INFORMATION TO USERS

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

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

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

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

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

MITOCHONDRIAL AND CHLOROPLAST GENOME EVOLUTION IN GREEN ALGAE

by

Aurora M. Nedelcu

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

at

Dalhousie University

Halifax, Nova Scotia

September, 1997

© Copyright by Aurora M. Nedelcu, 1997

National Library of Canada

Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services

Acquisitions et services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

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

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

0-612-36561-1

Canada

DALHOUSIE UNIVERSITY

FACULTY OF GRADUATE STUDIES

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "Mitochondrial and Chloroplast Genome Evolution Within the Green Algal Group"

by Aurora M. Nedelcu

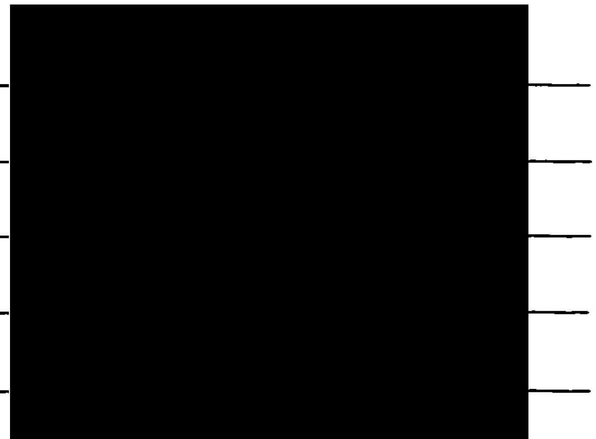
in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated: September 26, 1997

External Examiner

Research Supervisor

Examining Committee



DALHOUSIE UNIVERSITY

DATE: November 11, 1997

AUTHOR: AURORA M. NEDELCU

TITLE: MITOCHONDRIAL AND CHLOROPLAST GENOME EVOLUTION IN GREEN ALGAE

DEPARTMENT OR SCHOOL: BIOLOGY

DEGREE: PhD CONVOCATION: Spring YEAR: 1998

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

Signature of Author

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

The author attests that permission has been obtained for the use of any copyrighted material appearing in this thesis (other than brief excerpts requiring only proper acknowledgement in scholarly writing), and that all such use is clearly acknowledged.

To my parents,

Frusina and Constantin

Table of Contents

Page

Table of Contents.....v

List of Tables.....xiii

List of Figures.....xiv

Abstract.....xvii

List of Abbreviations.....xviii

Acknowledgments.....xix

Preface.....xx

General Introduction.....1

A phylogenetic framework of green algae.....1

Monophyletic versus polyphyletic origin of mitochondria and plastids: the green algal case.....7

Questions and Goals.....11

Chapter 1

Contrasting mitochondrial genome organizations and sequence affiliations among green algae: Potential factors, mechanisms, and evolutionary scenarios.....13

1.1. Introduction.....14

1.2. Two very distinct mitochondrial genome types among the known green algal lineages.....16

1.3. Green algal mitochondrial genomes resemble distinct counterparts.....20

1.4. Potential factors and mechanisms involved in the evolution of the chlorophycean mitochondrial genome......25

1.5. Potential evolutionary scenarios to explain the dichotomy in mitochondrial genome organization and affiliation among green algae......30

1.5.1. Distinct evolutionary origins......31

1.5.2. Early divergence......34

1.5.3. Different rates of evolution......37

1.5.3.1. Adaptive pressures......43

1.5.3.2. Life histories......44

1.6. Conclusions......46

Chapter 2

Discontinuous mitochondrial and chloroplast LSU rRNAs among green algae: Phylogenetic implications......48

2.1. Introduction......49

2.2. Materials and Methods......52

2.2.1. Algal cultures and growing conditions......52

2.2.2. Total RNA extraction and fractionation......52

2.2.3. Northern blot hybridization......54

2.2.4. Phylogenetic analyses......55

2.3. Results......55

2.3.1. Mitochondrial LSU rRNAs......56

2.3.2. Chloroplast LSU rRNAs....56

2.3.3. Reconstruction of the evolutionary history of three characters: flagellar apparatus configuration, discontinuous mitochondrial LSU rRNA, and discontinuous chloroplast LSU rRNA....61

2.4. Discussion....62

2.4.1. Continuous versus discontinuous mitochondrial and chloroplast LSU rRNAs among green algae....62

2.4.1.1. Mitochondrial LSU rRNAs....62

2.4.1.2. Chloroplast LSU rRNAs....67

2.4.2. Phylogenetic implications....70

2.5. Conclusions....72

Chapter 3

Evolution of discontinuous mitochondrial LSU and SSU rRNAs among green algae....74

3.1. Introduction....75

3.2. Material and Methods....75

3.2.1. Algal cultures and growing conditions....75

3.2.2. Total RNA extraction and fractionation....76

3.2.3. Northern blot hybridization....77

3.3. Results....77

3.3.1. Discontinuous mitochondrial LSU and SSU rRNAs among

chlorophycean green algae....77

3.4. Discussion....85

3.4.1. A gradual increase in the degree of fragmentation of the mitochondrial LSU and SSU rRNAs within the chlorophycean green algal group....85

3.5. Conclusions....89

Chapter 4

Discontinuous and scrambled mitochondrial rRNA coding regions among green algae: A model for their origin and evolution....90

4.1. Introduction....91

4.2. Materials and Methods....93

4.2.1. DNA sequence similarity analyses....92

4.2.2. Northern blot hybridizations....93

4.3. Results....93

4.3.1. *Scenedesmus obliquus* mitochondrial LSU rRNA coding region analysis....94

4.3.2. *Scenedesmus obliquus* mitochondrial $tRNA^{Ala}$ coding regioon analysis....101

4.4. Discussion....101

4.4.1. Fragmented and scrambled mitochondrial rRNA coding regions among chlorophycean green algae....102

4.4.2. Discontinuous then scrambled versus discontinuous and

scrambled mitochondrial rRNA coding regions......103

4.4.3. Recombination as a possible mechanism responsible for the mitochondrial rRNA gene rearrangements within the chlorophycean group......105

4.5. Conclusions......113

Chapter 5

Short repetitive sequences in green algal mitochondrial genomes: Potential roles in mitochondrial genome evolution......115

5.1. Introduction......116

5.2. Materials and Methods......119

5.3. Results......119

5.3.1. Short repetitive sequences within *Chlamydomonas* mitochondrial DNA sequences......119

5.3.1.1. Short direct repeated sequences......119

5.3.1.2. Short inverted repeated sequences......124

5.3.2. Short repetitive sequences within mitochondrial DNA sequences from other chlorophycean green algae......127

5.3.3. Genomic distribution of short repetitive sequences in green algal mitochondrial genomes......129

5.3.4. GC-rich repetitive sequences with similarity to origins of replication and recombination-related recognition sites......133

5.4. Discussion......137

5.4.1. Short repetitive sequences within green algal mitochondrial genomes....137

5.4.2. Potential roles of short repetitive sequences in the evolution of green algal mitochondrial genomes....140

5.4.2.1. Fragmentation and scrambling of rRNA coding regions....141

5.4.2.2. Genomic rearrangements....143

5.4.2.3. Replication....145

5.4.2.4. Deletions....146

5.4.2.5. Genome conformation....151

5.4.3. Hypothetical scenarios....154

5.5. Conclusions....157

Chapter 6

Introns and intronic *orf*s with potential involvement in the evolution of the chlorophycean mitochondrial genome....158

6.1. Introduction....159

6.2. Material and Methods....160

6.3. Results....161

6.3.1. A degenerate intron group II in the intronless mitochondrial genome of *Chlamydomonas reinhardtii*....161

6.3.2. An intronic *orf* with potential maturase/endonuclease/recombinase activity in the mitochondrial genome of *Chlamydomonas eugametos*....166

6.4. Discussion....171

6.4.1. Potential consequences of a group II intron acquisition for the evolution of the chlorophycean mitochondrial genome....171

6.4.2. Potential roles of intron-encoded polypeptides in the evolution of the chlorophycean mitochondrial genome....179

6.5. Conclusions....183

Chapter 7

Concerted modes and tempos of evolution of mitochondrial and chloroplast genomes in chlamydomonadalean green algae: A comparative analysis....185

7.1. Introduction....185

7.2. Evolution of mitochondrial and chloroplast genome size....187

7.2.1. Factors contributing to variation in genome size....190

7.2.1.1. Changes in intergenic spacer size....191

7.2.1.2. Changes in intron number....192

7.2.1.3. Changes in gene content....194

7.2.1.4. Changes in the amount of repeated DNA....198

7.2.2. Mechanisms possibly involved in the evolution of genome size....201

7.2.2.1. Length mutations....201

7.2.2.2. Intron mobility....204

7.2.2.3. Gene transfer....206

7.3. Evolution of mitochondrial and chloroplast genome organization 208
7.3.1. Mitochondrial and chloroplast genome structure..208
7.3.2. Mitochondrial and chloroplast gene order...209
7.3.2.1. Factors contributing to gene rearrangement...213
7.3.2.2. Mechanisms possiblly involved in gene rearrangement.......................215
7.4. Evolution of mitochondrial and chloroplast gene structure and organization....................217
7.4.1 Intron-containing coding regions..218
7.4.2. Fragmented coding regions...221
7.5. Evolution of mitochondrial and chloroplast DNA sequence 226
7.6. Conclusions..229
General conclusions..232
Appendix...233
References...238

List of Tables

Page

1. Mitochondrial genome size, map (structure), gene content and GC composition among green algal lineages.......17

2. Mitochondrial genome size, map (structure), gene content among *Chlamydomonas*-like, protist, fungal, metazoan, plant and *Prototheca*-like lineages.......22

3. Algal strains, sources and growing media.......53

4. The size and proposed cellular location of the hybridizing LSU rRNA species.......60

5. Characteristics of oligonucleotide probes 1 to 8.......78

6. Size (in nucleotides) of the hybridizing mitochondrial rRNAs detected by oligonucleotide probes 1 to 8.......84

7. Green algal taxa, their phylogenetic affiliation, and the GenBank accession numbers of the mtDNA sequences analyzed in this study.......120

8. Classes, families, and members of short direct repeated sequences in *C. eugametos* mtDNA.......121

9. Factors and their contribution to organellar genome evolution in land plants and *Chlamydomonas*.......231

List of Figures

Page

1. Phylogenetic analyses of nuclear, mitochondrial, and chloroplast small subunit ribosomal RNA (SSU rRNA) sequences....................4

2. Recombination events responsible for the loss of coding regions as well as the fragmentation and scrambling of rRNA genes during the evolution of chlorophycean mitochondrial genome....................29

3. Genetic maps of the *Chlamydomonas reinhardtii* mtDNA and linearized circular-mapping *Chlamydomonas eugametos* mtDNA....................39

4. Linearized genetic maps of *Chlamydomonas reinhardtii* and *Chlamydomonas moewusii* cpDNAs....................41

5. Electrophoretic patterns of total RNA extracted from the study group taxa....................57

6. Fragmentation patterns within the mitochondrial (mt) and chloroplast (cp) LSU rRNA of *Chlamydomonas*, and location of the target regions for oligonucleotide probes 5'-*C*, 3'-*S*, and 3'-*C*....................58

7. Northern blot hybridizations of green algal total RNA fractionated on 1.2% agarose gels, using three synthetic deoxyoligonucleotide probes....................59

8. Reconstruction of the evolutionary history of flagellar apparatus configuration, discontinuous mitochondrial LSU rRNA, and discontinuous chloroplast LSU rRNA....................63

9. The location of the target regions for oligonucleotide probes 1 to 8,

fragmentation patterns within the mitochondrial LSU and SSU rRNA of *C. reinhardtii* and *C. eugametos*, and the size and suggested coding capacity of the putative mitochondrial rRNAs identified....................79

10. Electrophoretic patterns and Northern blot analyses of green algal total RNA fractionated in 1.5% agarose....................81

11. Diagrammatic representation of a mitochondrial LSU rDNA sequence of *S. obliquus*....................95

12. Potential secondary structures of an LSU rRNA fragment and a tRNA (Ala, UGC) deduced from a *S. obliquus* mtDNA sequence....................98

13. A model for the fragmentation, scrambling and excision of mitochondrial rRNA coding regions involving recombinational events mediated by short inverted and direct repeated sequences....................108

14. Hypothetical pathway from conventional continuous mitochondrial rRNA genes to the gene arrangement described in *C. eugametos* mitochondrial genome....................111

15. The position of the short direct and inverted repeated sequences in the *C. eugametos* mtDNA....................123

16. Alignment of the L2/*nad6* intergenic spacer of the *C. moewusii* and *C. eugametos* showing the difference in the number of copies of short direct repetitive sequences between the two interfertile taxa....................125

17. *C. eugametos* mtDNA short inverted repeat sequences, and potential secondary structures of the *C. eugametos* mtDNA SIRSs....................126

18. *C. reinhardtii* short inverted repeated sequences......128

19. Short repeated sequences within a *Polytomella* spp. and a *S. obliquus* mtDNA sequence......130

20. Linear map and linearized map of the linear- and circular-mapping mtDNA of *C. reinhardtii* and *C. eugametos*......131

21. Origin of light strand replication-like structures in *C. reinhardtii* and *C. eugametos* mtDNA......135

22. Recombinatorial events via short inverted and direct repeats resulting in deletions, interchanges, inversions of the coding regions flanked by the repeats, or linearization of a circular genome......148

23. Hypothetical recombination events to account for some of the trends characterizing the green algal mitochondrial genome evolution......155

24. Alignment and secondary structure of the *C. reinhardtii* *rtl*/L_8 LSU rRNA intergenic spacer showing similarity to a group II intron......163

25. Alignment of the *C. eugametos* i5 intronic aminoacid sequence with those of other LAGLI-DADG introni- encoded polypeptides......167

26. Potential involvement of a group II intron in the fragmentation and scrambling of the mitochondrial LSU rRNA gene in *C. reinhardtii*..... 173

Abstract

The present study points out that, in contrast to those of land plants, the mitochondrial and chloroplast genomes in chlamydomonadalean green algae seem to exhibit concerted modes and tempos of evolution. Limited data on mitochondrial genome structure and organization in other green algal lineages suggest, however, contrasting evolutionary patterns among green algae. This work (i) indicates that the observed dichotomy in the organization of mitochondrial ribosomal RNA genes among green algae, i.e., continuous conventional ribosomal RNA genes in *Prototheca wickerhamii*, but highly fragmented and scrambled counterparts in *Chlamydomonas reinhardtii* and *Chlamydomonas eugametos*, is not limited to these lineages but, rather, extends to at least three green algal classes, the Prasinophyceae and Trebouxiophyceae (*sensu* Friedl 1995) on the one hand, and the Chlorophyceae on the other, (ii) suggests that the divergent evolutionary changes undergone by the mitochondrial lineages among green algae could be the result of both distinct genetic potentials as well as changes in the habitat and life history of their green flagellate ancestors, and (iii) proposes factors, mechanisms, and potential evolutionary scenarios to explain the distinct evolutionary patterns and apparent phylogenetic affiliations of the chlamydomonadalean mitochondrial lineages relative to other green algal and land plant counterparts. Potential factors that could be responsible for the distinct evolutionary series of changes undergone by the chlamydomonadalean mitochondrial genome include: (i) accumulation of short inverted and direct GC-rich repetitive sequences with recombinogenic properties, (ii) light strand DNA replication occurring at multiple sites, and (iii) acquisition of specific group I and II introns as well as intronic open reading frames. In addition, this study proposes models and evolutionary scenarios to explain the fragmentation and scrambling of the mitochondrial ribosomal RNA genes, the high level of gene rearrangement, the reduction in gene content, and the genome linearization that occurred in the chlorophycean mitochondrial lineage.

List of Abbreviations

A, adenine

bp, base pair

C, cytosine

°C, degree Celsius

DNA, deoxyribonucleic acid

EDTA, ethylenediaminetetraacetate

G, guanine

kb, kilobase

rRNA, ribosomal RNA

nt, nucleotide(s)

SDS, sodium dodecyl sulfate

T, thymine

Tris, tris(hydroxymethyl) aminomethane

U, uracil

Acknowledgements

I would like to thank my supervisor, Dr. Robert W. Lee, for his guidance and support during the course of this research and writing this thesis, as well as the members of my Supervisory Committee, Dr. M. W. Gray and Dr. L. Zouros, for their advice and encouragement. I also thank Eileen Denovan-Wright and Tamara Western for sharing the complete and partial *Chlamydomonas eugametos* and *Chlamydomonas moewusii* mitochondrial DNA sequences prior to release and publication, David Spencer and Murray Schnare for their suggestions concerning laboratory methodology and analysis of DNA sequence data, as well as Clayton Knight for assistance with generation of figures and illustrations.

This work was supported by a Natural Sciences and Engineering Research Council of Canada grant to R. W. Lee as well as an Isaac Walton Killam Memorial Scholarship and a Dalhousie Graduate Scholarship to A. M. Nedelcu.

Preface

Some of the data contained in Chapters 2, 3, and 4 have been previously published in the following papers:

Nedelcu, A. M. 1997. Fragmented and scrambled mitochondrial ribosomal RNA coding regions among green algae: A model for their origin and evolution. Mol. Biol. Evol. 14:506-517.

Nedelcu, A. M., D. F. Spencer, E. M. Denovan-Wright, and R. W. Lee. 1996. Discontinuous mitochondrial and chloroplast large subunit ribosomal RNAs among green algae: Phylogenetic implications. J. Phycol. 32:103-111.

Permission has been obtained to include data presented in these papers as part of this thesis.

General Introduction

Any attempt to assess features of organelle genome evolution in a given group requires a good understanding of the phylogeny of that group. On the other hand, a better understanding of phylogeny can grow from knowledge about organellar genome evolution. The first section of this general introduction, therefore, will present a phylogenetic framework of green algae focusing, however, only on the information necessary for understanding the phylogenetic position of the green algal lineages investigated in this work. The second section will address the issue of the mono- versus polyphyletic origin of mitochondria and plastids with reference to the green algal case.

A phylogenetic framework of green algae

The phylogeny of green algae is being progressively deciphered and the new information gathered through molecular approaches will probably trigger the reconsideration of the traditional green algal systematics (Chapman and Buchheim 1991). In 1984, Mattox and Stewart proposed a phylogenetic scenario in which they suggested the most important evolutionary events that occurred in the green flagellate ancestors of the four advanced green algal groups, namely, the Charophyceae, Ulvophyceae, Pleurastrophyceae, and Chlorophyceae. All the extant green flagellate taxa that retained ancestral-like features were placed in a distinct group, i.e. the Micromonadophyceae. The authors assumed that in the earliest green flagellates the flagellar apparatus had only one

of each kind of microtubular root. The swarmers of the Charophyceae are considered to be a minimal change from that ancestral condition. The flagellates with a cruciate root system with two of each kind of root (Ulvophyceae, Pleurastrophyceae, and Chlorophyceae) occurred by a doubling of one of the flagella (and the roots associated with it) of an early green flagellate. Flagellates of this type, with basal bodies in a counterclockwise orientation (CCW) gave rise to the Ulvophyceae with a few additional changes. Among other flagellates with a cruciate system, the fusion or interweaving of scales generated a theca, and this resulted in selection toward the phycoplast and collapsing interzonal spindle. The green flagellates that retained the primitive counterclockwise orientation of the basal bodies gave rise to the Pleurastrophyceae. Other CCW flagellates gave rise to the Chlorophyceae by further evolution to a directly opposed (DO) and clockwise (CW) orientation of the basal bodies in their flagellar apparatus.

Cladistic analyses of organismal data are consistent with the evolutionary hypotheses underlying Mattox and Stewart's classification system, and indicate that (i) the Chlorophyceae is a sister group to the Pleurastrophyceae; (ii) the Ulvophyceae is a sister group to the Chloro-/Pleurastrophyceae clade; and (iii) the charophycean group, the Ulvo-/Chloro-/Pleurastrophyceae clade, and the micromonadophycean taxa all emerge from an unresolved node (Kantz et al. 1990).

Furthermore, phylogenetic analyses based on nuclear rDNA sequence data confirm the presence of five main evolutionary lineages among green algae. Similar analyses indicate that the class Chlorophyceae itself consists of two distinct evolutionary lineages (Steinkötter et al. 1994, Friedl 1995) that are consistent with the two flagellar apparatus

configurations described among flagellate chlorophycean taxa, the DO and CW types; in addition, these analyses increasingly reveal more inconsistencies between the phylogenetic position and the polyphyly of many generic- and ordinal-level lineages on the one hand, and the traditional taxonomy on the other (Buchheim and Chapman 1992, Friedl 1995). For instance, phylogenetic analyses using nuclear and chloroplast rDNA sequences clearly show that the flagellate genus *Chlamydomonas* is not a natural assemblage of taxa: multiple lineages exist within the group, some of them containing both *Chlamydomonas* and non-*Chlamydomonas* taxa from distinct families or orders (Buchheim et al. 1990, Buchheim et al. 1996). Furthermore, while some members of the non-flagellate (autosporic) taxa, such as *Scenedesmus obliquus*, traditionally included in the chlorophycean order Chlorococcales, affiliate with DO flagellate chlorophycean taxa (e.g., *Neochloris aquatica*), other members, such as *Prototheca wickerhamii*, form a monophyletic group with advanced lineages of the class Pleurastrophyceae (sensu Mattox and Stewart 1984) (e.g., *Pleurastrum terrestre*) (Wilcox et al. 1992, Steinkötter et al. 1994, Friedl 1995) (Fig. 1A); this latter group was recently defined by Friedl (1995) as a new class, the Trebouxiophyceae. In addition, the primitive green flagellates grouped by Mattox and Stewart (1984) and Moestrup and Throndsen (1988) into the class Micromonadophyceae and Prasinophyceae, respectively, do not form a monophyletic assemblage of taxa (Steinkötter et al. 1994).

Despite the continuous reconsideration of the phylogenetic relationships among green algae, there is no question that the Chlorophyta (green algae) and Embryophyta (land plants) share a recent common ancestor (reviewed by McCourt 1995).

Figure 1. Phylogenetic analyses of A) nuclear, B) mitochondrial and C) chloroplast small subunit ribosomal RNA (SSU rRNA) sequences. Numbers at the nodes connecting the branches are bootstrap values; the bars indicate 10 substitutions per 100 nucleotides. A. Maximum parsimony nuclear SSU rRNA tree (modified from Steinkötter et al. 1994 with the permission of D. Bhattacharya); B. Neighbor-joining mitochondrial SSU rRNA tree (modified from Denovan-Wright et al. 1996 with the permission of R. W. Lee); C. Distance chloroplast SSU rRNA tree (from Gray and Spencer 1996 with the permission of D. F. Spencer).

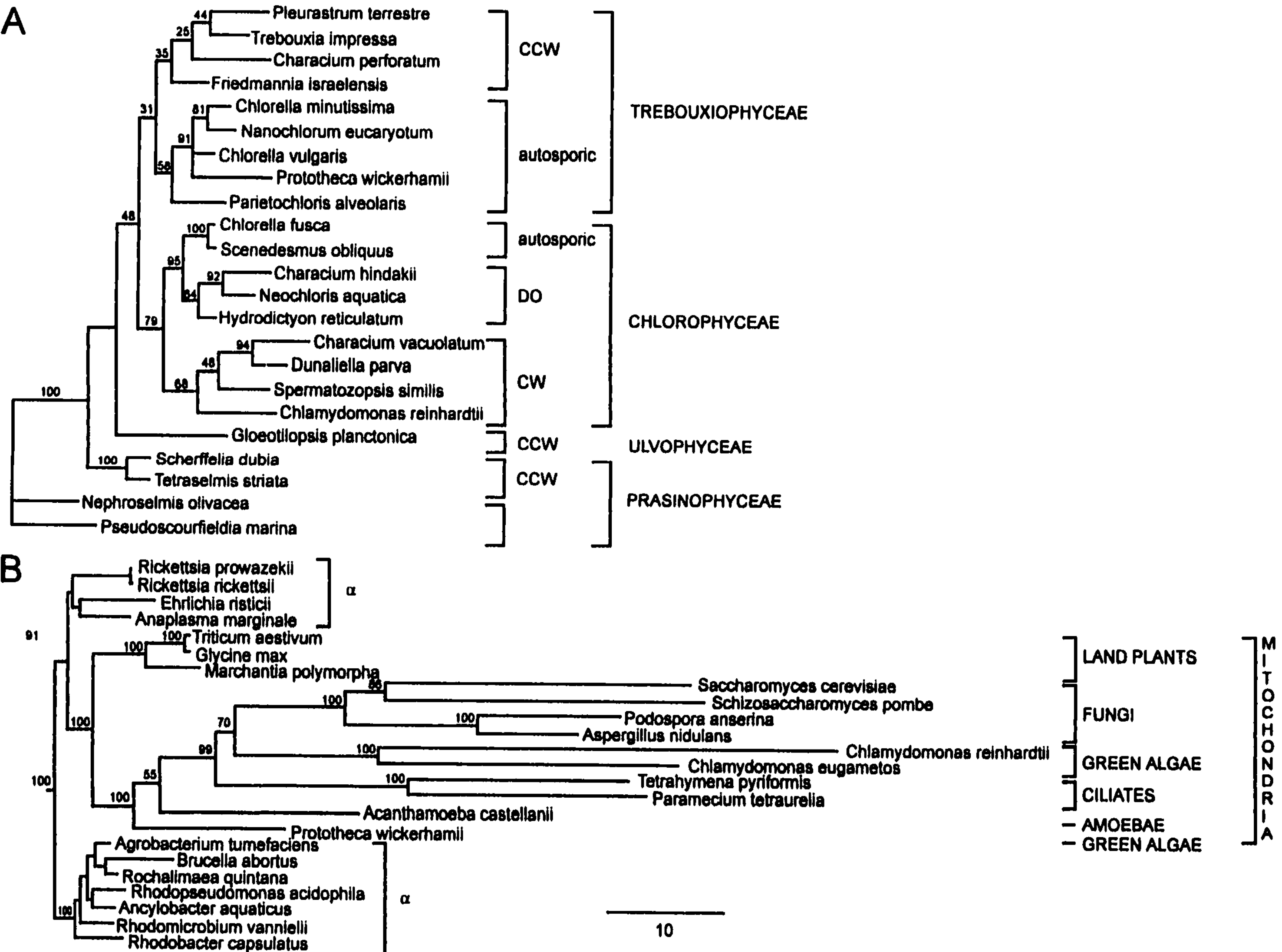


Figure 1

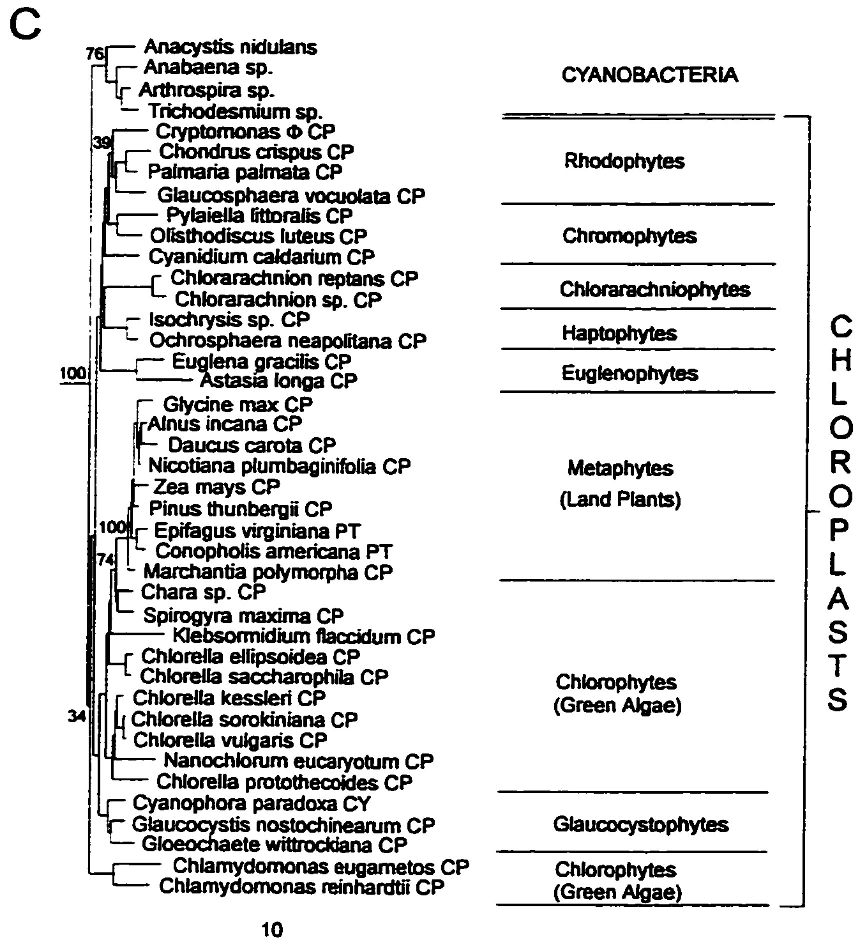


Figure 1

Nevertheless, phylogenetic analyses based on mitochondrial data, both rDNA nucleotide and cytochrome oxidase (COX1) amino acid sequences, reveal an unexpected dichotomy among green algae with respect to their relationships with land plants: although at least in COX1 phylogenetic analyses, some green algae most closely affiliate, as expected, with the land plant group, the chlamydomonadalean lineage branches inconsistently with ciliates, fungi, or animal counterparts in both rDNA and COXI trees (Wolff et al. 1993, Antamarian et al. 1996, Denovan-Wright et al. 1996) (Fig. 1B).

Monophyletic versus polyphyletic origin of mitochondria and plastids: the green algal case

Although the phylogenetic relationships among green algae are not fully understood, land plants appear to be the closest relatives of green algae. However, in an early archaebacterial-eubacterial-chloroplast-mitochondrial-nuclear phylogenetic tree inferred from small subunit ribosomal RNA gene sequence (SSU rDNA) data, the green algal nuclear and mitochondrial sequences (represented by those of *Chlamydomonas reinhardtii*)suggested different phylogenetic affiliations when compared to the land plant counterparts (Gray et al. 1989). In the nuclear subtree, *C. reinhardtii* formed a clade with the plant sequences (as it did also in the chloroplast subtree) and branched off at about the same point as animals and fungi. In contrast, in the mitochondrial subtree, *C. reinhardtii* branched with the ciliate/fungal/animal sequences, far away from land plants, which clustered very near the root, close to the α-proteobacterial clade. The affiliation of the nuclear SSU rRNA gene sequences of higher plants and *C. reinhardtii* was seen as

consistent with traditional phylogenies (Chapman and Ragan 1980, Chapman and Buchheim 1991), whereas the green algal/land plant dichotomy in the mitochondrial tree was interpreted as an anomaly. This anomaly in branching topology was, however, attributed to the plant rather than *C. reinhardtii* mitochondrial sequences and was considered not to be a "treeing artifact" due to the relatively rapid rate of sequence divergence of non-plant mitochondrial rRNA sequences (Gray et al. 1989).

