localized to the first three positions. Lack of base pairing at these positions would result in a 4-bp acceptor stem helix (Fig. 27). The derived secondary structures for the remaining three tRNAs of cluster 2, $tRNA^{Met1}$, $tRNA^{Ala}$ and $tRNA^{Met2}$, show a similar clustering of one or two mismatches localized to the first three positions of the acceptor stem (Fig. 28). According to the potential secondary structures, four of the five tRNAs of cluster 1 display a similar distribution of mismatches within the acceptor stems as that observed for the tRNAs of cluster 2 (Fig. 29).

Considering the importance of the acceptor stem in the normal functioning of tRNA (as described briefly in the introduction), experiments were performed to determine whether or not these structurally aberrant tRNA genes were expressed. Expression of the tRNA genes of cluster 2 was analyzed by northern hybridization. Using oligonucleotide probes specific for an internal region of each tRNA, northern hybridization detected a tRNA-sized species for each of the tRNAs of cluster 2 (Fig. 30). These experiments indicate that the five tRNA genes of cluster 2 are expressed and produce stable, tRNA-sized transcripts.

To determine whether these tRNA transcripts have the same sequence as that specified by the genes that encode them, the oligonucleotide probes from the northern experiments were used as primers for dideoxy sequencing (Fig. 31). As the oligonucleotides are complementary to an internal region of the tRNAs, sequence information is restricted to the 5'-half of each tRNA. For the four tRNAs of cluster 2 shown in Fig. 31, either one or two differences in the cDNA sequence relative to the

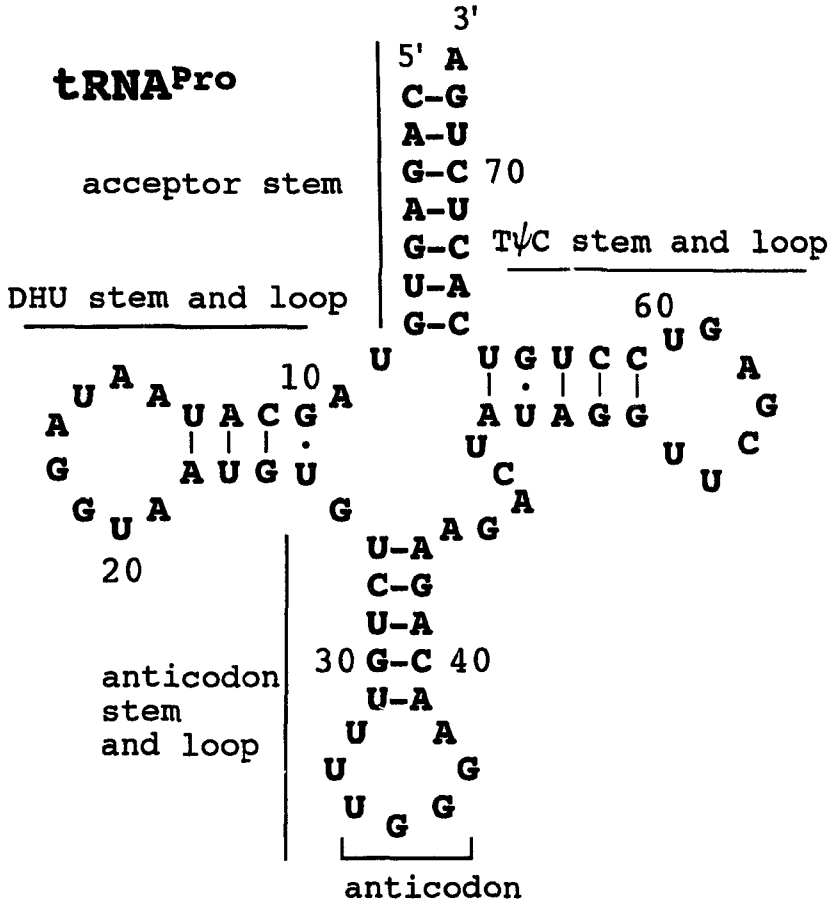


Figure 26. Potential secondary structure of *A. castellanii* mitochondrial proline tRNA as derived from the gene sequence. The 5' and 3' termini are indicated. Numbering of nucleotides is according to a standard convention (Sprinzl et al., 1989). Conserved secondary structural features of a conventional tRNA are indicated.

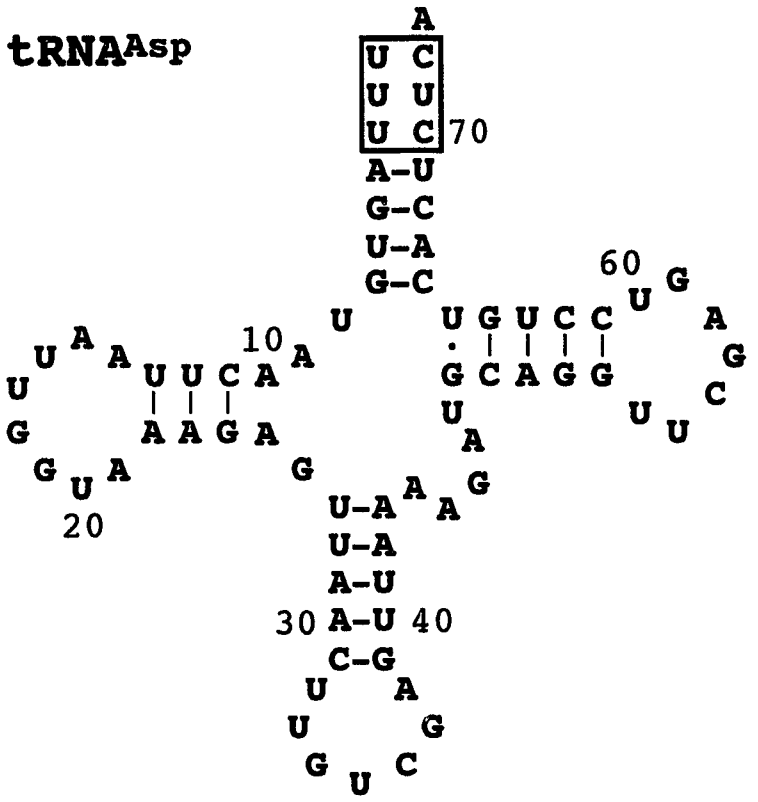


Figure 27. Potential secondary structure of *A. castellanii* mitochondrial aspartate tRNA as derived from the gene sequence. The positions of mismatch within the aminoacyl stem are indicated.

Figure 28. Potential secondary structures of cluster-2 tRNAs encoded by *A. castellanii* mtDNA, as derived from the gene sequence. The RNA-like strand is shown. Numbering of nucleotides is according to Sprinzl et al. (1989).

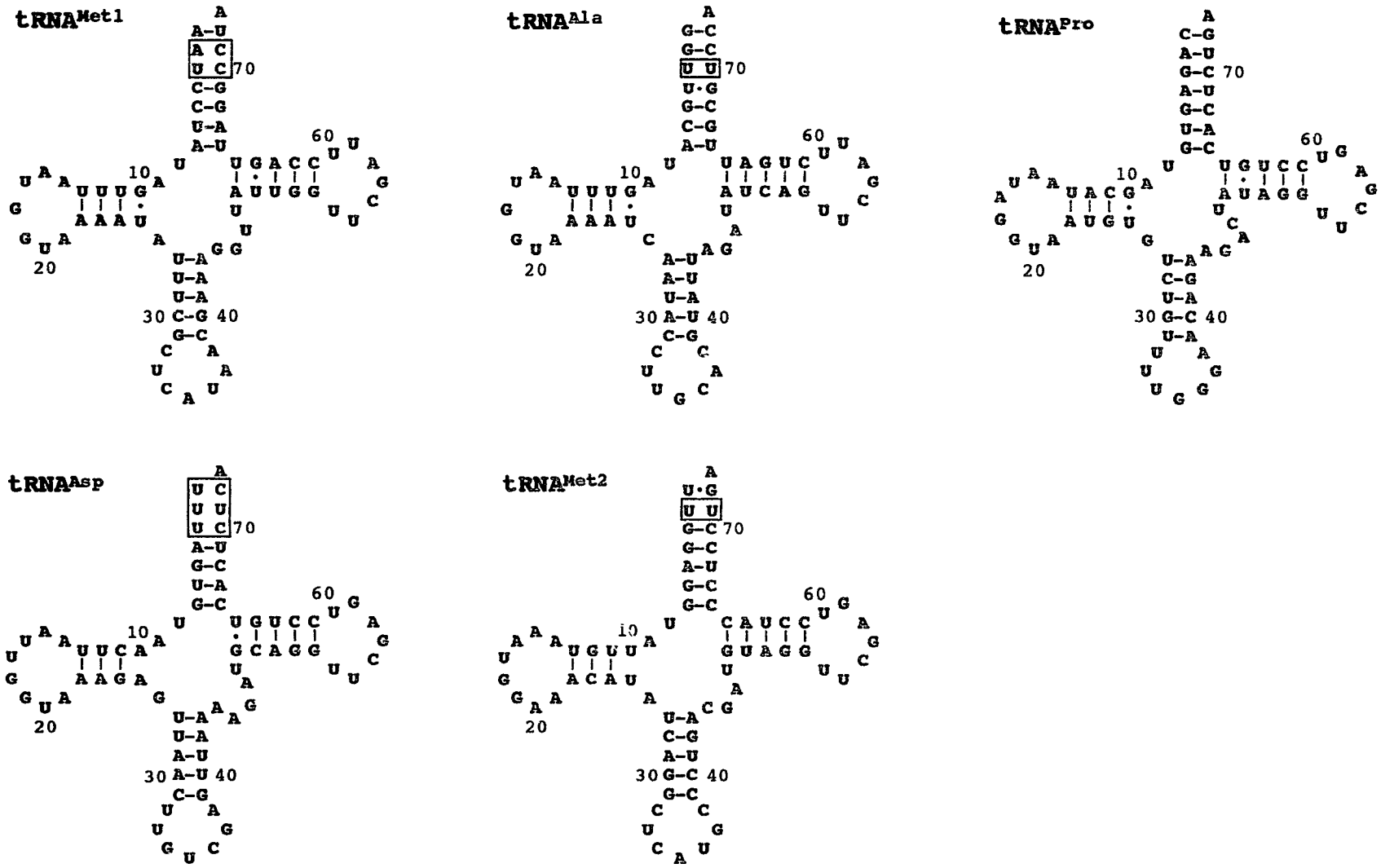
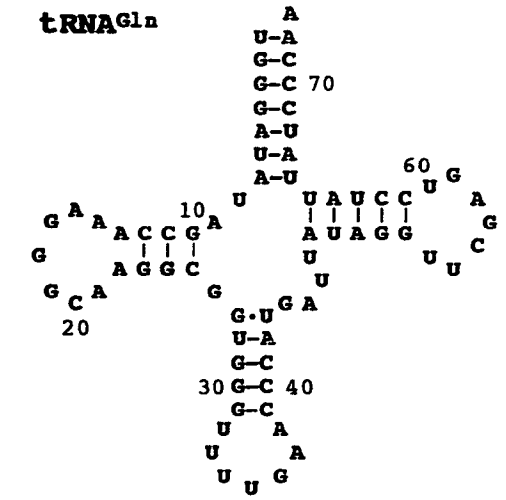
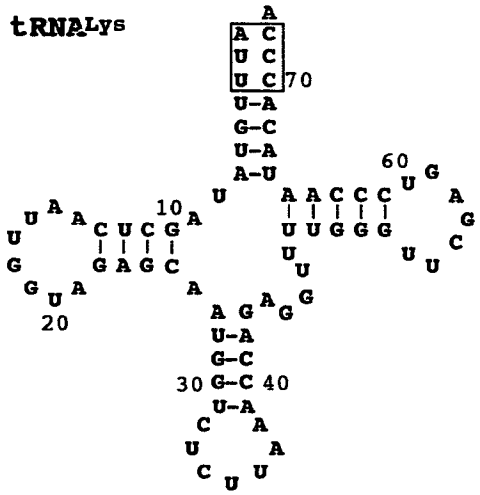
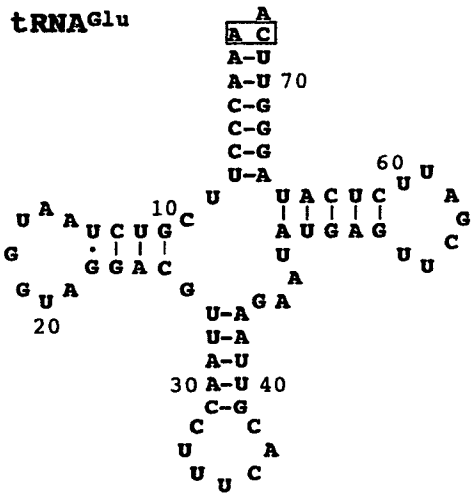


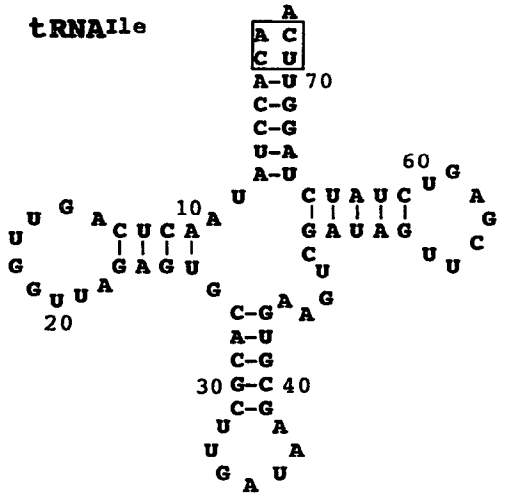
Figure 28

Figure 29. Potential secondary structures of *A. castellanii* mitochondrial cluster-1 tRNAs as derived from the gene sequence. The RNA-like strand is shown. Numbering of nucleotides is according to Sprinzl et al. (1989).









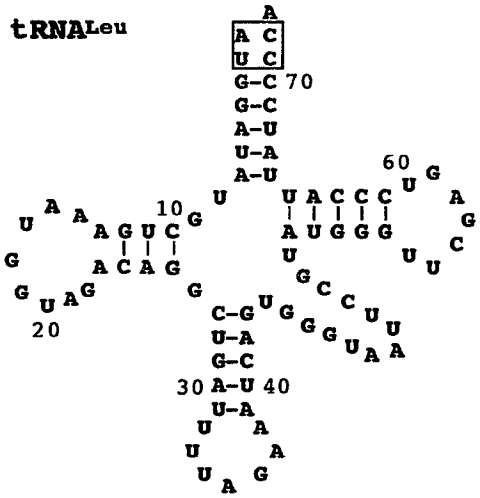


Figure 29

Figure 30. Autoradiograms showing the results of northern hybridization analysis of cluster-2 tRNA gene products using gene-specific oligonucleotide probes. A northern blot containing electrophoretically resolved mtRNA was probed with 5'-^{32}P-end-labeled oligonucleotides specific for either $tRNA^{Met1}$ (lane 1), $tRNA^{Ala}$ (lane 2), $tRNA^{Pro}$ (lane 3), $tRNA^{Asp}$ (lane 4), or $tRNA^{Met2}$ (lane 5). Size markers were the individual components of the cytosolic LSU rRNA of *Euglena gracilis* (Schnare and Gray, 1990); lengths (nt) and positions of migration are indicated to the left of the figure.

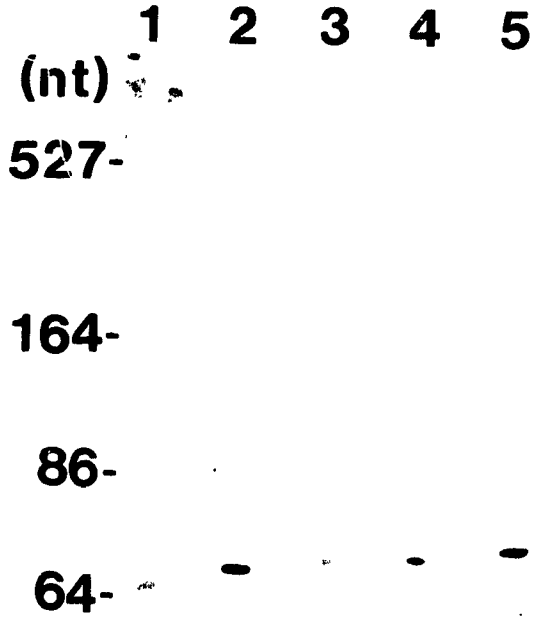


Figure 30

Figure 31. Comparison of tRNA gene sequence and tRNA mature transcript sequence. Autoradiograms of sequencing gels display DNA and RNA (cDNA) sequences for the 5'-terminal region of each gene and tRNA, plus a primer extension lane (E). Gene-specific oligonucleotide sequencing primers were (A) $tRNA^{Met1}$, (B) $tRNA^{Ala}$, (C) $tRNA^{Asp}$, and (D) $tRNA^{Met2}$. The nucleotide sequence of the RNA-like strand of the DNA and the nucleotide sequence of the RNA (the complements of the sequences actually determined) are shown, reading 3' to 5' to facilitate direct comparison with the secondary structure diagrams. The major band within lane E represents the 5'-terminal nucleotide of the tRNA. Arrowheads indicate differences between the sequences of each tRNA gene and its mature product.

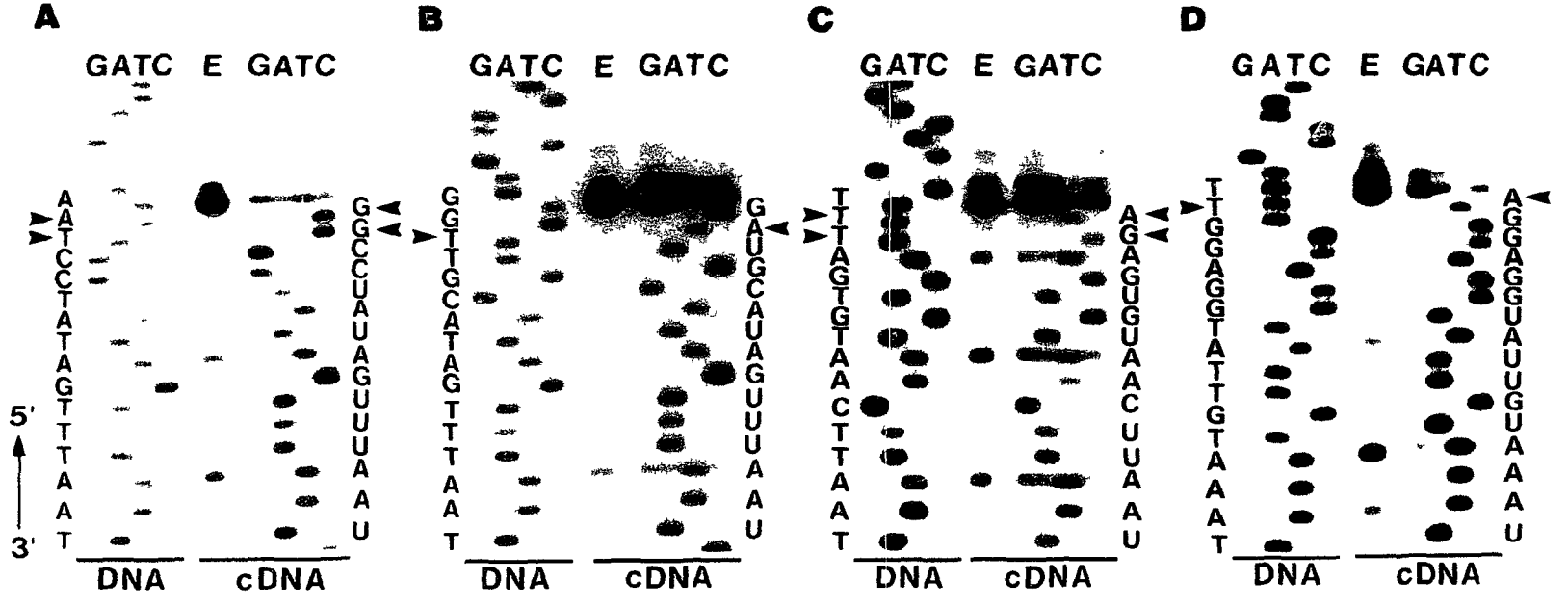

A
GATC E GATC
AATCCTATAGTTTAAT
5'
3'
GGCCUAUAGUUUAAU
DNA
cDNA
B
GATC E GATC
GGTTGCATAGTTTAAT
GAUGCAUAGUUUAAU
DNA
cDNA
C
GATC E GATC
TTTAGTGTAACTTAAT
AGAGUGUAACUUAAU
DNA
cDNA
D
GATC E GATC
TTGGAGGTATTGTAAAT
AGGAGGUAUUGUAAAU
DNA
cDNA

Figure 31

DNA sequence were observed. For example, TT in the RNA-like strand of the $tRNA^{Asp}$ gene reads as AG instead of UU in the RNA. These differences in the $tRNA^{Asp}$ primary sequence occur precisely at the positions of mismatch within the acceptor stem of the secondary structure (Fig. 32). The replacement of UU by AG results in complementary base pairing with the 3'-half of the acceptor stem at these positions.

For $tRNA^{Met1}$, AT in the RNA-like strand of the gene reads as GG in the RNA; for $tRNA^{Ala}$, T in the RNA-like strand of the DNA reads as A in the RNA; and for $tRNA^{Met2}$, T in the RNA-like strand of the gene reads as A in the RNA (Fig. 31). Once again, these differences in primary sequence between the transcript and gene occur only at the positions of mismatch within the acceptor stems of the secondary structure; in all cases, they allow standard base pairing with the 3'-half of the acceptor stem at these positions (Fig. 33). The $tRNA^{Pro}$ gene sequence predicts a 7-bp acceptor stem and in this case, no differences were found between the cDNA sequence of the transcript and the DNA sequence of the gene.

Before defining the sequence differences reported here as RNA editing, it is important to be absolutely certain that there are no genes within the cell that correspond exactly in primary sequence to the altered transcripts. Therefore, control experiments were performed to verify that the cDNA sequences from these experiments were generated from the transcripts of the mitochondrial tRNA genes of cluster 2, and do not represent the transcripts of nucleus-encoded tRNA genes or mitochondrial transcripts encoded elsewhere within the mitochondrial genome. For this purpose, the oligonucleotide primers from the sequencing reactions were used as probes for Southern hybridization analysis of nuclear and mitochondrial DNAs (Figs. 34 and 35). *A. castellanii* mtDNA is cleaved by *Bam*HI into four fragments (Fig. 7) and the $tRNA^{Ala}$-specific probe hybridized only to the completely sequenced 6.7 kbp fragment that encodes the five tRNA genes of cluster 2 (Fig. 34A, B). No other hybridizing

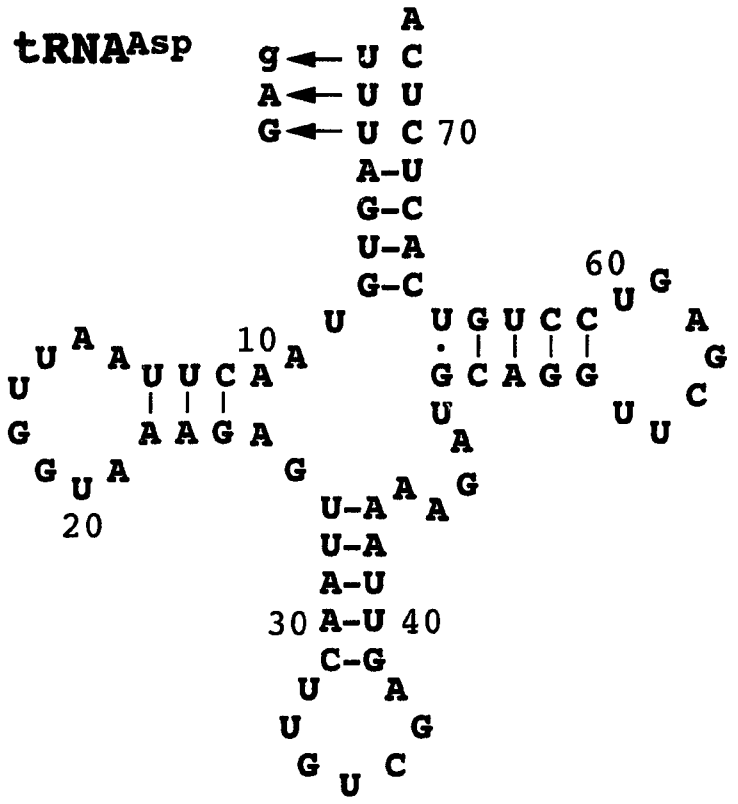


Figure 32. Potential secondary structure of *A. castellanii* mitochondrial aspartate tRNA showing deduced editing sites. Arrows indicate nucleotide changes induced by editing. The lowercase G indicates that editing at this position is by inference only due to the difficulties in determining the 5'-terminal nucleotide by dideoxy sequencing.

Figure 33. Potential secondary structures of *A. castellanii* mitochondrial cluster-2 tRNAs as derived from the gene sequence, showing changes induced by editing. The RNA-like strand is shown. Numbering of nucleotides is according to Sprinzl et al. (1989). Editing events are indicated by the arrows. The predicted editing of the 5'-terminal nucleotide of $tRNA^{Asp}$ is indicated by the lowercase G.

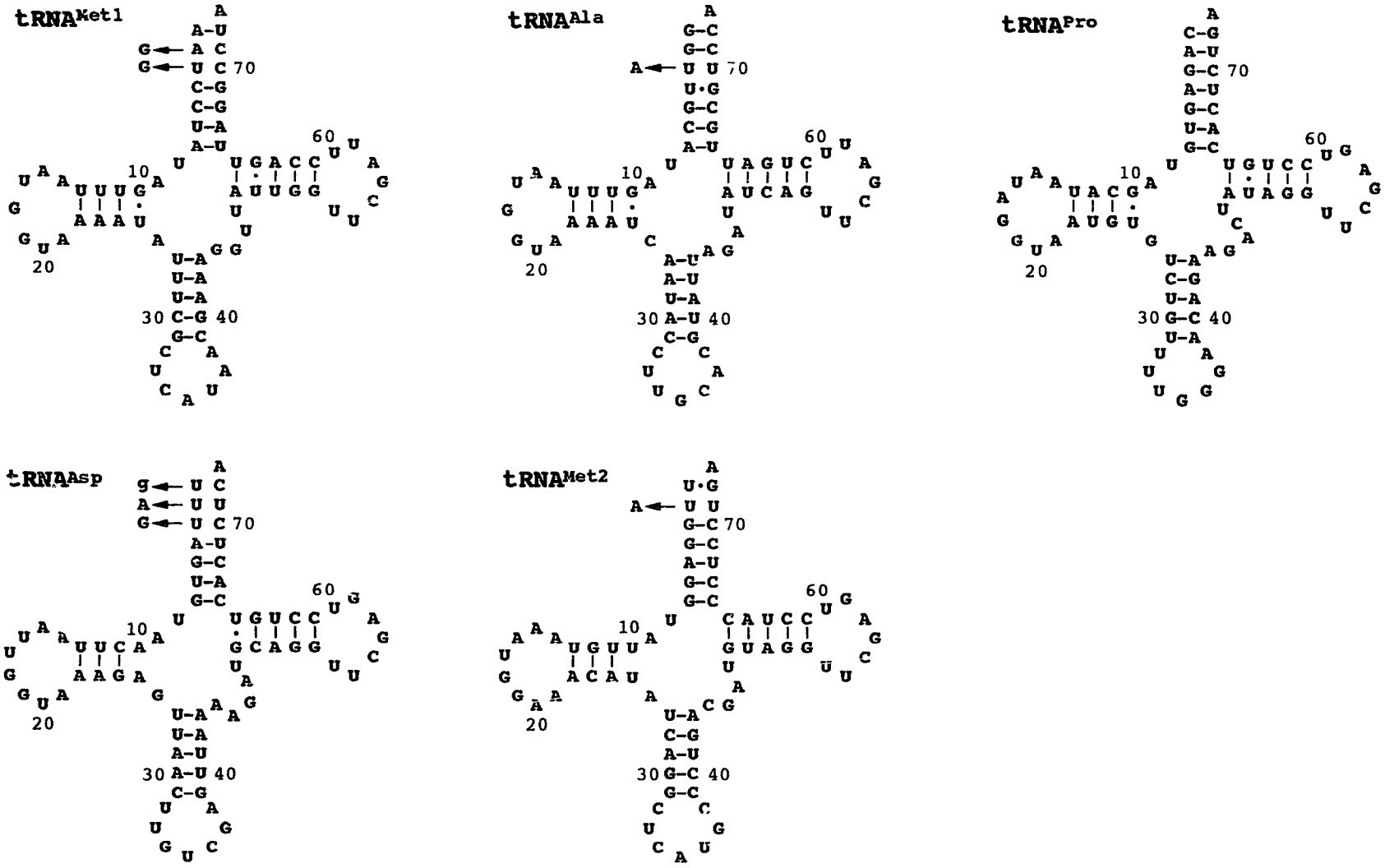


Figure 33

Figure 34. Southern hybridization analysis of the *A. castellanii* mitochondrial $tRNA^{Ala}$ gene of cluster 2 using a gene-specific oligonucleotide probe. A Southern blot containing electrophoretically separated *Bam*HI (B), *Mlu*I (M) and *Hind*III (H) digests of nDNA and mtDNA was probed sequentially with $5'$-^{32}P-end-labeled oligonucleotides specific for either the mitochondrial $tRNA^{Ala}$ gene (A and B) or the nuclear myosin IB (MIB) heavy chain gene (C) and the autoradiograms are shown. The blot in (B) is the same as that in (A) except that (B) was exposed threefold longer to permit detection of any weak nDNA hybridization signals. The migration positions of selected *Hind*III marker fragments of lambda DNA are indicated.

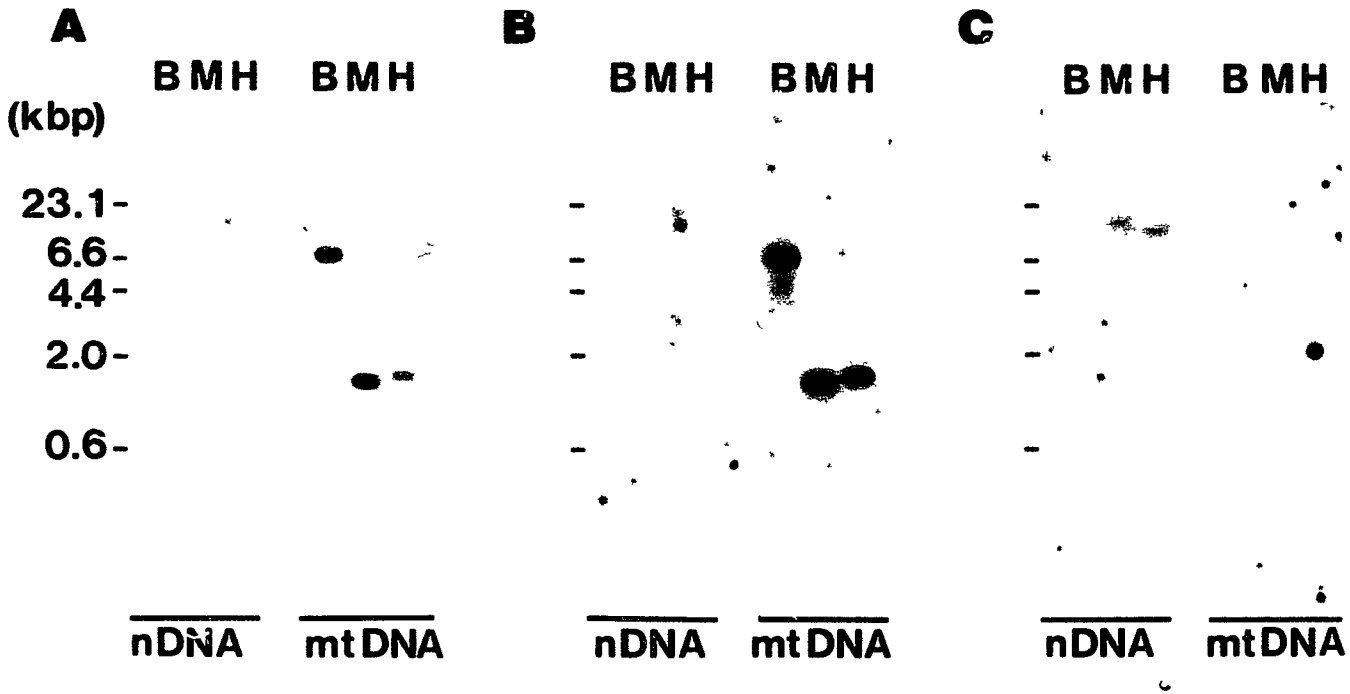

A
B
C
BMH
BMH
BMH
BMH
BMH
BMH
(kbp)
23.1
6.6
4.4
2.0
0.6
nDNA
mtDNA
nDNA
mtDNA
nDNA
mtDNA

Figure 34

species besides those expected were detected in *Mlu*I and *Hind*III digests of mtDNA (Fig. 34A, B). Only minor hybridization signals were detected in the nDNA and these could be attributed to residual trace contamination with mtDNA (Fig. 34B). The stripped blot was reprobed with an oligonucleotide specific for the myosin IB gene of *A. castellanii* nDNA; this experiment served as a positive control for the detection of a nuclear gene (Fig. 34C). These same experiments were repeated using oligonucleotide probes specific for the other three tRNAs of cluster 2 that showed signs of editing, with similar results (Fig. 35). Thus, the only genes corresponding to the tRNA transcripts sequenced in this study are those found within tRNA gene cluster 2 of the mitochondrial genome.

In summary, of the ten *A. castellanii* mitochondrial tRNA genes sequenced in this study, it seems likely that the transcripts of eight of these are altered by editing (Fig. 36).

There is evidence to suggest that the editing observed here occurs posttranscriptionally rather than cotranscriptionally. Given their proximity, it is likely that the tRNA genes of cluster 2 are co-transcribed and that the mature tRNAs are processed from an initial precursor; if so, such a precursor might be expected to be present in small amounts in the steady-state RNA population. The oligonucleotides used for the priming of cDNA synthesis initiated from within a mature tRNA also allow one to read the sequence of precursor RNA (after the removal of any contaminating mtDNA by DNase treatment). Because the tRNA genes are closely spaced, it is possible to read a sequence ladder that extends beyond the 5'-terminal nucleotide of a mature tRNA into and through the unprocessed tRNA precursor immediately upstream. To read the precursor $tRNA^{Met1}$ sequence, the $tRNA^{Ala}$-specific oligonucleotide was used as a sequencing primer and to read the precursor $tRNA^{Ala}$ sequence, the $tRNA^{Asp}$-specific oligonucleotide was used in the same way. For the precursor-$tRNA^{Met1}$ and precursor-$tRNA^{Ala}$, comparison of the cDNA and DNA sequences

Figure 35. Southern hybridization analysis of Met1, Asp and Met2 tRNA genes of *A. castellanii* mtDNA cluster 2 using gene-specific oligonucleotide probes. A Southern blot containing electrophoretically separated *Bam*HI (B), *Mlu*I (M), and *Hind*III (H) digests of nDNA and mtDNA was probed sequentially with $5'$-^{32}P-end-labeled oligonucleotides specific for $tRNA^{Met1}$ (A), $tRNA^{Asp}$ (B), or $tRNA^{Met2}$ (C) and the autoradiograms are shown. The migration positions of selected *Hind*III marker fragments of lambda DNA are indicated.

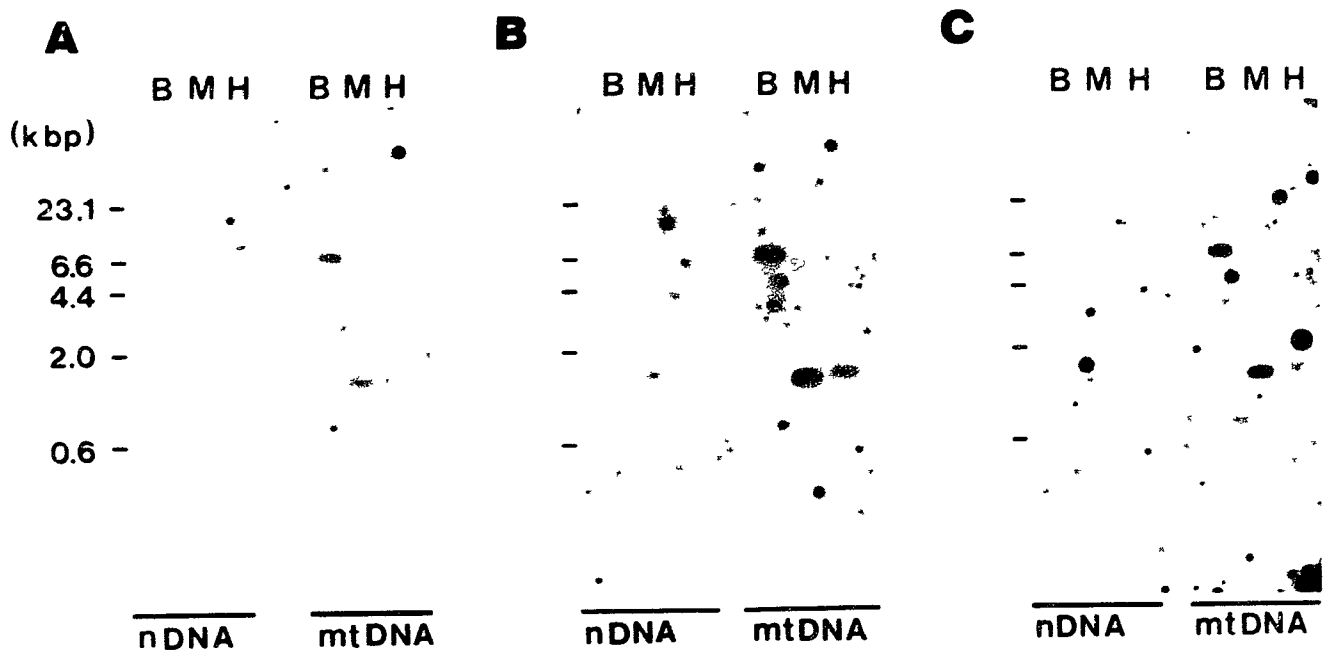

A
B M H
B M H
(kbp)
23.1
6.6
4.4
2.0
0.6
nDNA
mtDNA
B
B M H
B M H
nDNA
mtDNA
C
B M H
B M H
nDNA
mtDNA

Figure 35

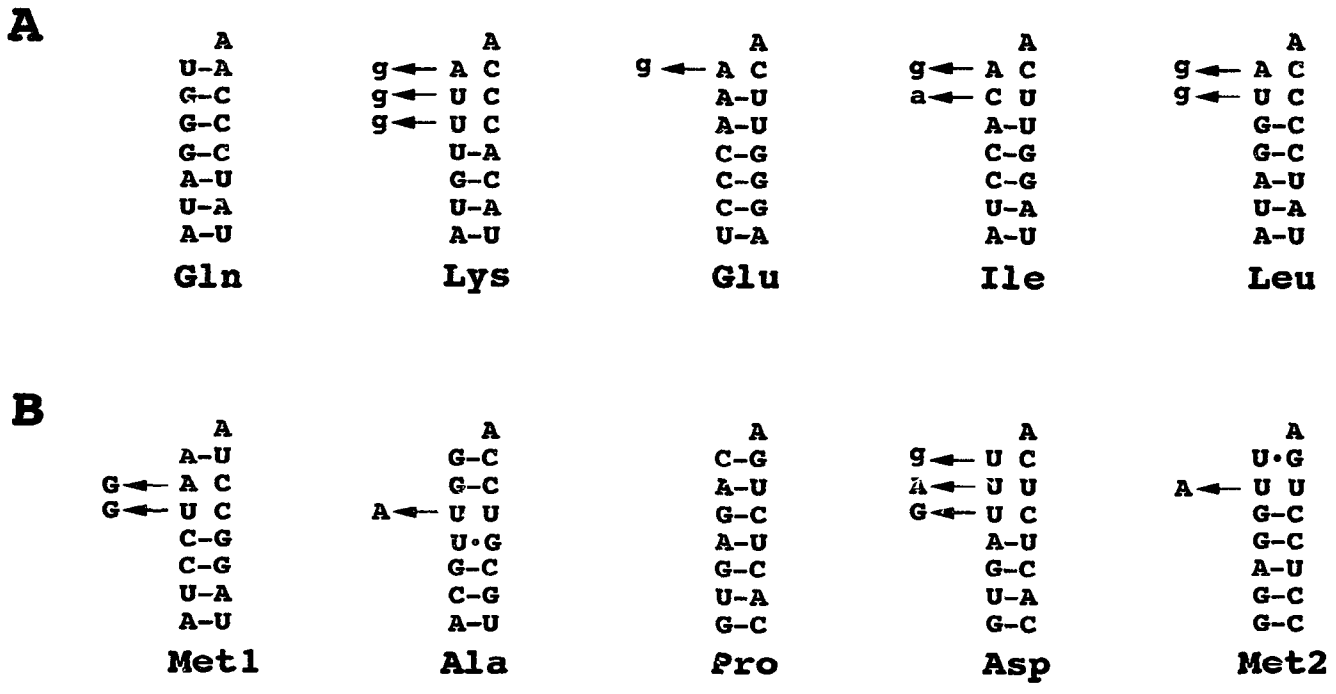


Figure 36. Summary of confirmed (uppercase) and predicted (lowercase) editing within the acceptor stems of tRNAs of cluster 1 (A) and cluster 2 (B) in *A. castellanii* mitochondria.

revealed no differences: i.e., positions that are edited in the mature tRNA appear not to be edited in the unprocessed precursor (Fig. 37).

E. Cytochrome Oxidase

Genes potentially encoding two subunits of cytochrome oxidase of complex IV of the mitochondrial respiratory chain, *cox*1 and *cox*2, are found within the region of *A. castellanii* mtDNA sequenced for this project (Fig. 7). Gene identification is based on comparative analysis of the derived amino acid sequences for *cox*1 (Fig. 38) and for *cox*2 (Fig. 39). These protein-coding genes of *A. castellanii* mitochondria follow the universal genetic code with the exception that TGA encodes tryptophan, as observed for almost all mitochondrial genomes, with the notable exception of land plants (Fox, 1987; Gray, 1992; Oda et al., 1992). Analysis of codon usage indicates that codons with A or especially U at the third position are used preferentially over those ending in G or C (Fig. 40). The *cox*1 gene potentially encodes a ~519 aa protein (56,474 Da) while the *cox*2 gene potentially encodes a 333 aa product (38,120 Da). It is noted that there are two regions (R) within the putative *cox*2 protein (aa coordinates: R1, 70 - 107; R2, 170 - 211) that are not present in the homologs from other sources (Fig. 39).

The size of the *cox*1 coding region is approximate due to the apparent absence of a translation termination codon between the 3'-coding region of the *cox*1 gene and the predicted ATG initiation codon of the downstream *cox*2 coding region (Fig. 8). From transcript studies, it was not possible to map a 5' terminus for the *cox*1 transcript, which appears to include an undefined portion of the upstream SSU rRNA (Fig. 41). Similarly, it was not possible to determine a 5' terminus for *cox*2 transcript, which apparently includes the 3'-terminal region of the upstream *cox*1 coding region (Fig. 42). These preliminary results on expression of the *cox*1 - *cox*2 gene region of *A. castellanii* mtDNA need to be extended with Southern and northern hybridization experiments in order to better define the size(s) of the *cox*1 - *cox*2 transcript(s), and to

Figure 37. Comparison of $tRNA^{Met1}$ and $tRNA^{Ala}$ gene sequences and the corresponding tRNA precursor sequences. Autoradiograms of sequencing gels display DNA and RNA (cDNA) sequences for selected portions of these genes and their pre-tRNAs, plus a primer extension lane (E). (A) Pre-$tRNA^{Met1}$, (B) pre-$tRNA^{Ala}$. The nucleotide sequence of the RNA-like strand of the DNA and the nucleotide sequence of the RNA (the complements of the sequences actually determined) are shown, reading in the 3' to 5' direction to facilitate direct comparison with the secondary structure diagrams. Arrowheads indicate positions of editing that occur in the mature tRNA transcripts.