To explain the different branching position of plants within the nuclear and mitochondrial lineages and to account for the strong eubacterial features of their mitochondrial rRNAs, Gray et al. (1989) suggested two possibilities: either (i) the mitochondrial rRNA genes of plants have diverged relatively little from the rRNA genes of the ancient eubacterial ancestor of all mitochondria (monophyletic origin) or (ii) the higher plant mitochondrial rRNA genes or the mitochondrion itself have been acquired more recently than those of other eukaryotic lineages (biphyletic origin). Because all the mitochondria investigated seemed to affiliate with only one subgroup of α-proteobacteria, the biphyletic or polyphyletic concept as used by Gray et al. (1989) did not, however, imply more than one primary original endosymbiosis but, rather, a more recent secondary endosymbiotic event for the land plant mitochondria or its rRNA genes.

In contrast to these results, Van de Peer et al. (1990) presented a phylogenetic tree based on SSU rRNA gene sequences of eukaryotic, archaebacterial, eubacterial, chloroplast, and mitochondrial origin and argued that mitochondria appeared polyphyletic: one cluster contained all the animal mitochondria; a second cluster, was formed by the *C. reinhardtii*, fungal and ciliate mitochondria; and the third cluster was comprised of the

land plant mitochondria and was embedded in the eubacterial cluster with the Proteobacteria α-subgroup as the closest relative.

The input of other green algal mitochondrial rDNA sequences, namely, of *Prototheca wickerhamii* (Wolff and Kück 1990, Wolff et al. 1993) and *Chlamydomonas eugametos* (Denovan-Wright et al. 1996) did not resolve *Chlamydomonas* and *P. wickerhamii* sequences as a green algal clade sharing a most recent common ancestor with the land plants to the exclusion of other groups (Fig. 1B). Phylogenetic trees based on COX1 amino acid sequence suggested, however, that the plant and green algal mitochondrial lineage including the trebouxiophycean (sensu Friedl 1995) taxon, *P. wickerhamii*, and the prasinophycean (sensu Moestrup and Throndsen 1988) flagellate, *Platymonas* (*Tetraselmis*) *subcordiformis*, do form a monophyletic group (Wolff et al. 1993, Kessler and Zetsche 1995). The expected congruency of nuclear, plastid, and mitochondrial phylogenetic trees appears thus verified in the case of trebouxiophycean and prasinophycean green algae and land plants, whereas the chlamydomonadalean taxa branch with land plants in nuclear and chloroplast trees but with very distantly related taxa (e.g., fungi or ciliates) in mitochondrial trees. To explain such findings, Wolff and Kück (1993) suggested a polyphyletic origin for the green algal mitochondria. Furthermore, Gray and Spencer (1996) considered that there is little or no evidence that the land plant and *Chlamydomonas* shared a common mitochondrial ancestor as recently as they shared a common chloroplast or nuclear ancestor; however, they proposed that the differences between the *Chlamydomonas* and *Prototheca*/land plant mitochondrial genome types are "best explained by a relatively rapid and extreme evolution" of the former

genome from the ancestral pattern represented by the more conservative genomes in the latter group.

Similarly, current evidence seems to favour the view of a primary monophyletic cyanobacterial origin of plastids followed by an early subsequent diversification of the accessory pigments (Gray 1993). Nevertheless, it appears that the rhodophyte, cryptophyte and chromophyte plastids are more closely related to each other than to their chlorophyte and land plant counterparts (Douglas 1994). Phylogenetic trees constructed from the small and large subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcS* and *rbcL*) amino acid sequences indicated that the plastids of non-green plants are most closely related to proteobacteria whereas the green algal and land plant counterparts are most closely related to cyanobacteria (Morden et al. 1992, Delwiche et al. 1995); potential explantions for thsi apparent dichotomy are discussed by Gray and Spencer (1996). Phylogenetic analyses of either SSU rDNA nucleotide sequences or translation elongation factors and other protein amino acid sequences are contradictory and do not provide confident support for either a monophyletic or polyphyletic origin of plastids (see Douglas 1994 and Gray and Spencer 1996 for a review). Although the chlorophyte/embryophyte connection is relatively well supported in most chloroplast phylogenetic analyses, it is noteworthy that in a recent phylogenetic analysis, *Chlamydomonas* SSU rDNA nucleotide sequences did not branch as expected with other green algae (Fig. 1A) but, rather, suggested a very early divergence of this lineage relative to all other chloroplast counterparts examined (Gray and Spencer 1996) (Fig. 1C).

Questions and Goals

The questions addressed in this study were: (i) Is the dichotomy in mitochondrial genome sequence and organization among green algae a consequence of distinct evolutionary origins or different evolutionary paths? (Chapter 1); (ii) What are the evolutionary origins and phylogenetic distributions of discontinuous mitochondrial and chloroplast large subunit ribosomal RNAs (LSU rRNAs) among green algae? (Chapter 2); (iii) How did the discontinuous mitochondrial LSU and SSU rRNAs among green algae evolve? (Chapter 3); (iv) What were the factors and mechanisms responsible for the fragmentation and scrambling of mitochondrial rRNA coding regions within the chlorophycean green algal group? (Chapter 4); (v) What was the potential involvement of the short repetitive sequences in the evolution of green algal mitochondrial genomes? (Chapter 5); (vi) What were the potential roles of introns and intronic open reading frames (*orf*s) in the evolution of green algal mitochondrial genomes? (Chapter 6); (vii) Are the modes and tempos of evolution of mitochondrial and chloroplast genomes in green algae similar or different? (Chapter 7).

The goal of this work is to (i) identify features of organelle genome structure, organization, and DNA sequence among green algal mitochondrial and chloroplast lineages that allow us to define evolutionary trends within the group, (ii) investigate distinctive traits among green algal mitochondrial and chloroplast genomes and point out potential factors and mechanisms responsible for their origin and evolution, and (iii)

suggest hypothetical evolutionary scenarios to explain the distinct patterns of mitochondrial genome evolution in the green algal group.

Chapter 1

Contrasting mitochondrial genome organizations and sequence affiliations among green algae: Potential factors, mechanisms, and evolutionary scenarios

1.1. Introduction

It is well accepted now that the eukaryotic cell is an associative system comprising at least two or three main subsystems with different evolutionary histories (Margulis 1981, Gray 1992). Well accepted also is the eubacterial (alpha-proteobacterial and cyanobacterial, respectively) endosymbiotic origin of at least two of the eukaryotic cell's organelles, namely, the mitochondria and plastids, although their mono- or polyphyletic origin is still debated (Dayhoff and Schwartz 1981, Stewart and Mattox 1984, Gray et al. 1989, Lockhart et al. 1992, Morden et al. 1992).

Single endosymbiotic events accounting for the origin of mitochondria and plastids, respectively, would imply that some common ancestral characters should be present in all the extant lineages, and that distinct derived traits should be developed within and shared among related lineages. Furtermore, monophyletic origins for the mitochondria and plastids, respectively, would also require that phylogenies based on organellar traits be consistent with the ones based on nuclear or nucleus-encoded features; in other words all the compartments within an eukaryotic cell should resemble their corresponding counterparts in the same compared lineage.

However, examples of lineages in which the organelles and the nucleo-cytosolic compartment do not suggest the same phylogenetic affiliations have been reported, involving either the plastids (e.g., the cryptomonads [Douglas 1992, McFadden and Gilson 1995] and the protist *Euglena gracilis* [Gibbs 1978, 1981, Morden et al. 1992]), or the mitochondria (e.g., land plants [Gray 1989], the protozoan *Acanthamoeba castellanii*

[Lonergan and Gray 1994], and the green alga *Chlamydomonas reinhardtii* [Gray 1992]). The question to be addressed in such cases is: Are these examples of incongruence between the nuclear and organelle phylogenies the result of very divergent evolutionary patterns among closely related lineages or, rather, of different evolutionary origins?

Secondary eukaryotic endosymbiotic events have been invoked to explain some of the observed incongruences between the phylogenetic affiliations suggested by the plastids on the one hand, and the nucleo-cytosolic compartment, on the other hand (Martin et al. 1992). For the mitochondria, a separate, more recent acquisition event was hypothesized for the land plant mitochondria or at least their mitochondrial rRNA genes (Gray et al. 1989).

An interesting case with which to address these types of evolutionary questions is represented by the green algae. The only three green algal mitochondrial genomes completely sequenced to date, namely those of *C. reinhardtii* (Boer and Gray 1988a, 1988b, 1988c, Gray and Boer 1988, Michaelis et al. 1990), *Chlamydomonas eugametos* (Denovan-Wright, Nedelcu and Lee in press), and *Prototheca wickerhamii* (Wolff et al. 1994) revealed very different mitochondrial genome organizations and sequence affiliations: the former two on the one hand, and the latter on the other, resemble more the ciliate/fungal/animal and plant mitochondrial types, respectively, than one another. In other words, there is an unexpected incongruence between the phylogenetic relationships suggested by the nucleo-cytosolic compartments of these two green algal lineages vis-a-vis their mitochondria.

The present chapter (i) analyzes the information available on other green algal

mitochondrial genomes; (ii) compares the mitochondrial genomes of green algae to the land plant and non-plant counterparts; and (iii) suggests factors, mechanisms and evolutionary scenarios to explain the two very distinct evolutionary patterns among the green algal mitochondrial genomes investigated to date.

1.2. Two very distinct mitochondrial genome types among the known green algal lineages

Information on green algal mitochondrial genomes, although rather limited and incomplete, suggests a fairly large range of genome sizes (from 15.7 kb in *C. reinhardtii* to 80 kb in *Chlorella pyrenoidosa*), and both linear or circular-mapping genomes (Table 1). The gene content is very reduced (i.e., thirteen genes) in *C. reinhardtii* (Michaelis et al. 1990) and *C. eugametos* (Denovan-Wright, Nedelcu and Lee in press), but much larger (i.e., fifty-three genes) in *P. wickerhamii* (Wolff et al. 1994) (Table 1). The genes coding for the subunits 2 and 3 of cytochrome *c* oxidase (*cox*2 and *cox*3) as well as subunits 1, 6, and 9 of the ATP synthase complex (*atp*1, *atp*6, and *atp*9) are missing in *C. reinhardtii* and *C. eugametos* but are present in *P. wickerhamii, P. subcordiformis*, and a *Chlorella*-like taxon (Michaelis et al. 1990, Waddle et al. 1990, Wolff et al. 1994, Kessler and Zetsche 1995, Denovan-Wright, Nedelcu and Lee in press). Only three tRNAs are mitochondrial-encoded in *C. reinhardtii* (Boer and Gray 1988c) and *C. eugametos* (Denovan-Wright, Nedelcu and Lee in press), whereas twenty-six are coded by the *P. wickerhamii* mitochondrial genome (Wolff et al. 1994).

Table 1. Mitochondrial genome size, map (structure), gene content and GC composition among green algal lineages. Cr = *Chlamydomonas reinhardtii*, Cp = *Chlamydomonas pitschmanii*, Cs = *Chlamydomonas smithii*, Pm = *Pandorina morum*, Ce = *Chlamydomonas eugametos*, Cm = *Chlamydomonas moewusii*, Pa = *Polytomella agilis*, Ps = *Platymonas subcordiformis*, So = *Scenedesmus obliquus*, Pw = *Prototheca wickerhamii*, C-l = *Chlorella*-like ex-symbiont, Cpy = *Chlorella pyrenoidosa* (circ=circular, lin=linear).

Mitochondrial traits	*Chlamydomonas*-like taxa								*Prototheca*-like taxa			
	Cr[a]	Cp[b]	Cs[c]	Pm[d]	Ce[e]	Cm[f]	Pa[g]	So[h]	Ps[i]	Pw[j]	C-l[k]	Cpy[l]
Size (kb)	15.7	16.5	17	20	22.9	22		45	42.8	55.3	76	80
Genome map	lin	circ	lin	lin	circ	circ		circ	circ	circ	circ	circ
Gene content												
cob	+		+		+	+			+	+	+	
cox1	+		+		+	+	+	+	+	+	+	
cox2	-		-		-	-			+	+	+	
cox3	-		-		-	-			+	+		
atp1	-		-		-	-			+	+	+	
atp6	-		-		-	-			+	+	+	
atp9	-		-		-	-				+	+	
nad1	+		+		+	+				+		
nad2	+		+		+	+			+	+		
nad3	-		-		-	-			+	+		
nad4	+		+		+	+	+		+	+		
nad4L	-		-		-	-				+		
nad5	+		+		+	+			+	+		
nad6	+		+		+	+			+	+		
rnl	+		+	+	+	+	+	+	+	+	+	
rns	+		+		+			+	+	+	+	
rrn5	-		-		-	-				+		
trn	3				3				7	22		

rpl+rps	0		+		0					2	11	
Fragmented rRNAs	+		+	+	+	+	+	+		-	-	
GC content (%)	47.5				34.6					25.8	32.5	40.0

[a] Grant and Chiang 1980, Michaelis et al. 1990, [b] Boudreau and Turmel 1995, [c] Boynton et al. 1987, [d] Moore and Coleman 1989, [e] Denovan-Wright, Nedelcu and Lee in press, [f] Lee et al. 1991, [g] Antamarian et al. 1996, [h] Kück 1989, [i] Kessler and Zetsche 1995, [j] Wolff et al. 1994, [k] Waddle et al. 1990, [l] Bayen an Rode 1973.

The mitochondrial rRNA coding regions are continuous in *P. wickerhamii*, *Platymonas subcordiformis* (Wolff et al. 1993, Kessler and Zetsche 1995), but fragmented into coding modules that are scrambled along the genome in *C. reinhardtii*, *C. eugametos*, and *Scenedesmus obliquus* (Boer and Gray 1988a, Denovan-Wright and Lee 1994, Nedelcu 1997). The presence (or absence) of a mitochondrial 5S rRNA and ribosomal protein genes is an additional distinctive feature among green algal lineages; *P. wickerhamii* has both a 5S rRNA gene and many ribosomal protein coding genes, *P. subcordiformis* has at least a few ribosomal protein genes, whereas *C. reinhardtii* and *C. eugametos* have neither ribosomal protein nor 5S rRNA genes (Michaelis et al. 1990, Wolff et al. 1994, Kessler and Zetsche 1995, Denovan-Wright, Nedelcu and Lee in press). The GC content varies from 25.8% in *P. wickerhamii* to 47.5% in *C. reinhardtii* (Table 1).

In pairwise rRNA nucleotide sequence comparisons the *P. wickerhamii* genes appear more related to their land plant homologs (84-86% sequence identity) than do the *Chlamydomonas* counterparts (only 65% sequence identity) (Gray 1995). Similarly, pairwise comparisons of an incomplete sequence (about 400 nucleotides within the 3'-half of the mitochondrial LSU rRNA) of another chlorophycean taxon, *S. obliquus*, show only 67% sequence identity of this algal sequence with the wheat or *P. wickerhamii* mitochondrial LSU rRNA counterparts (Gray 1995).

The current data suggest, therefore, that the known green algal mitochondrial genomes fall into two very distinct types: *Chlamydomonas*-like and *Prototheca*-like mitochondrial genomes (Table 1). The *Chlamydomonas*-like mitochondrial genomes are small, have a reduced gene content (no ribosomal protein and 5S rRNA genes, only few

protein-coding and tRNA genes) and fragmented and scrambled rRNA coding regions, whereas the *Prototheca*-like mitochondrial genomes are larger, have a larger set of protein-coding genes including ribosomal protein genes, a higher number of tRNA genes as well as continuous conventional SSU and LSU rRNA coding regions and a 5S rRNA gene (*rns*, *rnl*, and *rrn5*, respectively). Features of mitochondrial genome organization as well as the presence of discontinuous mitochondrial rRNAs in chlorophycean lineages, such as several *Chlamydomonas* taxa, *Polytomella agilis*, *Carteria crucifera*, *Carteria olivieri*, *Planophila terrestris*, *Hormotilopsis gelatinosa*, *Neochloris aquatica* and *Scenedesmus obliquus* (Denovan-Wright et al. 1996, Nedelcu et al. 1996, Nedelcu 1997), suggest that they most likely resemble the *Chlamydomonas* type. In contrast, the mitochondrial genome of prasinophycean and trebouxiophycean (sensu Friedl 1995) algae, such as *Platymonas subcordiformis*, *Pyramimonas parkae*, *Hafniomonas montana*, *Chlorella vulgaris* and *Pleurastrum terrestre* (Kessler and Zetsche 1995, Nedelcu et al. 1996), appear to share some of the distinctive features of the *Prototheca* mitochondrial genome type. It seems, therefore, that the previously observed great evolutionary distance between the mitochondrial genomes of *C. reinhardtii* and *P. wickerhamii*, extends in fact to at least two lineages, the chlorophycean and prasinophycean/trebouxiophycean, respectively.

1.3. Green algal mitochondrial genomes resemble distinct counterparts

The dichotomy in mitochondrial genome organization between the

Chlamydomonas-like and *Prototheca*-like types is also reflected in different degrees of resemblance to their land plant counterparts (Table 2), considered their closest relatives at the nucleo-cytosolic level (Ragan and Chapman 1978, Chapman and Buchheim 1992). Whereas the mitochondrial genomes of *C. reinhardtii* and *C. eugametos* (and, most likely, of all the chlorophycean taxa) share no specific features with their land plant counterparts, those of *P. wickerhamii* (and, most likely, of other trebouxiophycean taxa) display many of the embryophyte traits such as: (i) the presence of *atp*1 and *rrn5* (the only other non-plant taxa that posses *atp*1 or both *atp*1 and *rrn5* are the protozoans *Acanthamoeba castellanii* and *Reclinomonas americana*, respectively [Burger et al. 1995, Lang et al. 1996, Lang et al. 1997]), (ii) a virtually identical set of ribosomal protein genes arrayed in a similar manner, (iii) three *orf* homologs, (iv) conventional continuous rRNA genes, and (v) group I introns in the *cox*1 gene at the same positions as in the liverwort mitochondrial DNA (mtDNA) (Gray 1995).

The same molecular data that support the split between the land plant/*Prototheca*-like and *Chlamydomonas*-like mitochondrial genome types also emphasize the resemblance between the *Chlamydomonas* and non-plant mitochondrial lineages. The available data from kinetoplastids, apicomplexans, and ciliates (see Wolstenholme and Fauron 1995 for a review and references) suggest, surprisingly enough, that their mitochondrial genomes share with *Chlamydomonas* mitochondrial genome type traits that are absent both in the *Prototheca* and land plant counterparts.

Of the 23 features of mitochondrial genome organization compared in Table 2, thirteen are different between the *Chlamydomonas*-like and *Prototheca*-like genomes, and

Table 2. Mitochondrial genome size, map (structure), and gene content among *Chlamydomonas*-like, protist, fungal, metazoan, plant and *Prototheca*-like lineages (/ or [ ] indicatea that both traits, or only exceptions, respectively, have been reported; circ=circular, lin=linear).

Mitochondrial traits	Lineages					
	Chlamydomonas-like[a]	Protists[b]	Fungi[c]	Metazoans[d]	Plants[e]	*Prototheca*-like[f]
Size (kb)	15.7-45	6-69	19.4-100	14-42	200-2500	42.8-80
Genome map	circ/lin	circ/lin	circ[lin]	circ[lin]	circ	circ
Gene content						
cob	+	+	+	+	+	+
cox1	+	+	+	+	+	+
cox2	-	+[-]	+	+	+	+
cox3	-	+[-]	+	+	+	+
atp1	-	-[+]	-	-	+	+
atp6	-	-/+	+	+	+	+
atp9	-	-/+	+/-	-	+	+
nad1	+	+[-]	+[-]	+	+	+
nad2	+	+/-	+[-]	+	+	+
nad3	-	+/-	+[-]	+	+	+
nad4	+	+[-]	+[-]	+	+	+
nad4L	-	-[+]	+[-]	+	+	+
nad5	+	+[-]	+[-]	+	+	+
nad6	+	-[+]	+[-]	+	+	+
rnl	+	+	+	+	+	+
rns	+	+	+	+	+	+
rrn5	-	-[+]	-	-	+	+
trn	3	0-3[26]	25-28	22[1-2]	16-29	22
rpl+rps	0	0-4[53]	0-1	0	10	11

Fragmented rRNA coding regions	+	+/-	-	-	-	-

[a] Grant and Chiang 1980, Moore and Coleman 1989, Michaelis et al. 1990, Lee et al. 1991, Denovan-Wright and Lee 1994, Boudreau and Turmel 1995, Denovan-Wright, Nedelcu and Lee in press, [b] see Wolstenholme and Fauron 1995 for references, Burger et al. 1995, Lang et al. 1996, [c] see Wolstenholme and Fauron 1995 for references, [d] see Wolstenholme and fauron 1995 for references, Gjetvaj et al. 1992, [e] see Wolstenholme and Fauron 1995 for references, [f] Bayen and Rode 1973, Waddle et al. 1990, Coleman and Goff 1991, Wolff et al. 1994, Kessler and Zetsche 1995.

in each of these cases the character status present in *Chlamydomonas*-like mitochondrial genomes is also shared by at least one of the non-land plant lineages and absent from the land plant homologs (see Table 2). The *cox*2 and *cox*3 genes that are missing in the mitochondrial genomes of *C. reinhardtii* and *C. eugametos* but present in those of *Prototheca wickerhamii* and *Platymonas subcordiformis* are also missing in those of the apicomplexans (*Plasmodium falciparum*, *Theileria parva*) and the ciliates (*Paramecium aurelia*, *Tetrahymena pyriformis*) respectively; the apicomplexans and the ciliates are the only other reported examples in which the subunit 2 or 3 of the cytochrome *c* oxidase, respectively, is not mitochondrial-encoded. Whereas the mitochondrial genome of *P. wickerhamii* encodes three of the ATPase subunits, those of *C. reinhardtii, C. eugametos* and *Plasmodium falciparum* do not encode any of the ATPase subunits, and only one is encoded in the mitochondria of *Paramecium aurelia* and kinetoplastids. The gene coding for the subunit 4L of the NADH dehydrogenase (*nad4L)* is also missing in *C. reinhardtii, C. eugametos, Paramecium aurelia*, apicomplexans, and kinetoplastids, but present in *Prototheca wickerhamii* and land plants. Another striking resemblance is the lack of *atp1* and *rrn5* in the mitochondrial genome of *C. reinhardtii, C. eugametos, Paramecium aurelia*, apicomplexans, kinetoplastids, fungi and animals, in contrast to its presence in *Prototheca wickerhamii* and embryophytes. Only three tRNAs are encoded in the mitochondrial genomes of *C. reinhardtii, C. eugametos* and *Paramecium aurelia*, one or two in those of cnidarians and none in those of kinetoplastids, whereas *Prototheca wickerhamii* and liverwort mitochondrial genomes encode twenty-six and twenty-nine tRNAs, respectively. None of the mitochondrial ribosomal proteins is mitochondrially

encoded in *Chlamydomonas*, none or one in kinetoplastids, apicomplexans, fungi and animals, only 7 in *Paramecium aurelia* but thirteen and sixteen in *Prototheca wickerhamii* and liverwort, respectively. Fragmented and scrambled mitochondrial rRNA coding regions described in the *Chlamydomonas*-like genomes were also found in the ciliate *Tetrahymena pyriformis*, the apicomplexans *Plasmodium falciparum*, *Plasmodium sp*, and *Theileria parva* (see Nedelcu 1997 for references), but continuous rRNA genes are present in *Prototheca wickerhamii, Platymonas subcordiformis* and land plants.

In addition, the very abundant set of short GC-rich repeat clusters in the *C. reinhardtii* (Boer and Gray 1991) and *C. eugametos* (Denovan-Wright, Nedelcu and Lee in press, Nedelcu and Lee submitted) mitochondrial intergenic spacers and introns resembles in organization, base composition, and DNA sequence their counterparts in fungal mtDNAs (Nedelcu and Lee submitted and references therein) (discussed in Chapter 5). Moreover, the *C. reinhardtii* mitochondrial genome resembles the metazoan mitochondrial genomes in showing extensive physical and transcriptional linkage of genes, processing of long co-transcripts by discrete endonucleolytic scissions, and the absence of 5'-untranslated regions in the mature mRNAs (Gray et al. 1989).

1.4. Potential factors and mechanisms involved in the evolution of chlorophycean mitochondrial genome

If all the extant green algal mitochondria evolved from a single common ancestor (monophyletic origin), the questions to be addressed are: (i) Why are the

Chlamydomonas-like and *Prototheca*-like mitochondrial lineages so different in so many respects? and (ii) Why does the evolutionary trend in the *Chlamydomonas*-like mitochondrial lineage seem so similar at least as far as the end result (i.e. reduced genome size and gene content, similar gene organization and expression, DNA sequence) to the ciliate/fungal/animal lineage?

Similar evolutionary patterns could be consequences of either convergent evolution or common ancestry, thus the similar characters could be either analogous or homologous, respectively. In the case of convergent evolution, similar evolutionary patterns develop independently in distinct evolutionary lineages as the result of similar adaptative pressures or constraints acting on different genetic potentials. On the other hand, if the similarities observed between the *Chlamydomonas*-like and the ciliate/fungal/animal mitochondrial lineages are homologous rather than analogous, one has to presume that the mitochondria in *Chlamydomonas*-like lineage has a more recent common ancestry with the ciliate/fungal/animal mitochondrial lineage than it does with the *Prototheca*-like/land plant counterparts. In the possibility that the *Chlamydomonas*-like mitochondrial lineage had a different evolutionary origin than the other green algal lineages, the question would remain: Is this an example of secondary acquisition (as previously proposed by Gray et al. [1989] for the land plant mitochondrial lineage) or independent primary endosymbiosis (i.e. polyphyletic origin)?

Whether the two distinct green algal mitochondrial lineages, i.e. *Chlamydomonas*-like and *Prototheca*-like, do or do not share a more recent common ancestor with one another and with land plants than with other lineages remains to be answered. It is quite

obvious, however, that the two green algal mitochondrial lineages followed two very distinct evolutionary paths after their divergence from either the common green algal or the unknown ancestor. The questions to be then addressed are how and why these changes occurred and evolved.

The most distinctive features of the *Chlamydomonas*-like mitochondrial genome, relative to the *Prototheca*-like counterpart, include (i) a reduced gene content and (ii) the presence of fragmented and scrambled rRNA coding regions. Although more data are needed before the factors and mechanisms responsible for the origin and evolution of these two traits are deciphered, a few suggestions can be made.

The accumulation of short direct repeated sequences in the intergenic spacers of *Chlamydomonas*-like mitochondrial genomes has been proposed to have triggered intramolecular recombination events resulting in both the excision of subgenomic circles containing protein-, tRNA- or rRNA-coding regions as well as the fragmentation and scrambling of the rRNA genes in this lineage (Nedelcu 1997, Nedelcu and Lee submitted) (discussed in Chapter 4 and 5) (Fig. 2). Although short repeated sequences were found in the intergenic spacers of both *Chlamydomonas*-like and *P. wickerhamii* mitochondrial genomes, the repeated elements are mostly GC-rich in the first (Boer and Gray 1991, Denovan-Wright, Nedelcu and Lee in press, Nedelcu and Lee submitted) but highly AT-rich in the second (Wolff et al. 1994). If the different composition of the repeated elements in *Chlamydomonas*-like relative to *P. wickerhamii* mitochondrial genome (discussed in Chapter 5) is a feature potentially involved in their recombinogenic activity, the prediction would then be that repeated elements in mitochondrial genomes from one

or the other green algal lineage would reveal the same GC bias as in *Chlamydomonas* or *Prototheca*, respectively.

Nevertheless, the processes accounting for the transfer of genetic information into the nucleus remain to be deciphered. Theoretically, there are two ways one can envision such a transfer: at the DNA or RNA level. The transfer of a DNA molecule from one compartment to another could be comparable to the transfer of an episome from one eubacterial cell to another, following the excision, conjugation, and integration steps described in a transformation cycle. Alternatively, the genetic information transcribed into an RNA molecule could be reverse-transcribed into a DNA molecule either before or after leaving the organelle, and subsequently integrated into the nuclear DNA. It is noteworthy that gene transfer from the mitochondria to the nucleus seems to be an on-going process among flowering plants and at least in this group, the transfer seems to happen at the RNA level, because the nuclear copies resemble more the edited rather than unedited versions of the mitochondrial genes (Brennicke et al. 1993, Schuster and Brennicke 1994, Gray 1995).

Another distinctive feature potentially involved in the different evolutionary changes undergone by the *Chlamydomonas* and *Prototheca* mitochondrial genome types is the presence of a reverse transcriptase-like (*rtl*) coding region in the mitochondrial genome of *C. reinhardtii*. A putative role of such an enzyme in the evolution of the *Chlamydomonas*-like mitochondrial genomes would be supported by the presence of the corresponding gene in the *Chlamydomonas*-like mitochondrial genomes but not in *Prototheca*-like counterparts. To date, although rather few green algal mitochondrial

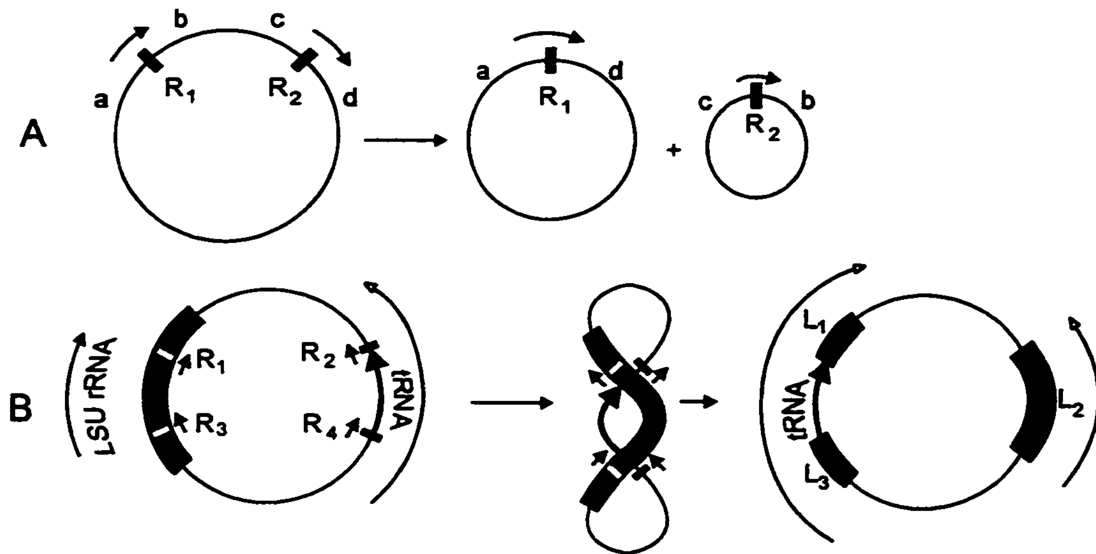


Figure 2. Recombination events between A) a two-copy (R_1/R_2) short direct repeat, and B) two sets of two-copy short inverted repeats (R_1/R_2 and R_3/R_4) resulting in the excision of a coding region and the fragmentation and scrambling of a LSU rRNA coding region, respectively; a, b, c, d are sequences flanking the repeats, and L_1, L_2, L_3 are LSU rRNA coding modules.

genomes have been investigated, a reverse transcriptase-like gene has been only found in *C. reinhardtii* (Boer and Gray 1988b), although its overexpressed product does not appear to have a reverse transcriptase activity (Faßbender et al. 1994). The finding that the *C. reinhardtii* mitochondrial *rtl*-like coding region might be in fact the remnant of an *orf* previously harboured by a group II intron (Chapter 6) argue for a potential involvement of mobile group II introns in the evolution of chlorophycean mitochondrial genomes (discussed in Chapter 6).