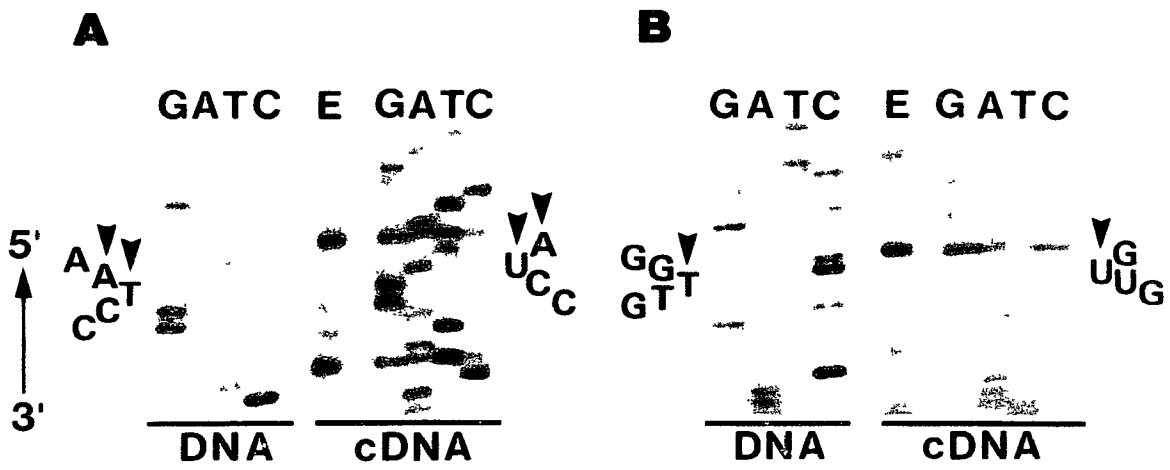

A
GATC E GATC
5'
3'
A A T
C C
A
U C C
DNA
cDNA
B
GATC E GATC
G G T
G T
G
U U G
DNA
cDNA

Figure 37

Figure 38. Amino acid sequence alignment of *A. castellanii* mitochondrial *cox*1, as inferred from the gene sequence. The *A. castellanii* sequence used for the alignment includes the first AUG codon downstream of the SSU rRNA coding region (assumed to be the translation initiation codon of the *cox*1 transcript) to the proposed translation initiation codon of the *cox*2 coding region. Sequences were aligned using the Clustal V Multiple Sequence Alignment program; asterisks indicate absolutely conserved amino acid residues while dots represent functionally conserved amino acid residues among all sequences. Abbreviations and references including GenBank accession numbers: A. castellanii, *Acanthamoeba castellanii* (this thesis); B. japonicum, *Bradyrhizobium japonicum* (X54800, Bott, 1990); R. sphaeroides, *Rhodobacter sphaeroides* (X62645, Shapleigh and Gennis, 1991); T. aestivum, *Triticum aestivum* (wheat) (Y00417, Bonen et al., 1987); D. melanogaster, *Drosophila melanogaster* (M57910, Satta and Takahata, 1990); N. crassa, *Neurospora crassa* (X14669, Field et al., 1989); S. cerevisiae, *Saccharomyces cerevisiae* (J01481, Bonitz et al., 1980).

```
                         |         |         |         |         |
A. castellanii     MINRLLNNLTSFFTDNR--------W---LFSTNHKDIGTLYLIFGGFSG   39
B. japonicum       MATSAAAHGDHAQDHGHDEHAHPTGWRRYVYSTNHKDIGTMYLIFAVIAG   50
R. sphaeroides     MAD------AAIHGHEHDRRGFFTRW---FMSTNHKDIGVLYLFTGGLVG   41
T. aestivum        MTNMV-----------R--------W---LFSTNHKDIGTLYFIFGAIAG   28
D. melanogaster    MSR-------------Q--------W---LFSTNHKDIGTLYFIFGAWAG   26
N. crassa          MSSI------SIWTE-R--------W---FLSTNAKDIGVLYLIFALFSG   32
S. cerevisiae      MVQ------------------------RXLYSTNAKDIAVLYFMLAIFSG   26
                   *                              *** ***. .*.  .   *

                         |         |         |         |         |
A. castellanii     IIGTIFSMIIRLELAAPGSQILSGNS------------------------   65
B. japonicum       VIGAAMSIAIRAELMYPGVQIFH-ET------------------------   75
R. sphaeroides     LISVAFTVYMRMELMAPGVQFMCAEHLESGLVKGFFQSLWPSAVENCTPN   91
T. aestivum        VMGTCFSVLIRMELARPGDQI-GGNH------------------------   54
D. melanogaster    MVGTSLSILIRAELGHPGA--_IGDD------------------------   50
N. crassa          LLGTAFSVLIRMELSGPGVQYIA-DN------------------------   57
S. cerevisiae      MAGTAMSLIIRLELAAPGSQYLHGNS------------------------   52
                   . .   .. .* **  **      .

                         |         |         |         |         |
A. castellanii     -QLYNVIITAHAFVMIF-FFVMPVMIGGFGNWFVPLMIGAPDMAFPRLNN   113
B. japonicum       -HTYNVFVTSHGLIMIF-FMVMPAMIGGFGNWFVPLMIGAPDMAFPRMNN   123
R. sphaeroides     GHLWNVMITGHGILMMF-FVVIPALFGGFGNYFMPLHIGAPDMAFPRMNN   140
T. aestivum        -QLYNVLITAHAFLMIF-FMVMPAMIGGFGNWFVPILIGAPDMAFPRLNN   102
D. melanogaster    -QIYNVIVTAHAFIMIF-FMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNN   98
N. crassa          -QLYNAIITAHAILMIF-FMVMPALIGGFGNFLLPLLVGGPDMAFPRLNN   105
S. cerevisiae      -QLFNLVV-GHAVLMIFCTLVMPALIGGFGNYLLPLIIGATDTAFPRINN   100
                    .  *  . .*. .*.*   *.* ..***** ..*. .*. * ****.**

                         |         |         |         |         |
A. castellanii     ISFWLLPPSLFLLLCSSLVEF-----GAGTGWTVYPPLSSIVAHSGGSVD   158
B. japonicum       ISFWLLPASFGLLLMSTFVEGEPGANGVGAGWTMYVPLSSS-GHPGPAVD   172
R. sphaeroides     LSYWLYVAGTSLAVASLFAPGGNGQLGSGIGWVLYPPLS--TSESGYSTD   188
T. aestivum        ISFWLLPPSLLLLLSSALVEV-----GSGTGWTVYPPLSGITSHSGGAVD   147
D. melanogaster    MSFWLLPPALSLLLVSSMVEN-----GAGTGWTVYPPLSAGIAHGGASVD   143
N. crassa          ISFWLLPPSLLLLVFSACIEG-----GAGTGWTIYPPLSGVQSHSGPSVD   150
S. cerevisiae      IAFXVLPMGLVCLVTSTLVE-----SGAGTGXTVYPPLSSIQAHSGPSVD   145
                   ... .   .    . *          * * *  .* ***   .. * . *

                         |         |         |         |         |
A. castellanii     LAIFSLHLAGISSLLGAINFITTIFNMRVPGLSMHKLPLFVWSVLITAFL   208
B. japonicum       FAILSLHLAGASSILGAINFITTIFNMRAPGMTLHKMPLFVWSILVTVFL   222
R. sphaeroides     LAIFAVHLSGAESILGAINMITTFLNMRAPGMTMHKVPLFAWSIFVTAWL   238
T. aestivum        LAIFSLHLSGISSILGSINFITTIFNMRGPGMTMHRLPLFVWSVLVTAFL   197
D. melanogaster    LAIFSLHLAGISSILGAVNFITTVINMRSTGISLDRMPLFVWSVVITALL   193
N. crassa          LAIFALHLSGVSSLLGSINFITTIVNMRTPGIRLHKLALFGWAVVITAVL   200
S. cerevisiae      LAIFALHLTSISSLLGAINFIVTTLNMRTNGMTMHKLPLFVXSIFITAFL   195
                   .**...**.. **.**..* * *  ***  *. .....**  .. .*  *

                         |         |         |         |         |
A. castellanii     LLFSLPVLAGAITMLLTDRNFNTSFFDPSGGGDPILYQHLFWFFGHPEVY   258
B. japonicum       LLLSLPVLAGAITMLLTDRNFGTTFFAPDGGGDPVLFQHLFWFFGHPEVY   272
R. sphaeroides     ILLALPVLAGAITMLLTDRNFGTTFFQPSGGGDPVLYQHILWFFGHPEVY   238
T. aestivum        LLLSLPVLAGAITMLLTDRNFNTTFFDPAGGGDPILYQHLFWFFGHPEVY   247
D. melanogaster    LLLSLPVLAGAITMLLTDRNLNTSFFDPAGGGDPILYQHLFWFFGHPEVY   243
N. crassa          LLLSLPVLAGAITMLLTDRNFNTSFFETAGGGDPILFQHLFWFFGHPEVY   250
S. cerevisiae      LLLSLPVLSAGITMLLLDRNFNTSFFEVAGGGDPILYEHLFXFFGHPEVY   245
                   .*..****...***** ***. *.**   *****.*..*.. ********
```

Figure 38 continues.

```
                          |         |         |         |         |
A. castellanii      ILILPAFGIVSQIIG-TFSNKSIFGYIGMVYAMLPIAVLGFIVWAHHMYT   307
B. japonicum        ILILPGFGMISQIVS-TFSRKPVFGYLGMAYAMVAIGGIGFVVWAHHMYT   321
R. sphaeroides      IIVLPAFGIVSHVIA-TFAKKPIFGYLPMVYAMVAIGVLGFVVWAHHMYT   337
T. aestivum         ILILPGFGIISHIVS-TFSRKPVFGYLGMVYAMISIGVLGFLVWAHHMFT   296
D. melanogaster     ILILPGFGMISHIISQESGKKETFGSLGMIYAMLAIGLLGFIVWAHHMFT   293
N. crassa           ILIIPGFGIISTTIS-AYSNKSVFGYIGMVYAMMSIGILGFIVWSHHMYT   299
S. cerevisiae       ILIIPGFGIISHVVS-TYSKKPVFGEISMVYAMASIGLLGFLVXSHHMYI   294
                    *...*.**..*  ..   . *  ** . * *** .*. .**.* .***.

                          |         |         |         |         |
A. castellanii      VGLDVDTRAYFTAATMMIAVPTGIKIFSWIATLWGGQIVRKTPLLFVIGF   357
B. japonicum        VGMSSATQAYFVAATMVIAVPTGVKIFSWIATMWGGSIEFRAPMIWAVGF   371
R. sphaeroides      AGLSLTQQSYFMMATMVIAVPTGIKIFSWIATMWGGSIELKTPMLWALGF   387
T. aestivum         VGLDVDTRAYFTAATMIIAVPTGIKIFSWIATMWGGSIQYKTPMLFAVGF   346
D. melanogaster     VGMDVDTRAYFTSATMIIAVPTGIKIFSWLATLHGTQLSYSPAILWALGF   343
N. crassa           VGLDVDTRAYFTAATLIIAVPTGIKIFSWLATCYGGSIRLTPSMLFALGF   349
S. cerevisiae       VGLDADTRAYFTSATMIIAIPTGIKIFSXLATIYGGSIRLATPMLYAIAF   344
                     *.    ..**  **..**.***.**** .**  *  .    ...  ..*

                          |         |         |         |         |
A. castellanii      LILFTLGGLTGIVLSNAGLDIMLHDTYYVVAHFHYVLSMGAVFAFFAGFY   407
B. japonicum        IFLFTVGGVTGVVLANAGVDRVLQETYYVVAHFHYVLSIGAVFAIFAGWY   421
R. sphaeroides      LFLFTVGGVTGIVLSQASVDRYYHDTYYVVAHFHYVMSIGAVFGIFAGST   437
T. aestivum         IFLFTIGGLTGIVLANSGLDIALHDTYYVVAHFHYVLSMGAVFALFAGFY   396
D. melanogaster     VFLFTVGGLTGVVLANSSVDIILHDTYYVVAHFHYVLSMGAVFAIMAGFI   393
N. crassa           VFMFTIGGLSGVVLANASLDIAFHDTYYVVAHFHYVLSMGAVFAMFSGWY   399
S. cerevisiae       LFLFTMGGLTGVALANASLDVAFHDTYYVVGHFYYVLSMGAIFSLFAGYY   394
                    ...**.**..*. *.....*   ..*****.** **.*.**.*.  .*

                          |         |         |         |         |
A. castellanii      YWFWKISGYTYNEMYGNVHFWLMFIGVNLTFFPMHFVGLAGMPRRIPDYP   457
B. japonicum        YWFPKMTGYMYNETLAKAHFWVTFIGVNLVFFPQHFLGLSGMPRRYVDYP   471
R. sphaeroides      SGIGKMSGRQYPEWAGKLHFWMMFVGANLTFFPQHFLGRQGMPRRYIDYP   487
T. aestivum         YWVGKIFGRTYPETLGQIHFWITFFGVNLTFFPMHFLGLSGMPRRIPDYP   446
D. melanogaster     HWYPLFTGLTLNNKWLKSHFIIMFIGVNLTFFPQHFLGLAGMPRRYSDYP   443
N. crassa           HWVPKILGLNYNMVLSKAQFWLLFIGVNLTFFPQHFLGLQGMPRRISDYP   449
S. cerevisiae       YXSPQILGLNYNEKLAQIQFXLIFIGANVIFFPMHFLGINGMPRRIPDYP   444
                            *        . .* . * * *. *** **.*  *****  ***

                          |         |         |         |         |
A. castellanii      DNYYYWNILSSFGSIISSVSVIVFFYLIYLAFNNNNTPK------LIKL-   500
B. japonicum        DAFAGWNLVSSVGSYISGFGVLIFLYCVIDAF--AKKVPAGDNPWGA---   516
R. sphaeroides      EAFATWNFVSSLGAFLSFASFLFFLGVIFYSLSGARVT--ANNYW-----   530
T. aestivum         DAYAGWNALSSFGSYISVVGIRRFFVVVAITSSSGKNQKCAESPWAVEQ-   495
D. melanogaster     DAYTTWNIVSTIGSTISLLGILFFFFII----------------WESLV-   476
N. crassa           DAFSGWNLISSFGSIVSVVASWLFLYIVYIQLVQGEYA--GRYPWSIPQF   497
S. cerevisiae       DAFAGXNYVASIGSFIATLSLFLFIYILYDQLVNNKSVIYAKAPSSSIEF   494
                    . .   * ... *. ..  .   *.  .

                          |         |         |         |         |
A. castellanii      -----------VHSIFAPYINTLSKNLLTFASIKSTSDSSFFKFSKFFIF   539
B. japonicum        -----------GATTLEWTLPS-PPPFHQFEVLP----------------   538
R. sphaeroides      ---------NEHADTLEWTLTSPPPE-HTFEQLPKREDWE----------   560
T. aestivum         -----------NPTTLEWLVQS-PPAFHTFGELPAVKETKS---------   524
D. melanogaster     -----------SQRQVIYPIQ-LNSSIEWYQNTPP---------------   499
N. crassa           YTDSLRALLNRSYPSLEWSISSPPKP-HSFVSLPLQSSSFFLSFFRLSSY   546
S. cerevisiae       -------LL------------TSPPAVHSFNTPAVQS-------------   512
                                                     .

                          |
A. castellanii      ----------F                                         540
B. japonicum        --------RVQ                                         541
R. sphaeroides      ------RAPAH                                         565
T. aestivum         -----------                                         524
D. melanogaster     -----------                                         499
N. crassa           GEQKEISGRQN                                         557
S. cerevisiae       -----------                                         512
```

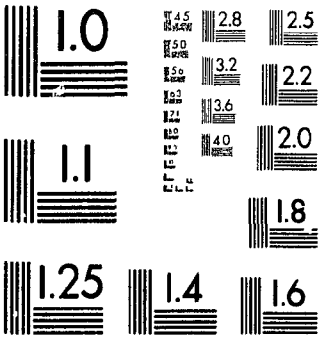
Figure 38 continued.

Figure 39. Amino acid sequence alignment of *A. castellanii* mitochondrial *cox*2, as inferred from the gene sequence. The *A. castellanii* sequence used for the alignment includes the first AUG codon downstream of the *cox*1 coding region (see Figure 35), assumed to be the translation initiation codon of the *cox*2 transcript, to the next in-frame translation termination codon. The amino acid sequence of *T. aestivum cox*2 has been altered to correct for confirmed RNA editing events (Covello and Gray, 1990). Sequences were aligned using the Clustal V Multiple Sequence Alignment program; astericks indicate absolutely conserved amino acid residues while dots represent functionally conserved amino acid residues among all sequences.

Abbreviations and references including GenBank accession numbers: A. castellanii, *Acanthamoeba castellanii* (this thesis); T. aestivum, *Triticum aestivum* (X52867, Covello and Gray, 1990); D. yakuba, *Drosophila yakuba* (X00924, Clary and Wolstenholme, 1983); B. taurus, *Bos taurus* (M10544, Young and Anderson, 1980); S. cerevisiae, *Saccharomyces cerevisiae* (V00706 J01482, Coruzzi and Tzagoloff, 1979).

3 of/de 3

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT



PRECISION[SM] RESOLUTION TARGETS

```
                        |         |         |         |         |
A. castellanii    MVSLSVLFIFYDSLLCLNDHTNSWKIGFQDPTTPIAYGIIKLHDHILFFL   50
T. aestivum       MI-LRSLSCRFLTIA-LCDAAEPWQLGFQDAATPMMQGIIDLHHDIXFFL   48
D. yakuba         MSTWANL-------------------GLQDSASPLMEQLIFFHDHALLIL   31
B. taurus         MAYPMQL--------------------GFQDATSPIMEELLHFHDHTLMIV   31
S. cerevisiae     M--LDLLRLQLTTFI-MNDVPTPYACYFQDSATPNQEGILELHDNIMFYL   47
                  *     *                    .**...*    .. .*..    .

                        |         |         |         |         |
A. castellanii    AVILFVVGYLLLSTYKKFYYGSLNNDLPESKRISLFDTLINTYKENLSFN   100
T. aestivum       ILILVFVLWMLVRALWHFN----EQTNPIPQRI-----------------   77
D. yakuba         VMITVLVGYLMF--------------------MLFFNNYVNRF-------   54
B. taurus         FLISSLVLYIIS--------------------LMLTTKLTHTS-------   54
S. cerevisiae     LVILGLVSWMLYTIVMTYS-------KNPIAYKY------------------   74
                   .*   *  ..

                        |         |         |         |         |
A. castellanii    VTNRTYNINHGTTIEIIWTILPAFILLFIAVPSFALLYAM-DEIIDPVLT   149
T. aestivum       -------V-HGTTIEIIWTIFPSVILLFIAIPSFALLYSMDGVLVDPAIT   119
D. yakuba         -------LLHGQLIEMIWTILPAIILLFIALPSLRLLYLL-DEINEPSVT   96
B. taurus         -------TMDPQEVETIWTILPAIILILIALPSLRILYMM-DEINNPSLT   96
S. cerevisiae     -------IKHGQTIEVIWTIFPAVILLIIAFPSFILLY-LCDEVISPAMT   116
                             .   .* ****.*. **..** **. .** . . .  * .*

                        |         |         |         |         |
A. castellanii    VKVIGHQWYWSYEYSDYSVVYSNRMLDYDSIDRFAAMEMMYKGMGYLKDR   199
T. aestivum       IKAIGHQWYWTYEYSDYNSS------------------------------   139
D. yakuba         LKSIGHQWYWSYEYSDF---- ----------------------------   113
B. taurus         VKTMGHQWYWSYEYTDY----- ---------------------------   113
S. cerevisiae     IKAIGYQWYWKYEYSDFIND------------------------------   136
                  .* .* **** ***.*.

                        |         |         |         |         |
A. castellanii    SLLSYLYIPMVIPETTIKFDSYMIHEAELNLGDLRLLKTDMPLFLPKNTH   249
T. aestivum       ------------DEQSLTFDSYMIPEDDLELGQLRLLEVDNRVVVPAKTH   177
D. yakuba         --------------NNIEFDSYMIPTNELAIDGFRLLDVDNRVILPMNSQ   149
B. taurus         --------------EDLSLDSYMIPTSELKPGELRLLEVDNRVVLPMEMT   149
S. cerevisiae     ------------SGETVEFESYVIPDELLEEGQLRLLDTDTSMVVPVDTH   174
                                  . ..**.*    *  . .***  *  . .*

                        |         |         |         |         |
A. castellanii    IRLLITSSDVLHSWAVPSFGVKVDAVPGRLNQTSLYLKNTGTFYGQCSEL   299
T. aestivum       LRMIVTSADVLHSWAVPSLGVKCDAVPGRLNLTSILVQREGXYYGQCSEI   227
D. yakuba         IRILVTAADVIHSWTVPALGVKVDGTPGRLNQTNFFINRPGLFYGQCS[illegible]   199
B. taurus         IRMLVSSEDVLHSWAVPSLGLKTDAIPGRLNQTTLMSSRPGLYYGQCSEI   199
S. cerevisiae     IRFVVTAADVIHDFAIPSLGIKVDATPGRLNQVSALIQREGVFYGACSEL   224
                  .* .... **.*  ..*..*.* *. *****         * .** ***.

                        |         |         |
A. castellanii    CGVNHAFMPIEVYVVNPVYFYNYVYIYFKNFNLI                   333
T. aestivum       CGTNHAFMPIVVEAVTLKDYADWVSNQLI-LQTN                   260
D. yakuba         CGANHSFMPIVIESVPVNNFIKWISS-----NNS                   228
B. taurus         CGSNHSFMPIVLELVPLKYFEKW-SA-----SML                   227
S. cervisiae      CGTGHANMPIKIEAVSLPKFLEWLNEQ-------                   251
                  **  *. *** .  *    .
```

Figure 39

A

MW : 56374 Dalton, 519 codons

TTT	phe	F	47	TCT	ser	S	17	TAT	tyr	Y	18	TGT	cys	C	1
TTC	phe	F	7	TCC	ser	S	7	TAC	tyr	Y	6	TGC	cys	C	-
TTA	leu	L	41	TCA	ser	S	4	TAA	OCH	Z	-	TGA	trp	W	9
TTG	leu	L	9	TCG	ser	S	-	TAG	AMB	Z	-	TGG	trp	W	4
CTT	leu	L	8	CCT	pro	P	11	CAT	his	H	11	CGT	arg	R	1
CTC	leu	L	-	CCC	pro	P	1	CAC	his	H	4	CGC	arg	R	-
CTA	leu	L	7	CCA	pro	P	7	CAA	gln	Q	5	CGA	arg	R	1
CTG	leu	L	-	CCG	pro	P	6	CAG	gln	Q	-	CGG	arg	R	-
ATT	ile	I	22	ACT	thr	T	17	AAT	asn	N	22	AGT	ser	S	6
ATC	ile	I	7	ACC	thr	T	3	AAC	asn	N	5	AGC	ser	S	4
ATA	ile	I	24	ACA	thr	T	5	AAA	lys	K	5	AGA	arg	R	5
ATG	met	M	21	ACG	thr	T	5	AAG	lys	K	4	AGG	arg	R	3
GTT	val	V	20	GCT	ala	A	18	GAT	asp	D	13	GGT	gly	G	30
GTC	val	V	3	GCC	ala	A	2	GAC	asp	D	-	GGC	gly	G	11
GTA	val	V	10	GCA	ala	A	7	GAA	glu	E	4	GGA	gly	G	4
GTG	val	V	2	GCG	ala	A	4	GAG	glu	E	-	GGG	gly	G	1

B

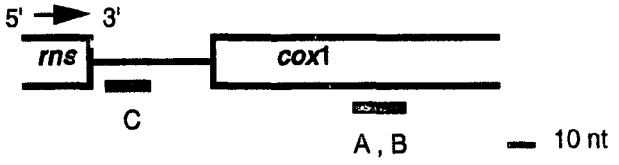
MW : 38120 Dalton, 333 codons

TTT	phe	F	17	TCT	ser	S	11	TAT	tyr	Y	21	TGT	cys	C	2
TTC	phe	F	4	TCC	ser	S	2	TAC	tyr	Y	7	TGC	cys	C	1
TTA	leu	L	25	TCA	ser	S	4	TAA	OCH	Z	-	TGA	trp	W	3
TTG	leu	L	1	TCG	ser	S	2	TAG	AMB	Z	-	TGG	trp	W	2
CTT	leu	L	11	CCT	pro	P	6	CAT	his	H	9	CGT	arg	R	2
CTC	leu	L	-	CCC	pro	P	2	CAC	his	H	-	CGC	arg	R	-
CTA	leu	L	4	CCA	pro	P	4	CAA	gln	Q	2	CGA	arg	R	-
CTG	leu	L	2	CCG	pro	P	2	CAG	gln	Q	2	CGG	arg	R	-
ATT	ile	I	15	ACT	thr	T	9	AAT	asn	N	15	AGT	ser	S	5
ATC	ile	I	2	ACC	thr	T	3	AAC	asn	N	5	AGC	ser	S	-
ATA	ile	I	12	ACA	thr	T	7	AAA	lys	K	11	AGA	arg	R	4
ATG	met	M	11	ACG	thr	T	1	AAG	lys	K	4	AGG	arg	R	2
GTT	val	V	16	GCT	ala	A	9	GAT	asp	D	17	GGT	gly	G	7
GTC	val	V	3	GCC	ala	A	2	GAC	asp	D	1	GGC	gly	G	4
GTA	val	V	3	GCA	ala	A	1	GAA	glu	E	6	GGA	gly	G	2
GTG	val	V	2	GCG	ala	A	-	GAG	glu	E	5	GGG	gly	G	1

Figure 40. Codon usage for *A. castellanii* mitochondrial *cox1* (A) and *cox2* (B) as determined from the gene sequences. Sequences were translated according to the universal genetic code, with the exception that TGA encodes tryptophan. Codon usage was determined by DNA Strider™.

Figure 41. Autoradiogram of cDNA sequence flanking the 5'-terminal coding region of the *cox*1 transcript from *A. castellanii* mitochondria. For A and B, an oligonucleotide complementary to a sequence within the 5'-terminal region of the *cox*1 transcript was used as a sequencing primer; for C, an oligonucleotide specific for the spacer region between the 5'-terminal coding region of *cox*1 and the upstream 3' terminus of the SSU rRNA was used as a sequencing primer (see below), where the 72 refers to the nt position upstream from the 3'-end of the SSU rRNA. The location of the proposed *cox*1 translation initiation codon is indicated by the dots in A. The location of the experimentally determined 3' terminus of the mature SSU rRNA transcript is indicated by the arrowheads in A and B. The samples in B are identical to those in A, but were electrophoresed longer. A portion of the RNA sequence of the *cox*1 transcript (i.e., the complement of the sequence determined) is shown in A.

Shown below: The oligonucleotide probes, used as primers for these experiments, are complementary to regions of the mitochondrial RNA as indicated by the bars shown below the genetic map. The 3'-terminal nucleotide of the oligonucleotide used as a primer for the cDNA sequencing reaction products shown in panel C binds 7 nt downstream of the 3'-end of the SSU rRNA (*rns*). The 3'-terminal nucleotide of the oligonucleotide used as a primer for the cDNA sequencing reaction products shown in panels A and B binds 65 nt downstream of the predicted AUG initiation codon of *cox*1, as indicated. The arrow above the map indicates the polarity of the RNA.



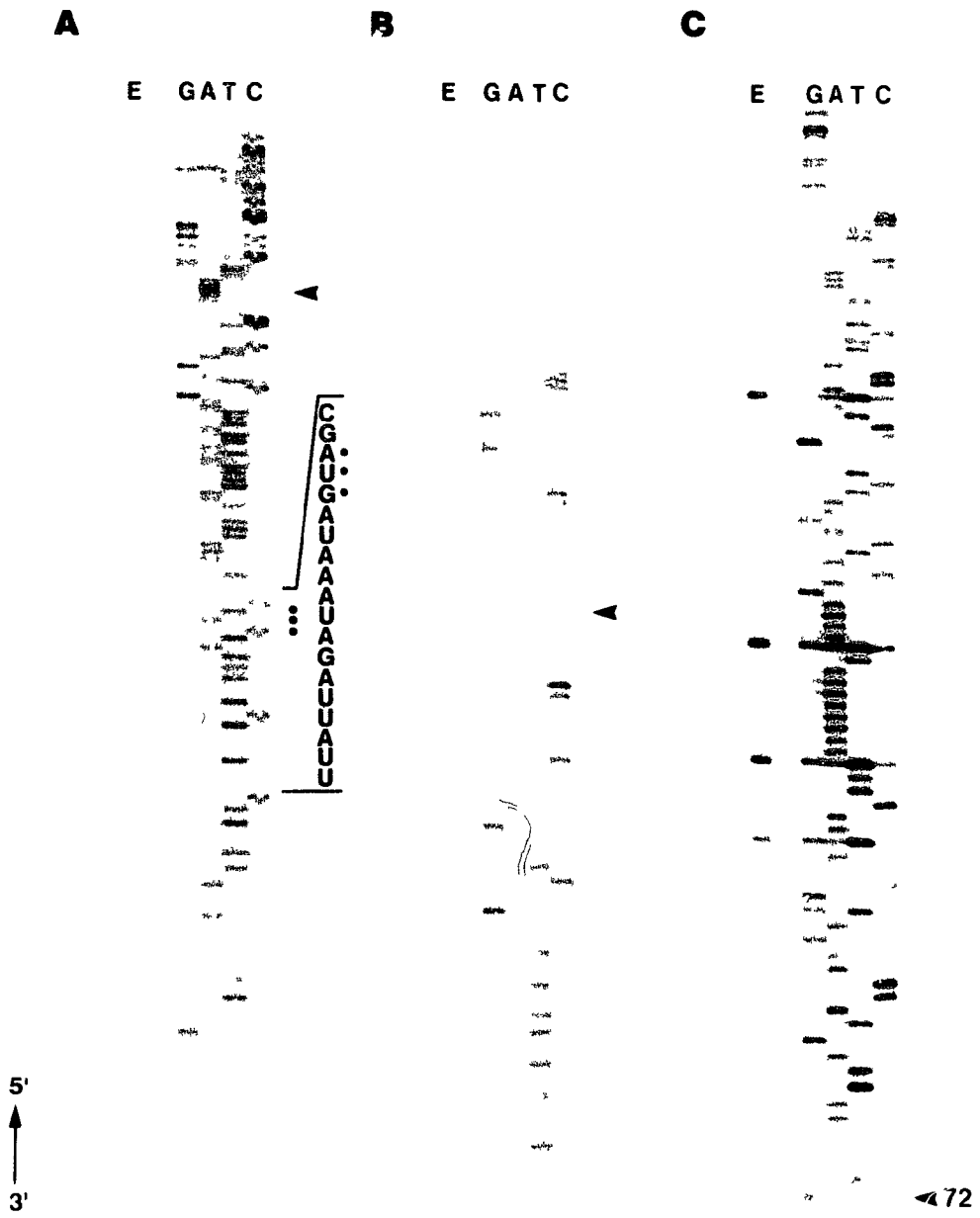
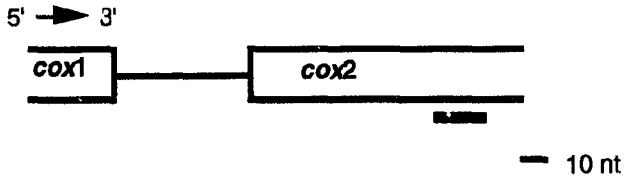


Figure 41

Figure 42. Autoradiogram of cDNA sequence flanking the 5'-terminal coding region of the *cox*2 transcript from *A. castellanii* mitochondria. The location of the proposed *cox*2 translation initiation codon is indicated by the dots. The approximate location of the 3' terminus of the *cox*1 coding region (as determined by amino acid sequence similarity to homologous sequences from other organisms) is indicated by the arrowhead. A portion of the RNA sequence of the *cox*2 transcript (i.e., the complement of the sequence determined) is shown.

Shown below: The oligonucleotide probe used as a primer for this experiment is complementary to a region of the mitochondrial RNA as indicated by the bar shown below the genetic map. The 3'-terminal nucleotide of the oligonucleotide binds 90 nt downstream of the predicted AUG initiation codon of *cox*2. The arrow above the map indicates the polarity of the RNA.



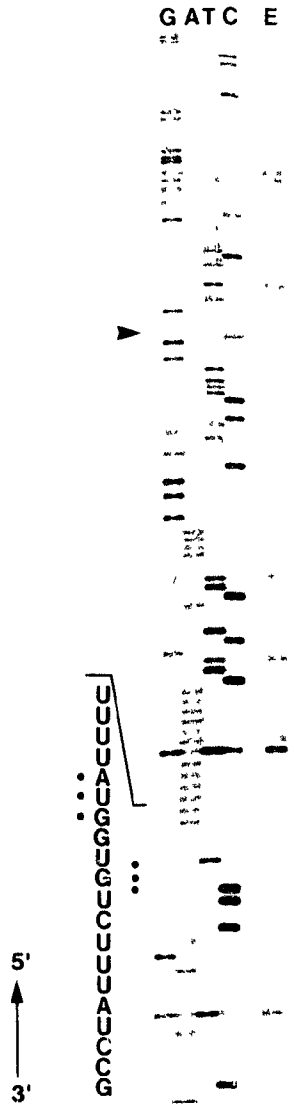


Figure 42

eliminate the possibility that the "real" *cox*1 and *cox*2 genes are in the nuclear genome in *A. castellanii.*

IV. Discussion

A. Evolutionary Affiliation of the Mitochondrial Genome of *Acanthamoeba castellanii*

The aim of the research project undertaken for this thesis was to provide information regarding the organization, mode of expression, structure and evolutionary affiliations of the mitochondrial rRNA genes in the non-photosynthetic protist, *Acanthamoeba castellanii.* This investigation was intended to provide further insights into the phylogenetic enigma, presented by the chlorophyte/metaphyte mitochondrial genomes, that questions the monophyletic origin of mitochondria. Gunderson et al. (1987) originally showed that *Acanthamoeba* occupies a position at the base of the chlorophyte/metaphyte clade in eukaryotic phylogenetic trees; being a non-photosynthetic organism, *Acanthamoeba* would represent an outgroup to this particular lineage. The specific question that was posed, therefore, was: Are the rRNA genes of *Acanthamoeba* mitochondria more similar to their homologs from plant mitochondria or to those from the mitochondrial genome of the chlorophyte, *Chlamydomonas reinhardtii*? It was felt that the answer would potentially help to evaluate whether a second endosymbiotic event might have occurred in the chlorophyte lineage or in the metaphyte lineage, after their divergence from *Acanthamoeba.* Since this project was initiated, sequence data for the mitochondrial SSU rRNA of another chlorophyte, *Prototheca wickerhamii,* have become available. These data indicate that *Prototheca* SSU rRNA has features distinct from its homolog in *C. reinhardtii,* raising concerns about variability of mtDNA within the chlorophyte lineage. In addition, whereas the apparent specific affiliation of *Acanthamoeba* with the chlorophyte/metaphyte lineage seemed reasonably convincing at the time this research was begun, it has recently been questioned (see below).

The organization and expression of the mitochondrial genome in *A. castellanii* shows some novel features. This genome is unique among presently characterized mitochondrial genomes in that the expression of tRNAs is subject to a very specific type of RNA editing (see section C of this Discussion). *A. castellanii* differs from both angiosperms and *C. reinhardtii* (but not from the bryophyte *M. polymorpha* or the chlorophyte *P. wickerhamii*) in containing group I introns in its mitochondrial LSU rRNA gene; these are, however, found at unique insertion sites in the *A. castellanii* mitochondrial rDNA (see section B of this Discussion). Preliminary results suggest that expression of the mitochondrial cytochrome oxidase genes has some intriguing peculiarities (see section D of this Discussion).

The rRNA genes in *Acanthamoeba* mtDNA are organized in a manner different from those of either plant mitochondria or chlorophyte mitochondria. In *Acanthamoeba* mitochondria, *rnl* is found upstream of *rns,* in contrast to the more common positioning of *rns* upstream of *rnl* in those mitochondrial genomes in which these genes are physically linked. In plant mitochondria (including both angiosperms and the bryophyte, *Marchantia*), a 5S rRNA gene is consistently found downstream of *rns*. Moreover, mitochondrial genes are not tightly clustered in angiosperms but they are in *Acanthamoeba* (Organelle Genome Megasequencing Project, M. W. Gray, personal communication); in this regard, *A. castellanii* mtDNA more closely resembles the mitochondrial genomes of *M. polymorpha* and *C.reinhardtii*, which have been completely sequenced (Oda et al, 1992; Boer and Gray, 1988).

It seems clear from DNA sequence analysis that in spite of these unique features of mitochondrial genome organization and expression in *A. castellanii*, its mitochondrial rRNA genes are more similar in many respects to those in land plants and the chlorophyte, *P. wickerhamii*, than they are to the mitochondrial rRNA genes in another chlorophyte, *C. reinhardtii*. The *A. castellanii* rRNA genes are conventional in the sense that they are arranged in a continuous manner, as opposed to the

fragmented organization of the rRNA genes in *C. reinhardtii* mitochondria (Boer and Gray, 1988). The rRNA genes of *A. castellanii* mitochondria are much more closely related, in terms of both potential secondary structure and primary sequence, to those of land plant mitochondria than they are to their homologs in *C. reinhardtii* mitochondria.

(i). Ribosomal RNA secondary structure analysis

The secondary structures of *A. castellanii* mitochondrial rRNAs resemble those of their eubacterial homologs to a notable extent. One of the more obvious examples of such similarity involves variable region V7 in SSU rRNA. In the mitochondria of all major groups of eukaryotes except *Acanthamoeba* and land plants, V7 is drastically truncated. The secondary structure of V6, another variable region in *A. castellanii* mitochondrial SSU rRNA, more closely resembles that of its homolog in *A. tumefaciens* (an α-Proteobacterium) than in *E. coli* (γ-Proteobacterium).

The pattern of modified nucleosides in *A. castellanii* mitochondrial SSU rRNA is also similar to that found in its eubacterial counterparts. It should be noted that the assignment of modifications at positions 25 and 43 from the 3'-end of *A. castellanii* mitochondrial SSU rRNA is by inference only and has not been directly confirmed. However, these preliminary data do suggest that modifications in *Acanthamoeba* mitochondrial SSU rRNA are more similar to those found in eubacterial SSU rRNA than is the case with the homologous rRNA in any other mitochondrial system.

(ii). Primary sequence comparisons

In agreement with the secondary structure comparative analysis described earlier in this thesis, primary sequence pairwise comparisons indicate that *Acanthamoeba* mitochondrial rRNAs display similarity to their eubacterial homologs to a much greater extent than is observed for fungal, ciliate protozoan or animal mitochondrial lineages. This resemblance is of the same order, although less pronounced than, that exhibited by land plant and *P. wickerhamii* mitochondrial rRNAs compared to their eubacterial counterparts (Tables 2 and 3). *Acanthamoeba* mitochondrial rRNAs are much more

similar to those in plant mitochondria than to those in *Chlamydomonas* or other mitochondria. These new data, which indicate that pronounced eubacterial features of mitochondria are not restricted to the chlorophyte/metaphyte lineage, suggest the likelihood that mitochondria from these particular taxa [with the possible exception of *C. reinhardtii* (see below)] all descended from a common ancestor. If plant mitochondria were acquired as the result of an independent, more recent endosymbiotic event, then one would not expect to see such strong similarity to eubacterial sequences in mitochondrial rRNAs outside of the chlorophyte/metaphyte clade (as observed here in the outgroup of *A. castellanii*) unless such a second endosymbiotic event had also included *Acanthamoeba.*

(iii). Ribosomal RNA phylogenetic analysis

Surprisingly, phylogenetic treeing analysis using both SSU and LSU rRNA sequences (Figs. 17 - 20) suggests that *Acanthamoeba* mitochondria may not be specifically affiliated with plant mitochondria, contrary to what might have been expected from the pairwise comparison analysis; rather, *Acanthamoeba* groups with, and is ancestral to, ciliates, fungi and animals in the mitochondrial lineage. The relatively extensive primary sequence similarity observed between *A. castellanii* and land plant mitochondrial rRNAs, as indicated from the pairwise comparison analysis, likely reflects a greater retention of primitive, eubacterial characters (for whatever reason, i.e., relatively less sequence divergence or a more recent acquisition) in plant and *A. castellanii* mitochondria, rather than a specific relationship shared between these two mitochondrial lineages. This interpretation emphasizes that overall primary sequence similarity, unless analyzed in the context of other sequences, i.e., in a phylogenetic context, may be misleading and may not accurately predict ***specific*** evolutionary relationships among the sequences. The finding that *Acanthamoeba* mitochondria do not specifically branch with plant mitochondria makes it less likely that plant and

Acanthamoeba mitochondrial genomes were acquired in a unique, more recent endosymbiotic event.

(iv). Phylogenetic treeing problems

As observed in this study for some of the taxa, there may be reason to doubt particular tree topologies, e.g., because of low bootstrap values or inconsistency between trees derived from the same database by different treeing methods. In such cases, one must consider possible sources for these ambiguities and evaluate the reliability of the analysis. Some of the problems associated with phylogenetic analysis in general are caused by variations in individual sequence characteristics, such as base composition and different rates of nucleotide substitutions among sequences, both of which can cause interpretation artifacts. Also, the number and choice of taxa in the database can influence the topology of the resulting tree (Olsen, 1987). The plant mitochondrial genome is the most slowly evolving mitochondrial genome in terms of rate of nucleotide substitution (Wolfe et al., 1987) while the animal mitochondrial genome is the most rapidly evolving one (Wilson et al., 1985). Mitochondrial genomes also differ greatly in G+C content; some, such as those of fungi, are A+T-rich whereas the plant mitochondrial genome has a much higher G+C content.

a. *A. castellanii*. The phylogenetic position of *A. castellanii* is consistent when either SSU or LSU rRNA sequences are analysed by any one of three different treeing methods (distance matrix, maximum likelihood, or maximum parsimony; Figs. 17 - 20). The main concern here is the low bootstrap values in the distance matrix and maximum parsimony trees. According to Felsenstein (1985), these topologies involving *Acanthamoeba* would not be reliable. One cannot completely exclude the possibility that *Acanthamoeba* and land plant mitochondria tree separately as a result of differences in base composition; if so, the interpretation that *Acanthamoeba* and plant mitochondria are not specifically related would be based on artifact. However, there is reason to doubt this possibility and to believe that the tree topology as regards the

relationship between plant and *Acanthamoeba* mitochondria is correct as presented. The point made here is that *Acanthamoeba* does not branch with plants in the mitochondrial subtree. This conclusion is based partially on the tree topologies presented in the figures but is also supported by analysis of alternative, less frequent topologies provided by the bootstrap analysis (not shown). For example, in the maximum parsimony tree shown in Fig. 20, *Acanthamoeba* mitochondrion is shown to be affiliated with ciliate and fungal mitochondria, to the exclusion of either plant or *C. reinhardtii* mitochondria, in 86% of the trees generated by the resampling method of bootstrap. Although the value of 86% is relativley low in terms of reliability for this type of analysis (95% is considered highly significant), one can ask: Where does *Acanthamoeba* LSU rRNA branch in the rest of the trees generated by the resampling? The answer is that it trees at variable places but no place in particular, i.e., it never trees exclusively with plant mitochondria. Similarly, in the distance matrix trees, although bootstrap values are low for *Acanthamoeba* branching with all other mitochondria except plants, only in 1% of the trees does *Acanthamoeba* specifically cluster with plants. Despite the structural conservatism of the rRNA sequences, there is little independent evidence to link the *A. castellanii* mitochondrial genome with land plant mtDNA. The absence of a mitochondrial 5S rRNA gene also distinguishes the *A. castellanii* mitochondrial genome from that of land plants.

b. ***P. wickerhamii.*** The LSU rRNA sequence of *P. wickerhamii* has not yet been published and therefore only phylogenetic analysis using the SSU rRNA is presented here (Fig. 17). According to this analysis, *P. wickerhamii* mitochondrion does not branch with plant mitochondria despite independent evidence to indicate that the mitochondrial genome of this chlorophyte is affiliated with the plant mitochondrial genome, e.g., the presence of a 5S rRNA gene in the same transcriptional orientation and physical linkage as that found in land plant mtDNA (Burger et al., 1993) and Cox1 phylogenetic data that place *Prototheca* and plants together (Wolff et al., 1993).

Moreover, *P. wickerhamii* mtDNA contains virtually the same complement of ribosomal protein genes as is found in *Marchantia* mtDNA (Oda et al., 1992; Burger et al., 1993). Thus, this treeing analysis does not agree with other independent taxonomic data. However, this particular tree is plagued by low bootstrap values, raising real concerns about the reliability of the tree. When alternate topologies, generated from resampling during the bootstrap analysis are examined, it is noted that *P. wickerhamii* trees exclusively with plants 22% of the time. Thus, although according to the consensus tree *P. wickerhamii* mitochondrion does not affiliate with plant mitochondria, there is reason to believe that under some conditions, *P. wickerhamii* does affiliate with plants in the mitochondrial lineage. When trees are constructed from a SSU rRNA database that includes additional nucleotide positions beyond the universal regions U1 - U8, again *P. wickerhamii* does not tree with plants in the mitochondrial subtree (D. F. Spencer, personal communication).

It is possible that these sequences have diverged to an extent that makes reconstruction of their evolutionary past very difficult. Another possible contributing factor to this ambiguity may be that the *Prototheca* mitochondrial SSU rRNA sequence, at 38.4 % G+C, is 'attracted' to the homologous fungal and *Acanthamoeba* sequences and away from the plant mitochondrial sequences partially due to the lower G+C contents of the former sequences compared to the latter (wheat, 53.6 %; *Oenothera*, 52.4 %).

c. *C. reinhardtii*. There are two problems concerning the phylogenetic analysis of *C. reinhardtii* mitochondrion; these are (i) low bootstrap values in all of the trees and inconsistencies between the SSU and LSU rRNA trees, and (ii) variant topologies with different treeing methods applied to the same LSU rRNA database. In SSU rRNA trees, *C. reinhardtii* trees at the base of the fungi and clearly there is no indication of this particular affiliation from the LSU rRNA treeing analysis. *C. reinhardtii* mitochondrial LSU rRNA affiliates with its homolog in plant mitochondria in the

maximium likelihood tree only, but an affiliation with plant mitochondria is suggested by alternative bootstrap values in both the distance matrix tree and the maximum parsimony tree. Another concern is that the mitochondrial genomes of the two related chlorophytes, *Chlamydomonas* and *Prototheca*, do not tree together. Three possible reasons for this are outlined below.