As the information from other green algal lineages becomes available, the comparisons among the distinct patterns of mitochondrial genome organization will hopefully provide insights into the mechanisms involved in the evolution of mitochondrial genomes in green algae. Distinguishing among the possible mechanisms involved in the evolution of the mitochondrial genomes in the *Chlamydomonas*-like lineage would indicate why only this lineage underwent such dramatic evolutionary changes, provided that in the *Prototheca*-like lineage those particular mechanisms were not functional, either because the substrate (recombinogenic repeats or reverse transcriptase-like genes, for example) or additional features required for such mechanisms (e.g., enzymes involved in recombination or facilitating the transfer of genetic information into the nucleus) were absent. Conversely, knowing more about the distinctive features of these two mitochondrial genome types would allow a better understanding of the mechanisms involved in the evolution of the mitochondrial genomes in green algae.

1.5. Potential evolutionary scenarios to explain the dichotomy in mitochondrial genome

organization and affiliation among green algae

1.5.1. Distinct evolutionary origins

Based on mitochondrial gene content and genome organization comparisons, Gray and Spencer (1996) concluded that "there is little or no evidence from these data that the land plant *M. polymorpha* and the green alga *C. reinhardtii* shared a common ancestor as recently as they shared a common chloroplast or nuclear ancestor." Nevertheless, a polyphyletic origin for the green algal mitochondria has not been favoured.

If the green algal mitochondria were monophyletic, the mitochondrial genome in the most recent common ancestor of the chlorophycean group should retain ancestral (symplesiomorphic) traits present in the common green algal ancestor but also share distinctive derived (synapomorphic) characters with members of the chlorophycean lineage, traits that are unique to the *Chlamydomonas* mitochondrial type. For example, the presence of fragmented mitochondrial rRNA genes is most likely a synapomorphic character, since it seems to be derived within, and therefore shared by the members of the chlorophycean lineages only (Nedelcu et al. 1996). The finding of a primitive green algal taxon whose mitochondrial genome both contains the *atp*1 and/or *rrn5* (that seem to be symplesiomorphic characters) and displays fragmented and scrambled mitochondrial rRNA genes whose nucleotide sequences affiliate the *Chlamydomonas*-like and *Prototheca*-like mitochondrial lineages would argue for the evolution of these two mitochondrial types from a recent common ancestor. If, however, that taxon does not

have any of the ancestral features of the green algal/embryophyte mitochondrial lineage and its rRNA sequences fail to connect the *Prototheca*-like and *Chlamydomonas*-like lineages to the exclusion of other non-green algal groups, a separate evolutionary origin for the *Chlamydomonas*-like mitochondria might have to be considered.

Although multiple independent endosymbiotic events leading to the organelle genetic diversity in the present eukaryotic lineages are thought to be less parsimonious, they should not be fully disregarded. The following arguments can be considered: (i) similarities in organelle genome organization among unrelated lineages, as well as striking differences among lineages within the same group; (ii) difficulty in explaining the distribution of many traits by means other than implying multiple independent losses or acquisitions of characters in distant evolutionary lineages; (iii) the diversity and dynamic of present endosymbiotic associations, not only in terms of eukaryotic host and prokaryotic symbiont but also of the interactions evolved (see Corliss 1990); (iv) the diversity and most likely genetic instability of prokaryotic life in ancient times. There is no obvious reason to assume that only one individual or even only one population of alpha-proteobacteria invaded only one particular protoeukaryotic cell and the present mitochondrial diversity is a consequence only of different selective pressures acting on the same genetic potential. It seems as likely that several populations of slightly different alpha-proteobacterial strains invaded distinct host populations, and the further interactions between the two components co-evolving were shaped by the distinct genetic potential carried by both partners previous to the association as well as by new adaptative pressures. For the chloroplast, such a polyphyletic origin from closely related

cyanobacteria has been considered to be "very difficult, if not impossible, to discern" from a monophyletic ancestry (Delwiche et al. 1995, cited by Gray and Spencer 1996).

The main argument against a polyphyletic origin of the mitochondria seems to be the fact that all the mitochondrial rRNA sequences available to date affiliate with only one subgroup of alpha-proteobacteria (Gray and Spencer 1996). Nevertheless, there are two aspects that could be challenged.

First, the fact that only one alpha-proteobacterial subgroup was found with which all the mitochondrial rRNA sequences cluster does not necessarily mean that another subgroup to which some rRNA sequences would more closely affiliate does or did not exist. Although considered as evidence in favour of the importance of slowly evolving sequences in properly positioning the more rapidly evolving rRNA sequences, it is noteworthy that the mitochondrial rRNA sequences from animals, fungi, and ciliates cluster with the alpha-proteobacterial counterparts only when their land plant homologs are included in the analysis (Gray 1995). Moreover, the inclusion of more mitochondrial and alpha-proteobacterial rRNA sequences as well as of the sequences from two ciliate alpha-proteobacterial endosymbionts, *Holospora* and *Caedibacter*, decreased the bootstrap value of the mitochondrial/rickettsial alpha-proteobacterial node from 95% to 73% and 41%, respectively (Gray and Spencer 1996). Theoretically, there is no reason to disregard the possibility that the inclusion of new alpha-proteobacterial sequences will decrease the support value of the rickettsial/mitochondrial node to the point that some of the mitochondrial sequences would affiliate more closely to another alpha-proteobacterial subgroup.

Secondly, the fact that all the mitochondrial rRNA gene sequences cluster together does not necessarily mean that the respective mitochondria evolved from one single alpha-proteobacterial ancestor; it may as well be that the present mitochondria are the descendants of a few slightly different invading alpha-proteobacterial strains that shared a common ancestor as recorded in their rRNA sequences but whose genomes already developed some of the distinctive features (error-prone replication mechanisms, deficient postreplication repair systems or copy-correction mechanisms, a novel gene or molecular mechanism, for instance) that triggered the changes (e.g., a high rate of nucleotide substitution or increased efficiency of recombination) responsible for the further evolution of their respective mitochondrial-to be genomes.

7.5.2. Early divergence

To explain the observed differences in mitochondrial genome organization and affiliation between the two green algal mitochondrial lineages, Gray (1995) suggested that a detailed characterization of mtDNA from other chlorococcalean taxa occupying a phylogenetic position between *C. reinhardtii* and *P. wickerhamii* is needed. Nevertheless, such an approach may not provide the information needed, at least because the two taxa belong most likely to two distinct evolutionary lineages, the chlorophycean and trebouxiophycean (sensu Friedl 1995), respectively, whose divergence is probably very old. Therefore, the answer is most likely going to come from members of the Micromonadophyceae class (sensu Mattox and Stewart 1984), a pool of primitive green

algae that are considered similar to forms from which higher green algae evolved (Mattox and Stewart 1984). Which micromonadophycean lineages might yield evidence about the phylogenetic relationships among the extant green algal lineages is difficult to assess, because the members of this group have different evolutionary origins and the phylogenetic relationships among taxa are not fully understood.

Mattox and Stewart (1984) proposed that: (i) the different lines of evolution of higher green algae are different from each other because they had independent origins from different types of green flagellates, and (ii) the most fundamental evolutionary changes occurred among green flagellates rather than after the origin of the higher green algal groups. There are at least two pieces of evidence that could argue for an early divergence of the *Chlamydomonas* lineage. First, phylogenetic analyses using nuclear 5S rRNA sequences suggested that *Chlamydomonas* may represent a very early divergence among green algae, basal to the ulvophycean and charophycean taxa (Hori et al. 1985, Hori and Osawa 1987). Moreover, based on the observation that the *Chlamydomonas* nuclear 5S rRNAs are sufficiently distinct in both primary and secondary structure from those of other green algae and plants (Darlix and Rochaix 1981) and diverge much deeper than expected, Devereux et al. (1990) proposed that *Chlamydomonas* may be diverged enough to be considered a major group of green algae. Despite the fact that the 5S rRNA trees constructed using various methods differed one from another considerably in terms of fine details and the authors accepted that 5S rRNA data lack the resolution to address all phylogenetic issues, all the tree-making methods consistently resolved *Chlamydomonas* as a very early lineage relative to other chlorophycean, charophycean and higher plant

taxa. These observations appear, however, to be in contrast to the usual interpretation of the mitotic and flagellar apparatus of *Chlamydomonas*, which are considered to be in a derived condition among green algae.

Secondly, phylogenetic analyses using different nuclear SSU and LSU rRNA data sets assigned chlamydomonadalean lineages to various positions among the other green algal taxa available without being able to definitely resolve their phylogenetic position (Buchheim et al. 1990, 1996, 1997, Chapman and Buchheim 1991). Moreover, comparisons among *Chlamydomonas* species revealed differences in their nuclear 18S rDNA sequences comparable to the sequence divergence between horsetail and maize (Jupe et al. 1988). Furthermore, since the position of some *Chlamydomonas* lineages was difficult to resolve, Buchheim and Chapman (1992) suggested that it is possible that these divergences have been ancient and rapid, and that the periods of shared ancestry have been too short, relative to the time since divergence, to be adequately recorded in the rRNA sequences.

It has also been proposed (Stewart and Mattox 1980) that all flagellate algal groups evolved from zooflagellate ancestors (e.g., Prymnesiophyceae from the unusual zooflagellate *Colponema*, Chrysophyceae probably from a *Bicoseca*-like protozoan); no known zooflagellate that can be closely linked to green algae has been found, although O'Kelly (1992) suggested that members of the genus *Jakoba* might share a distant common ancestry with the green algae. O'Kelly (1992) presented a speculative phylogenetic scenario suggesting that the green algae have an independent origin from all other photosynthetic organisms: a large predatory zooflagellate having mitochondria with

flattened cristae (which evolved from an amitochondrial phagotrophic zooflagellate) acquired a chlorophyll a,b-containing chloroplast and developed into the ancestral prasinophytes from which the green flagellate ancestors of the various green algal lineages evolved. However, one cannot exclude the possibility that not only the different flagellate algal groups but also some of the primitive green flagellates themselves might have evolved from distinct zooflagellate lineages. Such a scenario is consistent with the fact that the extant primitive green flagellates do not seem to cluster together in phylogenetic trees based on nuclear DNA-encoded rRNA sequences (Kantz et al. 1990, Steinkötter et al. 1994, Friedl 1995), and that in a recent chloroplast SSU rRNA tree, the *Chlamydomonas* sequences do not branch as expected with the other green algae, but they rather suggest a very early divergence relative to all plastid sequences (Gray and Spencer 1996).

1.5.3. Different rates of evolution

As the best explanation for the extreme differences between the mitochondrial genomes of *Marchantia*, *Prototheca* and *Acanthamoeba*, on the one hand, and *Chlamydomonas*, on the other, Gray and Spencer (1996) proposed a "relatively rapid and extreme evolution of the latter genome away from the ancestral pattern represented by the more conservative mtDNAs in the former three organisms."

No extensive studies on point mutation level in *Chlamydomonas* mitochondrial genes have yet been done. It was shown, however, that the number of substitutions in

Chlamydomonas mitochondrial SSU and LSU rRNA genes is severalfold higher than the accumulated substitutions in land plant mitochondrial counterparts (Denovan-Wright et al. 1996). Mitochondrial rRNA sequences of *P. wickerhamii*, however, also seem to have a high rate of nucleotide substitution and, together with the *Chlamydomonas*, ciliate, fungal and yeast counterparts, constitute a rapidly evolving group (associated with long branches in phylogenetic analyses), in marked contrast to the slowly evolving land plant mitochondrial rRNA sequences. Moreover, both *Chlamydomonas* (Fig. 3) and *Prototheca* mitochondrial genomes seem to have undergone extensive gene rearrangements, suggesting a rather high rate of genome evolution in both groups (discussed in Chapter 5 and 7). It is noteworthy that *Chlamydomonas* chloroplast genomes also display extensive sequence divergence (at least twice the range of sequence variation seen in all land plants) as well as gene rearrangement (Turmel et al. 1993, Boudreau et al. 1994) (Fig. 4); such observations led Nedelcu and Lee (in press) to suggest that in this group, in contrast to land plants, mitochondrial and chloroplast genomes exhibit concerted modes and tempos of evolution (discussed in Chapter 7).

The lack of knowledge as to the exact time of divergence of different green algal lineages makes it difficult to assess absolute rates of nucleotide substitutions in their mitochondrial genomes and to compare their tempo of DNA evolution. Moreover, substitution rates of protein-coding genes from all three genetic compartments from various lineages within the chlorophycean as well as pleurastrophycean group have to be

Figure 3. Genetic maps of the *Chlamydomonas reinhardtii* mtDNA and linearized circular-mapping *Chlamydomonas eugametos* mtDNA, showing the extent of gene rearrangement between the two mitochondrial genomes; only corresponding protein- and tRNA-coding regions are connected by lines. L1 to L2, S1 to S4, and L1 to L6, S1 to S3 are non-homologous LSU and SSU rRNA-coding regions in *C. reinhardtii* and *C. eugametos*, respectively; cob-cytochrome b; cox1-subunit 1 of cytochrome oxidase; nad1, 2, 4, 5, 6-subunits 1, 2, 4, 5, 6 of NADH dehydrogenase; DR-direct repeat; TIR-terminal inverted repeat; M_1, M_2, Q, W-coding regions for $tRNA^{Met-1}$, $tRNA^{Met-2}$, $tRNA^{Gln}$, and $tRNA^{Trp}$, respectively; thick arrows indicate the transcription orientation; solid, cross-hatched, and open blocks indicate coding regions/exons, introns, and intergenic spacers, respectively. The *C. reinhardtii* and *C. eugametos* maps are based on those presented by Boer and Gray (1991), and Denovan-Wright, Nedelcu and Lee (in press), respectively.

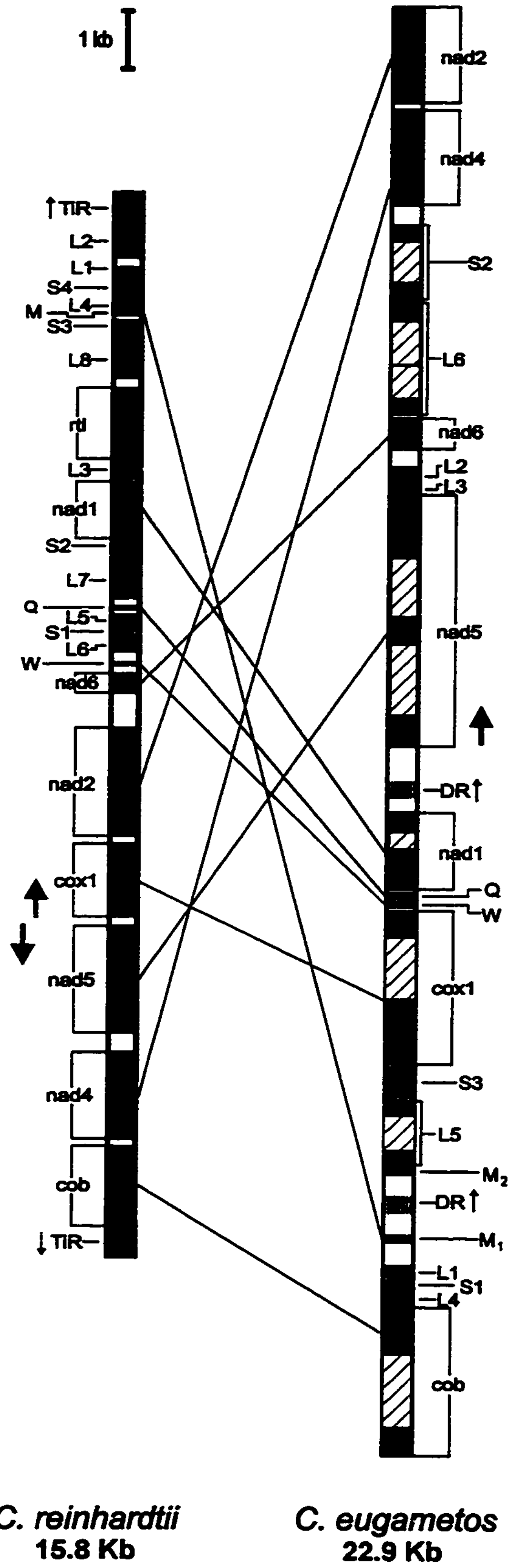


Figure 3

Figure 4. Linearized genetic maps of *Chlamydomonas reinhardtii* and *Chlamydomonas moewusii* cpDNAs (provided by E. Boudreau). Corresponding gene loci (dark areas) are connected by lines: conserved clusters (framed areas) are linked by solid lines, and the rest are connected by dashed lines; thick lines denote the inverted repeats; arrows denote the polarity of the DNA strand containing the coding region; gene abbreviations follow the nomenclature proposed by Hallick and Bairoch (1994 and references therein).

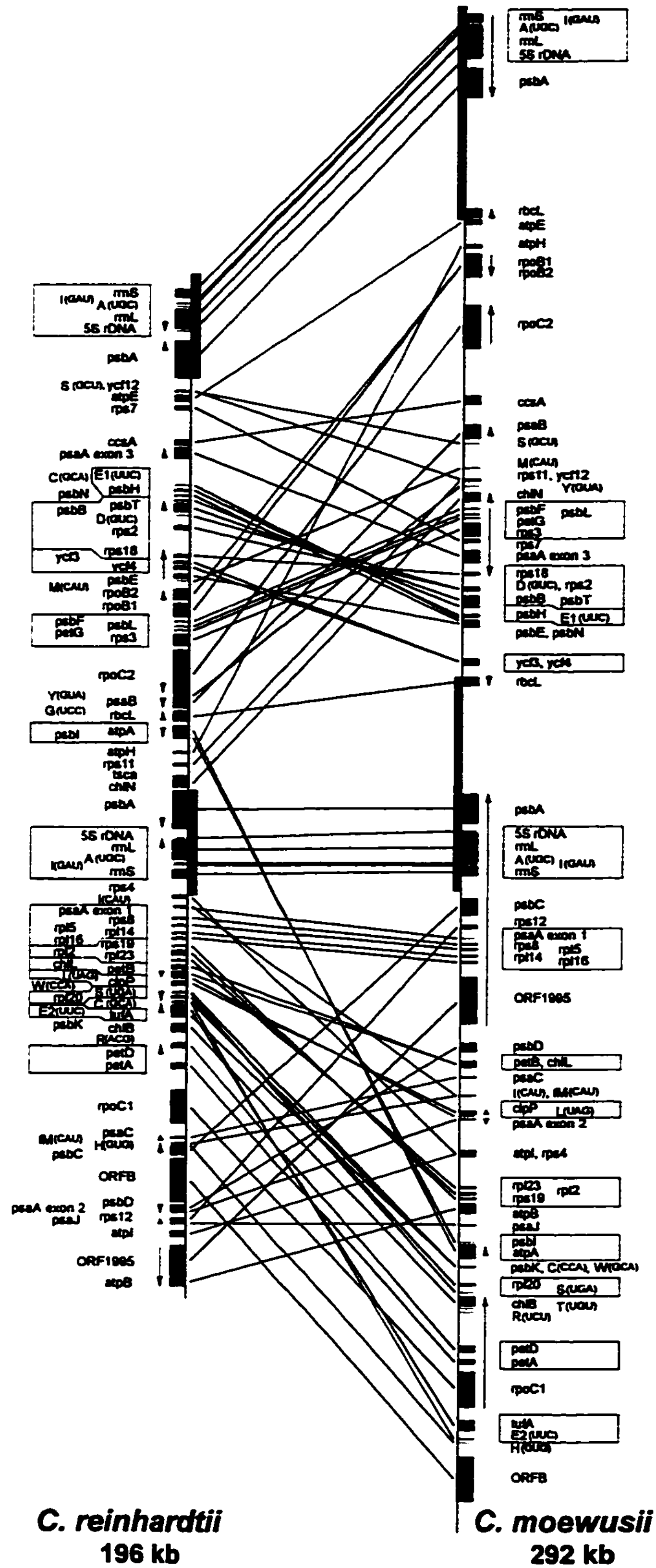


Figure 4

available before one can conclude that the mtDNA in *Chlamydomonas* evolved more rapidly than in other green algal lineages. Nevertheless, no indications have yet been made as to the causes and the mechanisms responsible for the postulated rapid evolution in the *Chlamydomonas* lineage.

1.5.3.1. Adaptive pressures

Although no suggestions of any adaptative pressure in the evolutionary history of green algae have been made, there is an aspect that may be noteworthy. Nedelcu et al. (1996) proposed that the changes in the mitochondrial rRNA gene organization leading to highly fragmented and scrambled rRNA coding regions in the *Chlamydomonas* mitochondrial lineage may have started around the time of the chlorophycean divergence from a prasinophycean-like green flagellate. It should be mentioned that all the prasinophycean taxa are primarily marine algae, whereas all the chlorophycean species are primarily freshwater algae. Thus, at some point around the time of the chlorophycean divergence, the prasinophycean ancestor had to switch from a marine to a freshwater habitat. Moreover, the advanced pleurastrophycean lineages to which *P. wickerhamii* affiliates are also predominantly freshwater algae, and the flagellate genus *Tetraselmis* (aka *Platymonas*), considered ancestral to the pleurastrophycean group (Mattox and Stewart 1984), is marine.

It is conceivable that the freshwater environment may have provided a strong adaptative pressure for the first explorers. It is interesting that land plant mitochondrial

genomes have been shown to respond to stress by increasing the frequency of recombination events mediated by small repeated sequences (Hartmann et al. 1994, Fauron et al. 1995, Benslimane et al. 1996). It is noteworthy, in this connection, that the level of mitochondrial rRNA fragmentation, which was suggested to be a consequence of recombination events (discussed earlier), is much lower in one of the very few chlorophycean marine species, *Chlamydomonas pulsatilla* (Nedelcu 1997), relative to freshwater *Chlamydomonas* counterparts. Moreover, the marine taxon *Hafniomonas montana* (aka *Pyramimonas montana*), previously known as a prasinophycean alga but currently considered related (and retaining ancestral-like features) to the Chlorophyceae, (Ettl and Moestrup 1980, O'Kelly et al. 1994) possesses continuous mitochondrial rRNAs (Nedelcu et al. 1996). Interestingly, it has been shown recently that among chlamydomonadalean taxa their habitat may be more closely linked to their natural history than are ultrastructural features such as the pyrenoid status (Buchheim et al. 1997).

1.5.3.2. Life histories

One of the obvious differences between the life histories of the freshwater chlorophycean and pleurastrophycean green algal groups is the presence of sexual reproduction involving a dormant zygote and zygotic meiosis in the first, and the absence of any type of sexual reproduction in the second. On the other hand, the life history of the predominantly marine ulvophycean green algae is characterized by (i) the complete absence of a dormant zygote, that is considered typical of sexual reproduction in

freshwater green algae and (ii) the presence of alternation of generations. Mattox and Stewart (1984) suggested that the more stable marine environment fostered the evolution of longer life cycles and alternation of generations in ulvophycean taxa. The authors also proposed that alternation of generations is an ancestral condition that was lost by the species that invaded freshwater secondarily (e.g., the case of some freshwater species of *Cladophora* [Graham 1982]), hence it is not particularly advantageous in the more rapidly changing freshwater habitats.

It appears, therefore, that the freshwater environment, at least due to its lesser stability relative to the marine counterpart, created adaptative pressures to which different "explorers" reacted differently, by developing distinct mechanisms to counteract the pressure (e.g., sexual reproduction and dormant zygote among chlorophyceans, and the loss of sexual reproduction and alternation of generations in pleurastrophycean and some ulvophycean green algae, respectively). The high rate of evolution in both nuclear as well as chloroplast and mitochondrial DNA sequences (Jupe et al. 1988, Turmel et al. 1993, Denovan-Wright et al. 1996) in the freshwater flagellate chlorophycean green algae might be correlated with the presence of sexual reproduction featuring a dormant zygote in the life cycle of these lineages. During the dormant stages mutational and recombination events are most likely to occur at high rates. It is known that during the dormant stages the mitochondria in the chlorophycean flagellate *Polytomella agilis* fuse into a single large mitochondrion unit (Burton and Moore 1974). Moreover, in the case of species undergoing sexual reproduction, the zygospore may contain mtDNA molecules acquired from both mating types, which could allow intermolecular recombination events. The

transfer of an intronic sequence during interfertile crosses between *Chlamydomonas smithii* and *C. reinhardtii* (Colleaux et al. 1990) suggests that recombination (in this case, an intermolecular endonuclease-facilitated recombination event) can occur in *Chlamydomonas* mitochondria during the sexual phase of its life cycle. It is interesting to mention that recombination mechanisms are even more active in the chloroplast of *Chlamydomonas* (Dürrenberger et al. 1996) and are thought to have been responsible for the high level of gene rearrangement observed among lineages.

It is also interesting that the mechanisms responsible for the evolution of the vascular plant mitochondrial genome probably developed after the bryophyte divergence, since the mitochondrial genome of the liverwort *M. polymorpha* does not share the distinctive characteristics of the land plant counterparts, such as a very large size, high frequency of recombination, and acquisition of foreign DNA. The distinct and relatively new evolutionary changes in the land plant mitochondrial lineage may have been related to changes in the land plant life history from a dominant haploid generation in bryophytes to a dominant diploid generation in flowering plants. The changes in the alternation of generation pattern may also be correlated to new adaptative pressures created by the transition from a wet to a dry environment.

1.6. Conclusions

The green algal mitochondrial genomes completely sequenced, namely of *Chlamydomonas reinhardtii, Chlamydomonas eugametos* and *Prototheca wickerhamii*

suggested an unexpected dichotomy with regard to mitochondrial genome organization and sequence affiliations among green algal lineages: those of the former two, on the one hand, and latter on the other, resemble more the ciliate/fungal/animal and land plant mitochondrial counterparts, respectively, than one another. This study points out that the other green algal mitochondrial genomes examined to date resemble either the *Chlamydomonas* or the *Prototheca* pattern of mitochondrial genome organization; the *Chlamydomonas*-like mitochondrial genomes are small, have a reduced gene content (no ribosomal protein or 5S rRNA genes, only few protein-coding and tRNA genes) and fragmented and scrambled rRNA coding regions, whereas the *Prototheca*-like mitochondrial genomes are larger, have a larger set of protein-coding genes including ribosomal protein genes, more tRNA genes as well as 5S rRNA and conventional continuous SSU and LSU rRNA coding regions. On this evidence, it appears that the evolutionary distance between not only the mitochondrial genomes of *C. reinhardtii* and *P. wickerhamii*, as previously thought, but rather of the two green algal mitochondrial types to which they belong is great enough to raise questions about the causes and mechanisms responsible for such a dichotomy. Moreover, this study suggests that a more integrative approach in explaining the occurrence of distinct evolutionary patterns and apparent phylogenetic affiliations among the known green algal mitochondrial lineages might be needed. The observed dichotomy could be the result of distinct genetic potentials differentiated during the previous evolutionary history of the flagellate ancestors, and/or consequent changes in habitat and life history of the more advanced green algal lineages.

Chapter 2

Discontinuous mitochondrial and chloroplast LSU rRNAs among green algae: Phylogenetic implications

2.1. Introduction

One of the most distinctive features shared by both mitochondria and chloroplasts of green algae is the presence of discontinuous LSU rRNAs. Generally, most known LSU rRNAs consist of single, continuous polyribonucleotide chains; however, an increasing number of examples of complexes comprised of split rather than covalently continuous LSU rRNAs has been reported in various lineages. The LSU rRNA pieces can interact by intermolecular basepairing and provide the same functional framework for translation as the conventional, continuous LSU rRNAs. Split LSU rRNAs have been reported for the mitochondrial, chloroplast, and nucleo-cytosolic compartments of eukaryotes as well as for eubacteria (see Denovan-Wright and Lee 1994 for references). The genes coding for discontinuous LSU rRNAs are fragmented into coding modules which can be interspersed with internal transcribed spacers, transfer RNA, or protein-coding genes.

Information on discontinuous mitochondrial and chloroplast LSU rRNAs is rather descriptive and limited; no exhaustive studies on the occurrence, evolution, and phylogenetic distribution of these phenotypes have been done. The green algae represent a good study group in this respect, because data on mitochondrial and chloroplast LSU rRNA gene organization as well as on phylogenetic relationships among lineages are available.

Two very distinct mitochondrial LSU rRNA gene organizations have been observed in the two green algal lineages investigated in this respect: conventional, eubacterial-like mitochondrial LSU rRNA coding regions in *Prototheca wickerhamii*

(Wolff et al. 1994), and highly fragmented and scrambled mitochondrial LSU rRNA coding regions in *Chlamydomonas* (Boer and Gray 1988, Denovan-Wright and Lee 1994). The mitochondrial LSU rRNAs of the two *Chlamydomonas* species examined so far, namely, *C. reinhardtii* and *C. eugametos* are encoded by eight and six coding modules, respectively, which are interspersed with each other and with small subunit rRNA coding regions, transfer RNA and protein-coding genes. Analysis of the differences in the mitochondrial LSU rRNA fragmentation pattern of the two *Chlamydomonas* species suggested that the last common ancestor of *C. eugametos* and *C. reinhardtii* had fragmented mitochondrial LSU rRNA genes (Denovan-Wright and Lee 1994). The only other information on green algal mitochondrial LSU rRNA gene organization refers to a partial sequence of a mitochondrial LSU rRNA coding region in *Scenedesmus obliquus* from which Kück et al. (1990) proposed a secondary structure for a continuous 3'-half of the mitochondrial LSU rRNA.