(v). Conclusions

a. Phylogenetic affiliation of *A. castellanii* mitochondria. In recently derived nuclear SSU rRNA trees (D. F. Spencer, personal communication), constructed from an extended database relative to those used to generate previously published trees (Sogin et al., 1989; Gunderson et al., 1987) and therefore constituting a more comprehensive study, *A. castellanii* consistently assumes a phylogenetic position ancestral to animals and fungi; more importantly, it does not appear to be specifically related to the chlorophyte/metaphyte clade. This topology of nuclear genomes is congruent with the topology of trees derived from mitochondrial SSU and LSU rRNAs as presented in this study, and in which the *A. castellanii* mitochondrial lineage is seen to be ancestral to the ciliate, fungal, and animal mitochondrial lineages. This phylogenetic analysis and the pairwise comparison data weaken the likelihood that plant mitochondria were derived via a second, more recent endosymbiosis. However, the data do strenghten the alternative possibility of a second symbiotic event having occurred in the branch leading to *C. reinhardtii* (see below). It seems more likely that the stronger resemblance of the plant mitochondrial rRNAs to the ancestral eubacterial rRNAs might in part be a reflection of the slower rate of nucleotide substitution in plant mitochondrial primary sequences (Wolfe et al., 1987), and that the more divergent mitochondrial lineages reflect faster rates of nucleotide substitution within specific clades (Wilson et al., 1985), as indicated by the very long branch lengths for the lineages leading to ciliate, fungal, animal, and, to a lesser degree, *Acanthamoeba* mitochondria.

b. Chlorophyte mitochondrial phylogeny. Due to conflicting data regarding tree topology, it is not possible at this time to reach a firm conclusion regarding the phylogeny of the mitochondrial genome of *C. reinhardtii.* However, some preliminary speculations can be made in this regard. It does appear that the mitochondrial genome of *C. reinhardtii* is very unusual in that it not only fails to branch with plant mitochondria in the SSU rRNA mitochondrial trees but, more significantly, it does not appear to be affiliated with the mitochondrial lineage of another chlorophyte, *P. wickerhamii.* Like *C. reinhardtii, P. wickerhamii* belongs to the Chlorophyceae. Not withstanding the problems associated with phylogenetic tree construction alluded to above, three alternative scenarios may be entertained to explain this unexpected topology of chlorophyte mitochondrial rRNAs: (1). The mitochondrial genome of *C. reinhardtii* may have been acquired by a second endoysmbiotic event. From the data presently available this would seem a more likely possibility than the original postulate that the plant mitochondrial genome was derived from an independent endosymbiotic event. Such an event would have had to have occurred more recently than the initial endosymbiotic event whereby mitochondria were acquired by a primitive amitochondriate host organism. If the endosymbiont was a eubacterium, then one would expect the mitochondrial genome of *C. reinhardtii* to be more eubacterial than it appears to be from primary sequence analysis. Perhaps the "endosymbiont" was another mitochondrion transmitted in some manner (other than phagocytosis of a whole cell) from another eukaryotic organism only distantly related to the chlorophtyte lineage.

(2). Both *P. wickerhamii* and *C. reinhardtii* mitochondrial lineages have diverged extensively since they last shared a common ancestor with land plants, with the *C. reinhardtii* mitochondrial lineage diverging very rapidly away from the ancestral pattern since its separation from *P. wickerhamii.* Long branch lengths in the phylogenetic trees do indeed indicate that the *C. reinhardtii* mitochondrial lineage has a

rapid rate of sequence divergence. This being the case, additional sequence data from algae closely related to both *C. reinhardtii* and *P. wickerhamii* should assist in filling in these phylogenetic gaps.

(3). It is possible that both the SSU and LSU rRNA genes of *C. reinhardtii* have been independently acquired via horizontal transfer, from different phylogenetic sources, thereby masking determination of their evolutionary history. This might explain the different topologies of *C. reinhardtii* in SSU and LSU rRNA phylogenetic trees. Perhaps the rRNAs of *C.reinhardtii* are scrambled not only in terms of arrangement, but also phylogenetically.

B. Evolution of Introns in *A. castellanii* Mitochondrial *rnl*

All three of the potential intron ORF-encoded protein products of *A. castellanii* mitochondrial *rnl* are relatively small in size compared to those encoded by intronic ORFs from other rDNAs and from fungal mitochondrial protein-coding genes (Dujon, 1989; Collins et al., 1988); the latter are often the products of fused 5' exon-intron reading frames (Michel et al., 1982; Lazowska et al., 1980). The smallest known intron ORF-encoded product, from the group I intron of the nuclear *rnl* of *Physarum polycephalum*, is a 15 kDa protein that has endonuclease activity; thus, the relatively small size (~16 kDa) of these *A. castellanii* mitochondrial ORF-encoded proteins does not imply lack of enzymatic activity.

Two of the *Acanthamoeba* introns, AcLSU.m2 and AcLSU.m3, belong to subgroup 1A3. Of the ~100 group I introns categorized to date (Michel and Westhof, 1990), the only other introns found thus far within subgroup IA3 are those from the chloroplast *rnl* (Turmel et al., 1991; Rochaix et al., 1985) and *rns* (Durocher et al., 1989) of *Chlamydomonas* species. As illustrated in Fig. 23, a substantial number of identical nucleotides are shared between the core regions of AcLSU.m3 and CrLSU.c, two introns found at corresponding sites in homologous genes. Significant amino acid

sequence similarity is also observed between the ORFs of these two introns (Fig. 22). The normalized alignment score (NAS)[2] was used as an estimate of amino acid sequence homology (Doolittle, 1986) and indicates that these two ORFs (39% identity) are within the range of certain homology. Thus, it would seem unlikely that the ORFs and the structural core regions of the two introns were acquired independently, as appears to have been the situation for some other introns. Within the Genbank database (FASTA search), the only other intron ORFs to which AcLSU.m3 or CrLSU.c show any degree of sequence similarity are intron ORFs of the *nad*3 gene of *Podospora anserina* mitochondria and intron 5 of *Chlamydomonas eugametos* chloroplast LSU rRNA gene. These similarities are in the range of ~23% over ~100 aa, much lower than that between AcLSU.m3 and CrLSU.c.

From the data presented in this thesis, it appears that a plastid intron from a chlorophyte and a mitochondrial intron from an amoeba are homologous. Having established an homologous relationship between AcLSU.m3 and CrLSU.c introns in their entirety, we may ask what this implies about their ancestry. That the CrLSU.c intron is frequently lost is shown by the study by Turmel et al. (1993), which reported a sporadic distribution of this intron within the *Chlamydomonas* lineage. Therefore, any attempt to retrace the path of intron evolution must consider the propensity for this intron to be lost from specific lineages. From the data available it is not possible to unequivocally explain the homology between AcLSU.m3 and CrLSU.c introns, but three alternative routes of evolution as outlined below are conceivable. The first possibility implies that vertical transmission alone is responsible for the intron evolution while the second and third scenarios suggest horizontal transfer as a likely means of intron propagation. I discuss the likelihood of each scenario and the predictions they make.

---

[2] $$\text{NAS} = \frac{\#\ \text{identical residues} + \#\ \text{cysteine residues} - 2.5 \times \#\ \text{gaps}}{\text{length of shortest ORF}} \times 1000$$

(1). Retention of an ancestral intron from a remote common ancestor of Proteobacteria and Cyanobacteria. It is possible that AcLSU.m3 and CrLSU.c introns represent divergent forms of an ancient intron that was present in the eubacterium ancestral to the Cyanobacteria and the Proteobacteria, and thus also present in both proto-mitochondrion and proto-plastid genomes. This is the simplest of the three proposals to explain the intron homology and is attractive in this regard. This scenario implies that a significant degree of nucleic acid and amino acid similarity has been maintained over an extensive evolutionary distance, and would offer strong support to the introns-ancient theory (Darnell and Doolittle, 1986). Thus far the only introns that have been traced back to a eubacterial ancestry are those found in chloroplast tRNAs of land plants. On the other hand, there is reason to believe that acquisition of intron ORFs is a relatively recent event (Belfort, 1990; Shub et al., 1988; Mota and Collins, 1988). In agreement with this, the only group I introns detected thus far in eubacteria do not contain ORFs (Reinhold-Hurek and Shub, 1992; Kuhsel et al., 1990). However, recent evidence indicates that reverse transcriptase-like ORFs are present in group II introns of Cyanobacteria and γ-Proteobacteria (Ferat and Michel, 1993). This raises the exciting possibility that ORFs may not be recently introduced components of introns. Perhaps further characterization of the group II intron ORF reported by Ferat and Michel (1993) will distinguish whether it is likely to represent a recent or ancient acquisition. If the belief that ORFs were recently introduced into introns is upheld despite these new data, this would suggest that the AcLSU.m3/CrLSU.c intron in its entirety, i.e., containing the ORF, is not likely to have been inherited from a remote eubacterial ancestor; rather, the ORF would have had to have been acquired independently within both the chloroplast and mitochondrial lineages. This scenario of the intron and ORF being acquired independently seems less likely than if the intron were acquired in its entirety.

(2). Intra-organismal transfer involving the mitochondrion and plastid in a unicellular photosynthetic ancestor of *Acanthamoeba* and *Chlamydomonas*, whereby the intron was acquired as the result of an intron homing event. As intron-encoded endonucleases appear to be at least partially responsible for the mobility of group I introns, and specifically for directing the intron to the specific target site within the gene, this would imply that the ancestral intronic ORF of AcLSU.m3 and CrLSU.c encoded an endonuclease. In fact, it has been demonstrated that the CrLSU.c intron has the ability to self-splice *in vitro* (Rochaix et al., 1991) and encodes endonuclease activity (Durrenberg and Rochaix, 1991). This scenario would also imply that *Chlamydomonas* and *Acanthamoeba* shared a recent common ancestor, an affiliation inferred from phylogenetic trees based upon nuclear SSU rRNA sequence data (Gunderson et al., 1987; Sogin, 1989). However, more recent phylogenetic analysis provides reason to doubt this association (D. F. Spencer, personal communication). In fact, it seems more likely that *Acanthamoeba* occupies a phylogenetic position in nuclear trees at the base of fungi and animals. This particular scenario of intron propagation is difficult to reconcile with this recent nuclear phylogenetic data. Also, one would have to question why the mitochondrial LSU rRNA gene of *C. reinhardtii* does not have this intron (in fact, the mitochondrial genome of *C. reinhardtii* does not have any introns). This may be a reflection of the activity of the mitochondrially-encoded reverse transcriptase-like protein of *C. reinhardtii* (Boer and Gray, 1988), which possibly assisted in the complete loss of introns from this genome.

(3). A third possibility is that of relatively recent inter-organismal transfer between specific ancestors of *C. reinhardtii* and *A. castellanii.* Considering that this particular intron has thus far been found only in *Acanthamoeba* and *Chlamydomonas* species, this scenario seems more likely than either (1) or (2). (However, one must consider the vast number of plastid and mitochondrial lineages that have not been yet investigated by molecular sequence approaches.) There is evidence to indicate that the CrLSU.c intron,

as well as other chloroplast LSU rDNA introns in other *Chlamydomonas* species, is of recent origin, i.e., that the intron originated within the *Chlamydomonas* lineage (Turmel et al., 1993). One is then left to speculate on the nature of an association that could have resulted in a transfer of genetic material between organisms. Phagocytosis, perhaps of liberated mitochondria or mtDNA from lysed cells, is a possible mechanism by which foreign DNA could be transferred between cells. Viruses could also act as vectors to transfer genetic material between cells. One problem with both this scenario and scenario (2) concerns the co-conversion of intron flanking sequences that has been shown to accompany intron homing via gene conversion in other systems (Muscarella & Vogt, 1989; Jacquier & Dujon, 1985; Quirk et al., 1989). In the present case, there is no evidence to indicate that a co-conversion of intron flanking sequences of the *A. castellanii* mitochondrial *rnl* to chloroplast-like *rnl* sequence (or vice versa) has occurred.

For the reasons outlined above, scenario (3), involving inter-organismal transfer of genetic information, is favored as an explanation for the homology between AcLSU.m3 and CrLSU.c, but with the reservations indicated. In theory the direction of transfer could have been either mitochondrion-to-chloroplast or chloroplast-to-mitochondrion. Codon usage in the CrLSU.c intron ORF differs in a number of respects from that of two *C. reinhardtii* chloroplast protein-coding genes, *rps*12 (small subunit ribosomal protein 12) and *pet*A (chloroplast gene for cytochrome *f*) (Fig. 43). These two *C. reinhardtii* chloroplast genes do not use codons ending in G or C in the four-codon families specifying valine, leucine, serine and threonine. In contrast, all of these codons are used in the CrLSU.c intron ORF. This difference in codon usage of the intron ORF may reflect its recent acquisition. Codon usage in AcLSU.m3 is not radically different from that of the *cox*1 and *cox*2 genes of *A. castellanii.* This single piece of data tentatively suggests that the intron was transferred from the mitochondrion to the chloroplast.

Figure 43. Codon usage in CrLSU.c intron ORF (163 codons, column A), and *C. reinhardtii* chloroplast protein-coding genes for *pet*A (317 codons, column B) and *rps*12 (134 codons, column C). Abbreviations: *pet*A, chloroplast gene for cytochrome f; *rps*12, small subunit ribosomal protein 12. GenBank accession numbers and references: CrLSU.c, X01977, Rochaix et al., 1985; *pet* A, X72917, X61464, Bueschlen et al., 1991; *rps*12, M29284, Liu et al., 1989.

	A	B	C		A	B	C
TTT phe F	5	7	1	TAT tyr Y	4	4	4
TTC phe F	3	5	-	TAC tyr Y	1	9	-
TTA leu L	9	23	6	TAA OCH Z	-	-	1
TTG leu L	1	-	-	TAG AMB Z	-	-	-
CTT leu L	1	5	3	CAT his H	2	-	-
CTC leu L	2	-	-	CAC his H	-	1	3
CTA leu L	6	-	-	CAA gln Q	8	19	5
CTG leu L	1	-	-	CAG gln Q	3	1	-
ATT ile I	6	14	5	AAT asn N	6	5	1
ATC ile I	2	3	1	AAC asn N	-	16	2
ATA ile I	1	-	-	AAA lys K	13	24	16
ATG met M	1	5	2	AAG lys K	6	1	1
GTT val V	5	24	9	GAT asp D	5	6	2
GTC val V	1	-	-	GAC asp D	5	2	-
GTA val V	2	13	6	GAA glu E	6	18	2
GTG val V	3	-	-	GAG glu E	1	2	-
TCT ser S	2	4	3	TGT cys C	1	2	2
TCC ser S	2	-	-	TGC cys C	-	-	-
TCA ser S	2	3	3	TGA OPA Z	-	-	-
TCG ser S	3	-	-	TGG trp W	3	-	-
CCT pro P	3	8	3	CGT arg R	3	5	11
CCC pro P	-	-	-	CGC arg R	-	2	-
CCA pro P	-	15	4	CGA arg R	-	-	2
CCG pro P	3	1	-	CGG arg R	1	-	-
ACT thr T	1	7	1	AGT ser S	2	4	2
ACC thr T	1	-	-	AGC ser S	3	-	-
ACA thr T	5	3	10	AGA arg R	2	-	-
ACG thr T	2	-	-	AGG arg R	1	-	1
GCT ala A	3	14	6	GGT gly C	4	25	9
GCC ala A	-	-	-	GGC gly G	-	1	1
GCA ala A	4	15	4	GGA gly G	-	-	1
GCG ala A	1	1	-	GGG gly G	2	-	1

Figure 43

There are two additional examples of group I introns found at the same position (but different from that at which the AcLSU.m3 and CrLSU.c introns are inserted) in the *rnl* of different organelles (Turmel et al., 1993). An intron inserted between positions 2449 and 2450 (*E. coli* co-ordinates) of domain V is found in fungal mitochondria (this includes the mobile omega intron of *Saccharomyces cerevisiae*), chlorophyte mitochondria (including *P. wickerhamii*, *C. eugametos* and *Chlamydomonas moewusii*), chlorophyte chloroplast (various *Chlamydomonas* species) and protozoan nucleus (*Physarum polycephalum*). An intron between positions 2500 and 2501, also within domain V, is found in the mitochondria and chloroplasts of various species of *Chlamydomonas* (Turmel et al., 1993). In this latter example, the intron sequences have not been published and therefore are not yet available for analysis. In the former example, sequence information is available for introns from each cellular compartment. The nuclear intron from *Physarum* is of a different subgroup (1C1) than the organelle introns and therefore is probably not homologous to them. The fungal mitochondrial introns and the chloroplast intron of *C. moewusii* all belong to subgroup 1A1 and primary sequence comparisons indicate similarity among these introns (Michel and Westhof, 1990). When additional *Chlamydomonas* sequences become available, it should be possible from a phylogenetic analysis to determine whether these introns were propagated by vertical or horizontal transfer.

It has been suggested that the ORFs of *P. ansernia nad*3 and *nad*4 introns can be transmitted independently of the intron core secondary structure (Cummings and Domenico, 1988). Primary sequence comparisons of AcLSU.m1 and AcLSU.m2 with Pa*nad*3 would support this suggestion, as similarities are found within the intron ORF (Fig. 24) but not within the intron core regions (Pa*nad*3 is a subgroup 1C2 intron). However, only a portion of the Pa*nad*3 intron ORF containing one of the two

dodecapeptides and downstream sequences appears to have resemblance to the AcLSU.m1 and AcLSU.m2 intron ORFs.

Mitochondrial LSU rRNA introns of *Neurospora* and *Aspergillus* and the second intron (r2) of *Podospora* LSU rRNA occur at the same insertion site. The *Neurospora* LSU rRNA intron ORF (and, perhaps those from the *Aspergillus* and *Podospora* LSU rRNA introns considering the extensive amino acid sequence similarity shared among these three intron ORFs) is believed to encode ribosomal protein S5; this intron ORF displays no amino acid sequence similarity to the *Acanthamoeba* intronic ORFs described here. These putative ribosomal protein ORFs do not have dodecapeptides P1 or P2. The ORF of the omega intron of *S. cerevisiae* mitochondrial LSU rRNA also encodes an endonuclease but shares no obvious amino acid similarity with the *Acanthamoeba* intronic ORFs. The 15 kDa intronic ORF-encoded product of *Physarum* nuclear LSU rRNA is an endonuclease; however, the protein product has not been fullly characterized.

C. Transfer RNA Editing

(i). Characteristics

Comparison of the tRNAs of cluster 2 and their genes (between LSU and SSU rRNA genes) provides evidence for an RNA editing system in *Acanthamoeba* mitochondria, one that alters the sequence of tRNAs in a very specific manner (Lonergan and Gray, 1993a). Like these genes in cluster 2, a second cluster of tRNA genes (cluster 1) upstream of the LSU rRNA gene also exhibits mismatches in the acceptor stems at the DNA level (Lonergan and Gray, 1993b). Considering the tRNA products of cluster 2, it is likely that canonical base pairing in the acceptor stems of these tRNAs of cluster 1 is also achieved by a similar editing process. The confirmed and predicted editing of *A. castellanii* mitochondrial tRNAs has a very distinct pattern and is different in many

respects from any other type of documented editing. This system has the following characteristics:

(1). The editing alters one or more of the first three nucleotides of the 5' half of the acceptor stem. Editing was not observed at any other positions within the 5' half of the molecule sequenced. It should be noted, however, that the method of sequencing does not provide information about the 3' half of the molecule. Although it cannot be ruled out that editing occurs within this portion, it does not seem likely because all changes necessary to establish acceptor stem base pairing have been identified within the 5' half of the tRNA. On the other hand, because mismatched base pairs are found at the first position of prokaryotic initiator methionine tRNAs, it is formally possible, although not considered likely, that in certain cases the 3' half of an acceptor stem could be edited to *preclude* base pairing at this position.

(2). The editing occurs opposite to a pyrimidine that is located in the 3' half of the acceptor stem.

(3). Both pyrimidine-to-purine and purine-to-purine changes are introduced by the editing. This feature mechanistically distinguishes (see below) the *A. castellanii* editing from other reported substitutional-type editings, which invariably involve purine-to-purine (Sommer et al., 1991) or pyrimidine-to-pyrimidine changes only (Gray and Covello, 1993; many examples cited in the Introduction).

(4). Standard base pairs (A-U or G-C) are formed as a result of the editing process. Although G:U base pairings occur elsewhere within the tRNA, none of the editing events in *A. castellanii* mitochondrial tRNAs introduces (or "correct") a G:U pairing.

(5). Restoration of complete base pairing within the acceptor stem imparts a more conventional tRNA secondary structure with the formation of the typical seven-base-paired acceptor stem (found in functional tRNAs). This suggests that the structural changes directed by this editing system are required for biological activity.

In certain cases, it is conceivable that the editing is necessary for the proper interaction of a particular tRNA with its cognate aminoacyl-tRNA synthetase, because in certain cases the first three base pairs of the acceptor stem are known to engage in sequence-specific contacts with the synthetase (Rould et al., 1989).

(ii). Mechanism

The editing described in *A. castellanii* mitochondria is unique in that the induced changes, with the exception of A to G changes, must involve either base or nucleotide replacement to account for the substitution of a pyrimidine base by a purine base. All types of substitutional editing described up to now involve either pyrimidine transitions or purine transitions and can be explained by some form of base conversion, such as deamination to convert C to U.

Base replacement in editing might be expected to involve the removal of a base by an RNA glycosylase, with subsequent substitution by another base (Wissinger et al., 1992). The polynucleotide chain is not broken in this process. In fact the presence of the hypermodified nucleoside queuosine, which occurs at the first (Wobble) position of the anticodon in certain mammalian and eubacterial tRNAs, occurs via base replacement (Okada and Nishimura, 1979; Okada et al., 1979). Nucleotide replacement involves excision of a nucleotide from the sugar-phosphate backbone, which then can be substituted by another nucleotide (Wissinger et al., 1992).

As in most other documented types of RNA editing, the editing described in *Acanthamoeba* appears to occur post-transcripitionally (in unprocessed precursor transcripts, RNA sequences resemble those of the DNA), which supports the notion that the 3'-half of the acceptor stem might be involved in directing the editing process. As no edited but unprocessed transcripts were detected, it is likely that editing and RNA processing, presumably involving an RNase P-like species to generate the 5' terminus, are tightly coordinated.

(iii). Specificity

Although it seems likely that the 3' half of the acceptor stem provides the information for the editing of *A. castellanii* mitochondrial tRNAs, the specificity cannot simply involve the detection of unpaired bases in stem regions of tRNAs. If this were the case, one would expect that unpaired nucleotides elsewhere in the tRNA molecules [in the DHU stem of $tRNA^{Asp}$ (A10:A25) and $tRNA^{Met2}$ (U10:U25)] would be edited; however, they are not.

The editing process specifically restores A-U pairing as opposed to G:U pairing. Thus the editing of $tRNA^{Ala}$ in *A. castellanii* presents a paradoxical situation. In eubacterial and eukaryotic $tRNA^{Ala}$, G3:U70 is absolutely conserved and serves as a major identity element. Charging experiments will therefore be necessary to determine if the edited *A. castellanii* $tRNA^{Ala}$ accepts alanine or some other amino acid.