Little is also known about chloroplast LSU rRNA gene organization among green algal lineages. The chloroplast LSU rRNA gene organization is similar in all of the 17 *Chlamydomonas* taxa investigated by Turmel et al. (1993); the LSU rRNA genes are interrupted at the same locations by three short ITSs that are post-transcriptionally excised to yield four rRNA fragments, named α, β, γ, and δ. Moreover, the chloroplast LSU rRNA gene of the chlorococcalean *Chlorella ellipsoidea* contains an insert designated as a new type of intron (Yamada and Shimaji 1987) at the same position as the *Chlamydomonas* ITS separating the coding modules γ and δ. However, no evidence is available that the mature chloroplast LSU rRNA in this taxon is intact, and Turmel et al.

(1991) suspected that this insert is, in fact, an ITS whose excision leads to a discontinuous chloroplast LSU rRNA.

With information from representatives of only two quite distant green algal lineages, namely, *Chlamydomonas* and *P. wickerhamii* for the mitochondria, and *Chlamydomonas* and *Chlorella ellipsoidea*, for the chloroplast, questions on the origin and evolution of fragmented mitochondrial and chloroplast LSU rRNA genes among green algae could not be addressed. This work was therefore designed to examine the occurrence and phylogenetic distribution of discontinuous mitochondrial and chloroplast LSU rRNAs among green algae. The present study addresses the following questions: (i) are the discontinuous mitochondrial and chloroplast LSU rRNAs among green algae confined to the *Chlamydomonas* group?; (ii) how deep in the evolutionary history of green algae can the discontinuous mitochondrial and chloroplast LSU rRNA traits be traced?; and (iii) can these traits constitute additional characters in assessing phylogenetic relationships among green algae?

To answer these questions, I surveyed LSU rRNA molecules of taxa that diverged early in the chlorophycean lineage, as well as of taxa included in the sister class Pleurastrophyceae and the basal class Micromonadophyceae (sensu Mattox and Stewart 1984). The experimental approach employed Northern blot analyses of total RNA preparations using synthetic oligodeoxynucleotide probes complementary to highly conserved regions within the LSU rRNA. The hybridizing RNA species in the study group taxa were assessed by comparisons with their respective electrophoretic patterns and their hybridizing mitochondrial and chloroplast RNA counterparts in *Chlamydomonas*

reinhardtii.

2.2. Material and Methods

2.2.1. Algal cultures and growing conditions

All strains used were checked for bacterial contamination and decontaminated, if necessary, by antibiotic treatment (Guillard 1973). Cultures were grown in the media indicated in Table 3 and supplied with 1% CO_2 in air. Illumination was provided by cool-white fluorescent lamps (50-80 $\mu mol \bullet m^{-2} \bullet s^{-1}$) on a 12:12h LD cycle. Cells were harvested when the culture density reached ca. 4×10^6 $cells \bullet mL^{-1}$.

2.2.2. Total RNA extraction and fractionation.

Total RNA was extracted according to Rochaix and Malnoë (1982). Glyoxalated total RNA (10 µg) from each taxa was fractionated by agarose (1.2%) gel electrophoresis (Sambrook et al. 1989) and transferred to Hybond-N Nylon membranes (Amersham) by vacuum blotting under the conditions recommended by the manufacturer (Pharmacia). RNA blots were stained with methylene-blue to check the efficiency of transfer (Herrin and Schmidt 1988). Total RNA (5 µg) was also fractionated by denaturing (7M urea) polyacrylamide (6%) gel electrophoresis (Schnare and Gray 1990).

Table 3. Algal strains, sources and growing media.

Taxa	Source[a] and strain	Medium[b]
Chlorophyceae		
Chlamydomonas reinhardtii Dangeard	GC wt 137c	MM
Polytomella agilis Prings.	UTEX LB 193	P
Carteria crucifera Korsch	UTEX 432	VE
Scenedesmus obliquus (Turp.) Kutz	UTEX 78	BM
Planophila terrestris Groover et Hofstetter	UTEX 1709	SE
Hormotilopsis gelatinosa Trainor et Bold	UTEX 104	SE
Uronema belkae Mattox et Bold	UTEX 1179	MM
Chlorella vulgaris Beij.	UTEX 259	MM
Prototheca wickerhamii Soneda et Tubaki	UTEX 1533	Malt
Hafniomonas montana Ettl & Moestrup	NZ c50 wcg/y	EM
Micromonadophyceae		
Pyramimonas parkae Norris et Pearson	UTEX LB 2287	EM
Pleurastrophyceae		
Pleurastrum terrestre Fritsch et John	UTEX 333	MM

[a] GC = Genetics Center at Duke University; UTEX = The Culture Collection of Algae at the University of Texas at Austin; NZ = New Zealand, strain isolated by Hans Preisig and kindly provided by Charles O'Kelly.

[b] BM = Basal Medium (Oh-Hama and Hase 1980); EM = Erdschreiber Medium (Starr and Zeikus 1993); Malt = Malt Medium (Wolff and Kück 1990); MM = Minimal Medium (Lemieux et al. 1980); P = Polytomella Medium (Burton and Moore 1974); SE = Soil Extract Medium (Starr and Zeikus 1993); VE = Volvox Medium (modified by McCracken et al. 1980) and Soil Extract Medium (5%).

2.2.3. Northern blot hybridization.

Northern blots were pre-hybridized for 3 h in the hybridization buffer: 5x SSPE (20x SSPE = 3.6M NaCl, 200 mM NaH_2PO_4, 20 mM EDTA, pH 7.4), 1X BLOTTO [10x BLOTTO = 5% Instant Skim Milk, 10% sodium dodecyl sulphate (SDS), pH 7.8], 50% formamide at 35°C. The probes were synthetic oligodeoxynucleotides complementary to highly conserved regions within the LSU rRNA. Probe 5'-*C* is a 27-mer (5'-GCTAGACCAGTGAGCTATTACGCTTTC-3') complementary to a region within the GTPase center in the 5'-half of the *C. reinhardtii* mitochondrial LSU rRNA. Probe 3'-*C* is a 29-mer (5'-AGGACGCGATGATCCAACATCGAGGTGCC-3') complementary to a region within the peptidyl transferase center in the 3'-half of the *C. reinhardtii* mitochondrial LSU rRNA. Probe 3'-*S* is a 27-mer (5'-GCTGATAAACCTGTTATCCCTAGCG TA-3') complementary to another region within the peptidyl transferase center in the 3'-half of the *Scenedesmus obliquus* mitochondrial LSU rRNA. *Escherichia coli* coordinates (Gutell et al. 1992) for the target regions of the three probes are: 1087-1113, 2494-2522 and 2438-2464, respectively. The oligonucleotide probes were 5'-end labelled using [γ-^{32}P] ATP and polynucleotide kinase (Pharmacia) at 37° C for 1 h and then purified on Microspin Columns S-200 HR (Pharmacia). Hybridization reactions were carried out at 35° C for 20 h. The blots were washed twice for 15 min at room temperature, once in 2X SSPE, 0.1% SDS, and once in 0.5X SSPE, 0.1% SDS.

2.2.4. Phylogenetic analyses

The evolutionary history of flagellar apparatus configuration, mitochondrial and chloroplast LSU rRNA was reconstructed using the MacClade program version 3 (Maddison and Maddison 1992) for tracing character evolution.

2.2.5. Results

In this work, I have investigated twelve species representing three classes of green algae (Table 3). The abundant cytosolic 5.8S (160 nt), SSU (1,800 nt) and LSU rRNAs (3,500 nt) as well as chloroplast SSU (1,500 nt) and LSU rRNA species (α=290 nt, δ=1,700 nt and γ=810 nt) in the *C. reinhardtii* total RNA served as size references for the identification of the rRNA counterparts in the study group taxa (Fig. 5). Mitochondrial rRNAs are not visible on ethidium bromide (EtBr)-stained gels or methylene-blue stained RNA blots due to their low abundance relative to their cytosolic and chloroplast counterparts.

Figure 6 shows the locations of the target regions for probes 5'-*C*, 3'-*C*, and 3'-*S* within the chloroplast and mitochondrial LSU rRNA of *Chlamydomonas*. Under the hybridization conditions employed, the probes annealed with chloroplast and/or mitochondrial but not with cytosolic LSU rRNA species. The mitochondrial or chloroplast nature of the hybridizing RNA species in the study group taxa was assessed

by comparisons with their electrophoretic patterns (Fig. 5A, B) and the hybridizing *Chlamydomonas reinhardtii* counterparts (Fig. 7A, B); hybridizing RNA species with visible correspondents on stained gels are not likely to have a mitochondrial origin. Table 4 provides information about the size of the hybridizing RNA species as well as their proposed mitochondrial or chloroplast location.

2.3.1. Mitochondrial LSU rRNAs

Probe 5'-*C* and 3'-*C* hybridized (Fig. 7A, B) as expected (Fig. 6) with the *C. reinhardtii* mitochondrial LSU rRNA L_5- and L_8-fragment, of 140 and 493 nucleotides (nt), respectively. Hybridizing RNAs (with no visible correspondents on stained gels, Fig. 5) of about 2,900-3,200 nt were identified in *Uronema belkae, Hafniomonas montana, Chlorella vulgaris, Prototheca wickerhamii, Pleurastrum terrestre* and *Pyramimonas parkae* (Fig. 7A-C). Hybridizing RNAs of about 1,300, 1,100, 500, and 400 nt were detected in *Scenedesmus obliquus, Hormotilopsis gelatinosa/Planophila terrestris, Carteria crucifera* and *Polytomella agilis*, respectively (Fig. 7B, C).

2.3.2. Chloroplast LSU rRNAs

Under the hybridization conditions used, probe 5'-*C* and probe 3'-*C* also hybridized (Fig. 7A, B) with the *C. reinhardtii* chloroplast LSU rRNA γ- fragment (810 nt) and δ-fragment (1,690 nt), respectively (Fig. 6). In addition, a 700 nt RNA species visible on

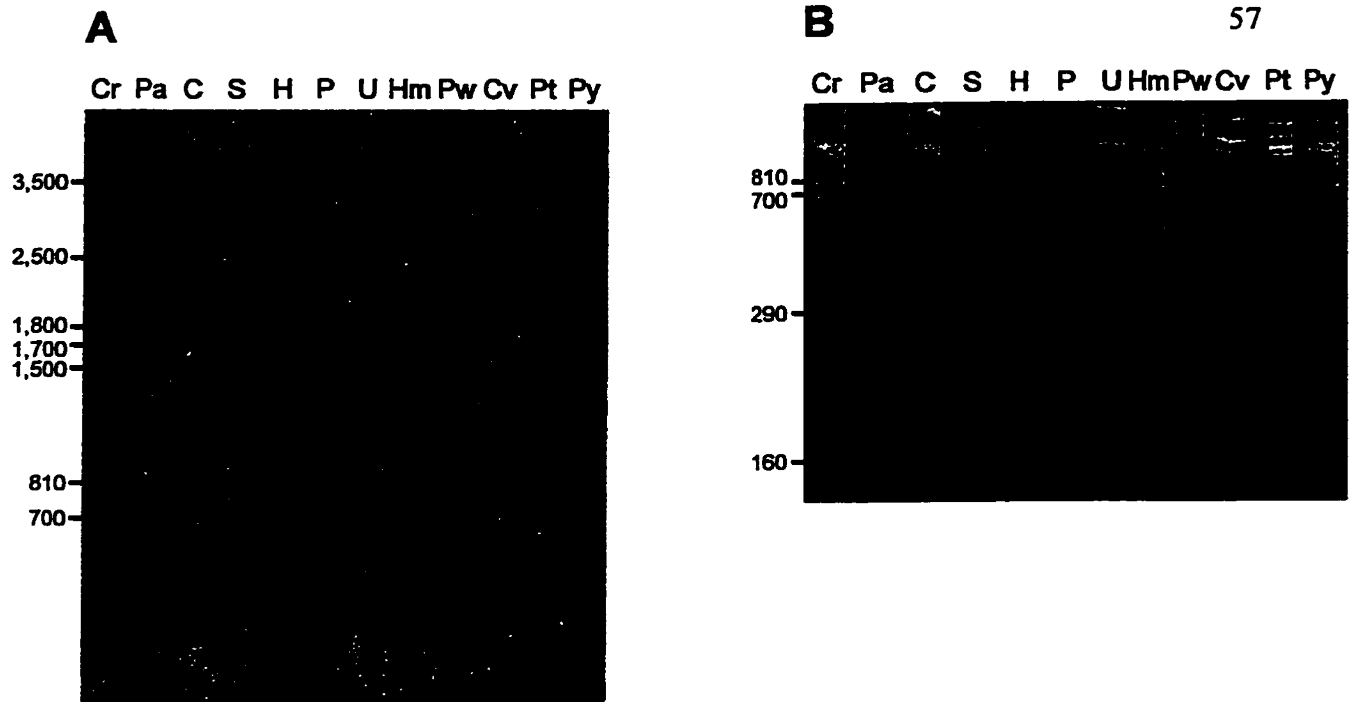


Figure 5. Electrophoretic patterns of total RNA extracted from the study group taxa. A Methylene-blue staining of fractionated RNA blotted from an 1.2% agarose gel and B Ethidium bromide staining of RNA fractionated on a 6% polyacrylamide gel. Cr = *Chlamydomonas reinhardtii;* Pa = *Polytomella agilis;* C = *Carteria crucifera;* S = *Scenedesmus obliquus;* H = *Hormotilopsis gelatinosa;* P = *Planophila terrestris;* U = *Uronema belkae;* Hm = *Hafniomonas montana;* Pw = *Prototheca wickerhamii;* Cv = *Chlorella vulgaris;* Pt = *Pleurastrum terrestre;* Py = *Pyramimonas parkae.* RNA molecular sizes are expressed in nucleotides. *C. reinhardtii* rRNA size references indicated correspond to the: cytosolic LSU rRNA (3,500 nt), cytosolic SSU rRNA (1,800 nt), chloroplast LSU rRNA δ-fragment (1,700 nt), chloroplast SSU rRNA (1,500 nt), chloroplast LSU rRNA γ-fragment (810 nt), chloroplast LSU rRNA α-fragment (290 nt) and cytosolic 5.8S rRNA (160 nt).

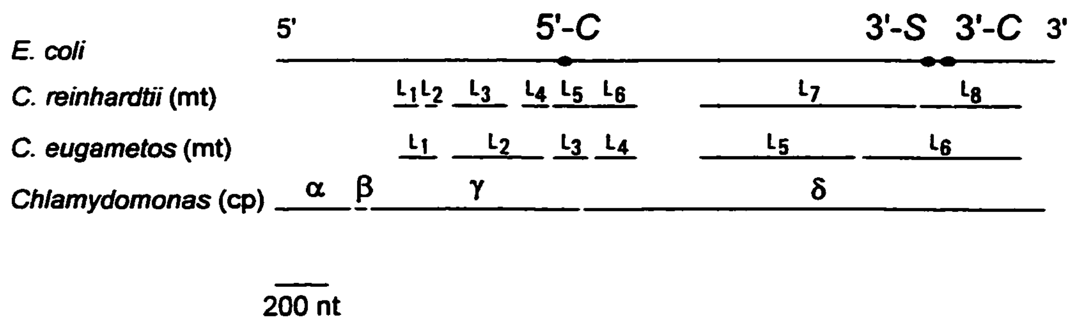


Figure 6. Fragmentation patterns within the mitochondrial (mt) and chloroplast (cp) LSU rRNA of *Chlamydomonas*, and location of the target regions for oligonucleotide probes 5'-*C*, 3'-*S*, and 3'-*C*, represented to the *Escherichia coli* LSU rRNA scale.

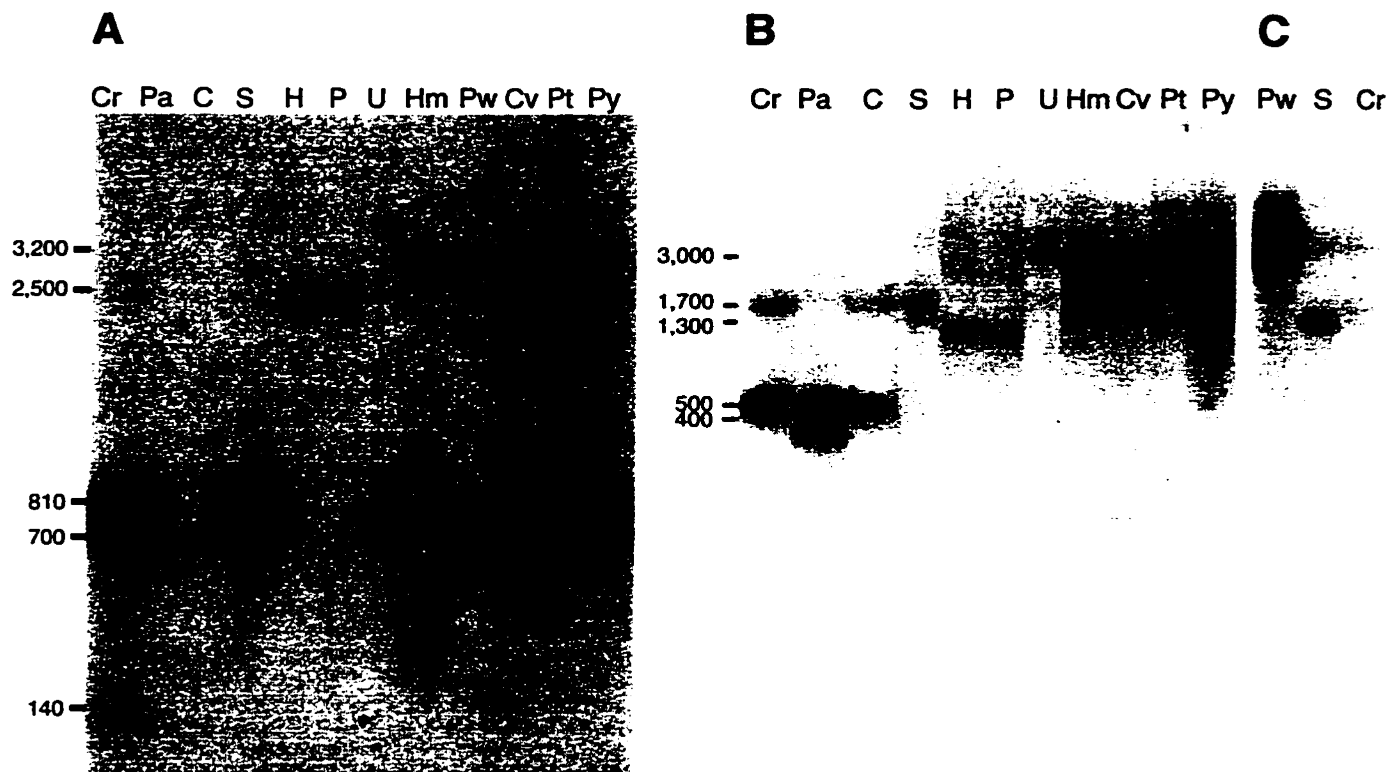


Figure 7. Northern blot hybridizations of green algal total RNA fractionated on 1.2% agarose gels, using three synthetic deoxyoligonucleotide probes, respectively: A) 5'-*C*, B) 3'-*C* and C) 3'-*S* ; only data for informative taxa shown. Abbreviations as in Figure 5. RNA molecular sizes are expressed in nucleotides. *C. reinhardtii* hybridizing rRNAs correspond to the: mitochondrial LSU rRNA L_5- and L_8- fragments (140 and 500 nt, respectively) and chloroplast LSU rRNA γ- and δ- fragments (810 and 1,700 nt, respectively). The size and proposed cellular location of the rest of the hybridizing LSU rRNA species are indicated in Table 4.

Table 4. The size and proposed cellular location of the hybridizing LSU rRNA species. Cr = *Chlamydomonas reinhardtii*; Pa = *Polytomella agilis*; C = *Carteria crucifera*; S = *Scenedesmus obliquus*; H = *Hormotilopsis gelatinosa*; P = *Planophila terrestris*; U = *Uronema belkae*; Hm = *Hafniomonas montana*; Pw = *Prototheca wickerhamii*; Cv = *Chlorella vulgaris*; Pt = *Pleurastrum terrestre*; Py = *Pyramimonas parkae*.

Probe / Proposed cellular origin	Taxa											
	Cr	Pa	C	S	H	P	U	Hm	Pw	Cv	Pt	Py
5'-C												
Mitochondria	140	-	-	-	-	-	-	3,200	3,000	3,100	3,200	3,200
Chloroplast	810	-	-	800	-	-	800	800	-	800	-	800
	700	-	-	700	-	-	700	700	-	700	-	700
	-	-	-	-	2,500	2,500	-	2,800	-	2,700	2,800	2,700
					2,300	2,300						
3'-C and/or 3'-S												
Mitochondria	493	400	500	1,300	1,100	1,100	2,900	3,200	3,000	3,100	3,200	3,200
Chloroplast	1,690	-	1,700	1,700	-	-	1,700	1,700	-	1,700	-	1,700

an EtBr-stained polyacrylamide gel (Fig. 5B) was detected by probe 5'-*C* (Fig. 7A). Counterpart RNA species of about 700, 800 and 1700 nt were identified only in *Carteria crucifera*, *Scenedesmus obliquus*, *Uronema belkae*, *Hafniomonas montana*, *Chlorella vulgaris* and *Pyramimonas parkae* (Fig. 7A, B). Large RNA molecules were detected in *Hafniomonas montana* (2,800 nt), *Chlorella vulgaris* (2,700 nt), *Planophila terrestris/Hormotilopsis gelatinosa* (2,300 and 2,500 nt) *Pyramimonas parkae* (2,700 nt), and *Pleurastrum terrestre* (2,800 nt) (Fig. 7A, B). RNAs of about 240-290 nt, counterparts of the *Chlamydomonas reinhardtii* chloroplast LSU rRNA α-fragment (290 nt), were identified on an EtBr-stained polyacrylamide gel (Fig. 5B) in all the above species. No plastid LSU rRNA transcripts were detected in the colourless taxa *Polytomella agilis* and *Prototheca wickerhamii* (Fig. 7A, C).

2.3.3. Reconstruction of the evolutionary history of three characters: flagellar apparatus configuration, discontinuous mitochondrial LSU rRNA, and discontinuous chloroplast LSU rRNA

To compare the phylogenetic distribution of discontinuous mitochondrial and chloroplast LSU rRNAs to the flagellar apparatus configuration of the green algal lineages investigated, I reconstructed the evolutionary history of flagellar apparatus configuration (Fig. 8A), mitochondrial LSU rRNA (Fig. 8B) and chloroplast LSU rRNA (Fig. 8C), using the MacClade program version 3 (Maddison and Maddison 1992) for tracing character evolution. Figure 8 provides a phylogenetic framework for the data presented

here; the scheme is based on phylogenetic relationships previously suggested by molecular (Conner et al. 1989, Buchheim and Chapman 1992, Wilcox et al. 1992, Friedl and Zeltner 1994, Steinkötter et al. 1994) and ultrastructural data (Lembi 1975, Brown et al. 1976, Ettl and Moestrup 1980, Mattox and Stewart 1984, O'Kelly and Floyd 1984, Watanabe and Floyd 1989, O'Kelly 1992, O'Kelly et al. 1994).

2.4. Discussion

2.4.1. Continuous versus discontinuous mitochondrial and chloroplast LSU rRNAs among green algae

2.4.1.1. Mitochondrial LSU rRNAs

The continuous or discontinuous nature of the mitochondrial LSU rRNAs in the study group taxa has been assessed based upon the size of their hybridizing RNA molecules and the *C. reinhardtii* counterparts. The 3,000 nt RNA detected by two probes directed to both the 5'-half and 3'-half of the mitochondrial LSU rRNA of *P. wickerhamii* is presumably the predicted mature, continuous mitochondrial LSU rRNA in this species, as the mitochondrial LSU rRNA gene of *P. wickerhamii* is 3,890 nt in length and contains two group I introns, of about 374 and 499 nt, respectively (Wolff et al. 1994). The detection of mitochondrial LSU rRNA molecules of about 3,000-3,200 nt in *Hafniomonas montana, Chlorella vulgaris, Pleurastrum terrestre* and *Pyramimonas parkae* implies the

Figure 8. Reconstruction of the evolutionary history of three characters: A) flagellar apparatus configuration, B) discontinuous mitochondrial LSU rRNA and C) discontinuous chloroplast LSU rRNA, using MacClade program version 3 (Maddison and Maddison 1992) for tracing character evolution. The phylogenetic scheme on which the three characters were mapped is based on phylogenetic relationships previously proposed (see references in text).

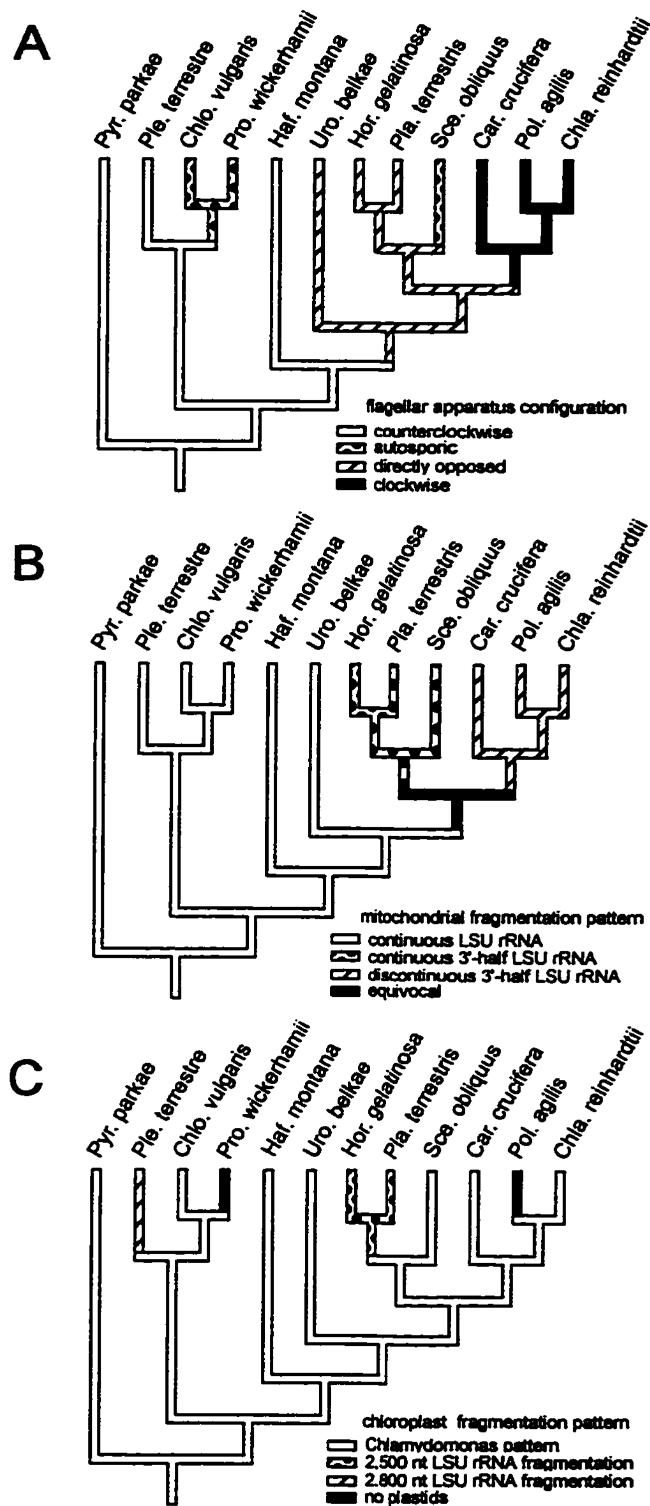


Figure 8

presence of large, continuous mitochondrial LSU rRNAs in these taxa. The 2,900 nt RNA species detected in *Uronema belkae* most likely represents a smaller continuous mitochondrial LSU rRNA in this lineage.

The 1,300 nt RNA molecule identified in *Scenedesmus obliquus* was detected by both probes directed to the 3'-half of the mitochondrial LSU rRNA. Kück et al. (1990) determined the sequence of a mitochondrial clone isolated from a wild strain (KS3/2) of *S. obliquus* and proposed a secondary structure for a continuous 3'-half mitochondrial LSU rRNA comprised of 1,264 nt. I analyzed the remaining nucleotides in the 5'-end of the reported clone (Nedelcu 1997) and noticed that: (i) only the next 40 nt show sequence similarity with the 3'-end of the 5'-half of other LSU rRNAs, and (ii) the region in which the nucleotide sequence does not show sequence similarity with other LSU rRNAs corresponds with a variable region within many 5'-half LSU rRNAs, but also represents a common region of discontinuity in the continuous mitochondrial LSU rRNAs of *C. reinhardtii* (Boer and Gray 1988a) and *C. eugametos* (Denovan-Wright and Lee 1994) (Fig. 6) (discussed in Chapter 4). Therefore, I consider that the 1,300 nt RNA species detected in this study accounts for a mitochondrial LSU rRNA fragment comprised of the 3'-half of about 1,260 nt, as well as 40 nt in the 5'-half of its LSU rRNA.

The identification of corresponding hybridizing mitochondrial RNA species similar in size (1,100 nt) to the mitochondrial 3'-half LSU rRNA fragment of *Scenedesmus obliquus* (1,300 nt) suggests the presence of discontinuous mitochondrial LSU rRNAs also in *Planophila terrestris* and *Hormotilopsis gelatinosa*. Moreover, the size of these LSU rRNAs is similar to the size of a putative non-fragmented mitochondrial 3'-half LSU

rRNA in *Chlamydomonas* (Fig. 6); the two LSU rRNA fragments added together would lead to a 1,172 and 1,120 nt LSU rRNA species in *C. reinhardtii* and *C. eugametos*, respectively (the continuous mitochondrial 3'-half LSU rRNA of *Chlamydomonas pulsatilla* is about 1,200 nt [data shown in Chapter 3]). Considering the size of the RNA species accounting for the 3'-half of the mitochondrial LSU rRNA in *Planophila terrestris* and *Hormotilopsis gelatinosa*, as well as the observation that the sites of discontinuity are always located in variable regions of the rRNA mature transcripts (Gray and Schnare 1995), it is likely that *Planophila terrestris, Hormotilopsis gelatinosa, Scenedesmus obliquus*, and *Chlamydomonas* spp. share a discontinuity site within the same corresponding variable region between the 3'-half and the 5'-half of their mitochondrial LSU rRNAs.