D. Cytochrome Oxidase

Transcript analysis of the cytochrome oxidase genes is preliminary and remains to be investigated in more detail. The cDNA sequence presented in Figs. 41 (for *cox*1) and 42 (for *cox*2) can be translated into amino acid sequences that contain all of the highly conserved regions of Cox1 and Cox2 proteins, strongly suggesting that the transcript detected is functional. Hybridization experiments (using *A. castellanii* mitochondrial *cox*1 as a probe) detected no additional *cox*1 sequences (not shown) in the mtDNA. However, further hybridization analysis is required to ensure that there are no copies of these genes in the nucleus. This would confirm that these mitochondrial genes are in fact functional.

The inability to detect the 5' termini for transcripts corresponding to the *cox1*/*cox*2 coding region suggests two alternative possibilities:

(1). The two genes are co-transcribed without subsequent RNA processing. In this case, it is likely that the genes would also be co-translated due to the absence of a translation termination codon between the two coding regions, as deduced from the

DNA sequence. (The cDNA sequencing suggests that a translation termination codon is not generated by RNA editing or other post-transcriptional processing events.) The resulting translation product might function as a fusion protein or, alternatively, might be proteolytically cleaved to generate separate Cox1 and Cox2 polypeptides;

(2). The 5' terminus of the *cox*2 transcript is located somewhere within the *cox*1 gene sequence (beyond the resolution of the primer extension analysis and cDNA sequencing reported here). The 3' terminus of the *cox*1 mRNA could be generated by RNA processing events, perhaps mediated by the stretch of U nucleotides immediately upstream of the *cox*2 translation initiation codon. This putative *cox*1 transcript might then be polyadenylated to generate a translation termination codon, UAA, as occurs in animal mitochondria. This latter possibility would imply that the mature *cox*1 and *cox*2 transcripts are generated from separtate precursor transcripts by alternative processing. Transcript analysis by northern hybridization should distinguish between these two possibilities.

A. castellanii Cox2 is predicted to be ~80 aa residues longer than other known homologs. This increase in length is due to the presence of two extra sequence motifs located on either side of a major hydrophobic, trans-membrane domain (Fig. 44; Raitio et al., 1987; Holm et al., 1987). It is suspected that these extra regions of the Cox2 sequence are present in the mRNA and thus in the potential protein, as they do not disrupt the reading frame. The corresponding nucleotide sequences are too short to represent either group I or group II introns, and they probably do not constitute group III introns, although this has not been verified experimentally. It is not likely that they represent protein introns, as the conserved dodecapeptides characteristic of protein intervening sequences (PIVS) (Shub and Goodrich-Blair, 1992) are not present in these Cox2 sequences; moreover, PIVS are usually in the order of several hundred amino acid residues in length whereas the sequences reported here are much shorter. It is unlikely that these sequences would have catalytic activity as they are found only in

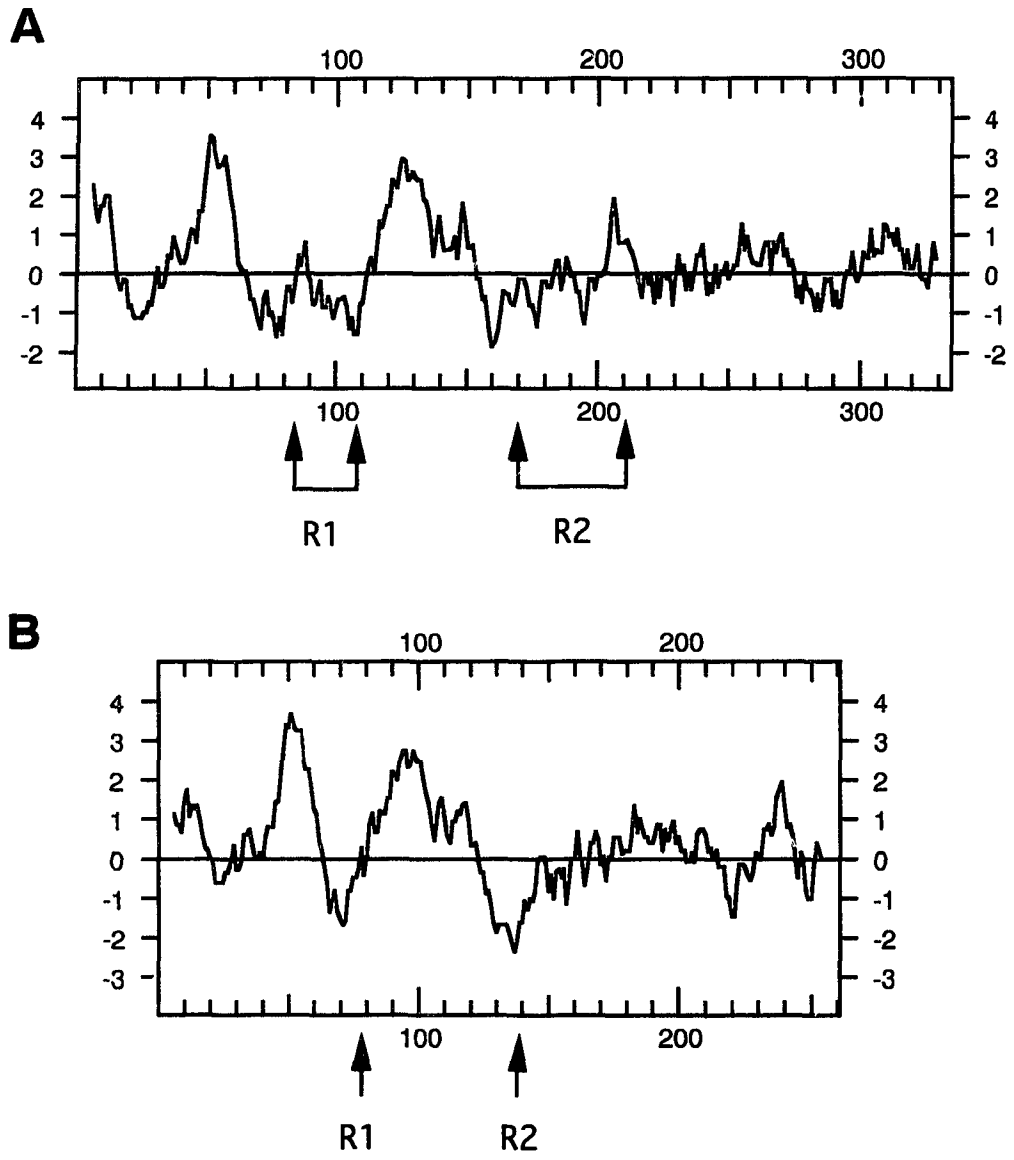


Figure 4. Hydropathy (Kyte Doolittle) plots of *A. castellanii* Cox2 (A), and *Triticum aestivum* (wheat) Cox2 (B). The arrows in A refer to the location of the extra sequences in *A. castellanii* Cox2 [R1, region 1 (25 aa); R2, region 2 (42 aa)] relative to the homolog from other sources. The arrows in B refer to the insertion points of the extra sequences of *A. castellanii* Cox2 relative to the wheat sequence.

Acanthamoeba and therefore could not be essential for activity. Neither does it seem likely that they would disrupt the catalytic activity of Cox2 as they are located some distance upstream of the copper binding site (Cu_A) and proposed cytochrome *c* binding site of Cox2 (Holm et al., 1987).

Cox2 proteins from *Paramecium* (Pritchard et al., 1990) and *Tetrahymena* (Organelle Genome Megasequencing Project, M. W. Gray, personal communication) are unusual in that the entire 5' half of the molecule shows no sequence similarity with non-ciliate homologs with the exception of the invariant aromatic sequence, WYW.Y, possibly involved in electron transfer from either Cu_A or from cytochrome *c* to Cox1 (Capaldi, 1990). The second region of extra sequence within the *A. castellanii* Cox2 (R2, Fig. 44) and the variable region found within Cox2 of the ciliates are present at similar locations, but the region in the ciliates is >200 aa longer than that in *A. castellanii.*

E. Future Directions

(i). Ribosomal RNAs of A. castellanii mitochondria

a. A monophyletic origin of plant mitochondria. Any additional support for a monophyletic origin of plant mitochondria would be encouraging and bolster confidence in the interpretations derived from the data presented in this thesis. The finding of other eubacteria-like mitochondrial genomes outside of the chlorophyte/metaphyte clade would help in this respect. Hence, additional sequence data are needed, of both rRNA gene and entire mitochondrial genomes. It is predicted that other mitochondrial genomes from organisms closely related to *Acanthamoeba*, as judged by nuclear rRNA-based phylogenies or other taxonomic characteristics, will have retained eubacteria-like mitochondrial genomes. Characterization of such genomes and the rRNAs they encode should help to fill in the wide gaps between

certain mitochondrial lineages, thereby providing a more accurate survey of mitochondrial genome evolution.

b. Problems with chlorophytes. It is disturbing that the chlorophyte *P.wickerhamii* consistently fails to branch with plants despite other independent data to the contrary, including phylogenetic trees derived from protein-coding genes (Wolff et al., 1993) and other genome characteristics such as the presence of a mitochondrial 5S rRNA gene. It may be that the rapid evolution of most mitochondrial rRNAs necessitates the inclusion of additional sequences from more closely related relatives than those currently in the database used for the study presented here. Additional sequences from organisms phylogenetically positioned between *A. castellanii* and *P. wickerhamii* and between *P. wickerhamii* and land plants might help to form a specific linkage of metaphytes and chlorophytes by these treeing procedures. As the charophytes are the algae most closely related to land plants, sequence analysis of mitochondrial genomes from this group should prove to be particularly useful in this regard.

c. *Chlamydomonas reinhardtii.* It seems increasingly likely that a more complete understanding of mitochondrial genome evolution will not be achieved until the issue of the phylogenetic position of the *C. reinhardtii* mitochondrial rRNAs is settled. The real phylogenetic position of *C. reinhardtii* can be more confidently assessed only when different treeing methods indicate at least a similar if not an identical branching position for this organism within the mitochondrial lineage. This will only be accomplished by the addition of mitochondrial rRNA sequence data from other chlorophytes more closely related to *Chlamydomonas* than is *Prototheca.* Such an approach would initially involve determining whether the peculiar characteristics observed for the mitochondrial genome of *C. reinhardtii* are representive of the "*Chlamydomonas*" genus as a whole. Therefore, rRNA sequence data from other species of *Chlamydomonas* should be included in future phylogenetic analyses. Secondly, the addition of still more rRNA sequence information from within the

chlorophycean algae should 'bridge the gap' between *Chlamydomonas* and *Prototheca*, if indeed these algae share a common mitochondrial heritage. If at some point after using the appropriate sequences (as suggested by nuclear phylogenetic trees or other taxonomic data) this cannot be established, then it could be more confidently stated that the *Chlamydomonas* mitochondrial genome originated via an independent, second endosymbiotic event. In essence, there is no short cut to obtaining additional sequence information to resolve this particular phylogenetic issue. This is essentially the same approach that has been used in this thesis in an attempt to explore whether the plant mitochondrial genome was acquired as a result of an independent endosymbiosis.

(ii). Further investigation of group I introns of A. castellanii mitochondria

With the data available, it is not possible to distinguish between the three possible scenarios outlined above to account for the homology between the AcLSUm.3 and CrLSU.c introns. Some experiments that might help in this regard are briefly discussed.

As the CrLSU.c intron ORF encodes an endonuclease, and considering the homologous relationship between the two potential proteins, it is suspected that AcLSUm.3 ORF also encodes an endonuclease. This could be tested experimentally by *in vitro* assays such as those developed by Muscarella and Vogt (1989) and Wenzlau et al. (1989). It would be interesting to investigate mitochondrial *rnl* intron distribution in other strains of *A. castellanii* and other species of *Acanthamoeba*. If the intron is actively mobile, one might expect to observe a sporadic distribution among *Acanthamoeba* mitochondrial *rnl* similar to that observed for group I chloroplast *rnl* introns within *Chlamydomonas* species (Turmel et al., 1993). This survey could be performed by direct sequencing of the mitochondrial LSU rRNA gene. PCR primers could be designed such that the sequence generated would either span the intron insertion sites of *A. castellanii* within domains IV and V or alternatively read into any intron sequence present. Ribosomal RNA domain-specific oligonucleotides might be

suitable for a more extensive survey of intron distribution as the introns are found within highly conserved regions of the rRNA. Intron and/or intron ORF-specific primers will probably not suffice to determine the phylogenetic distribution of intron and ORF beyond the *Acanthamoeba* lineage as the degree of similarity among these molecules (even between *Acanthamoeba* and *Chlamydomonas*) is not very high in terms of what would permit hybridization under standard conditions.

Introns at the identical position of the AcLSU.m3 and CrLSU.c introns have been found in other *Chlamydomonas* species (Turmel et al., 1993). Although these intron sequences have not yet been made available for analysis, they should prove to be informative in regards to the evolution of this intron. It has been suggested by Turmel et al. (1993) that this intron was introduced into the *Chlamydomonas* lineage because it has thus far not been detected in other more ancestral chlorophytes. Phylogenetic treeing analysis of these cognate introns (i.e., introns at the same insertion site in homologous genes) from various sources is necessary to provide evidence either for (if the phylogeny of the resident introns differs from that of the organismal phylogeny) or against (if the phylogenies of the resident introns and the organismal phylogeny are congruent) horizonal transfer (Smith et al., 1992; Sogin et al., 1986).

(iii). Transfer RNA editing

It appears that there is no editing of either rRNAs or of protein-coding transcripts in *A. castellanii* mitochondria. How or why such an editing system would have evolved in tRNAs is perplexing. It is an interesting thought that the editing may be developmentally regulated if one considers encystment to be a developmental process. One could imagine that cessation of editing could act as a signal for initiation of encystment. Although this is a purely speculative suggestion, it is nonetheless an attractive possibility that might be worth further investigation.

Experiments designed to investigate the mechanism and specificity of tRNA editing in *A. castellanii* mitochondria should be pursued. This could be approached

through the establishment of an *in vitro* tRNA editing system such as that developed for plant mitochondrial (Araya et al., 1992) and mammalian apoB (Driscoll et al., 1989) mRNAs. In this strategy, the mitochondrial tRNA gene(s) could be cloned into a transcription vector so that unedited synthetic transcripts could be generated *in vitro.* Mitochondrial extracts would be prepared from *A. castellanii* and tested for editing activity using these unedited tRNA substrates. RNA editing could be assayed by cDNA synthesis primed from the gene-specific oligonucleotides using the *in vitro*-edited transcript mixture directly or after cDNA synthesis, to generate either PCR amplification products of the *in vitro*-edited transcripts or to generate individual cloned products that could be sequenced. Alternatively, oligonucleotide probing of *in vitro*-edited products might be specific enough to distinguish between edited and non-edited transcripts, especially for those tRNAs containing multiple editing sites such as $tRNA^{Asp}$. Upon characterization of the editing extract by such criteria as protease and nuclease sensitivity, it should eventually be possible to isolate active components of the editing activity through standard means of protein and/or RNA purification.

Once an active *in vitro* editing system has been established, it would then be feasible to investigate the editing mechanism, i.e., does editing occur by base replacement, nucleotide replacement or by some other unpredicted mechanism? If the editing is enhanced by the addition of exogenous nucleotides, this would imply that either base or nucleotide replacement is operative in the editing process. It should be possible to distinguish between these two possible mechanisms by monitoring (i) the incorporation or (ii) the release of radiolabelled nucleotides during editing. For example, suppose that unedited transcripts are generated in the presence of unlabelled NTPs and that editing is performed in the presence of radiolabeled [α–^{32}P]GTP or [α–^{32}P]ATP (depending on the tRNA being assayed). If label were incorporated into the tRNA specifically at the edited site (this could be assayed by hydrolysis of the edited RNA to mononucleotides followed by tlc analysis), this would implicate nucleotide

replacement in the editing mechanism. Alternatively, [α-^{32}P]UTP could be incorporated into the tRNA synthetic transcript (to give ^{32}P-UMP) which, after incubation with the editing extract, would then be assayed to determine if [α-^{32}P]GMP and/or [α-^{32}P]AMP (depending on the particular tRNA examined) is present at the edited site. If positive, this would support an editing mechanism involving either base replacement (transglycosylation) or base conversion to the exclusion of nucleotide excision and replacement.

Specificity of the editing process likely resides in the tRNA itself and one can ask: Do the specificity determinants of tRNA editing include positional effects and/or nucleotide (primary sequence) identity elements? Again, these issues could be investigated through the use of an *in vitro* editing system. For example, primary sequence or nucleotide requirements for editing can be studied by experiments employing *in vitro* site-directed mutagenesis of nucleotides at or surrounding the editing site of the unedited transcripts. The identity of the pre-edited nucleotide could be altered while maintaining the mismatch with the specific nucleotide of the 3' half of the acceptor stem. The question would be: Are these altered transcripts substrates for editing? Alternatively, or in combination, the identity of the mismatched nucleotide within the 3' half of the acceptor stem could be changed, for example from a pyrimidine to a purine, to determine if nucleotide identity at this position is important for recognition in the editing process. Mismatches could be created within the first three positions of the acceptor stem in those tRNAs (such as tRNA^{Pro}) having a standard cloverleaf secondary structure with a 7-base-paired acceptor stem, and these synthetic transcripts could then be assayed for editing. If mismatches occurred beyond the first three base pairs of the acceptor stem, would these be edited? Positional effects could be investigated by generating deletions or insertions within the acceptor stem or other parts of the tRNA secondary structure and assaying for any effect on editing specificity or efficiency.

A method frequently used to gain insight into mechanism of editing in various systems, and one that does not depend upon the establishment of an *in vitro* editing system, involves the isolation of intermediates in the editing process by cDNA cloning. cDNAs could be generated from the mitochondrial tRNA transcripts of *A. castellanii* and individual clones analyzed by sequencing. If nucleotide replacement is operative in this editing, then one might expect to see nucleotide deletions (or truncation of cDNAs) at the editing positions in a proportion (representing intermediates) of the cDNAs.

An attempt should be made to embark upon a survey to detect similar types of RNA editing in other mitochondrial systems, so that some conclusions regarding the evolution of this editing can be made. As this type of editing should be readily predictable from tRNA gene sequences, the Megasequencing Project should prove useful in this regard also.

Any relationship between processing (to generate the mature 5' and 3' termini of the tRNA) and editing should also be investigated. Considering that 5'-terminal processing of tRNAs by RNase P normally requires a 7-base-paired acceptor stem (McClain et al., 1987; Carrara et al., 1989) and that no edited but unprocessed transcripts are detectable by the methods used in this study, it is suspected that the generation of base pairing within the acceptor stem by editing is tightly coordinated with tRNA processing.

As a final point in this discussion, it would be interesting to confirm whether the unedited versus edited tRNAs are aminoacylated. Particularly relevant in this regard would be the charging of $tRNA^{Ala}$.

(iv). Cytochrome oxidase

As mentioned above, Southern hybridization is necessary to confirm that the mitochondrial *cox*1 and *cox*2 genes represent the only cellular copies of these genes, which would imply that these are functional genes. Northern hybridization analysis will be necessary to determine whether the precursor transcript is processed into

separate messages, since primer extension experiments were unable to map 5' termini of *cox*1 and *cox*2 transcripts. If Cox1 and Cox2 are encoded by a single transcript, western blotting would be appropriate to determine whether or not a functional Cox1-Cox2 fusion protein exists.

V. References

Abdel-Malik, M. S., Court, D. A., Cheng, C. K. and Bertrand, H. 1990. Nucleic Acids Res. 18, 7440-7440.

Adler, B. K., Harris, M. E., Bertrand, K. I. and Hajduk, S. L. 1991. Mol. Cell. Biol. 11, 5878-5884.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. and Young, I. G. 1981. Nature 290, 457-465.

Araya, A., Domec, C., Begu, D. and Litvak, S. 1992. Proc. Natl. Acad. Sci. USA 89, 1040-1044.

Assali, N.-E., Martin, W. F., Sommerville, C. C. and Loiseaux-de Göer, S. 1991. Plant Mol. Biol. 17, 853-863.

Baldauf, S. L., Manhart, J. R. and Palmer, J. D. 1990. Proc. Natl. Acad. Sci. USA 87, 5317-5321.

Baldauf, S. L. and Palmer, J. D. 1990. Nature 344, 262-265.

Banroques, J., Delahodde, A. and Jacq, C. 1986. Cell 46, 837-844.

Basavappa, R. and Sigler, P. B. 1991. EMBO J. 10, 3105-3111.

Beagley, C. T., Okada, N. A. and Wolstenholme, D. R. 1992. in preparation.

Bégu, D., Graves, P.-V., Domec, C., Arselin, G., Litvak, S. and Araya, A. 1990. Plant Cell 2, 1283-1290.

Beier, H., Lee, M. C., Sekiya, T., Kuchino, Y. and Nishimura, S. 1992. Nucl. Acids Res. 20, 2679-2683.

Belfort, M. 1990. Annu. Rev. Genet. 24, 363-385.

Belfort, M. 1991. Cell 64, 9-11.

Benne, R., Van Den Burg, J., Brakenhoff, J. P., Sloof, P., Van Boom, J. H. and Tromp, M. C. 1986, Cell 46, 819-826.

Beresten, S., Jahn, M. and Söll, D. 1992. Nucl. Acids Res. 20, 1523-1530.

Bibb, M. J., Van Etten, R. A., Wright, C. T., Walberg, M. W. and Clayton, D. A. 1981. Cell 26, 167-180.

Binder, S., Marchfelder, A., Brennicke, A. and Wissinger, B. 1992. J. Biol. Chem. 267, 7615-7623.

Bjork, G. R., Ericson, J. U., Gustafsson, C. E. D., Hagervall, T. G., Jonsson, Y. H. and Wikstrom, P. M. 1987. Annu. Rev. Biochem. 56, 263-287.

Blum, B., Bakalara, N. and Simpson, L. 1990. Cell 60, 189-198.

Blum, B., Sturm, N. R., Simpson, M. A. and Simpson, L. 1991. Cell 65, 543-550.

Boer, P. H. and Gray, M. W. 1988. Cell 55, 399-411.

Bonen, L., Boer, P. H., McIntosh, J. E. and Gray, M. W. 1987. Nucleic Acids Res. 15, 6734-6734.

Bonen, L., Cunningham, R. S., Gray, M. W. and Doolittle, W. F. 1977. Nucleic Acids Res. 4, 663-671.

Bonen, L. and Doolittle, W. F. 1976. Nature 261, 669-673.

Bonitz, S. G., Coruzzi, G., Thalenfeld, B., Tzagoloff, A and Macino, G. 1980. J. Biol. Chem. 255, 11927-11941.

Bostrom, K., Garcia, Z., Paksay, K. S., Johnson, D. F., Lusis, A. J. and Innerarity, T. L. 1990. J. Biol. Chem. 265, 22446-22452.

Bostrom, K., Lauer, S., Poksay, K. S., Garcia, Z., Taylor, J. M. and Innerarity, T. L. 1989. J. Biol. Chem. 264, 15701-15708.

Bott, M. 1990. Mol. Microbiol. 4, 2147-2157.

Brennicke, A., Moeller, S. and Blanz, P. A. 1991. Mol. Gen. Genet. 198, 404-410.