The detection of mitochondrial RNA species of about 400 nt in *Polytomella agilis* and 500 nt in *Carteria crucifera*, which are similar in size to the mitochondrial LSU rRNA L_8-fragment counterpart of *Chlamydomonas reinhardtii* (493 nt), suggests that the mitochondrial 3'-half LSU rRNA is discontinuous in these two lineages. The mitochondrial 3'-half LSU rRNA in the two *Chlamydomonas* species examined so far (Boer and Gray 1988a, Denovan-Wright and Lee 1994) is not interrupted in corresponding variable regions (Fig. 6) and consists of two LSU rRNA fragments whose sizes are different in *C. reinhardtii* and *C. eugametos*. Moreover, a study of other *Chlamydomonas* taxa from both *C. reinhardtii* and *C. eugametos* clades, as defined by nuclear and chloroplast sequence affiliations (Buchheim et al. 1990, Turmel et al. 1993), revealed that taxa placed in each clade also share similar fragment sizes within their mitochondrial 3'-

half LSU rRNA with *C. reinhardtii* or *C. eugametos*, respectively (Denovan-Wright et al. 1996). Considering the similarity in size between the mitochondrial LSU rRNA fragments of *Polytomella agilis, Carteria crucifera* (400 and 500 nt, respectively), and their counterpart *C. reinhardtii* mitochondrial LSU rRNA L_8-fragment (493 nt), *P. agilis* and *C. crucifera* might share with species in the *C. reinhardtii* clade a similar discontinuity site within their mitochondrial 3'-half LSU rRNA.

The results discussed here suggest that the mitochondrial LSU rRNAs are discontinuous in *Polytomella agilis, Carteria crucifera, Scenedesmus obliquus, Planophila terrestris* and *Hormotilopsis gelatinosa* but continuous in *Uronema belkae, Prototheca wickerhamii, Hafniomonas montana, Chlorella vulgaris, Pleurastrum terrestre* and *Pyramimonas parkae*. Among the mitochondrial discontinuous LSU rRNAs investigated in this study, the fragmentation pattern is not identical. Lineages considered closely related to *Chlamydomonas reinhardtii* (i.e., *Polytomella agilis* and *Carteria crucifera)* have a discontinuous mitochondrial 3'-half LSU rRNA, as indicated by the identification of a small 3'-half LSU rRNA fragment counterpart of the mitochondrial LSU rRNA L_8-fragment of *C. reinhardtii*, whereas lineages diverging earlier relative to *C. reinhardtii* (e.g., *Scenedesmus obliquus, Planophila terrestris,* and *Hormotilopsis gelatinosa*) have a continuous mitochondrial 3'-half LSU rRNA, as indicated by a large LSU rRNA fragment accounting for the 3'-half of their LSU rRNA.

2.4.1.2. Chloroplast LSU rRNAs

The hybridizing chloroplast RNAs in the study group taxa were assessed by comparisons with their electrophoretic patterns and *C. reinhardtii* counterparts. I consider that the 800 and 1,700 nt RNA species identified in *Carteria crucifera, Scenedesmus obliquus, Uronema belkae, Hafniomonas montana, Chlorella vulgaris*, and *Pyramimonas parkae*, as well as the 280 nt RNA visible on EtBr-stained polyacrylamide gels, represent the homologous counterparts of the γ- and δ- and α-fragments of *Chlamydomonas reinhardtii* chloroplast LSU rRNA, respectively (Fig. 6), thereby suggesting similar fragmentation patterns within the LSU rRNA complexes of the above mentioned taxa. The absence of an 800 nt hybridizing RNA (visible on EtBr-stained gels) in *Carteria crucifera* might be due to a higher level of sequence divergence in this region. The additional presence of hybridizing large RNAs of about 2,700-2,800 nt in *Hafniomonas montana, Chlorella vulgaris* and *Pyramimonas parkae* might be related to processing of ITS in primary chloroplast LSU rRNA transcripts. The 700 nt RNA detected in *C. reinhardtii* and all of the taxa displaying discontinuous chloroplast LSU rRNAs was also identified as a minor band in a *Chlamydomonas moewusii* total RNA hybridized with a cloned fragment spanning the 5'-half of the *C. eugametos* chloroplast LSU rRNA gene (Turmel et al. 1988) and is most likely of chloroplast origin.

Only large chloroplast LSU rRNA species (2,300, 2,500 and 2,800 nt) were detected in *Planophila terrestris, Hormotilopsis gelatinosa*, and *Pleurastrum terrestre*. However, the presence of small RNAs of about 240-290 nt, presumably counterparts of the 290 nt α-fragment of the *Chlamydomonas reinhardtii* chloroplast LSU rRNA suggests that at least one ITS corresponding to a *C. reinhardtii* ITS is present and excised from the

primary transcript of these taxa. I consider that the absence of one or two ITS or the inability to process them accounts for the larger LSU rRNA fragments detected in the above mentioned taxa. The identification of chloroplast LSU rRNA precursors of about 2,700-2,800 nt in lineages with a fragmentation pattern similar to that of *Chlamydomonas*, indicates that the processing of the primary LSU rRNA transcripts can be incomplete. A theoretical LSU rRNA transcript obtained by excising only the ITS between the α- and β-fragment from the primary LSU rRNA transcript can be as small as 2,583 nt in *C. reinhardtii* or as large as 2,815 nt in *Chlamydomonas starii*, due to differences in the size of their ITS (Turmel et al. 1993). This variation might explain the difference in size between the 2,500 nt RNA detected in the *Planophila terrestris/Hormotilopsis gelatinosa* lineage and the 2,800 nt RNA identified in *Pleurastrum terrestre*. Alternatively, the 2,500 nt RNA present in the *Planophila terrestris/Hormotilopsis gelatinosa* lineage could represent the counterpart of a theoretical transcript comprised of the nonprocessed γ- plus δ-fragment, which in *Chlamydomonas reinhardtii* would account for a 2,511 nt RNA fragment. The simultaneous presence of another chloroplast LSU rRNA species of about 2,300 nt in *Planophila terrestris/Hormotilopsis gelatinosa* lineage could be related to the simultaneous presence of a 700 and 800 nt RNA in all taxa displaying the *Chlamydomonas reinhardtii* chloroplast LSU rRNA fragmentation pattern.

The data presented in this study suggest that the discontinuous chloroplast LSU rRNA phenotype among green algae is more widespread than the discontinuous mitochondrial phenotype. Electrophoretic fractionation and Northern blot analyses of total RNA extracted from *Carteria crucifera*, *Scenedesmus obliquus*, *Uronema belkae*,

Hafniomonas montana, Chlorella vulgaris, and *Pyramimonas parkae* revealed chloroplast LSU rRNA species similar in size to their counterparts in *Chlamydomonas*, suggesting a similar fragmentation pattern within their LSU rRNAs. In *Planophila terrestris, Hormotilopsis gelatinosa,* and *Pleurastrum terrestre,* the fragmentation pattern appears different from the rest of the taxa examined and is most likely due to the absence of one or two ITSs or the inability to process them. Further characterization of these chloroplast LSU rRNA genes is needed to test the inferences presented here.

2.4.2. Phylogenetic implications

Discontinuous mitochondrial LSU rRNA can be used as an additional phylogenetic character in assessing relationships among groups of green algae (Fig. 8A, B): (i) zoosporic chlorophycean species with a clockwise (CW) configuration in their flagellar apparatus, e.g., *Chlamydomonas reinhardtii, Polytomella agilis,* and *Carteria crucifera,* appear to have discontinuous mitochondrial LSU rRNAs; (ii) chlorophycean taxa whose quadriflagellate zoospores exhibit a directly opposed (DO) configuration in their flagellar apparatus, e.g., *Neochloris aquatica* (data shown in Chapter 3), *Hormotilopsis gelatinosa,* and *Planophila terrestris,* as well as autosporic chlorococcalean species phylogenetically related to this group (e.g., *Scenedesmus obliquus* [Wilcox et al. 1992]), also have discontinuous mitochondrial LSU rRNAs, but the fragmentation pattern is different from the one present in the CW clade; (iii) chaetophoralean species, e.g., *Uronema belkae,* considered a separate chlorophycean lineage (O'Kelly et al. 1994), whose zoospores have

a flagellar apparatus including directly opposed upper basal bodies and lower basal bodies in the CW orientation (Floyd et al. 1980, Watanabe and Floyd 1989), have continuous mitochondrial LSU rRNAs; and (iv) zoosporic green algal taxa with a counterclockwise (CCW) orientation in their flagellar apparatus, e.g., *Hafniomonas montana, Pyramimonas parkae, Pleurastrum terrestre*, and autosporic chlorococcalean species phylogenetically related to them (e.g., *Chlorella vulgaris* and *Prototheca wickerhamii* [Steinkötter et al. 1994]), also have continuous mitochondrial LSU rRNAs.

The distribution of the discontinuous mitochondrial LSU rRNA trait within Chlorophyceae overlaps with other molecular and ultrastructural characters, reinforcing phylogenetic relationships previously suggested, i.e.: (i) CW and DO clades are sister groups; (ii) Chaetophorales is a separate lineage within Chlorophyceae and diverged at or near the origin of Chlorophyceae itself; (iii) *Hafniomonas montana* is ancestral to other Chlorophyceae; (iv) autosporic chlorococcalean group is polyphyletic (some taxa, e.g., *Scenedesmus obliquus*, affiliate with DO lineages, whereas other taxa, e.g., *Prototheca wickerhamii* and *Chlorella vulgaris*, affiliate with CCW lineages); and (v) the CW group evolved from CCW lineages.

Discontinuous chloroplast LSU rRNA does not represent an informative phylogenetic character because this trait is present as a shared ancestral character among green algae (Fig. 8C). Fragmentation patterns appear different in some lineages most likely as a consequence of independent events that occurred in phylogenetically distant lineages.

The presence of discontinuous mitochondrial LSU rRNAs in both CW and DO

taxa, but not in *Uronema belkae* and *Hafniomonas montana*, indicates that this trait appeared at or near the point of divergence into the two distinct chlorophycean lineages, namely, the CW group and DO group. Moreover, a trend can be identified in the evolution of the chlorophycean mitochondrial LSU rRNAs, that is, a tendency to increase the extent of discontinuity, from a continuous mitochondrial LSU rRNA in *Hafniomonas montana*, to at least two fragments in *Scenedesmus obliquus, Hormotilopsis gelatinosa, Planophila terrestris, Carteria crucifera, Polytomella agilis,* six in *Chlamydomonas eugametos* and eight in *C. reinhardtii.*

The presence of discontinuous chloroplast LSU rRNAs in *Pyramimonas parkae*, a micromonadophycean species, traces this trait back into the pool of primitive green algae from which all the green algal lineages evolved (Mattox and Stewart 1984), suggesting that this feature developed very early in the evolution of green algae.

2.5. Conclusions

Discontinuous mitochondrial and chloroplast LSU rRNAs are quite widespread among green algae. Mitochondrial LSU rRNAs appear discontinuous in zoosporic chlorophycean lineages displaying a clockwise or directly opposed configuration in their flagellar apparatus, as well as in chlorococcalean autosporic taxa phylogenetically related to them, but are continuous among zoosporic green algal lineages with a counterclockwise flagellar apparatus configuration, as well as among chlorococcalean autosporic taxa phylogenetically related to them. Chloroplast LSU rRNAs appear discontinuous in all of

the lineages investigated. Discontinuous mitochondrial LSU rRNA represents a molecular trait which might have originated at or near the base of Chlorophyceae, whereas discontinuous chloroplast LSU rRNA might have developed very early in the evolutionary history of green algae. I suggest, therefore, that the presence of discontinuous mitochondrial but not chloroplast LSU rRNA can be used as an additional character in assessing phylogenetic affiliations among green algae.

Chapter 3

Evolution of discontinuous mitochondrial LSU and SSU rRNA among green algae

3.1. Introduction

In the previous chapter I have shown that discontinuous mitochondrial LSU rRNAs are not confined to *Chlamydomonas* species but are, rather, a unifying feature for chlorophycean taxa with a CW or DO flagellar configuration as well as chlorococcalean taxa phylogenetically related to them. Moreover, the data suggested a trend in the evolution of this trait, that is, a tendency toward an increase in the degree of discontinuity, from a continuous mitochondrial LSU rRNA in *Hafniomonas montana* to highly fragmented mitochondrial LSU rRNAs in *C. eugametos* and *C. reinhardtii*.

This chapter provides additional information on fragmentation patterns of mitochondrial LSU as well as SSU rRNAs among chlorophycean lineages and provides insight into the mode of evolution of discontinuous mitochondrial rRNAs in this group. I have surveyed the mitochondrial LSU and SSU rRNAs in three taxa representing both the CW (i.e., *Chlamydomonas pulsatilla*) and DO (i.e., *Neochloris aquatica* and *Scenedesmus obliquus*) evolutionary lineages within the chlorophycean green algal group.

3.2. Materials and Methods

3.2.1. Algal cultures and growing conditions

The algal strains, sources, and growing media used were as follows: *Chlamydomonas reinhardtii* Dangeard (GC wt137c, Genetics Center at Duke University),

minimal medium (Lemieux et al. 1980); *Chlamydomonas eugametos* (UTEX 9, The Culture Collection of Algae at the University of Texas at Austin), minimal medium; *Chlamydomonas pulsatilla* Wollenweber (UTEX 2534, The Culture Collection of Algae at the University of Texas at Austin), artificial sea water (Starr and Zeikus 1993); *Neochloris aquatica* Starr (UTEX 138, The Culture Collection of Algae at the University of Texas at Austin), soil extract medium (Starr and Zeikus 1993); *Scenedesmus obliquus* (Turp.) Kutz (UTEX 78, The Culture Collection of Algae at the University of Texas at Austin), basal medium (Oh-Hama and Hase 1980); and *Prototheca wickerhamii* Soneda and Tubaki (UTEX 1533, The Culture Collection of Algae at the University of Texas at Austin), malt medium (Wolff and Kück 1990). Cultures were supplied with 1% CO_2 in air. Illumination was provided by cool-white fluorescent lamps (50-80 $\mu mol{\cdot}m^{-2}{\cdot}s^{-1}$) on a 12:12 h light-dark cycle. Cells were harvested when the culture density reached ca. $4x10^6$ $cells{\cdot}mL^{-1}$.

3.2.2. Total RNA extraction and fractionation

Total RNA from six green algal species was extracted according to Rochaix and Malnoë (1982). Glyoxalated total RNA (ca. 10 µg) from each taxon was fractionated by agarose (1.5%) gel electrophoresis (Sambrook et al. 1989) and transferred to Hybond-N nylon membranes (Amersham) by vacuum blotting under the conditions recommended by the manufacturer (Pharmacia).

3.2.3. Northern blot hybridization

RNA blots were prehybridized at 37°C for 4 h in the hybridization buffer: 5 x SSPE (20 x SSPE = 3.6 M NaCl, 200 mM NaH_2Po_4, 20 mM EDTA, pH 7.4), 1 x BLOTTO (10 x BLOTTO = 5% instant milk, 10% sodium dodecyl sulphate [SDS], pH 7.8), 50% formamide. To detect putative mitochondrial rRNAs in the species investigated, synthetic oligodeoxynucleotide probes complementary to regions within the mitochondrial LSU and SSU rRNA of *C. reinhardtii* and *S. obliquus* were used. The characteristics of the probes are summarized in Table 5, and the locations of their target regions are indicated in Figure 9IA and IIA. The oligonucleotide probes were 5'-end-labelled using [γ-^{32}P] ATP and polynucleotide kinase (Pharmacia) at 37°C for 1 h and then purified on Microspin Columns S-200 HR (Pharmacia). Hybridization reactions were carried out at 37°C for 21 h. The blots were washed twice for 15 min each time at room temperature, first in 2 x SSPE, 0.1% SDS, then in 0.5 x SSPE, 0.1% SDS.

3.3. Results

3.3.1. Discontinuous mitochondrial LSU and SSU rRNAs among chlorophycean green algae

Total RNA was extracted from *C. reinhardtii, C. eugametos, C. pulsatilla, N. aquatica, S. obliquus,* and *P. wickerhamii.* Figure 10A shows the electrophoretic patterns

Table 5. Characteristics of oligonucleotide probes 1 to 8.

Probe	Sequence (5 ' to 3')	complementary to	*E. coli* coordinates (Gutell et al. 1992)
1	CGGGACTATCACCCTCTTTGGTTTCC (26-mer)	5'-half LSU rRNA of *S. obliquus*	313-338
2	CACAGGACAACGGTGGCCCTTCTT (24-mer)	5'-half LSU rRNA of *S. obliquus*	478-500
3	GACTCGCTCACTCATGTTGCAAAAG GC (27-mer)	5'-half LSU rRNA of *S. obliquus*	563-589
4	CCGAACTTGATTGGCCTTTCACCCCT AGCCAC (32-mer)	5'-half LSU rRNA of *C. reinhardtii*	768-799
5	GCTAGACCAGTGAGCTATTACGCTT TC (27-mer)	5'-half LSU rRNA of *C. reinhardtii*	1087-1113
6	GCTGATAAACCTGTTATCCCTAGCG TA (27-mer)	3'-half LSU rRNA of *S. obliquus*	2438-2464
7	AGGACGCGATGATCCAACATCGAG GTGCC (29-mer)	3'-half LSU rRNA of *C. reinhardtii*	2494-2522
8	GGGTCTCTAATCCGGTTCGCTACCCA (26-mer)	SSU rRNA of *C. reinhardtii*	772-797

Figure 9. IA and IIA, The location of the target regions for oligonucleotide probes 1 to 8 represented to the *E. coli* LSU and SSU rRNA scale, respectively. IB and IIB, Fragmentation patterns within the mitochondrial (Mt) LSU and SSU rRNA, respectively, of *C. reinhardtii* (*Cr*) and *C. eugametos* (*Ce*); solid blocks indicate the smallest corresponding coding units that are continuous in the mitochondrial rRNAs of both *C. reinhardtii* and *C. eugametos*; stippled blocks indicate variable regions that are interrupted (arrows) in the two *Chlamydomonas* taxa; cross-hatched blocks indicate regions that are missing in both *Chlamydomonas* taxa. IC and IIC, The size and suggested coding capacity of the putative mitochondrial rRNAs identified in the taxa investigated; *Cp* = *Chlamydomonas pulsatilla*, *Na* = *Neochloris aquatica*, *So* = *Scenedesmus obliquus*, *Pw* = *Prototheca wickerhamii*.

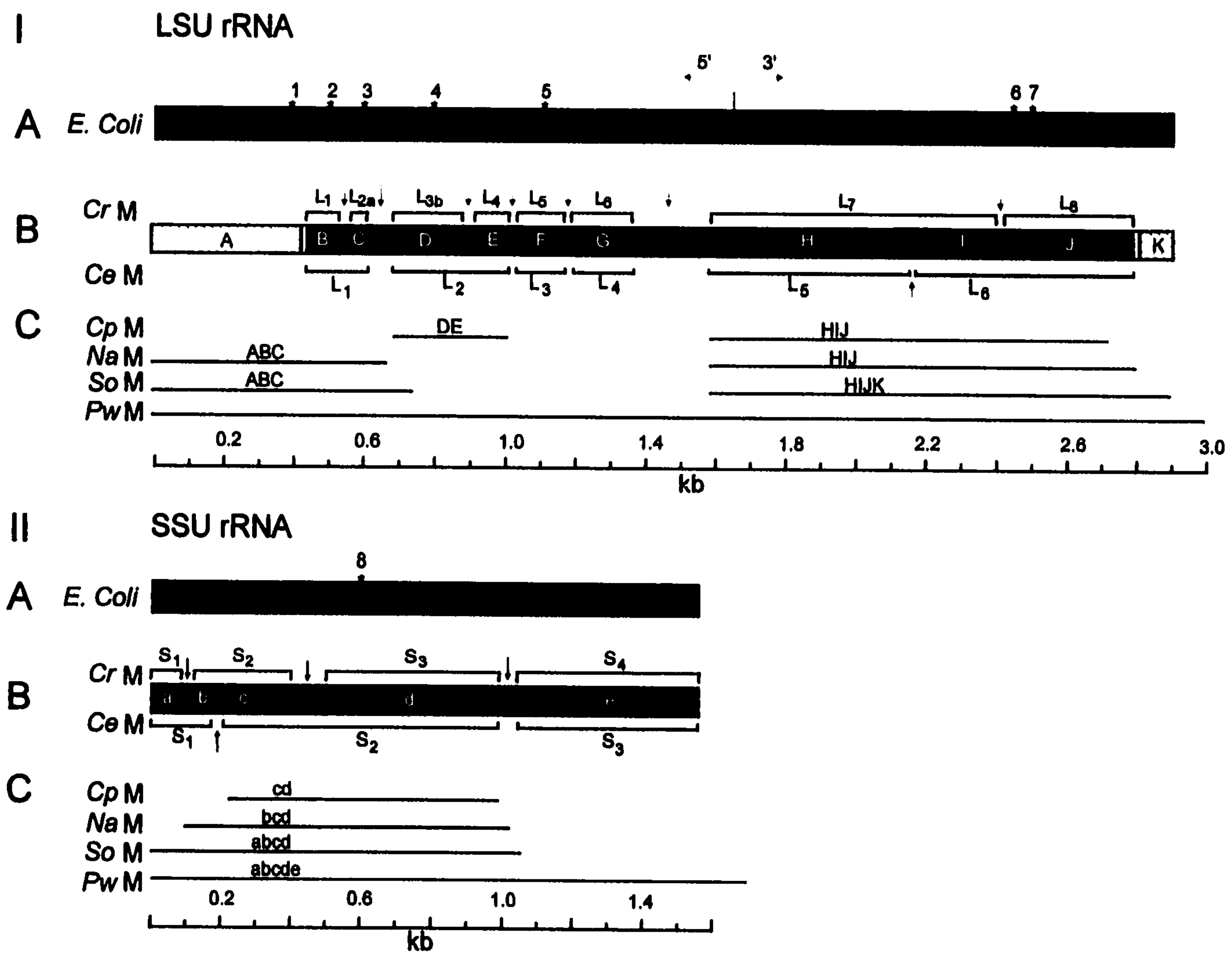


Figure 9

Figure 10. A. Electrophoretic patterns of green algal total RNA fractionated in 1.5% agarose gels; abbreviations as in Figure 9. RNA molecular sizes are expressed in nucleotides. *C. reinhardtii* rRNA size references indicated are: cytosolic LSU rRNA (3500 nt), cytosolic SSU rRNA (1800 nt), chloroplast LSU rRNA δ- fragment (1700 nt), chloroplast SSU rRNA (1500 nt), and chloroplast LSU rRNA γ-fragment (820 nt). B, Northern blot analyses of total RNA fractionated in 1.5% agarose gels and hybridized with oligonucleotide probes (1 to 8). Abbreviations are as in panel A). Only data for informative taxa shown. RNA molecular sizes are expressed in nucleotides. Stars, triangles, and circles denote rRNAs of mitochondrial, chloroplast, and cytosolic origin, respectively.

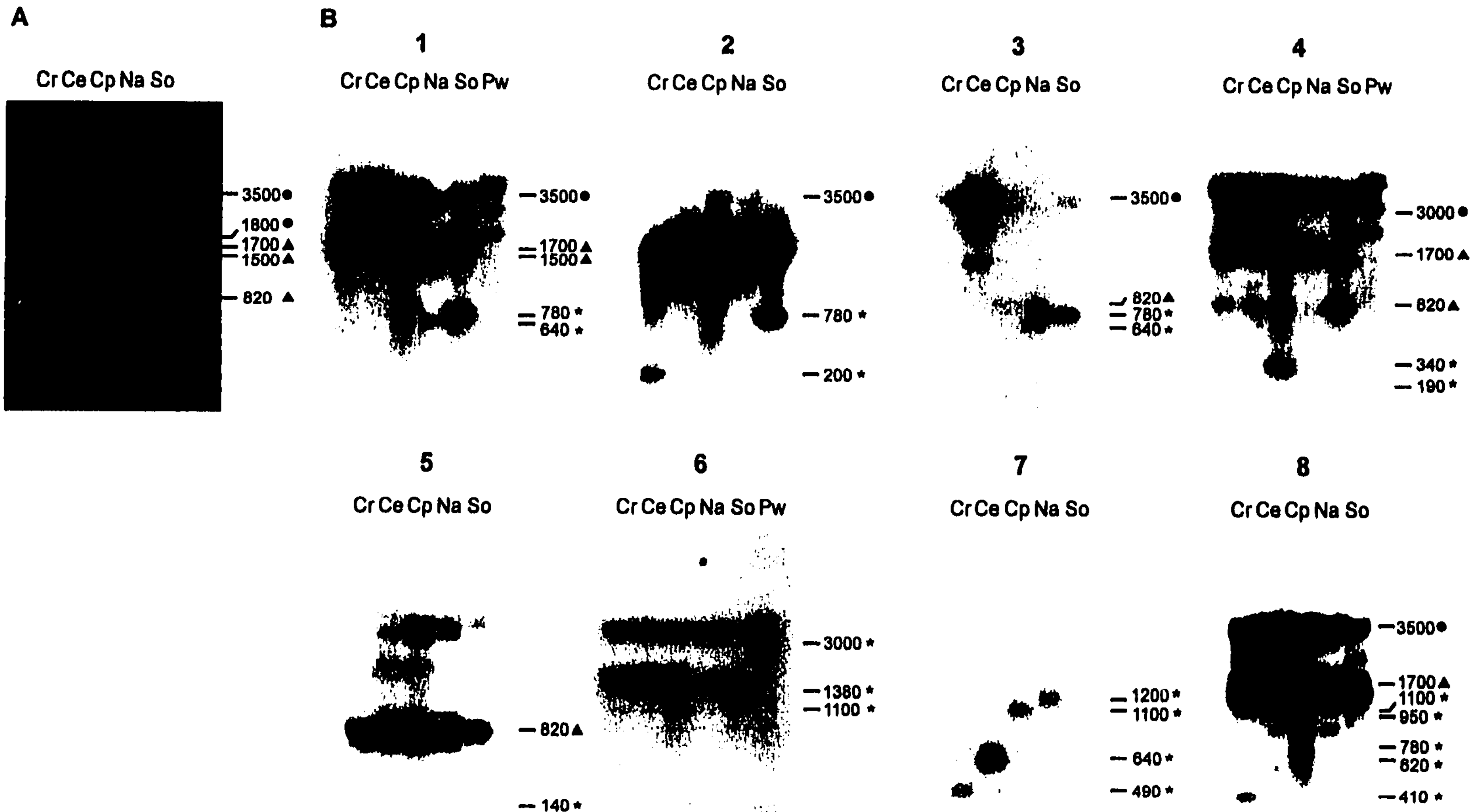


Figure 10

of the green algal total RNAs fractionated in 1.5% agarose gels and stained with ethidium bromide (EtBr). The abundant cytosolic SSU (1800 nt) and LSU rRNA (3500 nt), as well as the chloroplast SSU (1500 nt) and γ- and δ-LSU rRNA fragments (1700 and 820 nt, respectively) of *C. reinhardtii* and *C. eugametos* served as references for the identification of the rRNA counterparts in the taxa investigated. Mitochondrial rRNAs are not visible by EtBr-staining due to their low abundance in total RNA preparations.

Figure 9 indicates: (i) the location of the target regions for oligonucleotide probes 1 to 8 represented according to the *E. coli* LSU and SSU rRNA scale (Fig. 9-IA and 9-IIA), (ii) the corresponding mitochondrial LSU and SSU rRNA fragments of *C. reinhardtii* and *C. eugametos* (Fig. 9-IB and 9-IIB), and (iii) the putative mitochondrial rRNA counterparts identified in the species examined (Fig. 9-IC and 9-IIC). Information about the size of mitochondrial rRNAs hybridizing with oligonucleotide probes 1 to 8, as well as their locations within the mature rRNA complexes, i.e., SSU rRNA, 5'-half or 3'-half LSU rRNA, respectively, is summarized in Table 6.

Under the hybridization conditions used, the probes annealed not only with mitochondrial rRNAs but also with their cytosolic and/or chloroplast counterparts. Hybridizing RNAs with no visible correspondents on EtBr-stained gels and no hybridizing counterparts among the cytosolic and chloroplast rRNAs of *C. reinhardtii* or *C. eugametos* are most likely of mitochondrial origin (Fig. 10B). The oligonucleotide probes directed to regions within the 5'-half of the LSU rRNA identified small RNAs of about 340, 640, and 780 nt in *C. pulsatilla*, *N. aquatica*, and *S. obliquus*, respectively. The 200-, 190-, and 140-nt rRNAs corresponding to the *C. reinhardtii* mitochondrial LSU rRNA fragments

Table 6. Size (in nucleotides) of the hybridizing mitochondrial rRNAs detected by oligonucleotide probes 1 to 8.

Taxa	Probe directed to rRNA							
	5'-half LSU					3'-half LSU		SSU
	1	2	3	4	5	6	7	8
C. reinhardtii	-	200	-	190	140	-	490	410
C. eugametos	-	-	-	-	-	-	640	620
C. pulsatilla	-	-	-	340	-	1100	1100	780
N. aquatica	640	-	640	-	-	-	1200	950
S. obliquus	780	780	780	-	-	1380	-	1100
P. wickerhamii			3000			3000		

L_1, L_{3b}, and L_5, respectively, as well as the 3,000-nt rRNA transcript in *P. wickerhamii*, are presented as positive controls. The two oligonucleotide probes targeted to the peptidyl-transferase center within the 3'-half of the LSU rRNA hybridized with RNAs of about 1,100, 1,200, and 1,380 nt in *C. pulsatilla*, *N. aquatica*, and *S. obliquus*, respectively. The 490- and 640-nt RNAs correspond to the *C. reinhardtii* and *C. eugametos* mitochondrial LSU rRNA L_8 and L_6 fragments, respectively, and the 3000-nt RNA represents the large continuous mitochondrial LSU rRNA of *P. wickerhamii*. The oligonucleotide probe directed to the SSU rRNA detected RNAs of about 780, 950, and 1,100 nt in *C. pulsatilla*, *N. aquatica*, and *S. obliquus*, respectively. Positive controls are the S_3 SSU rRNA fragment in *C. reinhardtii* (410 nt) and the S_2 SSU rRNA fragment in *C. eugametos* (640 nt).