Bridge, D., Cunningham, C. W., Schierwater, B., De Salle, R. and Buss, L. W. 1992. Proc. Natl. Acad. Sci. USA 89, 8750-8753.

Brosius, J., Dull, T. J. and Noller, H. F. 1980. Proc. Natl. Acad. Sci. USA 77, 201-204.

Brosius, J., Palmer, M. L., Kennedy, P. J. and Noller, H. F. 1981. J. Mol. Biol. 148, 107-127.

Brown, W. M., Prager, E. M., Wang, A. and Wilson, A. C. 1982. J. Mol. Evol. 18, 225-239.

Brown, T. A., Ray, J. A., Waring, R. B., Scazzocchio, C. and Davies, R. W. 1984. Curr. Genet. 8, 489-492.

Brown, T. A., Waring, R. B., Scazzocchio, C. and Davies, R. W. 1985. Curr. Genet. 9, 113-117.

Bueschlen,S., Choquet,Y., Kuras,R. and Wollman,F.A. 1991. FEBS Lett. 284, 257-262.

Burger, G., Lang, B. F., Wolff, G. and Kück, U. 1993. In *Genetic Maps-1992*, (S. J. O'Brien, ed.) 6th ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y., Book 3, p 172-173.

Burke, J. M., Belfort, M., Cech, T. R., Davies, R. W., Schweyen, R. J., Shub, A., Szostak, J. W. and Tabak, H. F. 1987. Nucleic Acids Res. 15, 7217-7221.

Burke, J. M. and RajBhandary, U. L. (1982). Cell 31, 509-520.

Byers, T. J. 1986. Int. Rev. Cytol. 99, 311-341.

Capaldi, R. A. 1990. Annu. Rev. Biochem. 59, 569-596.

Carignani, G., Groudinsky, O., Frezza, D., Schiavon, E., Bergantino, E. and Slonimski, P. P. 1983. Cell 35, 733-742.

Carrara, G., Calandra, P., Fruscoloni, P., Doria, M. and Tocchini-Valentini, G. P. 1989. Cell 58, 37-45.

Cattaneo, R., 1991. Annu. Rev. Genet. 25, 71-88.

Cavalier-Smith, T. 1987. Annu. N.Y. Acad. Sci 503, 17-54; 55-71.

Cech, T. R. 1988. Gene 73, 259-271.

Cech, T. R. 1990. Annu. Rev. Biochem. 59, 543-568.

Cech, T. R. 1991. Cell 64, 667-669.

Chapdelaine, Y. and Bonen, L. 1991. Cell 65, 465-472.

Chen, S. H., Habib, G., Yang, C. Y., Gu, Z. W., Lee, B. R., Weng, S. A., Silberman, S. R., Cai, S. J., Deslypere, J. P., Rosseeneu, M., Goto, A. M., Li, W. H. and Chan, L. 1987. Science 238, 363-366.

Chen, S. H., Li, X., Liao, W. S. L., Wu, J. H. and Chan, L. 1990. J. Biol. Chem. 265, 6811-6816.

Chen, J.-Y. and Martin, N. C. 1988. J. Biol. Chem. 263, 13677-13682.

Christopher, D. A. and Hallick, R. B. 1989. Nucleic Acids Res. 17, 7591-7608.

Chung, C. T. and Miller, R. H. 1988. Nucleic Acids Res. 16, 3580-3580.

Chung, C. T., Niemela, S. L. and Miller, R. H. 1989. Proc. Natl. Acad. Sci. USA 86, 2172-2175.

Clark-Walker, G. D. 1989. Proc. Natl. Acad. Sci. USA 86, 8847-8851.

Clark-Walker, G. D. 1991. Curr. Genet. 20, 195-198.

Clark-Walker, G. D. 1992. Int. Rev. Cytol. 141, 89-121.

Clary, D. O. and Wolstenholme, D. R. 1983. Nucleic Acids Res. 11, 4211-4227.

Clary, D. O. and Wolstenholme, D. R. 1985. J. Mol. Evol. 22, 252-271.

Clayton, D. A. 1992. Int. Rev. Cytol. 141, 217-232.

Colleaux, L., d'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Galibert, F. and Dujon, B. 1986. Cell 44, 521-533.

Collings, J. C. 1991. Ph.D. Dissertation, Dalhousie University, Halifax, Nova Scotia.

Collins, R. A., Reynolds, C. A. and Olive, J. 1988. Nucleic Acids Res. 16, 1125 1134.

Conklin, P. L., Wilson, R. K. and Hanson, M. R. 1991. Genes and Dev. 5, 1407-1415.

Copertino, D. W., Christopher, D. A. and Hallick, R. B. 1991. Nucleic Acids Res. 19, 6491-6497.

Copertino, D. W. and Hallick, R. B. 1991. EMBO J. 10, 433-442.

Copertino, D. W., Shigeoka, S. and Hallick, R. B. 1992. EMBO J. 11, 5041-5050.

Coruzzi, G. and Tzagoloff, A. 1979. J. Biol. Chem. 254, 9324-9330.

Covello, P. S. and Gray, M. W. 1989. Nature 341, 662-666.

Covello, P. S. and Gray, M. W. 1990. Nucleic Acids Res. 18, 5189-5195.

Covello, P. S. and Gray, M. W. 1992. EMBO J. 11, 3815-3820.

Cozens, A. L. and Walker, J. E. 1987. J. Mol. Biol. 194, 359-383.

Cozens, A. L., Walker, J. E., Phillips, A. L., Huttly, A. K. and Gray, J. C. 1986. EMBO J. 5, 217-222.

Crothers, T., Seno, T. and Söll, D. G. 1972. Proc. Natl. Acad. Sci. USA 69, 3063-3067.

Cummings, D. L. 1992. Int. Rev. Cytol. 141, 1-64.

Cummings, D. J. and Domenico, J. M. 1988. J. Mol. Biol. 204, 815-839.

Cummings, D. J., Domenico, J. M. and Nelson, J. 1989a. J. Mol. Evol. 28, 242-255.

Cummings, D. L., Domenico, J. M., Nelson, J. and Sogin, M. L. 1989b. J. Mol. Evol. 28, 232-241.

Cummings, D. L., McNally, K. L., Domenico, J. M. and Matsuura, E. T. 1990. Curr. Genet. 17, 375-402.

Cummings, D. J., Michel, F. and McNally, K. L. 1989c. Curr. Genet. 16, 381-406.

Cummings, D. J., Michel, F. and McNally, K. L. 1989d. Curr. Genet. 16, 407-418.

Cunningham, P. R., Weitzmann, C. J., Negre, D., Sinning, J. G., Frick, V., Nurse, K. and Ofengand, J. 1990. In *The Ribosome: Structure, Function and Evolution.*

eds. W. E. Hill, A. Dahlberg, R. A. Garrett, P. B. Moore, D. Schlessinger and J. R. Warner, American Society for Microbiology, Washington, D. C., pp 243-252.

Dahlberg, A. E. 1989. Cell 57, 525-529.

Darnell, J. E. and Doolittle, W. F. 1986. Proc. Natl. Acad. Sci. USA 83, 1271-1275.

Davies, M. S., Wallis, S. C., Driscoll, D. M., Wynne, J. K., Williams, G. W. 1989. J. Biol. Chem. 264, 13395-13398.

Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A. and Scazzocchio, C. 1982. Nature 300, 719-724.

Davila-Aponte, J. A., Huss, V. A. R., Sogin, M. L. and Cech, T. R. 1991. Nucleic Acids Res. 19, 4429-4436.

Dayhoff, M. O. and Schwartz, R. M. 1981. Annu. N. Y. Acad. Sci. 361, 92-104.

DeBorde, D. C., Naeve, C. W., Herlocher, M. L. and Maassab, H. F. 1986. Anal. Biochem. 157, 275-282.

de la Cruz, V. F., Lake, J. A., Simpson, A. M. and Simpson, L. 1985. Proc. Natl. Acad. Sci. USA 82, 1401-1405.

Delahodde, A., Goguel, V., Becam, A. M., Creusot, F., Perea, J., Banroques, J. and Jacq, C. 1989. Cell 56, 431-441.

Delihas, N. and Fox, G. E. 1987. Annu. N. Y. Acad. Sci. 503, 92-102.

Denovan-Wright, E. M. and Lee, R. W. 1992. Curr. Genet. 21, 197-202.

De Rijk, P., Neefs, J.-M., Van de Peer, Y. and De Wachter, R. 1992. Nucleic Acids Res. 20, 2075-2089.

Deutscher, M. P. 1984. Crit. Rev. Biochem. 17, 45-71.

De Zamaroczy, M. and Bernardi, G. 1986. Gene 41, 1-22.

Diaconis, P. and Efron, B. 1983. Sci. Amer. 268, 116-129.

Diamond, A. M., Montero-Puerner, Y., Lee, B. J. and Hatfield, D. 1990. Nucleic Acids Res. 18, 6727.

Dickerson, R. E. 1980. Nature 283, 210-212.

Donis-Keller, H., Maxam, A. M. and Gilbert, W. 1977. Nucleic Acids Res. 4, 2527-2538.

Doolittle, R. F. 1986. *Of URFS and ORFS: A Primer on How to Analyze Derived Amino Acid Sequences.* University Science Books, Mill Valley, CA.

Douglas, S. E. and Doolittle, W. F. 1984. Nucleic Acids Res. 12, 3373-3386.

Douglas, S. E., Murphy, C. A., Spencer, D. F. and Gray, M. W. 1991. Nature 350, 148-151.

Douglas, S. and Turner, S. 1991. J. Mol. Evol. 33, 267-273.

Driscoll, D. M. and Casanova, E. 1990. J. Biol. Chem. 265, 21401-21403.

Dryden, S. C. and Kaplan, S. 1990. Nucleic Acids Res. 18, 7267-7277.

Dubin, D. T., HsuChen, C. C. and Tillotson, L. E. 1986. Curr. Genet. 10, 701-707.

Dubin, D. T., Timko, K. D. and Baer, R. J. 1981. Cell 23, 271-278.

Dujardin, G., Jacq, C. and Slonimski, P.P. 1982. Nature 298, 628-632.

Dujon, B. 1980. Cell 20, 185-197.

Dujon, B. 1989. Gene 82, 91-114.

Dujon, B., Belfort, M., Butow, R. A., Jacq, C., Lemieux, C., Perlman, P. S. and Vogt, V. M. 1989. Gene 82, 115-118.

Durocher, V., Gauthier, A., Bellemare, G. and Lemieux, C. 1989. Curr. Genet. 15, 277-282.

Durrenberger, F. and Rochaix, J.-D. 1991. EMBO J. 10, 3495-3501.

Elliott, M. S. and Trewyn, R. W. 1984. J. Biol. Chem. 259, 2407-2410.

Eperon, I. C., Anderson, S. and Nierlich, D. P. 1980. Nature 286, 460-467.

Eperon, I. C., Janssen, J. W., Hoeijmakers, J. H. and Borst, P. 1983. Nucleic Acids Res. 11, 105-125.

Falconet, D., Sevignac, M. and Quetier, F. 1988. Curr. Genet. 13, 75-82.

Feagin, J. E., Abraham, J. M. and Stuart, K. 1988. Cell 53, 413-422.

Feagin, J. E., Jasmer, D. P. and Stuart, K. 1987. Cell 49, 337-345.

Fearnley, I. M. and Walker, J. E. 1986. EMBO J. 5, 2003-2008.

Felsenstein, J. 1985. Evolution 39, 783-791.

Felsenstein, J. 1988. Annu. Rev. Genet. 22, 521-565.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) Version 3.5 (Computer software package and manual distributed by the author, Dept. Genetics, University of Washington, Seattle, WA).

Ferat, J.-L. and Michel, F. 1993. Nature 364, 358-361.

Field, D. J., Sommerfield, A., Saville, B. J. and Collins, R. A. 1989. Nucleic Acids Res. 17, 9087-9099.

Fitch, W. M. 1977. Am. Nat. 111, 223-257.

Fox, T. D. 1987. Annu. Rev. Genet. 21, 67-91.

Fox, T. D. and Leaver, C. J. 1981. Cell 26, 315-323.

Gampel, A. and Cech, T. R. 1991. Genes Dev. 5, 1870-1880.

Gantt, J. S., Baldauf, S. L., Calie, P. J., Weeden, N. F. and Palmer, J. D. 1991. EMBO J. 10, 3073-3078.

Garrett, R. A., Dalgaard, J., Larsen, N., Kjems, J. and Mankin, A. S. 1991. Trends Biochem. Sci. 16, .

Garriga, G. and Lambowitz, A. M., 1986. Cell 46, 669-680.

Gauthier, A., Turmel, M. and Lemieux, C. 1991. Curr. Genet. 19, 43-47.

Geliebter, J. 1987. Focus 9:1, 5-8.

Gibbs, S. 1981. Annu. N. Y. Acad. Sci. 361, 193-207.

Giovannoni, S. J., Turner, S., Olsen, G. J., Barns, S., Lane, D. J. and Pace, N. R. 1988. J. Bacteriology 170, 3584-3592.

Goguel, V., Delahodde, A. and Jacq, C. 1992. Mol. Cell. Biol. 12, 695-705.

Goldbach, R. W., Borst, P., Bollen-DeBoer, J. E. and van Bruggen, E. F. J. 1978. Biochim. Biophys. Acta 521, 169-186.

Goldschmidt-Clermont, M., Choquet, Y., Girard-Bascou, J., Michel, F., Schirmer-Rahire, M. and Rochaix, J.-D. 1991. Cell 65, 135-143.

Gray, M. W. 1989. Trends Genet. 5, 294-299.

Gray, M. W. 1992. Int. Rev. Cytol. 141, 233-357.

Gray, M. W., Cedergren, R., Abel, Y. and Sankoff, D. 1989. Proc. Natl. Acad. Sci. USA 86, 2267-2271.

Gray, M. W. and Covello, P. S. 1993. FASEB J. 7, 64-71.

Gray, M. W. and Doolittle, W. F. 1982. Microbiol. Rev. 46, 1-42.

Gray, M. G., Hanic-Joyce, P. J. and Covello, P. S. 1992. Annu. Rev. Plant Physiol. Plant Mol. Biol. 43, 145-175.

Gray, M. W., Sankoff, D. and Cedergren, R. J. 1984. Nucleic Acids Res. 12, 5837-5852.

Gray, M. W. and Spencer, D. F. 1983. FEBS Lett. 161, 323-327.

Green C. J., Stewart, G. C., Hollis, M. A., Vold, B. S. and Bott, K. F. 1985. Gene 37, 261-266.

Greer, C. L. and Abelson, J. 1984. Trends Biochem. Sci. 9, 139-141.

Greeve, J., Navaratnam, N. and Scott, J. 1991. Nucleic Acids Res. 19, 3569-3576.

Gruissem, W. 1989. Cell 56, 161-170.

Gualberto, J. M., Bonnard, G., Lamattina, L. and Grienenberger, J. M. 1991. Plant Cell 3, 1109-1120.

Gunderson, J. H., Elwood, H., Ingold, Kindle, K. and Sogin, M. L. 1987. Proc. Natl. Acad. Sci. USA 84, 5823-5827.

Gutell, R. R. 1993. Nucleic Acids Res. 21, 3051-3054.

Gutell, R. R., Gray, M. W. and Schnare, M. N. 1993. Nucleic Acids Res. 21, 3055-3074.

Gutell, R. R., Larsen, N. and Woese, C. R. 1992a. In *Ribosomal RNA: Structure, Evolution, Gene Expression and Function in Protein Synthesis* (Zimmermann, R. A. and Dahlberg, A. E., eds.). CRC Press, boca Raton, FL, in press.

Gutell, R. R., Schnare, M. N. and Gray, M. W. 1990. Nucleic Acids Res. 18, Suppl., 2319-2330.

Gutell, R. R., Schnare, M. N. and Gray, M. W. 1992b. Nucleic Acids Res. 20, Suppl., 2095-2109.

Gutell, R. R., Weiser, B., Woese, C. R. and Noller, H. F. 1985. Progr. Nucleic Acid Res. Mol. Biol. 32, 155-216.

Guthrie, C. 1991. Science 253, 157-163.

Hancock, K. and Hajduk, S. L. 1990. J. Biol. Chem. 265, 19208-19215.

Hanic-Joyce, P. J. and Gray, M. W. 1990. J. Biol. Chem. 265, 13782-13791.

Hanley-Bowdoin, L. and Chua, N.-H. 1987. Trends Biochem. Sci. 12, 67-70.

Hanson, M. R. and Folkerts, O. 1992. Int. Rev. Cytol. 141, 128-172.

Hartman, R. K. and Erdmann, V. A. 1989. J. Bacteriol. 171, 2933-2941.

Hartman, R. K., Wolters, J., Kroeger, B., Schultze, S., Specht, T. and Erdmann, V. A. 1989. Syst. Appl. Microbiol. 11, 243-249.

Hayase, Y., Jahn, M., Rogers, M. J., Sylvers, L. A., Koizumi, M., Inoue, H., Ohtsuka, E. and Söll, D. 1992. EMBO J. 11, 4159-4165.

Heidecker, G., Messing, J. and Gronenborn, B. 1980. Gene 10, 69-73.

Heinonen, T. Y. K., Schnare, M. N. and Gray, M. W. 1990. J. Biol. Chem. 265, 22336-22341.

Heinonen, T. Y. K., Schnare, M. N., Young, P. G. and Gray, M. W. 1987. J. Biol. Chem. 262, 2879-2887.

Henikoff, S. 1984. Gene 28, 351-359.

Hensgens, L. A. M., Bonen, L., de Haan, M., van der Horst, G. and Grivell, L. A. 1983. Cell 32, 379-389.

Hiesel, R., Schobel, W., Schuster, W. and Brennicke, A. 1987. EMBO J. 6, 29-34.

Higgins, D. G., Bleasby, A. J. and Fuchs, R. 1992. CABIOS 8 (2), in press.

Hoch, B., Maier, R. M., Appel, K., Igloi, G. L. and Kössel, H. 1991. Nature 353, 178-180.

Hodges, P. E., Navaratnam, N., Greeve, J. C. and Scott, J. 1991. Nucleic Acids Res. 19, 1197-1201.

Hodges, P. and Scott, J. 1992. Trends Biochem. Sci. 17, 77-81.

Hoepfl, P., Ludwig, W., Hartman, R. K., Ulbrich, N. and Schleifer, K. H. 1988. Nucleic Acids Res. 16, 9043-9043.

Hoepfl, P., Ludwig, W., Schleifer, K. H. and Larsen, N. 1989. Eur. J. Biochem. 185, 355-364.

Hoffmann, R. J., Boore, J. L. and Brown, W. M. 1992. Genetics 131, 397-412.

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R. and Zamir, A. 1965. Science 147, 1462-1465.

Holm, L., Saraste, M. and Wikström, M. 1987. EMBO J. 6, 2819-2823.

Hou, Y.-M., Francklyn, C. and Schimmel, P. 1989. Trends Biochem. Sci. 14, 233-237.

Hou, Y.-M. and Schimmel, P. 1988. Nature 333, 140-145.

Huh, T. Y. and Gray, M. W. 1982. Plant Mol. Biol. 1, 245-249.

Hunt, M. D. and Newton, K. J. 1991. EMBO J. 10, 1045-1052.

Huss, V. A. R. and Sogin, M. L. 1990. J. Mol. Evol. 31, 432-442.

Huttenhofer, A., Sakai, H. and Weiss-Brummer, B. 1988. Nucleic Acids Res. 16, 8665-8674.

Iwami, M., Muto, A., Yamao, F. and Osawa, S. 1984. Mol. Gen. Genet. 196, 317-322.

Jacquier, A. and Dujon, B. 1983. Mol. Gen. Genet. 192, 487-499.

Jacquier, A. and Dujon, B. 1985. Cell 41, 383-394.

Janke, A. and Pääbo, S. 1993. Nucleic Acids Res. 21, 1523-1525.

John, P. 1987. Annu. N.Y. Acad. Sci. 503, 140-150.

Joyce, P. M. B. and Gray, M. W. 1989. Nucleic Acids Res. 17, 5461-5476.

Kiesewetter, S., Ott, G. and Sprinzl, M. 1990. Nucleic Acids Res. 18, 4677-4682.

Kim, S.-H. 1978a. J. Mol. Biol. 123, 607-630.

Kim, S.-H. 1978b. Adv. Enzymol. 46, 279-315.

Kim, S.-H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H. J., Seeman, N. C. and Rich, A. 1974. Science 185, 435.

Kittle, J. D., Jr., Mohr, G., Gianelos, J. A., Wang, H. and Lambowitz, A. M. 1991. Genes Dev. 5, 1009-1021.

Knoop, V., Schuster, W., Wissinger, B. and Brennicke, A. 1991. EMBO J. 10, 3483-3493.

Kohchi, T., Umesono, K., Ogura, Y., Komine, Y., Nakahigashi, K., Komano, T., Yamada, Y., Ozeki, H. and Ohyama, K. 1988. Nucleic Acids Res. 16, 10025-10036.

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. and Cech, T. R. 1982. Cell 31, 147-157.

Kück, U., Godehardt, I. and Schmidt, U. 1990. Nucleic Acids Res. 18, 2691-2697.

Kudla, J., Igloi, G. L., Metzlaff, M., Hagemann, R. and Kössel, H. 1992. EMBO J. 11, 1099-1103.

Kuhsel, M. G., Strickland, R. and Palmer, J. D. 1990. Science 250, 1570-1573.

Küntzel, H. and Köchel, H. G. 1981. Nature 293, 751-755.

Lagerkvist, U. 1981. Cell 23, 305-306.

Lambowitz, A. M. 1989. Cell 56, 323-326.

Lang, B. F. 1984. EMBO J. 3, 2129-2136.

Lang, B. F., Cedergren, R. and Gray, M. W. 1987. Eur. J. Biochem. 169, 527-537.

Lazowska, J., Claisse, M., Gargouri, A., Kotylak, Z., Spyridakas, A. and Slonimski, P. P. 1989. J. Mol. Biol. 205, 275-290.

Lazowska, J., Jacq, C. and Slonimski, P. P. 1980. Cell 22, 333-348.

Lemieux, C., Boulanger, J., Otis, C. and Turmel, M. 1989. Nucleic Acids Res. 17, 7997-7997.

Levings III, C. S. and Brown, G. G. 1989. Cell 56, 171-179.

Levitt, M. 1969. Nature 224, 759-763.

Li, W.-H. and Graur, D. 1991. *Fundamentals of Molecular Evolution*, Sinauer Associates, Inc., Publishers, Sunderland, Massachusetts.

Li, M., Tzagoloff, A., Underbrink-Lyon, K. and Martin, N. C. 1982. J. Biol. Chem. 257, 5921-5928.

Liesack, W., Hoepfl, P. and Stackebrandt, E. 1988. Nucleic Acids Res. 16, 5194-5194.

Liu, X.-Q., Gillham, N. W. and Boynton, J. E. 1989. J. Biol. Chem. 264, 16100-16108.