3.4. Discussion

3.4.1. A gradual increase in the degree of fragmentation of the mitochondrial LSU and SSU rRNAs within the chlorophycean green algal group

The Chlorophyceae is one of the five green algal classes (sensu Mattox and Stewart 1984) and consists of two evolutionarily distinct lineages with respect to the position of the basal bodies within the flagellar apparatus; these can be in either a DO or CW orientation. It has been proposed that the CW configuration evolved from an ancestral CCW configuration, and the DO configuration represents an intermediate stage

in the evolution of this trait (O'Kelly 1992). Taxa investigated in this study are representative of both CW and DO chlorophycean lineages. As a member of the *Chlamydomonas* genus, *C. pulsatilla* most likely belongs to the CW clade (no phylogenetic data are available for this taxon), whereas *N. aquatica* is a member of the DO clade. *S. obliquus* is an autosporic taxon (i.e., no flagellate stage in its life cycle) phylogenetically affiliated at the nuclear rRNA sequence level with taxa within the DO clade (Wilcox et al. 1992, Steinkötter et al. 1994).

In all known discontinuous rRNAs, the break points are confined to the variable regions of the rRNA molecules (Gray and Schnare 1996). Figures 9-IB and 9-IIB indicates the variable regions of the mitochondrial LSU and SSU rRNAs that are interrupted in both or only one of the two *Chlamydomonas* taxa, namely *C. reinhardtii* and *C. eugametos*. Letter designations were assigned to the smallest corresponding coding units that are continuous or missing in the mitochondrial rRNAs of both *Chlamydomonas* taxa; a-e and A-K refer to the SSU and LSU rRNAs, respectively, in the 5'-3' transcriptional order of their counterparts in conventional rRNAs. Among the variable regions of the *Chlamydomonas* mitochondrial rRNAs, five are interrupted in both taxa and seven are unique to either *C. reinhardtii* or *C. eugametos*. Denovan-Wright and Lee (1994) suggested that: (i) some or all of the common breakpoints in corresponding variable regions in the two *Chlamydomonas* taxa were inherited from their last common ancestor, and (ii) the unique breakpoints were derived independently after the divergence of the lineages leading to *C. reinhardtii* and *C. eugametos*. The interruption of additional variable regions in *C. reinhardtii* mitochondrial rRNAs relative to their *C. eugametos*

counterparts resulted in an increase in the degree of fragmentation and a decrease in the coding capacity of the corresponding coding modules in *C. reinhardtii*. For example, *C.eugametos* L_1 LSU rRNA coding module incorporates two adjacent coding units, i.e., B and C, whereas the *C. reinhardtii* mitochondrial L_1 LSU rRNA coding module consists only of coding unit B; in other words, the coding modules are not necessarily homologous between taxa, but the coding units are.

The previous chapter provided evidence that discontinuous mitochondrial LSU rRNA is not a feature unique to *Chlamydomonas* taxa but, rather, a unifying characteristic for the CW and DO green algal taxa. The data also suggested a trend in the evolution of the chlorophycean mitochondrial LSU rRNAs, i.e., a gradual increase in the extent of discontinuity, from a continuous mitochondrial LSU rRNA molecule in *H. montana* to six and eight LSU rRNA fragments in *C. eugametos* and *C. reinhardtii*, respectively. Moreover, the presence of continuous mitochondrial LSU rRNAs in *H. montana*, a taxon that retained ancestral-like features relative to other chlorophycean lineages (Ettl and Moestrup 1980, O'Kelly et al. 1994), indicated that discontinuous mitochondrial LSU rRNAs may have developed at or near the base of the Chlorophyceae.

This work provides additional data supporting a gradual increase in the degree of fragmentation of the mitochondrial LSU as well as SSU rRNA coding regions among chlorophycean green algae. The oligonucleotide probes used here are complementary to highly conserved regions within the mitochondrial rRNA coding units A, B, C, D, F, J, and d as designated in Figures 9-IB and 9-IIB. Extending the suggestion made by Denovan-Wright and Lee (1994), i.e., that the common breakpoints in the mitochondrial

rRNAs of *C. reinhardtii* and *C. eugametos* were inherited from their last common ancestor, I consider it most likely that the variable regions interrupted in both *Chlamydomonas* taxa as well as in *S. obliquus* were already interrupted in their most recent common ancestor. Note that the DO clade including *N. aquatica* and *S. obliquus* shares a common ancestor with the CW clade containing the *Chlamydomonas* taxa (Steinkötter et al. 1994). Consequently, the variable regions that are interrupted in both *C. reinhardtii* and *C. eugametos* as well as in *S. obliquus* mitochondrial rRNAs would most likely also be interrupted in *C. pulsatilla* and *N. aquatica*. Based on the size of the hybridizing mitochondrial rRNAs, the corresponding and the adjacent coding units in *C. reinhardtii* and *C. eugametos*, and the locations of variable regions likely to be interrupted, I suggest in Figure 9-IC and 9-IIC the coding units corresponding to the rRNA fragments identified in the species investigated. However, more detailed analyses have to be done to fully confirm these inferences. The identification of mitochondrial rRNA fragments corresponding to coding modules containing a higher number of coding units in *S. obliquus* (abcd, ABC, HIJK), *N. aquatica* (bcd, ABC, HIJ), and *C. pulsatilla* (cd, DE, HIJ) relative to *C. eugametos* (ab, cd, e, BC, DE, F, G, H, IJ) and *C. reinhardtii* (a, bc, d, e, B, C, D, E, F, G, HI, J) definitely suggests an evolutionary trend toward an increase in the degree of rRNA-coding-module fragmentation within the chlorophycean green algal group. The lower degree of fragmentation of mitochondrial rRNAs in *C. pulsatilla* compared to those of *C. reinhardtii* or *C. eugametos* (e.g., HIJ in *C. pulsatilla* compared to HI/J or H/IJ in *C. reinhardtii* or *C. eugametos*, respectively) indicates that this lineage may have diverged earlier than the most recent common ancestor of *C. reinhardtii* and *C.*

eugametos.

The results presented here also suggest that some of the breakpoints (i.e., d/e, C/D, G/H) shared by the three *Chlamydomonas* taxa are older than the most recent common ancestor of this group. This observation contradicts the previous suggestion of Denovan-Wright and Lee (1994) that the processes of mitochondrial rRNA fragmentation began in the *Chlamydomonas* lineage after its divergence from other chlorophycean species with conventional rRNAs but before the last common ancestor of *C. reinhardtii* and *C. eugametos.*

3.5. Conclusions

This chapter provides additional information on fragmentation patterns of mitochondrial SSU and LSU rRNAs that strongly supports the concept of a gradual increase in the extent of discontinuity of mitochondrial rRNAs among chlorophycean green algae.

Chapter 4

Discontinuous and scrambled mitochondrial rRNA coding regions among green algae:

A model for their origin and evolution

4.1. Introduction

Ribosomal RNAs are essential components for both the structure and function of ribosomes in all prokaryotic, eukaryotic, and organellar genetic systems. Most known LSU and SSU rRNAs are rather strongly conserved in size and secondary structure within their respective type and evolutionary lineage (Gutell 1992). However, in some lineages, unconventional rRNAs have been described, including the rRNA complexes composed of split rather than single, covalently continuous polyribonucleotide chains. The genes coding for the discontinuous rRNAs deviate from the conventional structure in that they are fragmented into coding modules that can be interspersed with either internal transcribed spacers, or protein-coding and/or tRNA genes. Moreover, in some mitochondrial genetic systems (e.g., *Tetrahymena pyriformis* [Heinonen et al. 1987], *Chlamydomonas reinhardtii* [Boer and Gray 1988a], *Chlamydomonas eugametos* [Denovan-Wright and Lee 1994], *Plasmodium sp.* [Vaidya et al. 1989, Feagin et al. 1992] and *Theileria parva* [Kairo et al. 1992]) the rRNA coding modules no longer follow the 5'-3' transcriptional order of their counterparts within conventional continuous genes; rather, they are scrambled within the genome. In addition, in *Plasmodium falciparum* and *T. parva*, mitochondrial rRNA gene pieces are not coded on the same DNA strand.

It has been proposed that all split rRNA genes are derived and evolved from continuous homologs (Gray and Schnare 1996). However, the processes responsible for the fragmentation of rRNA coding regions are not fully understood in any of the described genetic systems. Fragmented and scrambled mitochondrial rRNA genes have

been reported in green algae, ciliates, and apicomplexans and it seems likely they are the result of independent events that occurred in evolutionarily distant lineages.

The green algae represent a good study group with which to address questions about the origin and evolution of fragmented and scrambled mitochondrial rRNA coding regions. Two very distinct mitochondrial rRNA gene organizations have been described for the three green algal species investigated in this respect: conventional continuous rRNA genes in *Prototheca wickerhamii* (Wolff et al. 1994) and highly fragmented and scrambled rRNA coding regions in *C. reinhardtii* (Boer and Gray 1988a) and *C. eugametos* (Denovan-Wright and Lee 1994).

Although mitochondrial rRNA genes are highly fragmented and scrambled in both *C. reinhardtii* and *C. eugametos*, the distributions of the coding information among their coding modules, as well the order of these modules within the genome, are different between the two species (Denovan-Wright and Lee 1994). Calculations of the minimal number of transpositions required to convert hypothetical ancestral rRNA gene organizations to the arrangements present in the two *Chlamydomonas* taxa, as well as a limited survey of the size of mitochondrial LSU rRNAs in other *Chlamydomonas* species, led Denovan-Wright et al. (1996) to propose that the last common ancestor of *Chlamydomonas* algae possessed fragmented mitochondrial rRNA genes whose coding modules were nearly colinear with their counterparts in conventional continuous rRNA genes. The authors presented a model in which the fragmentation and scrambling of the coding modules were assumed to be separate, consecutive events, and the rearrangement of the rRNA gene pieces was limited to transpositional events. However, no specific

mechanism has been proposed to explain either the fragmentation or the scrambling of the resulting gene pieces.

This study (i) suggests a mechanism that may have been responsible for both the fragmentation and scrambling of the mitochondrial rRNA genes within the chlorophycean green algal group, and (ii) presents a hypothetical pathway for converting continuous mitochondrial rRNA genes to the highly fragmented and scrambled rRNA coding regions of *Chlamydomonas*.

4.2. Material and Methods

4.2.1. DNA sequence similarity analysis

The mitochondrial DNA sequence (EMBL X17375) of *Scenedesmus obliquus* (KS3/2) was analyzed for sequence similarity using the BLAST algorithm (Altschul et al. 1990).

4.2.2. Northern blot hybridizations

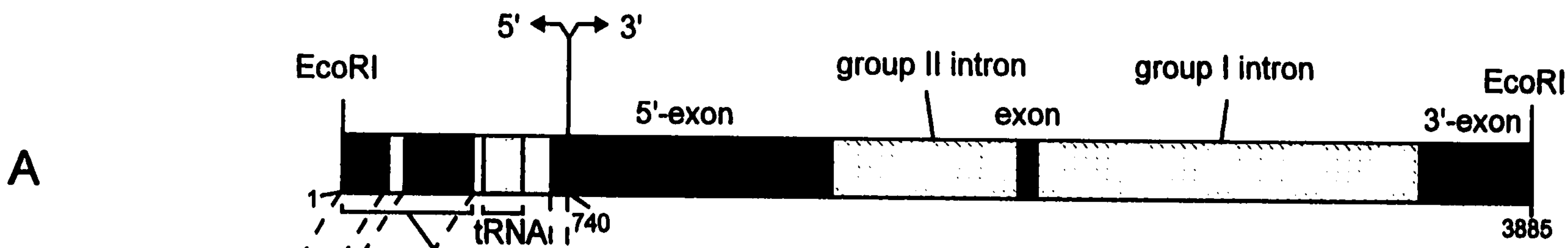
S. obliquus total RNA was extracted, fractionated and blotted as described in Chapter 3. Hybridization conditions were as indicated in Chapter 3.

4.3. Results

4.3.1. *Scenedesmus obliquus* mitochondrial LSU rRNA coding regions analysis

A 3,885-bp mtDNA sequence (EMBL X17375) of a wild strain of *S. obliquus* was analyzed for sequence similarity with the BLAST algorithm (Altschul et al. 1990). New coding regions in addition to the 3'-half LSU rRNA coding region previously reported (Kück et al. 1990) were identified. Figure 11 provides a diagrammatic representation of the 3'-region of the LSU rRNA gene of *S. obliquus* as described by Kück et al. (1990), as well as the locations of the new tRNA- and rRNA-coding regions identified here (Fig. 11A), and compares the mitochondrial LSU rRNA coding regions of *S. obliquus* to those of *P. wickerhamii* (Fig. 11B). Within the first 480 nucleotides of the 3,885-bp *S. obliquus* mtDNA sequence, short regions showed 70-89% similarity to sequences within the 5'-distal part of the 5'-half of eubacterial, mitochondrial, chloroplastic, and cytosolic LSU rRNAs. In addition, the region between coordinates 500 and 600 of the DNA sequence displayed 70-85% sequence similarity to tRNA coding regions from various eubacterial, mitochondrial, and chloroplast genomes. Finally, starting around coordinate 700 of the *S. obliquus* mtDNA sequence, the first 40 nucleotides showed high sequence similarity with the very 3'-end of the 5'-half LSU rRNA and the following nucleotides with the adjacent 3'-half of various LSU rRNAs as proposed by Kück et al. (1990). No sequence similarities with the central part of the 5'-half LSU rRNA domain were found, but the functional significance and its ubiquitous presence in most counterparts strongly suggest that the corresponding LSU rRNA coding region is present at a different location

Figure 11. A. Diagrammatic representation of a mitochondrial LSU rDNA sequence of *Scenedesmus obliquus* illustrating the 3'-half LSU rRNA coding region (solid blocks) and two introns (cross-hatched blocks) as reported by Kück et al. (1990), as well the new coding regions identified within its 5'-region (stippled blocks). B. Comparison between the mitochondrial LSU rRNA gene of *Prototheca wickerhamii* and that of *S. obliquus* indicating the location of corresponding coding regions (stippled blocks) within the 5'-half of their LSU rRNA genes.



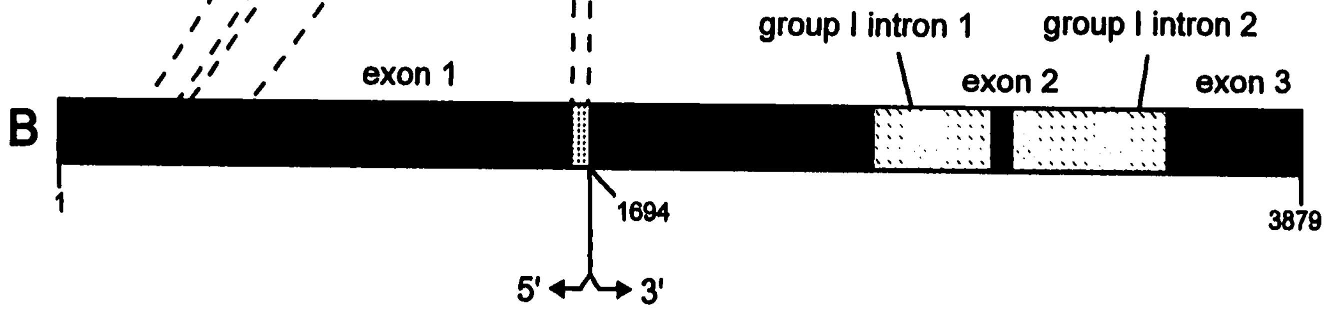


Figure 11

along the mitochondrial genome. On this evidence, I therefore conclude that in *S. obliquus*, the mitochondrial LSU rRNA gene is fragmented and the coding regions are scrambled. This is the first example of fragmented and scrambled mitochondrial LSU rRNA coding regions in a non-*Chlamydomonas* green algal taxon.

To determine if the observed sequence similarities of the new rRNA-coding regions identified in the *S. obliquus* mtDNA sequence are biologically meaningful, the DNA sequence was used to model the secondary structures of the potential transcripts (Fig. 12A). The first 480 nt of the DNA sequence can be folded in a conventional eubacteria-like secondary structure corresponding to the region between *E. coli* coordinates 270 and 660. An additional 70-nt sequence is present around the 370 *E. coli* coordinate; no other LSU rRNAs have such a sequence in this region, but in higher plant mitochondrial LSU rRNAs, a longer extra sequence is present 100 nt downstream of this location. A 98-nt sequence not present in other LSU rRNA counterparts (with the exception of the mitochondrial LSU rRNA of *Saccharomyces cerevisiae* that contains a 120-nt sequence) was also reported in one of the variable regions in the 3'-half of the *S. obliquus* mitochondrial LSU rRNA (Kück et al. 1990). The putative LSU rRNA fragment corresponds to the *C. reinhardtii* mitochondrial L_1 and L_{2a} LSU rRNA fragments and extends 180 nt into the first 450 nt of the 5'-end LSU rRNA, a region that is completely missing from both *C. reinhardtii* and *C. eugametos* mitochondrial LSU rRNAs.

To rule out the possibility that the LSU rRNA coding region identified is a pseudogene and to determine if the coding region is transcribed as one single transcript, a series of Northern blot hybridizations were performed using total RNA and three

Figure 12. A. Potential secondary structure of an LSU rRNA fragment deduced from the DNA sequence of the *S. obliquus* mtDNA analyzed. Unbracketed and bracketed numbers represent the coordinates in the mtDNA sequence (EMBL X17375) of *S. obliquus* and LSU rRNA of *E. coli*, respectively. Nucleotides that are identical in the eubacterial LSU rRNA of *Flavobacterium odoratum* are in bold. Arrows indicate the breakpoints in the mitochondrial LSU rRNA of *C. reinhardtii*. Boxed region indicates the LSU rRNA part that is missing in *Chlamydomonas*. B. Potential secondary structure of a tRNA (Ala, UGC) deduced from the tDNA-like sequence identified in the *S. obliquus* mtDNA analyzed. Nucleotides are numbered according to Sprinzl et al. (1996). Bold and stars mark nucleotide identical in *Rickettsia* tRNA (Ala, UGC) and different in *Rickettsia* but identical in *Anacystis* tRNA (Ala, UGC), respectively. C. Alignment between the mitochondrial tDNA (Ala, UGC) sequences of *S. obliquus* (S.o.) and *P.wickerhamii* (P.w.). Nucleotides in bold indicate the anticodon sequence. D. Comparison between the potential secondary structures and tertiary interactions of *S. obliquus* and *P. wickerhamii* tRNA^{Ala}. Bold characters denote invariable nucleotides, continuous and dotted lines indicate potential and missing tertiary interactions, respectively, relative to the standard tertiary structure of other tRNAs. The stars in the *S. obliquus* tRNA^{Ala} mark the nucleotides that are identical between *S. obliquus* and *P. wickerhamii* counterparts.

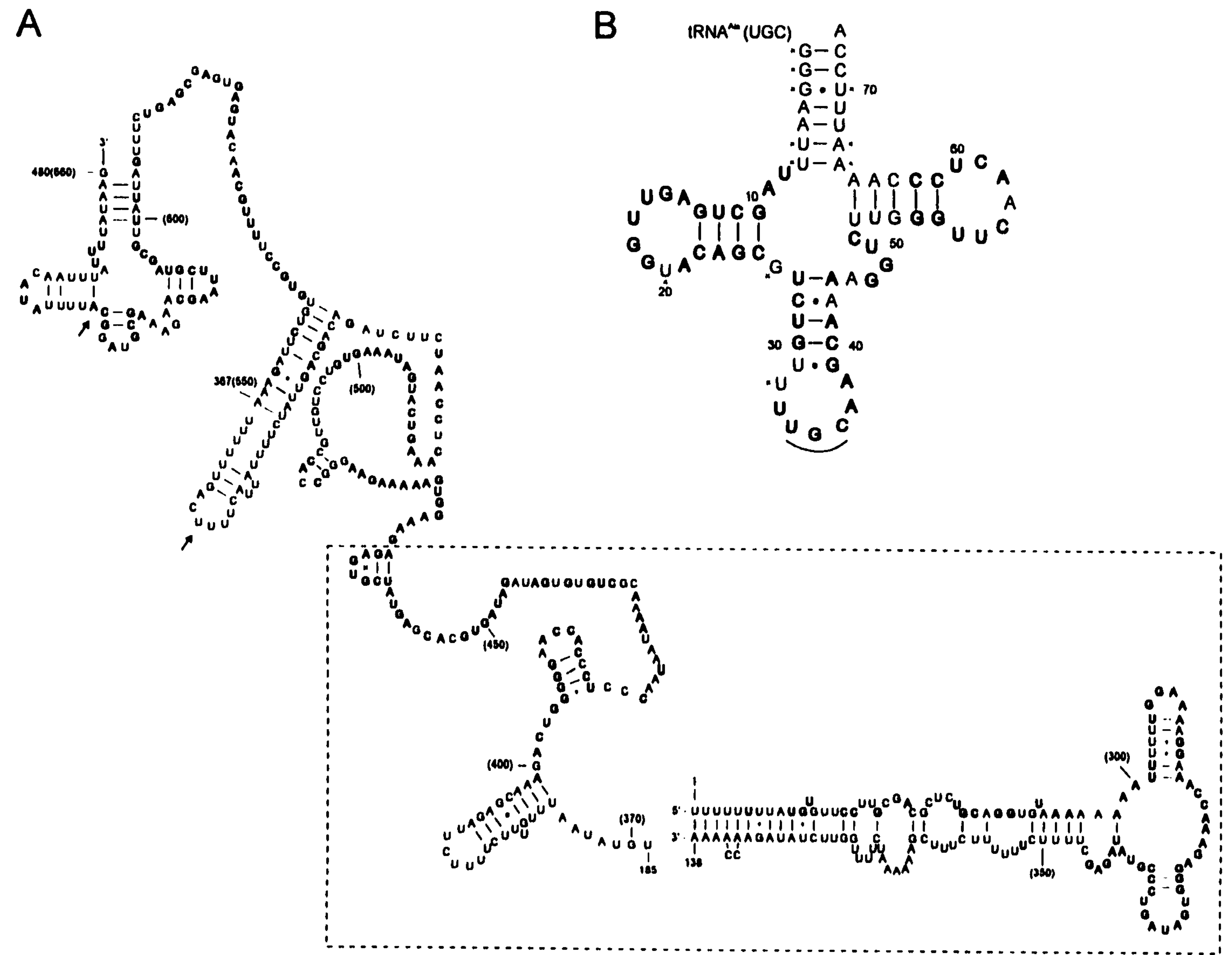


Figure 12

C

```
So: GGGAATTTAGCTGAGTTGGTACAGCGTCTGTTTTGCAAGCAAAAGGTCTTGGGTTCAACTCCCAAAATTTCCA
Pw: GGGGATGTAGTTCAGTTGGTAGAACGCATGCTTTGCACGCATGAGGTCATCGGTTCGATCCCGTTCGTCTCCA
    *** ** *** * ******** * **  ** ****** ***  ***** * ***** *  **      * ****
```

D

S. obliquus tRNA[Ala]

P. wickerhamii tRNA[Ala]

Figure 12

synthetic deoxyoligonucleotide probes complementary to regions within the 5'-end LSU rRNA coding region of *S. obliquus*. All three oligonucleotide probes (i.e., 1, 2, and 3) directed to the potential 5'-end mitochondrial LSU rRNA transcript hybridized with a 780-nt RNA molecule (Fig. 10B). The size of this transcript is larger than that of the rRNA predicted based on the coding region analyzed (480 nt), indicating that the coding region extends outside the 5'-end of the sequenced clone and is transcribed into a single transcript.

4.3.2. *Scenedesmus obliquus* mitochondrial tRNA coding region analysis

The *S. obliquus* tRNA-like DNA sequence can be folded in a conventional tRNA structure with the normal pattern of invariant and semi-invariant nucleotides as well as potential tertiary interactions described in other conventional tRNAs and (Fig. 12D). The UGC triplet at the anticodon position suggests a tRNA (Ala, UGC) role for the potential transcript. The tDNA sequence shows 68.5% identity to the *P. wickerhamii* $tRNA^{Ala}$ counterpart (Fig. 12C). Moreover, the *S. obliquus* tDNA sequence is 60% identical to the tRNA (Ala, UGC) from *Rickettsia*; in addition, at 11 positions that are different between the *S. obliquus* and *Rickettsia* $tRNA^{Ala}$ coding regions, the nucleotides in the former are identical to those at homologous positions in *Anacystis* $tRNA^{Ala}$ (Fig. 12B).

4.4. Discussion

4.4.1. Fragmented and scrambled mitochondrial rRNA coding regions among chlorophycean green algae

The analysis of the 3,885-nt mtDNA sequence of *S. obliquus* revealed that the mitochondrial LSU rRNA gene in this species is not only fragmented, as previously suggested (Chapter 2 and 3) (Nedelcu et al. 1996), but also scrambled. The two LSU rRNA coding regions identified through nucleotide sequence and secondary structure modelling comparisons consist of the coding modules ABC and HIJ, respectively, (Fig. 11) and are separated by a tRNA coding region. It should be noted that the mitochondrial LSU rRNA gene in the ciliate *Tetrahymena pyriformis* is also interrupted by a tRNA gene and the two LSU rRNA coding modules are rearranged (Heinonen et al. 1987). The internal part of the LSU rRNA gene corresponding to the D, E, F, and G coding units, which is highly conserved, functionally important, and ubiquitously present in various counterparts, is presumably present in the *S. obliquus* mitochondrial genome at another location. Both of the breakpoints within the mitochondrial LSU rRNA coding region of *S. obliquus* (i.e., C/D and G/H) are also shared by all three *Chlamydomonas* taxa (Fig. 9). On the other hand, the mitochondrial LSU rRNA gene of *S. obliquus* contains the two coding units A and K, which are missing in *Chlamydomonas* but are present in *P. wickerhamii* and land plants. This finding suggests that these two coding units were present in the mitochondrial LSU rRNA gene in the common ancestor of the DO and CW lineages.

The tDNA sequence interrupting the rRNA coding region of *S. obliquus* mtDNA

can be folded in a conventional $tRNA^{Ala}$ secondary structure (Fig. 12B). Such a tRNA is missing in both of the *Chlamydomonas* mtDNAs investigated to date but is present in the *P. wickerhamii* mitochondrial genome. It is noteworthy that the potential *S. obliquus* tRNA (Ala, UGC) reveals the normal pattern of invariant and semi-invariant nucleotides present in most conventional tRNAs, such as: (i) the invariant U8, A14, G18, U33, and a purine at position 26, (ii) many Watson-Crick basepairing in the cloverleaf stems, and (iii) the presence of many tertiary interactions (Fig. 12D). It should be mentioned that although many *P. wickerhamii* tRNAs deduced from mt DNA sequences have been reported to have unorthodox features, they are expected to be functional (Wolff et al. 1994). It is therefore conceivable that the $tRNA^{Ala}$ as deduced from the *S. obliquus* mitochondrial tDNA sequence is functional and the tRNA coding region is not a pseudogene.

4.4.2. Discontinuous then scrambled versus discontinuous and scrambled mitochondrial rRNA coding regions

The mitochondrial rRNA genes of *C. reinhardtii* (Boer and Gray 1988a) and *C. eugametos* (Denovan-Wright and Lee 1994) show different degrees and patterns of fragmentation and scrambling of the mitochondrial rRNA coding regions. Denovan-Wright et al. (1996) used the DERANGE program (Sankoff et al. 1992) to deduce the structure and organization of the mitochondrial rRNA coding regions in the last common ancestor of *C. reinhardtii* and *C. eugametos*. Denovan-Wright et al. assumed that the

mitochondrial rRNA coding regions were altered to produce individual transcripts prior to the gene piece rearrangements that were limited to transpositional events. The model presented by the authors suggests that (i) in the last common ancestor of *C. reinhardtii* and *C. eugametos*, the mitochondrial rRNA coding regions were fragmented in gene pieces colinear with their counterparts in continuous rRNA genes with the exception of a small coding module situated upstream of the SSU rRNA gene; and (ii) the fragmented and scrambled mitochondrial rRNA coding regions in this ancestor had evolved from already fairly fragmented but not scrambled rRNA coding regions in a single evolutionary step (i.e., one transpositional event). This model predicts that in taxa basal to the *Chlamydomonas* group, mitochondrial rRNA genes should be fragmented but not scrambled.

The lower degree of fragmentation of the mitochondrial rRNAs in *Chlamydomonas pulsatilla* relative to *C. reinhardtii* and *C. eugametos* (discussed in the previous chapter) suggests that *C. pulsatilla* might have diverged before the most recent common ancestor of these two *Chlamydomonas* species. Therefore, it would be of interest to find out whether the mitochondrial rRNA coding regions of *C. pulsatilla* are colinear with their counterparts in continuous genes or scrambled within the genome. However, the presence of fragmented and scrambled mitochondrial LSU rRNA coding regions in *S. obliquus*, an autosporic species phylogenetically related to zoosporic DO taxa (Steinkötter et al. 1994), indicates that scrambling may have already been present in the most recent common ancestor of the DO and CW lineages, much earlier than previously proposed (Denovan-Wright et al. 1996). The suggested low degree of fragmentation of the *S. obliquus*

mitochondrial LSU rRNA gene, as well as the presence of features that are absent in *Chlamydomonas* but present in *P. wickerhamii* mitochondria (e.g., the 5'-end and 3'-end LSU rRNA regions), indicates that the mitochondrial LSU rRNA gene in this taxon represents an early stage in the evolution of discontinuous and scrambled mitochondrial rRNA genes within Chlorophyceae. It appears, therefore, that the mitochondrial rRNA genes in the chlorophycean green algae may have undergone rearrangements before they became highly fragmented. The presence of a tRNA gene between the two mitochondrial LSU rRNA coding regions in *S. obliquus* may be a consequence of a rearrangement event. Similar observations regarding the presence of tRNA or tRNA-like genes in the proximity of the endpoints of rearranged sequences in land plant and *Chlamydomonas* chloroplast genomes have led to the hypothesis that tRNA gene sequences may be implicated in gene shuffling (Boudreau and Turmel 1996 and references therein). In conclusion, the scrambling of the mitochondrial rRNA coding regions may have developed at an early stage in the evolution of chlorophycean mitochondrial rRNA genes, most likely in parallel with the fragmentation events.