Lonergan, K. M. and Gray, M. W. 1993a. Science 259, 812-816.

Lonergan, K. M. and Gray, M. W. 1993b. Nucleic Acids Res., in press.

Lonsdale, D. M., Hodge, T. P. and Fauron, C. M.-R. 1984. Nucleic Acids Res. 12, 9249-9261.

MacKay, R. M., Spencer, D. F., Doolittle, W. F. and Gray, M. W. 1980. Eur. J. Biochem. 112, 561-576.

Mahendran, R., Spottswood, M. R. and Miller, D. L. 1991. Nature 349, 434-438.

Maier, R. M., Hoch, B., Zeltz, P. and Kössel, H. 1992. Plant Cell 4, 609-616.

Manam, S. and Van Tuyle, G. C. 1987. J. Biol. Chem. 262, 10272-10279.

Maniatis, T., Fritsch, E. F. and Sambrook, J. 1982. *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y.

Manna, E. and Brennicke, A. 1985. Curr. Genet. 9, 505-515.

Marck, C. 1989. Service de Biochimie, Departement de Biologie, Institut de Recherche, Fondamentale, CEA, France.

Marechal-Drouard, L., Weil, J.-H. and Guillemaut, P. 1988. Nucleic Acids Res. 16, 4777-4788.

Margulis, L. 1981. *Symbiosis in Cell Evolution*, W. H. Freeman and Company, San Francisco, CA, USA.

Martin, W., Somerville, C. C. and Loiseaux-de Göer, 1992. J. Mol. Evol. 35, 385-404.

Maslov, D. A. and Simpson, L. 1992. Cell 70, 459-467.

McClain, W. H. 1993. FASEB J. 7, 72-78.

McClain, W. H., Foss, K., Jenkins, R. A. and Schneider, J. 1991. Proc. Natl. Acad. Sci. USA 88, 9272-9276.

McClain, W. H., Guerrier-Takada, C. and Altman, S. 1987. Science 238, 527-530.

McPheeters, D. S. and Abelson, J. 1992. Cell 71, 819-831.

Meier, F., Suter, B., Grosjean, H., Keith, G. and Kubli, E. 1985. EMBO J. 4, 823-827.

Mery-Drugeon, E., Crouse, E. J., Schmitt, J. M., Bohnert, H.-J. and Bernardi, G. 1981. Eur. J. Biochem. 114, 577-583.

Messing, J. 1983. Meth. Enzymol. 101, 20-78.

Michel, F. and Dujon, B. 1983. EMBO J. 2, 33-38.

Michel, F., Jacquier, A. and Dujon, B. 1982. Biochimie 64, 867-881.

Michel, F. and Lang, B. F. 1985. Nature 316, 641-643.

Michel, F., Netter, P., Xu, M-Q. and Shub, D. 1990. Genes Dev. 4, 777-788.

Michel F., Umesono, K. and Ozeki, H. 1989. Gene 82, 5-30.

Michel, F. and Westhof, E. 1990. J. Mol. Biol. 216, 585-610.

Monnerot, M., Solignac, M. and Wolstenholme, D. R. 1990. J. Mol. Evol. 30, 500-508.

Mota, E. M. and Collins, R. A. 1988. Nature 332, 654-656.

Muramatsu, T., Nishikawa, K., Nemoto, F., Kuchino, Y., Nishimura, S., Miyazawa, T. and Yokoyama, S. 1988. Nature 336, 179.

Muscarella, D. E. and Vogt, V. M. 1989. Cell 56, 443-454.

Musier-Forsyth, K., Usman, N., Scaringe,S., Doudna, J., Green, R. and Schimmel, P. 1991. Science 253, 784-786.

Nakagawa, K., Morishima, N. and Shibata, T. 1991. J. Biol. Chem. 266, 1977-1984.

Neefs, J.-M., Van de Peer, Y., De Rijk, P., Goris, A. and De Wachter, R. 1991. Nucleic Acids Res. 19, 1987-2001.

Neff, R. J., Ray, S. A., Benton, W. F. and Wilborn, M. 1964. Methods Cell Physiol. 1, 55-83.

Netzker, R., Kochel, H. G., Basak, N. and Küntzel, H. 1982. Nucleic Acids Res. 10, 4783-4794.

Nichols, J. L. and Lane, B. G. 1966. Biochim. Biophys. Acta 119, 649-651.

Nishimura, S. 1979. In *Transfer RNA: Strucure, Properties and Recognition*, ed. P. R. Schimmel, D. Söll, J. N. Abelson, pp. 59-79. New York: Cold Spring Harbor Laboratory.

Nixon, P. J., Gounaris, K., Coomber, S. A., Hunter, C. N., Dyer, T. A. and Barber, J. 1989. J. Biol. Chem. 264, 14129-14135.

Noller, H. F. 1984. Annu. Rev. Biochem. 53, 119-162.

Noller, H. F. 1991. Annu. Rev. Biochem. 60, 191-227.

Noller, H. F. 1993. FASEB J. 7, 87-89.

Noller, H. F., Hoffarth, V. and Zimniak, L. 1992. Science 256, 1416.

Nugent, J. M. and Palmer, J. D. 1991. Cell 66, 473-481.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T. and Ohyama, K. 1992. J. Mol. Biol. 223, 1-7.

Ohta, E., Oda, K., Yamato, K., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Ohyama, K. and Michel, F. 1993. Nucleic Acids Res. 5 (in press).

Ojala, D., Montoya, J. and Attardi, G. 1981. Nature 290, 470-474.

Okada, N. and Nishimura, S. 1979. J. Biol. Chem. 254, 3061-3066.

Okada, N., Noguchi, S., Kasai, H., Shindo-Okada, N., Ohgi, T., Goto, T. and Nishimura, S. 1979. J. Biol. Chem. 254, 3067-3073.

Okimoto, R., Chamberlin, H. M., Macfarlane, J. L. and Wolstenholme, D. R. 1991. Nucleic Acids Res. 19, 1619-1626.

Okimoto, R., MacFarlane, J. L., Clary, D. O. and Wolstenholme, D. R. 1992. Genetics 130, 471-498.

Okimoto, R. and Wolstenholme, D. R. 1990. EMBO J. 9, 3405-3411.

Olsen, G. J. 1987. Cold Spring Harbor Symp. Quant. Biol. 52, 825-837.

Pace, N. R. 1992. Science 256, 1402-1403.

Pace, N. R., Olsen, G. J. and Woese, C. R. 1986. Cell 45, 325-326.

Palenik, B. and Haselkorn, 1992. Nature 355, 265-267.

Palmer, J. D. 1985. Annu. Rev. Genet. 19, 325-354.

Palmer, J. D. and Herbon, L. A. 1988. J. Mol. Evol. 28, 87-97.

Palmer, J. D. and Shields, C. R. 1984. Nature 307, 437-440.

Pearson, W. R. and Lipman, D. J. 1988. Proc. Natl. Acad. Sci., USA 85, 2444-2448.

Peattie, D. 1979. Proc. Natl. Acad. Sci., USA 76, 1760-1764.

Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A. and Cheng, H. L. 1986. Cell 44, 213-223.

Pelet, T., Currran, J. and Kolakofsky, D. 1991. EMBO J. 10, 443-448.

Pernodet, J.-L., Boccard, F., Alegre, M.-T., Gagnet, J. and Guerineau, M. 1989. Gene 79, 33-46.

Perret, V., Garcia, A., Grosjean, H., Ebel, J. P., Florentz, C. and Giege, R. 1990. Nature 344, 787.

Piccirilli, J. A., McConnell, T. S., Zaug, A. J., Noller, H. F. and Cech, T. R. 1992. Science 256, 1420.

Powell, L. M., Wallis, S. C., Pease, R. J. , Edwards, Y. H., Knott, T. J. and Scott, J. 1987. Cell 50, 831-840.

Pritchard, A. E., Seilhamer, J. J., Mahalingam, R., Sable, C. L., Venuti, S. E. and Cummings, D. J. 1990. Nucleic Acids Res. 18, 173-180.

Pruitt, K. D. and Hanson, M. R. 1989. Curr. Genet. 16, 281-291.

Pütz, J., Puglisi, J. D., Florentz, C. and Giegé, R. 1991. Science 252, 1696-1699.

Queen, C. and Korn, L. J. 1984. Nucleic Acids Res. 12, 581-599.

Quirk, S. M., Bell-Pedersen, D. and Belfort, M. 1989. Cell 56, 455-465.

Raff, R. A. and Mahler, H. R. 1972. Science 177, 575-582.

Raitio, M., Jalli, T. and Saraste, M. 1987. EMBO J. 6, 2825-2833.

Randerath, E., Agrawal, H. P. and Randerath, K. 1981a. Biochem. Biophys. Res. Commun. 103, 739-744.

Randerath, K., Agrawal, H. P. and Randerath, E. 1981b. Biochem. Biophys. Res. Commun. 100, 732-737.

Raué, H. A., Musters, W., Rutgers, C. A., Van T Riet, J. and Planta, R. J. 1990. In *The Ribosome: Structure, Function and Evolution.* eds. W. E. Hill, A. Dahlberg, R. A. Garrett, P. B. Moore, D. Schlessinger and J. R. Warner, American Society for Microbiology, Washington, D. C., pp 217-235.

Raven, P. H. 1970. Science 169, 641-646.

Recipon, H., Perasso, R., Adoutte, A. and Quetier, F. 1992. J. Mol. Evol. 34, 292-303.

Regensburger, A., Ludwig, W., Frank, R., Bloecker, H. and Schleifer, K. H. 1988a. Nucleic Acids Res. 16, 2343-2343.

Regensburger, A., Ludwig, W., Frank, R., Bloecker, H. and Schleifer, K. H. 1988b. Nucleic Acids Res. 16, 2344-2344.

Reinhold-Hurek, B and Shub, D. A. 1992, Nature 357, 173-176.

Rochaix, J. D., Rahire, M. and Michel, F. 1985. Nucleic Acids Res. 13, 975-984.

Rould, M. A., Perona, J. J., Söll, D. and Steitz, T. A. 1989. Science 246, 1135-1142.

Rould, M. A., Perona, J. J. and Steitz, T. A. 1991. Nature 352, 213-218.

Rubin, G. M. 1973. J. Biol. Chem. 248, 3860-3875.

Ruby, S. W. and Abelson, J. 1991. Trends Genet. 7, 79-85.

Saitou, N. and Nei, M. 1987. Mol. Biol. Evol. 4, 406-425.

Saldanha, R., Mohr, G., Belfort, M. and Lambowitz, A. M. 1993. FASEB J. 7, 15-24.

Sampson, J. and Uhlenbeck, O. C. 1988. Proc. Natl. Acad. Sci. USA 85, 1033-1037.

Sampson, J. R., DiRenzo, A. B., Behlen, L. S. and Uhlenbeck, O. C. 1989. Science 243, 1363-1366.

Sanger, F., Nicklen, S. and Coulson, A. R. 1977. Proc. Natl. Acad. Sci., USA 74, 5463-5467.

Satta, Y and Takahata, N. 1990. Proc. Natl. Acad. Sci. USA 87, 9558-9562.

Sawada, M., Muto, A., Iwami, M., Yamao, F. and Osawa, S. 1984. Mol. Gen. Genet. 196, 311-316.

Schimmel, P. R., Söll, D., Abelson, J. N. eds., 1979, Cold Spring Harbour Laboratory, N. Y., pp. 518-519, *Transfer-RNA: Structure, Properties and Recognition.*

Schmelzer, C. and Schweyen, R. J. 1986. Cell 46, 557-565.

Schnare, M. N. 1984. Ph.D. Dissertation, Dalhousie University, Halifax, Nova Scotia.

Schnare, M. N. and Gray, M. W. 1982. Nucleic Acids Res. 10, 3921-3932.

Schnare, M. N. and Gray, M. W. 1990. J. Mol. Biol. 215, 73-83.

Schnare, M. N., Heinonen, T. Y. K., Young, P. G. and Gray, M. W. 1986. J. Biol. Chem. 261, 5187-5193.

Schulman, L. H. 1991. Progr. Nucleic Acid Res. Mol. Biol. 41, 23-87.

Schuster, W. and Brennicke, A. 1987. EMBO J. 2857-2863.

Schuster, W., Ternes, R., Knoop, V., Hiesel, R., Wissinger, B. and Brennicke, A. 1991. Curr Genet 20, 397-404.

Schwartz, R. M. and Dayhoff, M. O. 1981. Annu. N. Y. Acad. Sci. 361, 260-272.

Seilhamer, J. J., Gutell, R. R. and Cummings, D. J. 1984. J. Biol. Chem. 259, 5175-5181.

Seilhamer, J. J., Olsen, G. J. and Cummings, D. J. 1984. J. Biol. Chem. 259, 5167-5172.

Shapleigh, J. P. and Gennis, R. B. 1991. Mol. Microbiol. 6, 635-642.

Sharp, P. 1991. Science 254, 663.

Shaw, J. M., Feagin, J. E., Stuart, K. and Simpson, L. 1988. Cell 53, 401-411.

Shimizu, M., Asahara, H., Tamura, K., Hasegawa, T. and Himeno, H. 1992. J. Mol. Evol. 35, 436-443.

Shub, D. A. and Goodrich-Blair, H. 1992. Cell 71, 183-186.

Siemeister, G., Buchholz, C. and Hachtel, W. 1990. Mol. Gen. Genet. 220, 425-432.

Simpson, L. 1990. Science 250, 512-513.

Simpson, A., M., Suyama, Y., Dewes, H., Campbell, D. A. and Simpson, L. 1989. Nucleic Acids Res. 17, 5427-5445.

Sloof, P., van den Burg, J., Voogd, A., Benne, R., Agostinelli, M., Borst, P., Gutell, R. R. and Noller, H. 1985. Nucleic Acids Res. 13, 4171-4190.

Smith, H. C., Kuo, S.-R., Backus, J. W., Harris, S. G., Sparks, C. E. and Sparks, J. D. 1991. Proc. Natl. Acad. Sci. USA 88, 1489-1493.

Sogin, M. L. and Edman, J. C. 1989. Nucleic Acids Res. 17, 5349-5359.

Sogin, M. L., Elwood, H. J. and Gunderson, J. H. 1986. Proc. Natl. Acad. Sci. USA 83, 1383-1387.

Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A. and Peattie, D. A. 1989. Science 243, 75-77.

Sogin, M. L., Ingold, A., Karlok, M., Nielsen, H. and Engberg, J. 1986. EMBO J. 5, 3625-3630.

Somerville, C. C., Jouannic, S. and Loiseaux-de Göer, S. 1992. J. Mol. Evol. 34, 246-253.

Sommer, B., Kohler, M., Sprengel, R. and Seeburg, P. H. 1991. Cell 67, 11-19.

Sor, F. S. and Fukuhara, H. 1983. Nucleic Acids Res. 11, 339-348.

Spencer, D. F., Schnare, M. N. and Gray, M. W. 1984. Proc. Natl. Acad. Sci. USA 81, 493-497.

Sprinzl, M., Hartmann, T., Weber, J., Blank, J. and Zeidler, R. 1989. Nucleic Acids Res. 17, rl.

Stackebrandt, E. 1991. submitted to Genbank.

Steinmetz, A., Gubbins, E. J. and Bogorad, L. 1982. Nucleic Acids Res. 10, 3027-3037.

Steinmuller, K., Ley, A. C., Steinmetz, A. A., Sayre, R. T. and Bogorad, L. 1989. Mol. Gen. Genet. 216, 60-69.

Stern, D. B. and Palmer, J. D. 1984. Proc. Natl. Acad. Sci. USA 81, 1946-1950.

Stewart, C.-B. 1993. Nature 361, 603-607.

Stuart, K. 1991a. Annu. Rev. Microbiol. 45, 327-344.

Stuart, K. 1991b. Trends Biochem. Sci. 16, 68-72.

Stuart, K. and Feagin, J. E. 1992. Int. Rev. Cytol. 141, 65-88.

Sturm, N. R., Maslov, D. A., Blum, B. and Simpson, L. 1992. Cell 70, 469-476.

Suyama, Y. 1986. Curr. Genet. 10, 411-420.

Swofford, D. L. Laboratory of Molecular Systematics, MRC 534, MSC, Smithsonian Institution, Washington, DC, 20560, USA

Taylor, F. J. R. 1987. Annu. N. Y. Acad. Sci. 503, 1-16.

Tennyson, G. E., Sabatos, C. A., Kiguchi, K., Meglin, N. and Brewer, H. B. 1989. Proc. Natl. Acad. Sci. USA 86, 500-504.

Thomas, S. M., Lamb, R. A. and Paterson, R. G. 1988. Cell 54, 891-902.

Tomioka, N. and Sugiura, M. 1983. Mol. Gen. Genet. 191, 46-50.

Toschka, H. Y., Hoepfl, P., Ludwig, W., Schleifer, K. H., Ulbrich, N. and Erdmann, V. A. 1987. Nucleic Acids Res. 15, 7182-7182.

Toschka, H. Y., Hoepfl, P., Ludwig, W., Schleifer, K. H., Ulbrich, N. and Erdmann, V. A. 1988. Nucleic Acids Res. 16, 2348-2348.

Turmel, M., Boulanger, J., Schnare, M. N., Gray, M. W. and Lemieux, C. 1991. J. Mol. Biol. 218, 293-311.

Turmel, M., Gutell, R. R., Mercier, J.-P., Otis, C. and Lemieux, C. 1993. J. Mol. Biol. in press.

Turner, S., Burger-Wiersma, T., Giovannoni, S. J., Mur, L. R. and Pace, N. R. 1989. Nature 337, 380-382.

Twigg, A. J. and Sherratt, D. 1980. Nature 283, 216-218.

Tzeng, C.-S., Hui, C.-F., Shen, S.-C. and Huang, P. C. 1992. Nucleic Acids Res. 20, 4853-4858.

Urbach, E., Robertson, D. L. and Chisholm, S. W. 1992. Nature 355, 267-270.

van Charldorp, R., Heus, H. A. and van Knippenberg, P. H. 1981. Nucleic Acids Res. 9, 2717-2725.

Van der Boogaart, P., Samallo, J. and Agsteribbe, E. 1982. Nature 298, 187-189.

van der Horst, G. and Tabak, H. F. 1985. Cell 40, 759-766.

van der Veen R., Arnberg, C. A., van der Horst, G., Bonen, L., Tabak, H. F., Grivell, L. A. 1986. Cell 44, 225-234.

Van Etten, R. A., Bird, R. A. and Clayton, D. A. 1983. J. Biol. Chem. 258, 10104-10110.

Vawter, L. and Brown, W. M. 1986. Science 234, 194-196.

Wahleithner, J. A., Macfarlane, J. L. and Wolstenholme, D. R. 1990. Proc. Natl. Acad. Sci. USA 87, 548-552.

Wallace, D. C. 1986. Hospital Practice October 15, 77-92.

Wallace, D. C. 1993. Trends Genet. 9, 128-133.

Walseth, T. F. and Johnson, R. A. 1979. Biochim. Biophys. Acta 562, 11-31.

Waring, R. B. and Davies, R. W. 1984. Gene 28, 277-291.

Waring, R. B., Davies, R. W., Scazzocchio, C. and Brown, T. A. 1982. Proc. Natl. Acad. Sci., USA 79, 6332-6336.

Warrior, R. and Gall, J., 1985. Arch. Sci. 38, 439-445.

Weiller, G., Schueller, C. M. E. and Schweyen. 1989. Mol. Gen. Genet. 218, 272-283.

Weisburg, W. G., Dobson, M. E., Samuel, J. E., Dasch, G. A., Mallavia, L. P., Mandelco, L., Sechrest, J. E., Weiss, E. and Woese, C. R. 1989. J. Bacteriol. 171, 4202-4206.

Weiss-Brummer, B., Rodel, G., Schweyen, R. and Kaudewitz, F. 1982. Cell 29, 527-536.

Wenzlau, J. M., Saldanha, R. J., Butow, R. A. and Perlman, P. S. 1989. Cell 56, 421-430.

Westhof, E., Dumas, P. and Moras, D. 1985. J. Mol. Biol. 184, 119-145.

Whatley, F. R. 1981. Annu. N. Y. Acad. Sci. 361, 330-340.

Wilcox, L. W., Lewis, L. A., Fuerst, P. A. and Floyd, G. L. 1992. Mol. Biol. Evol. 9, 1103-1118.

Wild, M. A. and Gall, J. G. 1979. Cell 16, 565-573.

Wilson, A. C., Cann, R. L., Carr, S. M., George, M., Gyllensten, U. B., Helm-Bychowski, K. M., Higuchi, R. G., Palumbi, S. R., Prager, E. M., Sage, R. D. and Stoneking, M. 1985. Biol. J. Linn. Soc. 26, 375-400.

Winkler, M. and Kück, U. 1991. Curr. Genet. 20, 495-502.

Wissinger, B., Brennicke, A. and Schuster, W. 1992. Trends Genet. 8, 322-328.

Wissinger, B., Hiesel, R., Schuster, W. and Brennicke, A. 1988. Mol. Gen. Genet. 212, 56-65.

Wissinger, B., Schuster, W and Brennicke, A. 1991. Cell 65, 473-482.

Woese, C. R. and Fox, G. E. 1977. J. Mol. Evol. 9, 369-371.

Woese, C. R., Gutell, R., Gupta, R. and Noller, H. F. 1983. Microbiol. Rev. 47, 621-669.

Wolfe, K. H., Li, W.-H. and Sharp, P. M. 1987. Proc. Natl. Acad. Sci. USA 84, 9054-9058.

Wolff, G., Burger, G., Lang, B. F. and Kück, U. 1993. Nucleic Acids Res. 21, 719-726.

Wolff, G. and Kück, U. 1990. Curr. Genet. 17, 347-351.

Wolstenholme, D. R. 1992. Int. Rev. Cytol. 141, 173-216.

Woo, N. H., Roe, B. A., and Rich, A. 1980. Nature 286, 346.

Woodson, S. A. and Cech, T. R. 1989. Cell 57, 335-345.

Yang. D., Oyaizu, Y., Oyaizu, H., Olsen, G. J. and Woese, C. R. 1985. Proc. Natl. Acad. Sci. USA 82, 4443-4447.

Yokoyama, S., Watanabe, T., Murao, K., Ishikura, H., Yamaizumi, Z., Nishimura, S. and Miyazwa, T. 1985. Proc. Natl. Acad. Sci. USA 82, 4905-4909.

Young, I. G. and Anderson, S. 1980. Gene 12, 257-265.

Young, I. G., Rogers, B. L., Campbell, H. D., Jaworowski, A. and Shaw, D.C. 1981. Eur. J. Biochem. 116, 165-170.

Zeff, R. A. and Geliebter, J. 1987. Focus 9:2, 1-2.

Zennaro, E., Francisci, S., Ragnini, A., Frontali, L. and Bolotin-Fukuhara, M. 1989. Nucleic Acids Res. 17, 5751-5764.

Zhang, H., Scholl, R., Browse, J. and Somerville, C. 1988. Nucleic Acids Res. 16, 1220.