4.4.3. Recombination as a possible mechanism responsible for the mitochondrial rRNA gene rearrangements within Chlorophyceae

The mechanisms responsible for either the fragmentation or the scrambling of the mitochondrial rRNA coding regions in *Chlamydomonas* are not known yet, although several suggestions have been made. The GC-rich repeat clusters identified in *C.*

reinhardtii mtDNA were suspected to have contributed to the extensive rRNA gene arrangements through a mechanism analogous to bacterial transposition (Boer and Gray 1991). Denovan-Wright and Lee (1994), however, assumed that the unusual gene structure in *Chlamydomonas* mitochondria arose from conventional, continuous rRNA genes by two separate, consecutive processes: the introduction of processing signals and the scrambling of coding regions defined by these signals. The absence of a reverse-transcriptase-like open reading frame in *C. eugametos* mtDNA led the authors to favour the view that the mitochondrial rRNA coding regions in *Chlamydomonas* became scrambled by recombination between nonhomologous regions of mtDNA molecules such as the dispersed repeated elements found in *C. reinhardtii* (Boer and Gray 1991) and *C. eugametos* (their unpublished results) rather than by reverse transcription (Boer and Gray 1988a).

The presence of scrambled but not highly fragmented mitochondrial LSU rRNA coding regions in *S. obliquus* suggests that scrambling may have developed at an early stage in the evolution of discontinuous and scrambled rRNA genes, probably complementing the fragmentation events. Considering this possibility, I propose a model that would disrupt and scramble a coding region in a single step via an intramolecular homologous recombinatorial event between two sets of two-copy inverted repeats. Generally, such an event would result in an interchange of the sequences situated between the two sets of inverted repeats (Fig. 4C). Figure 13 illustrates how recombination events such as the one suggested above, as well as those shown to be responsible for deletions in fungal mtDNAs (Jamiet-Vierny et al. 1997 and references therein) (discussed in

Chapter 5), could have been involved in the evolution of chlorophycean mitochondrial rRNA genes. The model requires that short inverted repeat sequences be present within the variable regions of an rRNA gene as well as flanking a non-rRNA coding region; in addition, the two coding regions to be exchanged must be present in oppositely oriented transcriptional units. In this way the rRNA gene becomes fragmented and scrambled simultaneously, and the exchanged coding regions could be incorporated within the opposite transcriptional unit and transcribed accordingly. The rRNA coding module transferred into a new transcriptional unit could be released from the new polycistronic transcript following the processing of the adjacent protein-coding or tRNA sequences from the transcript, and could subsequently interact by intermolecular basepairing with the other two rRNA fragments to restore the conserved rRNA secondary structure.

Such a scenario has the advantage that an rRNA gene would become fragmented and scrambled as a result of a single rearrangement event; there is no need to suggest two distinct mechanisms to explain the observed rearrangements. In addition, by proposing recombination as a mechanism involved in the evolution of mitochondrial rRNA coding regions within the chlorophycean group, not only the fragmentation and scrambling of the rRNA coding region, but also the loss of rRNA and non-rRNA coding regions during the evolution of the mitochondrial genome of green algae could be explained.

Short repeated sequences have been reported in the fungal and plant mitochondrial genome and shown to be involved in intramolecular recombination events (discussed in Chapter 5). Short GC-rich repeat clusters have been described in the *C. reinhardtii* mitochondrial genome and suggested to be reminiscent of the *Pst*I palindromes in

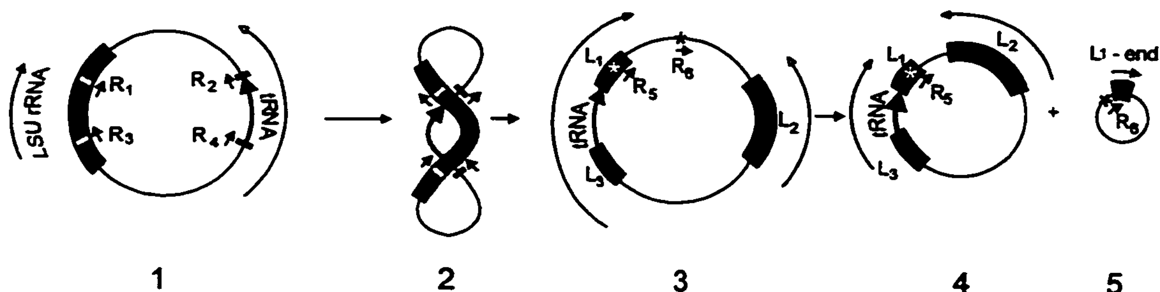


Figure 13. A model for the fragmentation, scrambling and excision of mitochondrial rRNA coding regions within a chlorophycean mitochondrial genome involving recombinational events mediated by short inverted and direct repeated sequences. Two sets of short inverted repeats, i.e., R_1/R_2 and R_3/R_4 (small solid blocks), present within the variable regions of a LSU rRNA gene (grey block) and flanking a tRNA gene (solid arrow), respectively (step 1), recombine simultaneously (step 2) and generate a new molecule in which the sequences between the two sets of repeats are interchanged (step 3). Note that the LSU rRNA and tRNA genes belong to transcription units transcribed in opposite directions. A recombination event between a two-copy direct repeat, i.e., R_5 and R_6 (stars), one of which is situated within the L_1-LSU rRNA coding region (step 4), would result in the excision of the sequences between the repeats, including a small LSU rRNA coding region (step 5).

Neurospora crassa mtDNA or analogous to the GC clusters in *Saccharomyces cerevisiae* mtDNA (Boer and Gray 1991 and references therein). I have analyzed the mtDNA sequence of *C. eugametos* mtDNA and identified a similar but larger and more complex set of GC-rich direct and inverted repeats within spacer regions as well as introns (data will be presented and discussed in Chapter 5). Likewise, in the intergenic spacers between the tRNA gene and the LSU rRNA coding regions of the *S. obliquus* mtDNA sequence analyzed, I have identified remnants of the complex AT-rich repetitive motifs flanking the tRNA genes in the mitochondrial genome of *P. wickerhamii* (Wolff et al. 1994), as well as GC-rich short inverted repeats similar to those present in the mitochondrial genome of *C. reinhardtii* (discussed in Chapter 5) (Nedelcu and Lee submitted). Comparisons among the locations of these elements within the mitochondrial genome of *C. reinhardtii*, *C. eugametos*, and the available sequence of *S. obliquus* revealed similarities regarding the positions of these repeats relative to the rRNA coding units within the respective genomes (see Fig. 14). I therefore suspect that the fragmented and scrambled mitochondrial rRNA coding regions in the chlorophycean green algal group could have been generated through multiple recombination events triggered by the accumulation of short repeated sequences within the variable regions of the rRNA genes and the intergenic spacers of these mitochondrial genomes. It is noteworthy in this connection that recombination events between short dispersed repeats have also been proposed to account for the various rearrangements described in the *Chlamydomonas* chloroplast genome (discussed in Chapter 7).

To illustrate how recombination events similar to those presented here could be

entirely responsible for the extensive mitochondrial rRNA gene rearrangements, a hypothetical pathway to gradually convert conventional continuous mitochondrial rRNA genes to the rRNA gene arrangement described in *C. eugametos* is presented in Figure 14. Fragmentation patterns similar to those suggested for *S. obliquus*, *N. aquatica*, and *C. pulsatilla* were incorporated into the pathway. Note that the pathway is not necessarily the most parsimonious solution and is not intended to suggest the succession of events or the ancestral genomic organization of any of the taxa indicated at the termini of the pathway. The scenario also shows that recombinatorial events could have been involved not only in the rRNA gene rearrangements but also in the removal of some rRNA as well as non-rRNA coding regions to the extent observed in the mitochondrial genomes of *C. reinhardtii* and *C. eugametos*.

The hypothetical pathway presented in Figure 14 starts from a circular mitochondrial genome as described for most of the green algal lineages investigated to date, including the primitive-like green flagellate *Platymonas subcordiformis*. In addition, in this presumptive ancestral mitochondrial genome the genes are considered to have been organized in more than one transcriptional unit, as reported in *P. subcordiformis* (Kessler and Zetsche 1995), and the SSU and LSU rRNA genes are assumed to have been transcribed in opposite directions, as described in *P. wickerhamii* (Wolff et al. 1994). The rRNA-coding units and tRNA- or protein-coding genes to be interchanged are always from opposite transcriptional units. The presence of more than one gene copy for the rRNA genes was not taken into account since all of the green algal mitochondrial genomes investigated to date encode single copies of all their genes. This scenario shows

Figure 14. Hypothetical pathway from conventional continuous mitochondrial rRNA genes to the gene arrangement described in *C. eugametos* mitochondrial genome, and the location of short repeated sequences within the *C. eugametos, C. reinhardtii*, and *S. obliquus* mtDNA. Diagrams are not drawn to scale. Letters indicate rRNA coding units (a-e refer to SSU rRNA, and A-K refer to LSU rRNA); blocked letters indicate a coding module. Thick solid arrows on the circles designate non rDNA sequences (i.e., tRNA or protein-coding genes) and indicate the transcriptional direction of that sequence. Thin solid arrows outside the circle indicate the transcriptional direction of that region. Interrupted arrows indicate recombinatorial events as follows: 1 to 9 are recombinatorial events between two sets of two-copy inverted repeats resulting in interchanges of sequences from opposite transcriptional units; 10 is a recombinatorial event between one set of a two-copy inverted repeats resulting in changing the transcriptional orientation of the regions flanked by the repeats; 11 to 14 are recombinatorial events between one set of two-copy direct repeats followed by the excision of small subgenomic circles containing the sequence situated between the repeats. Small solid circles on the *C. reinhardtii, C. eugametos* and *S. obliquus* gene maps denote the position of intergenic clusters of repeats.

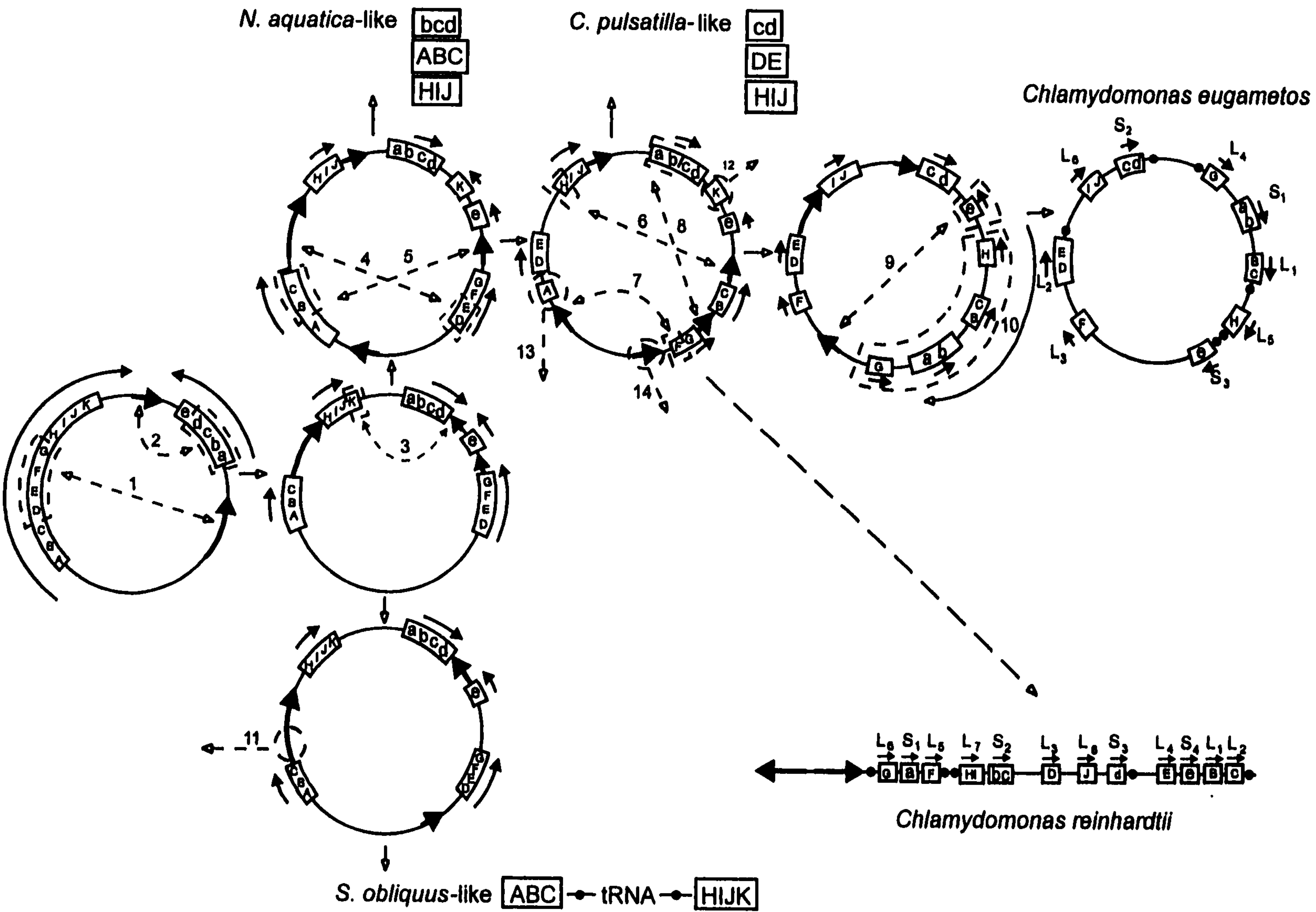


Figure 14

that, starting from a circular genome containing more than one transcriptional unit and continuous conventional LSU rRNA and SSU rRNA genes transcribed in opposite directions, a genome organization similar to that of *C. eugametos* (i.e., nine rRNA coding modules scrambled along the genome and transcribed in one direction) can be reached through a series of 12 recombinatorial events between two sets of two-copy inverted repeats (9 events), one set of two-copy direct repeats (2 events), and one set of a two-copy inverted repeat (1 event).

The phylogenetic position, i.e., in both CW and DO chlorophycean lineages, of the taxa examined in this study allows one to speculate that evolutionary changes in rRNA gene organization started around the chlorophycean divergence from the pool of ancestral green flagellate algae and have continued since. These changes may have involved both a gradual fragmentation and loss of mitochondrial rRNA coding regions as well as a decrease in size and coding capacity of the entire mitochondrial genome to the dramatic extent observed in *C. reinhardtii* and *C. eugametos*. To further argue for such a scenario, more data on mitochondrial genomes of other chlorophycean green algae have to become available.

4.5. Conclusions

This chapter reports the first example of fragmented and scrambled mitochondrial LSU rRNA coding regions in a green algal taxon outside the *Chlamydomonas* group. Scrambling of the mitochondrial rRNA coding regions most likely occurred early in the

evolution of fragmented and scrambled mitochondrial rRNA genes within the chlorophycean green algal group, most likely in parallel with the fragmentation events. Recombination might have been involved in the evolution of the fragmented and scrambled mitochondrial rRNA genes within the chlorophycean green algal group. Continuous mitochondrial rRNA genes could be converted into highly fragmented and scrambled rRNA coding regions of *Chlamydomonas* through a small number of recombinatorial events between short repeated sequences.

Chapter 5

Short repetitive sequences in green algal mitochondrial genomes: Potential roles in mitochondrial genome evolution

5.1. Introduction

Current data on green algal mitochondrial genomes suggest a rather unexpected dichotomy within the group with respect to genome structure, organization, and sequence affiliations, thus indicating very distinct modes and tempos of evolution in its mitochondrial lineages (discussed in Chapter 1). The green algal mitochondrial genomes investigated to date resemble either those of *Chlamydomonas reinhardtii* (Michaelis et al. 1990) and *Chlamydomonas eugametos* (Denovan-Wright and Lee 1992, Denovan-Wright and Lee 1994, Denovan-Wright, Nedelcu and Lee in press) on the one hand, or that of *Prototheca wickerhamii* (Wolff et al. 1994) on the other, the only three green algal taxa whose mitochondrial genomes are completely sequenced. The *Chlamydomonas*-like mitochondrial genomes are small, either linear- or circular-mapping, have a reduced gene content (no ribosomal protein or 5S rRNA genes and only a few protein-coding and tRNA genes) and fragmented and scrambled rRNA coding regions, whereas the *Prototheca*-like mitochondrial genomes are larger, circular-mapping, have an expanded set of protein-coding genes including ribosomal protein genes, more tRNA genes as well as 5S rRNA and conventional continuous SSU and LSU rRNA coding regions (Nedelcu in press). Whereas the former mitochondrial genome type is most likely confined to one of the five green algal classes (sensu Friedl 1995), i.e., the Chlorophyceae, the latter characterizes at least two of the remaining four classes, namely the Prasinophyceae and Trebouxiophyceae.

There is, however, one feature that seems to be shared by both green algal

mitochondrial genome types as well as their higher plant counterparts, that is, a rather extensive level of gene rearrangement among lineages (discussed in Chapter 7) (Nedelcu and Lee in press). Among land plant mitochondrial genomes, abundant gene rearrangements, and most often complex and dynamic multipartite genomic structures are thought to be consequences of homologous recombination events whose frequency has been shown to be related to the abundance of recombinogenic repeated sequences (Fauron et al. 1995). Large (0.7-120 kb) repeated sequences are considered responsible for the multipartite structure of the genome, whereas smaller repeats recombine in response to stress and can generate mutations, both deletions and duplications, as well as gene rearrangements (see Fauron et al. 1995 for a review).

Fungal, and more recently animal mitochondrial genomes are also thought to have been undergone genomic rearrangements mediated by short repeated sequences that acted as hot spots for recombination (e.g., Jamiet-Vierny et al. 1997 and references therein, Hunt and Hyman 1997). Moreover, the susceptibility to gene rearrangement of chloroplast genomes among *Chlamydomonas* taxa was shown to be correlated with the abundance of short repeated sequences in their intergenic spacers (Boudreau and Turmel 1996). It becomes more obvious, therefore, from both circumstantial and experimental data, that the short repetitive sequences have played an important role in land plant and fungal mitochondrial as well as green algal chloroplast genome evolution.

The factors and mechanisms responsible for the two apparent contrasting evolutionary patterns among the known green algal mitochondrial genomes are not fully understood, although a few suggestions have been made. Recombination via short

repeated sequences has been invoked to explain the two most extreme differences in mitochondrial genom organization among green algae: (i) the reduced gene content of *Chlamydomonas*-like mitochondria genomes relative to *Prototheca*-like and land plant counterparts has been suggested to be a consequence of recombination events between short direct repeated sequences resulting in the excision of the coding regions located between the repeats (Nedelcu 1997, Nedelcu and Lee in press), and (ii) the presence of fragmented and scrambled rRNA coding regions in *Chlamydomonas*-like mitochondrial genomes has been proposed to be the result of transposition (Boer and Gray 1991) or recombination events (Denovan-Wright and Lee 1994, Nedelcu 1997) mediated by short inverted repeats (discussed in Chapter 4).

The questions to be then addressed are these: (i) is there any correlation between the abundance, base composition and distribution of the short repetitive sequences in green algal mitochondrial genomes and the apparent divergent evolutionary patterns in the mitochondrial lineage within the group? and (ii) what are the potential involvements of the short repetitive sequences in the evolution of green algal mitochondrial genomes? In this work I (i) characterize the short repeated sequences in the *Chlamydomonas*-like mtDNA sequences available, (ii) analyze the genomic distribution of these repeated sequences, (iii) discuss their potential roles in the evolution of mitochondrial genomes within the chlorophycean green algal group, and (iv) propose hypothetical evolutionary scenarios to explain the differences in mitochondrial genome organization among green algae.

5.2. Materials and Methods

The green algal taxa, their phylogenetic affiliation as well as the GenBank accession numbers of the complete or partial mtDNA sequences that have been investigated in this study are summarized in Table 7. The DNA sequences were analyzed using GeneRunner version 3 and Clustal V (Higgins et al. 1992).

5.3. Results

5.3.1. Short repetitive sequences (SRSs) within *Chlamydomonas* mitochondrial DNA sequences

5.3.1.1. Short direct repeated sequences (SDRSs)

I have analyzed the complete mtDNA sequence of *C. eugametos* and identified more than 80 repetitive elements that range in size from 6 to 17 bp, and are dispersed throughout the intergenic regions as well as within several introns. Considering their base composition, two classes can be defined: GC-rich and AT-rich SDRSs (Table 8).

The GC-rich direct repeated sequences (9 to 14 bp) can be grouped into four families (Table 8). Each family contains two to six closely related members (one to three base differences) and a total number of copies per family ranging from two to 30. The general consensus sequence, **CGAGTCG**, derived from the consensus sequences of

Table 7. Green algal taxa, their phylogenetic affiliation, and the GenBank accession numbers of the mtDNA sequences analyzed in this study.

Class (sensu Mattox and Stewart 1984)	Taxon	Accession number
Chlorophyceae	*Chlamydomonas eugametos*	AF008237
	Chlamydomonas reinhardtii	U03843
	Chlamydomonas moewusii	unpublished[a]
	Chlorogonium elongatum	Y07814[a]
	Polytomella spp.	U31972[a]
	Scenedesmus obliquus	X17375[a]
Pleurastrophyceae	*Prototheca wickerhamii*	U02970
	Platymonas (*Tetraselmis*) *subcordiformis*	Z47795-Z47797[a]

[a] partial mtDNA sequence

Table 8. Classes, families, and members of short direct repeated sequences in the *C. eugametos* mtDNA.

	Classes						
	GC-rich				AT-rich		
	Families						
Members	1	2	3	4	5	6	7
a	CGAGTCGCATG(5)	CGACTCGAC(18)	CGCGTCGCA(7)	GTTTAGCGGGGCGT(2)	CTTTGT(5)	TGCAATTGT(5)	TATAGTATAGGATCTAG(2)
b	CGAGTCGCAaG(2)	CGACTCGcC(1)	CGCGTCGCg(2)		CTaTGT(1)	TGCAATTGg(1)	cATAGTATAGGATCTgG(1)
c	CGAGTCGCgTG(2)	CGACgCGAC(8)	CGCGaCGCg(2)		CTTcGg(1)		cATAGTATAGGATCTAG(1)
d	CGAGTCGCtTG(7)	CGACgCGcC(2)	CGCGTtGCg(1)				TATAGTcTAGtcgCTAG(1)
e	CGcaTCGCATG(1)	CGACTgcAC(1)	CGCGTCGCc(1)				TATAcTATActcgTCTAG(1)
f	CGAGgCGtATa(1)		tcCGTCGCA(1)				

Number of copies per family member are indicated in parentheses.

Nucleotide indicated in lower case denote differences relative to the consensus sequence of the family (represented by the first sequence in each column).

families **1**, **2**, and **3**, i.e., CGAGTCGCATC, CGACTCGAC, and CGCGTCGC, respectively, is identical to the first 7 bp in the consensus sequence of family **1**, suggesting that families **2** and **3** have evolved from family **1**. It is interesting to note that the consensus sequence is a palindrome in itself: i.e., CGA/G/TCG. Moreover, although the consensus sequences of family **2** and **3** differ from that of family **1** at three and one positions, respectively, the changes (in lower case) do not affect the formation of palindromes: i.e., CGA/c/TCGac and CGcG/T/CGC.

Copies of the same member or closely related members of the same family can be found tandemly repeated, e.g., **2a-2a-2a** or **2a-2c**, **2a-2a-2c**, **3a-3c** (Figure 15). In most instances, members of different families are clustered together with no additional sequences between them (e.g., **1d-2a-2a-2a** or **1b-2a-2a-2c**). In only a few cases individual copies of a repetitive sequence are present apart from a cluster. Another interesting feature is represented by the fact that several combinations of repetitive sequences from different families (e.g, **1d** and **2a**, **1a** and **2a**, **3a** and **3c**) result in palindromic structures (Figure 15); for example, the repetitive sequences **1a** and **2a** (in capital and lower case, respectively, in the sequence below) form the palindromic element CGAGTCGCA/TGcgactcg. In addition, it is noteworthy that most of the repeat clusters contain a core comprised of at least two adjacent members of family **1** and **2** (Figure 15) such that a larger repeated unit (i.e., more than 20 bp) is formed.

AT-rich repetitive sequences have also been found in the *C. eugametos* mtDNA. These sequences are however less abundant than the GC-rich counterparts and have been grouped into three families containing two to five members and a total copy number per family of six or seven (Table 8). The lack of sequence similarity among the three families of AT-rich repetitive sequences suggests independent origins for these families.

```
nad2/cob                                                  1d-    2a-    2a-    2a
                                                          ->     <-
intron 2 L6 (orf)                                         1d-    2a-    2c-(4bp)-1f
                                                          ->     <-
L2/nad6        5a-    1a-    5a-    1a-    5a-            1a-    2a-    2a-    2a-(20bp)-
                                                          ->     <-
                                                                               3a-3c
                                                                                     -> <-
intron 1 L6    1d-    2b-(6bp)-     1e-    5a-            1b-    2a-    2a-    2c
                                                          ->     <-
intron nad1    3d-    (22bp)-       1a-    5c-            1a-    2a-    2a
                                                          ->     <-
intron L5                                                 1a-    2a-(43bp)-  3f-(3bp)-2c-
                                                          ->     <-
                      -(112bp)-     3a-    6a-            1d-    2a-    2d
                                                          ->     <-
intron S2 (orf)                     3a-    6a-            1d-    2a-    2c-(259bp)-
                                    -->                   ->     <-     <--
                                                                               3b-(2bp)-2c
                                                                               --->  <---
L1/M1                               3a-    6a-            1d-    2a-    2c-(278bp)-
                                                          ->     <-
                                                                               4a(58bp)-4a
                                                                                  -><-
L5/S3                        5a-    3a-    6a-            1d-    2a-    2c
                                                          ->     <-
S3                           5a-    3a-    6a-            (3bp)-        2c
                             ->     ->                                  <-
cob/L4                       5b-    3a-    6b-            (3bp)-        2c
                             ------------------------------><-------------
M1/M2                        DR-    (231bp)-              1c-    2a-(62bp)-  3b-(2bp)-2d
                                                          ->     <-          -->     <--
nad1/nad5      3f-(197bp)-DR-       (39bp)-               1c-    2e
                                                          ->     <-
S2/nad4               7e-    7c-    7d-    7b-            7a-    7a-(13bp)-  3c
                      ->     <-
```

Figure 15. The position of the short direct and inverted repeated sequences in the *C. eugametos* mtDNA; opposite arrows designate palindromic sequences; gene abbreviations are as in Figure 20A.

I have also analyzed a partial mtDNA sequence from a taxon that is interfertile with *C. eugametos*, namely *C. moewusii*, and found both GC- and AT-rich sequences identical to the repetitive sequences of *C. eugametos* family **1**, **2** and **5**. However, whereas *C. eugametos* contains in the L6-LSU rRNA/*nad6* intergenic spacer three tandemly repeated **5a-1a** units adjacent to three tandemly repeated **2a** sequences, *C. moewusii* revealed at the same location a number of four **5a-1a** units followed by seven **2a** repetitive sequences, with no spacers between them (Fig. 16).

5.3.1.2. Short inverted repeat sequences (SIRSs)

In the *C. eugametos* complete mtDNA sequence I have identified 20 SIRSs dispersed throughout the intergenic spacers as well as within several introns (Fig. 17A). Seventeen of the inverted repeats are the result of combinations of SDRSs from different GC-rich families (Fig. 15) and belong to one of the two closely related consensus sequences: **GTCGAGTCGC**/--/GCGACTCGAC or **GTCGcGTCGC**/--/GCGACgCGAC. The single nucleotide difference (in lower case in the sequences above) between the two sequences is compensated in the inverted copy such that the formation of a palindromic element is not impeded (Fig. 17B). Interestingly, each copy of the inverted repeat in the first consensus sequence contains a palindrome in itself: i.e., CGA/G/TCG, whereas each copy of the inverted repeat in the second sequence is in fact a two-copy direct repeat: i.e., GTCGC|GTCGC.

The **GTCGC** conserved sequence motif (in bold in the sequences above) present

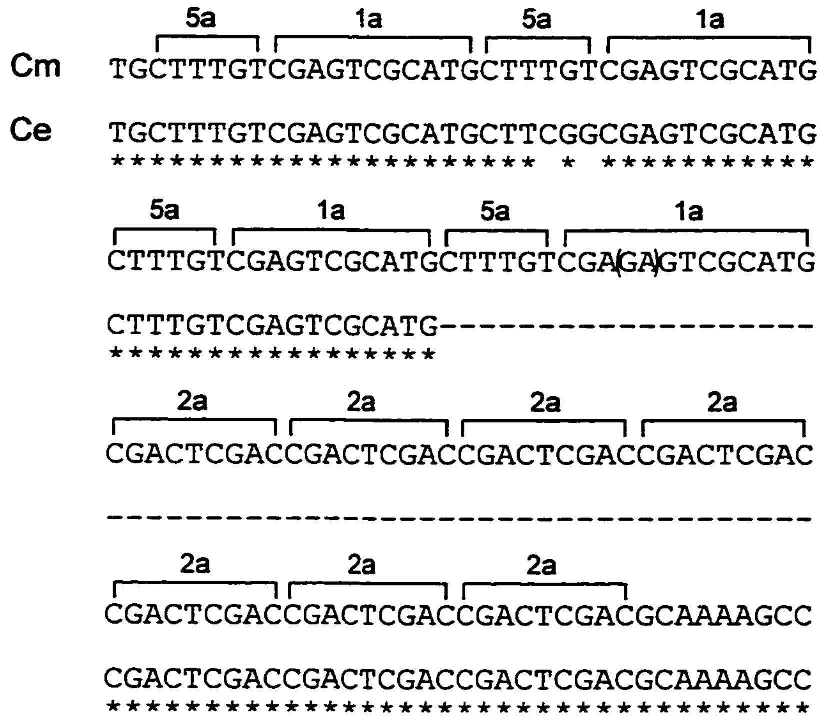


Figure 16. Alignment of the L_6/*nad6* intergenic spacer of the *C. moewusii* (upper line) and *C. eugametos* (lower line) showing the difference in the number of copies of short direct repetitive sequences (as defined in Table 8) between the two interfertile taxa.

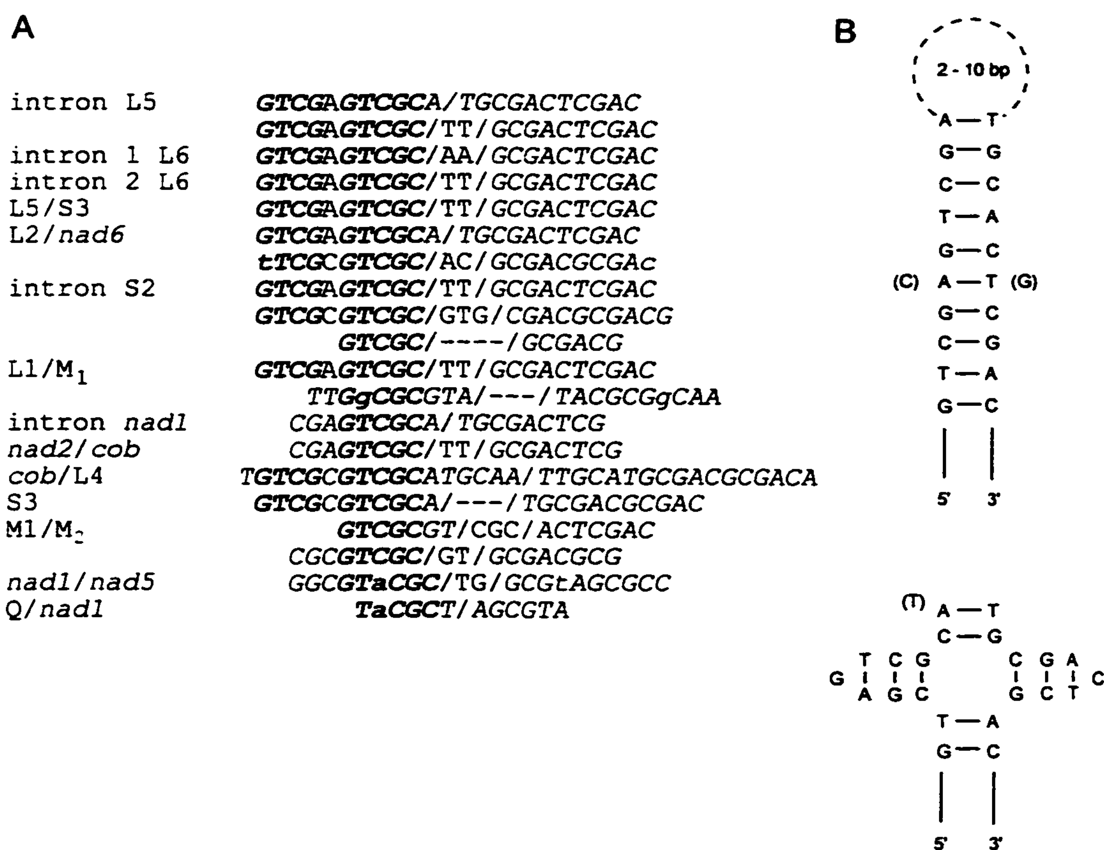


Figure 17. A. *C. eugametos* mtDNA short inverted repeat sequences. Italicized characters denote the inverted repeat sequences and nonitalicized letters or dashes stand for the sequences between the inverted repeats; gene abbreviations are as in Figure 20A. Two highly conserved sequence motifs are in bold; substituted nucleotides relative to the conserved sequence motifs are in lower case. B. Potential secondary structures of the *C. eugametos* mtDNA SIRSs: stem-loop and cruciform structures.

in most of the inverted repeats in the mitochondrial genome of *C. eugametos* is similar to the **CTCGG** motif present in the short inverted repeat sequences previously described in the mitochondrial genome of *C. reinhardtii* (Boer and Gray 1991). Moreover, I have found that some of the SRSs described by Boer and Gray (1991) contain sequence motifs of the *C. eugametos* SDRSs family **1**, **2** and **3** (Fig. 18). In addition, I analyzed the DNA sequence of the long terminal inverted repeats of *C. reinhardtii* mtDNA and identified additional copies of the sequence motifs of the *C. eugametos* mtDNA SDRS family **1** and **2** (Fig. 18).

5.3.2. Short repetitive sequences within mtDNA sequences from other chlorophycean green algae

The only other chlorophycean mtDNA sequences available in GenBank are two partial sequences from the chlamydomonadalean algae *Polytomella* spp. and *Chlorogonium elongatum*, and one from the chlorococcalean taxon, *Scenedesmus obliquus* (Table 7). I analyzed the short mtDNA sequence that contains *cox1* and its flanking regions from *Polytomella* spp. and identified sequence motifs that correspond to both SDRS family **1** and **2** as well as the SIRSs of *C. eugametos* mtDNA (Fig. 19A). In the available partial *C. elongatum* mtDNA sequence, the spacer between *nad5* and *cob* contains eight tandemly repeated copies of the 13-bp GC-rich sequence, CCTTTCGGGCTCG, that was not found in either *C. reinhardtii* or *C. eugametos* mtDNA. In addition, I searched for SRSs within a partial mtDNA sequence coding for two rRNA-

```
5' - nad4      TGCTCGT/---/CGACGAC
                           Ceu 2
3' - nad2        CTCGG/---/CCGAC
W/L6             CTCGG/---/CCGAG
               GTACTCGG/---/CCGAGTAC
                 Ceu 2      Ceu 1
L5/Q          GGGACTCGG/---/CCGAGTACCC
                 Ceu 2      Ceu 1
Q/L7             CTCGG/---/CCGAGCCGCGT
                                 Ceu 3
L3a              CTCGG/---/CCGAG
S3/M           GTACTCGG/---/CCGAGTAG
                 Ceu 2      Ceu 1
                 CTCGG/---/CCGAGT
                            Ceu 1
                 CTCGG/---/CCGAG
L2b/TIR(R)       ATCGG/---/CCGAT
               CGACTCGC/---/CGAGTCG
                 Ceu 2      Ceu 1
TIR(L)         CGACTCGC
                 Ceu 2
```

Figure 18. *C. reinhardtii* short inverted repeated sequences. The conserved sequence motifs as described by Boer and Gray (1991) are in bold; the sequence motifs present in the SDRS class **1**, **2** and **3** of *C. eugametos* (*Ceu*) mtDNA as described in Table 8 are underlined; TIR(L) and TIR(R) are the long terminal inverted repeats of *C. reinhardtii* mtDNA, left and right, respectively; gene abbreviations are as in Figure 20A.

and a tRNA-coding region of *S. obliquus* and found four closely related AT-rich sequences (consensus sequence: TTTTATAGAAGT) tandemly repeated in the rRNA/tRNA intergenic spacer, and several short GC-rich inverted repeats in the tRNA/rRNA spacer (Fig. 19B). However, more mtDNA sequences from these taxa have to become available before one can asses the frequency and distribution of the short repetitive sequences in these mitochondrial genomes.

5.3.3. Genomic distribution of short repetitive sequences in green algal mitochondrial genomes

Comparing the distribution of SRSs between the mitochondrial genomes of *C. reinhardtii* and *C. eugametos*, several differences can be noted (Fig. 20A). Of the 19 intergenic spacers of *C. eugametos* mtDNA, eight contain complex sets of short direct and inverted repeat sequences; in addition, five out of the nine intervening sequences contain SRSs identical to those dispersed throughout the intergenic regions (Figures 15 and 20A). On the other hand, of the 22 intergenic spacers in the mitochondrial genome of *C. reinhardtii*, only four are populated by SRSs (Fig. 20A). The intergenic regions of the *C. eugametos* mitochondrial genome contain eleven SRSs, the intronic sequences host eight SIRSs and only one was found in an exonic sequence, namely an SSU rRNA coding region (Fig. 17A). On the other hand, since there are no introns in the *C. reinhardtii* mitochondrial genome (Michaelis et al. 1990), most of the SRSs are located in intergenic regions; however, two short inverted repeats have been found in the untranslated regions

A

LSU rRNA/*cox1* - CCGAGT - *26bp* - ACGACG - 45bp - *GTCGA* -
Ceu **1** *Ceu* **2** *Ceu* SIRS.

cox1/nad4 - *GTCGC* -
Ceu SIRS

B
LSU rRNA-5'/tRNA^{Ala}

5' -
TTTTATACATTTTATTTTATAAGAAGTTTTTTTATATAATTTTTTTTTATAGAAGTAAT-
3'

d **b** **c** **a**

TTTTATA-GAAACT-*Pwi* Dβ-motif
TTTTATA-GAAGT-**a**
TTTTATAAGAAGT-**b**
TTTTATA-TAATT-**c**
TTTTATA-CATTT-**d**

tRNA^{Ala}/LSU rRNA-3'

1 1' 2 2'
<---- ----> <----- ----->
5'-AGCAG GTTTATTTTAA-*36bp*-CTGCT-*5bp*-AAAGGCAGCCTTTGCCT-3'
<---- ---->
3 3'

-GTTTATTTTAA- -AAAGGCA-
Pwi Aγ-motif *Pwi* Dγ-motif

Figure 19. A. Short repeated sequences within a *Polytomella* spp. mtDNA sequence containing *cox1* and its flanking intergenic spacers. *C. eugametos* (*Ceu*) sequence motifs within the SDRS class **1** and **2,** as well as SIRSs, as described in Table 8 and Figure 17, are underlined. B. Short repeated sequences within an *S. obliquus* mtDNA sequence containing rRNA and tRNA coding regions. *Pwi*=*Prototheca wickerhamii*; Dβ, Aγ, Dγ are sequence motifs in the *P. wickerhamii* mtDNA repetitive sequences as defined by Wolff et al. (1994); the sequences resembling the *P. wickerhamii* sequence motifs are underlined; arrows designate short inverted repeats: 1/1', 2/2', 3/3'.

Figure 20. A. Linear map and linearized map of the linear- and circular-mapping mtDNA of *C. reinhardtii* and *C. eugametos*, respectively, aligned at the *nad5* position (vertical dashed line). S1-S4 and L1-L8 are SSU rRNA and LSU rRNA coding modules, respectively; cob and cox1 are coding regions for cytochrome b and subunit 1 of cytochrome oxidase, respectively; nad1, 2, 4, 5, and 6 are the genes coding for the subunit 1, 2, 4, 5, and 6 of the NADH dehydrogenase; M_1, M_2, Q, W are coding regions for $tRNA^{Met-1}$, $tRNA^{Met-2}$, $tRNA^{Gln}$, and $tRNA^{Trp}$, respectively; solid, cross-hatched and open boxes denote coding regions, introns and intergenic spacers, respectively; DR and TIR denote large direct and terminal inverted repeats, respectively; thick horizontal arrows indicate the direction of transcription; thin vertical arrows indicate the position of the short repetitive sequences; flags indicate the position and orientation of small repeated sequences in the mtDNA of *C. reinhardtii*. B. Gene order of *nad5*, *nad4*, and *nad2* in *C. eugametos* (Ce), *C. reinhardtii* (Cr), *P. wickerhamii* (Pw), *P. subcordiformis* (Ps), and *M. polymorpha* (Mp); the arrow indicates the direction of transcription.

A

1 kb

C. reinhardtii
15.8 Kb

C. eugametos
22.9 Kb

B

Ce	nad2 - nad4 ( ) nad5
Cr	nad4 - nad5
Pw	nad4 - nad5
Ps	nad4 nad5
Mp	nad2 - nad4 nad5

Figure 20

of two protein-coding genes and one is part of an LSU rRNA coding module (Boer and Gray 1991).

It is interesting that the *C. eugametos* SRSs are present in all four of the rRNA intronic sequences but in only one of the five introns found in protein-coding genes, namely, in the only intron that does not contain an *orf* (Denovan-Wright, Nedelcu and Lee in press). Moreover, in the two of the four rRNA introns that contain short *orfs* the SRSs are precisely located within the *orf* sequence; none of these *orfs*, however, code for potential endonuclease/maturase proteins, whereas the deduced amino acid sequences of the intronic *orfs* found in the *C. eugametos* protein-coding genes show similarity to such proteins. Within the two rRNA introns lacking *orfs*, the SRSs are located in regions of secondary structure (data not shown).

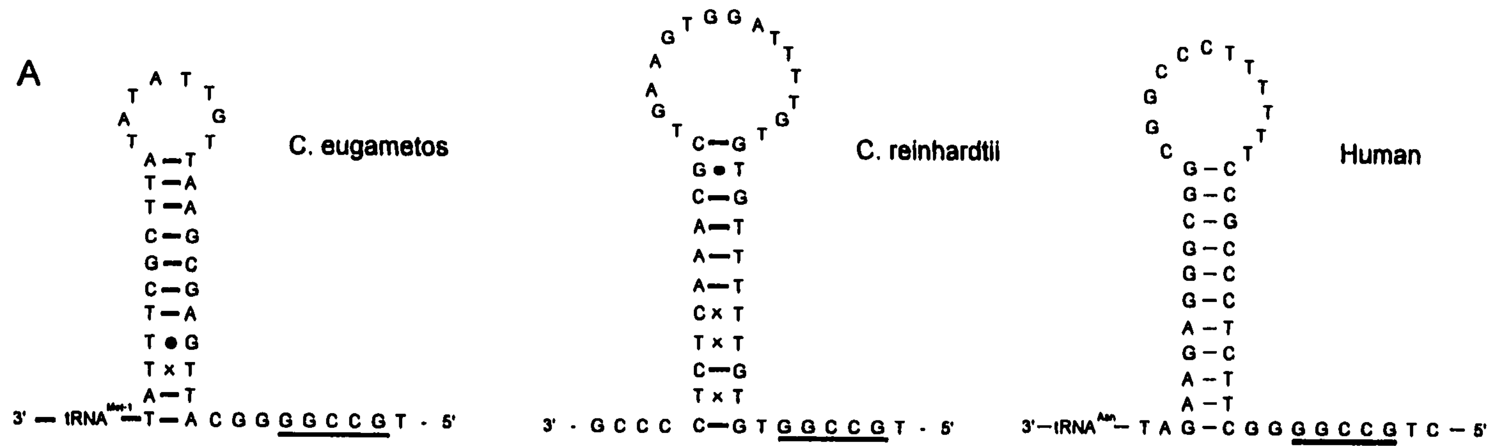
5.3.4. GC-rich sequences with similarity to origins of replication and other recombination-related recognition sites

I have searched the complete mtDNA sequences of *C. eugametos*, *C. reinhardtii* and *P. wickerhamii* for the conserved sequence 3'-GGCCG-5' (in the heavy strand) that has been found to be part of the origin of light strand replication (O_L) in most vertebrate mtDNA (Macey et al. 1997) and shown to be required for *in vitro* replication of human mtDNA (Hixson et al. 1986). Of the one, three and eleven 3'-GGCCG sequences found in *P wickerhamii*, *C. eugametos* and *C. reinhardtii* mtDNA, respectively, none was found in the intergenic spacers of *P. wickerhamii* mtDNA, but two are located in intergenic

spacers in each of the two *Chlamydomonas* taxa. Surprisingly, the two O_L-like sequences in *C. eugametos* mtDNA are flanking (nucleotide 11404 and 11953) one of the two copies of a large direct repeat (DR), and one of the sequences is located in the close proximity of the $tRNA^{Met-1}$ coding region. It is noteworthy that although O_L-like sequences have not been identified through the computer search in the flanking regions of the second copy of the large direct repeat in *C. eugametos* mtDNA, similar sequences with one or two mismatches are present at similar positions relative to the DR. Furthermore, the two O_L-like sequences in the *C. reinhardtii* mtDNA (nucleotide 133 and 15623) are each situated in one of the two long terminal inverted repeats, about 130 bp upstream from the 3'-end of the linear DNA molecule. The sequences upstream of these conserved sequences I have found in the two *Chlamydomonas* mtDNAs form imperfect stem-loop structures that are similar to the stem-loop structures associated with the O_L in vertebrate mtDNA; moreover, they are located in regions of potential secondary structure represented by the $tRNA^{Met-1}$ in *C. eugametos* mtDNA, or a complex set of palindromes in *C. reinhardtii* counterpart (Fig. 21).

Furthermore, I have searched for the sequence 5'-GGAGGGG-3', which is complementary to the sequence 5'-CCCCUCC-3' recognized by the mouse mitochondrial RNA-processing endoribonuclease (RNAase MRP) that cleaves the RNA primer at the origin of heavy-strand replication (Chang and Clayton 1989), and identified a unique sequence flanking the second large repeat in *C. eugametos* mtDNA (nucleotide 18806 on the light strand) and two copies of the same sequence in the two long terminal inverted repeats (nucleotide 239 and 15514) of *C. reinhardtii* mtDNA, at about 100 bp upstream

Figure 21. A. Origin of light strand replication-like structures (the sequence is of the heavy strand) in *C. reinhardtii* and *C. eugametos* mtDNA, and O_L in human mtDNA (as described by Wong and Clayton 1985); the underlined GGCCG is a sequence conserved in vertebrate mitochondrial O_L which is required for the *in vitro* light-strand synthesis in human mtDNA. B. Potential foldings of the 5'-region flanking the O_L-like structure in the *C. reinhardtii* mtDNA (the sequence is of the light strand).



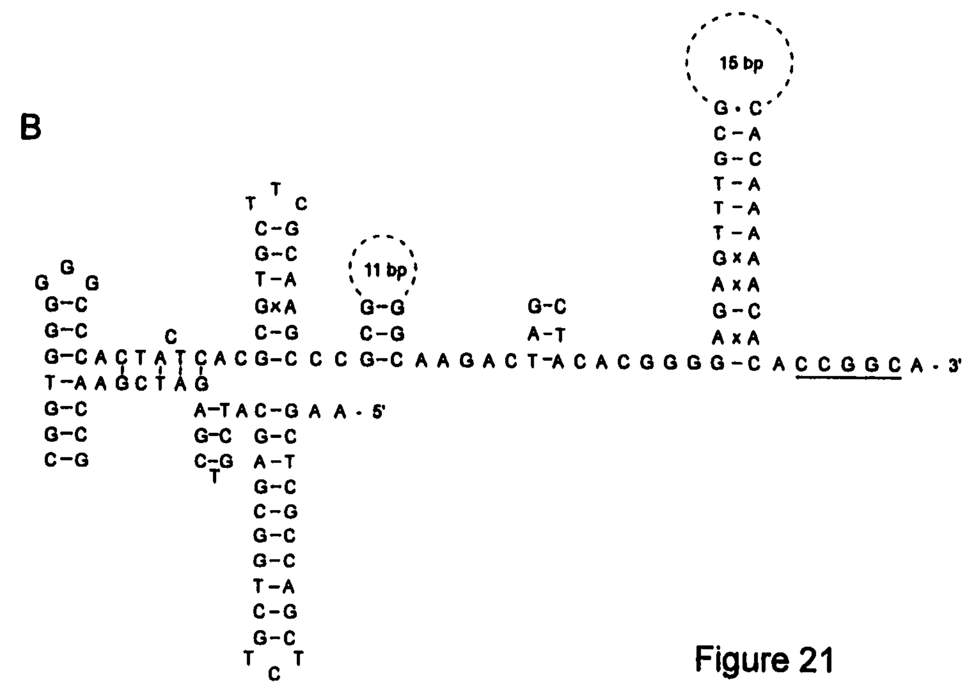


Figure 21

of the O_L-like sequences. It is intriguing that I found the complement of this sequence on the heavy strand of the reverse-transcriptase-like coding region of *C. reinhardtii*, a region in which I have also identified an O_L-like sequence.

Finally, I have searched the mtDNAs of *C. eugametos*, *C. reinhardtii* and *P. wickerhamii* for conserved sequences associated with recombinational processes in other biological systems. A sequence similar to the bacterial *Chi* site sequence, 5'-GCTGGTGG-3', has been found in the first exon of *P. wickerhamii cox1* and a complementary sequence has been identified in the reverse transcriptase-like coding region of *C. reinhardtii*, but none in the *C. eugametos* mtDNA. Furthermore, I found a recognition site of the *Drosophila* topoisomerase topo-II in the vicinity of the 5'-end of the L1-LSU rRNA coding region of *C. eugametos*.

5.4. Discussion

5.4.1. Short repetitive sequences within green algal mitochondrial genomes

My analysis of SRSs in green algal mitochondrial genomes indicates many similarities in various respects with their fungal counterparts. (i) Short repeated sequences, both GC- and AT-rich are very abundant in yeast mtDNA (e.g., Hüttenhofer, et al. 1988, Ragnini and Fukuhara 1988, Zinn et al. 1988, Weiller et al. 1989) but mostly GC-rich sequences have been found in other fungal mtDNAs (Yin et al. 1981, Koll et al. 1996). It is interesting that mostly GC-rich sequences have been found in

Chlamydomonas mtDNAs (Boer and Gray 1991, this study), both AT- and GC-rich sequences have been identified in a partial mtDNA sequence of *S. obliquus* analyzed in this study, but only AT-rich repetitive sequences have been previously described in the *P. wickerhamii* mtDNA (Wolff et al. 1994). It is noteworthy that two of the AT-rich sequences in *C. eugametos* mtDNA seem to resemble sequence motifs found among the *P. wickerhamii* AT-rich repetitive elements: the *C. eugametos* **5a** -CTTTGT resembles a region of the *P. wickerhamii* Aα/β motif, and members of family **7** contain the TATAGTATA sequence which is part of the Dβ repetitive motif in *P. wickerhamii* mtDNA. Moreover, the AT-rich sequences I identified in the intergenic spacers of the *S. obliquus* partial mtDNA sequence (Fig. 19) seem to resemble parts of the AT-rich repetitive motifs described in the intergenic spacers of *P. wickerhamii* mtDNA (Wolff et al. 1994). However, there are no GC-rich repetitive sequences in *P. wickerhamii* mtDNA that resemble those found in the *Chlamydomonas*-like mitochondrial genomes. Such observations suggest that although the AT-rich sequences could have been present in the mitochondrial genome of the most recent common ancestor of green algae, an accumulation of GC-rich repetitive sequences might have taken place in the *Chlamydomonas*-like but not *Prototheca*-like lineage.

(ii) The abundance and complexity of repetitive sequences I have found in the *C. eugametos* mtDNA resembles the situation described in several fungal counterparts. The yeast GC clusters have also been grouped into several families among which some include up to 30 very closely related members (Weiller et al. 1989). It is interesting that in the mtDNA of the fungus *Podospora anserina*, copies of the same member or members

of different SRS families are found tandemly repeated in a direct (with combinations from different classes resulting in palindromic sequences) or inverted orientation (Turker et al. 1987) in a complex manner that is very similar to that of the SRSs in *C. eugametos* mtDNA.

(iii) Although most of the SRSs in green algal as well as yeast mtDNA occur in non-coding regions, short repetitive sequences have been found inserted in the rRNA- and protein-coding regions as well as in open reading frames in both green algal (Boer and Gray 1991, Wolff et al. 1994, this work) (Fig. 15 and 20) and fungal mtDNAs (e.g., Sor and Fukuhara 1982,1983, Hudspeth et al. 1984). The limited distribution of *C. eugametos* SRSs to introns lacking *orf*s or whose *orf*s do not encode potential endonucleases or maturases suggests that they might play a role in the mobility and/or processing of these introns. However, half of the *P. anserina* SRSs, which are called mitochondrial ultra-short elements (MUSEs), that are present in intronic sequences are localized within *orf*s (Koll et al. 1996).

(iv) An additional feature shared by the *Chlamydomonas* (this study) and fungal SRSs (Koll et al. 1996 and the references therein) is their contribution to the mitochondrial genome polymorphism within the genus. Although the mitochondrial genomes of the two interfertile taxa, *C. eugametos* and *C. moewusii*, are colinear (Denovan-Wright and Lee 1992), there are considerable differences within their intergenic spacers in the number of short repeated sequences (Fig. 16), which add to the genomic polymorphism due to differences in number and position of intronic sequences between these mtDNAs (Denovan-Wright and Lee 1992, unpublished data).

5.4.2. Potential roles of short repetitive sequences in the evolution of green algal mitochondrial genomes

Although the insertion of short repeated sequences in mitochondrial genomes was mostly considered associated with deleterious or degenerative phenomena, it is increasingly being suggested that some of these events might actually be neutral, introduce genetic variability, or even confer evolutionary advantages, and thus have a significant impact on the evolution of mitochondrial genomes (Koll et al. 1996). The functions (if any) of the SRSs in green algal mitochondrial genomes have not yet been investigated and their potential roles in the mitochondrial genome evolution have only superficially been suggested (Boer and Gray 1991, Nedelcu 1997). I think that the correlation between the observed differences in abundance, base composition and distribution of short repetitive sequences in green algal mitochondrial genomes on the one hand, and the dichotomy in mitochondrial genome structure and organization among green algae on the other (discussed below), may not be fortuitous. If true, it is conceivable that the accumulation of GC-rich SRSs limited to *Chlamydomonas*-like mitochondrial genomes might have triggered some of the evolutionary events responsible for the divergent evolutionary changes undergone by the *Chlamydomonas*-like and *Prototheca*-like mitochondrial genomes. The similarity in base composition, nucleotide sequence, abundance and mode of organization I have observed between the SRSs present in *Chlamydomonas*-like and fungal mitochondrial genomes might extend to some of the

roles that the SRSs have been shown to have in the latter. In the next five sections I discuss such similarities and suggest additional potential involvements of the SRSs in the evolution of *Chlamydomonas*-like mitochondrial genomes.

5.4.2.1. Fragmentation and scrambling of rRNA coding regions

The SRSs in *Chlamydomonas* mitochondrial genomes have been previously suggested to have contributed to the extensive rearrangement of the rRNA coding regions through either a mechanism analogous to bacterial transposition (Boer and Gray 1991) or recombination (Denovan-Wright and Lee 1994, Nedelcu 1997). Nedelcu (1997) proposed that the accumulation of short inverted repeated elements within the intergenic regions as well as within some variable regions in the rRNA coding regions of the chlorophycean mitochondrial genomes might have triggered a series of recombinational events responsible for the fragmentation and scrambling of the mitochondrial rRNA coding regions to the extent observed in *Chlamydomonas* (discussed in Chapter 4). The distribution of SRSs in the mtDNA sequences investigated in this study indicates that they are mostly associated with rRNA coding regions. All or all but one of the intergenic spacers populated by SRSs in the *C. reinhardtii* and *C. eugametos* mtDNA, respectively, are flanking a rRNA-coding region.

It is noteworthy that although SRSs are present both in mitochondrial genomes with continuous (i.e., *P. wickerhamii*) as well as in those with fragmented and scrambled rRNA coding regions (i.e., *Chlamydomonas*), they are highly AT-rich in the former but

GC-rich in the latter. In addition, our analysis of the available mtDNA sequence of another green algal taxon whose mitochondrial rRNA coding regions seem to be continuous, *Platymonas* (*Tetraselmis*) *subcordiformis* (Kessler and Zetsche 1995), indicated that the intergenic spacers are very AT-rich and do not seem to contain GC-rich repetitive sequences. On the other hand, the SRSs are more GC-rich in the chlamydomonadalean alga *Polytomella* spp. than in the chlorococcalean taxon *S. obliquus* (Fig. 19), which is consistent with the higher degree of fragmentation of the mitochondrial rRNA coding regions in the former taxon relative to the latter (Nedelcu et al. 1996, Nedelcu 1997).

A correlation between the base composition of the short repeated sequences and their recombinogenic properties has been previously observed in other mitochondrial systems. In yeast mtDNA the GC clusters are favoured over AT spacers as both excision sequences (de Zamaroczy et al. 1983) and sites for intramolecular and intermolecular recombinational processes (e.g., Dieckmann and Gandy 1987). As to why the GC clusters would be better promoters for recombinatorial events, few suggestions have been made. (i) Mitochondria might contain factors that recognize the primary sequence of these GC clusters. Such proteins could either be involved in DNA replication, or breaking the DNA strand at these GC sites. Both processes result in the generation of free ends that could invade a double-helix at a site homologous to their ends (Orr-Weaver et al. 1981). Short repeated GC-rich sequences such as the MUSE1 sequences in *P. anserina* and the GC clusters in *S. cerevisiae* mitochondria have been suggested to be the actual binding sites for specific proteins involved in mitochondrial recombination processes (Jamet-Vierny et

al. 1997). If so, it is interesting to note that the MUSE1 sequence, GGCGCAAGCTC, is very similar to the central part of some of the palindromic sequences in the *C. eugametos* mtDNA, i.e., GtCGCAAGCga (the nucleotide differences are in lower case). Moreover, I found in the *C. reinhardtii* mtDNA reverse-transcriptase-like coding region the sequence CCACCAGC which is the complement of the *Chi* site involved in the bacterial chromosomal recombination.

(ii) Mitochondria might contain factors that bind to the stem-loop structures that the GC clusters form. It is very interesting that some of the palindromic sequences in *C. eugametos* mtDNA not only form stem-loop structures but their sequences can be folded in cruciform structures (Fig. 17B). It should be noted that cruciform structures formed by the individual strands within a GC element have been suggested to be recognized by the enzymes involved in the resolution of recombination intermediates such as those known to resolve Holliday junctions (Dieckman and Gandy 1987).

5.4.2.2. Genomic rearrangements

The level of gene rearrangement among both green algal (Nedelcu and Lee in press) and fungal (e.g., Cardazzo et al. 1997) mitochondrial genomes is very high. There is no identical gene cluster between the *C. eugametos* and *C. reinhardtii* mtDNAs (Fig. 20A) and only one, the *nad5-nad4* gene cluster, is common to both *Prototheca*-like and *Chlamydomonas*-like genomes. Given that (i) the *nad5-nad4* gene cluster, which is the only one shared by *P. wickerhamii* and *P. subcordiformis* as well as *C. reinhardtii* and the

liverwort *Marchantia polymorpha*, has been broken in the *C. eugametos* lineage, and (ii) the rRNA coding modules in *C. eugametos* are more interspersed with protein-coding genes than they are in *C. reinhardtii*, it seems likely that the mitochondrial genome of *C. eugametos* has undergone additional rearrangements relative to its *C. reinhardtii* counterpart. Moreover, based on the available partial mtDNA sequence of *C. elongatum*, a taxon that is considered to be more closely related to *C. eugametos* than *C. reinhardtii*, the *nad5-nad4* cluster appears also broken, which is consistent with this gene rearrangement event occurring after the divergence of the two major *Chlamydomonas* evolutionary lineages.

It is thus interesting to note that the higher level of gene rearrangement in the *C. eugametos* relative to the *C. reinhardtii* lineage is positively correlated with a more abundant and complex set of repetitive elements in the former relative to the latter (Fig. 15, 19 and 20A). Furthermore, there are no short repeated sequences flanking either the six adjacent protein-coding genes or the two dispersed ones in the mitochondrial genome of *C. reinhardtii*, whereas short repeated sequences are flanking four out of the seven protein-coding genes in the *C. eugametos* counterpart. It is noteworthy in this connection that the abundance of SRSs throughout the *Chlamydomonas* chloroplast genomes was also considered to be directly correlated with the susceptibility of these genomes to gene rearrangement (Boudreau and Turmel 1996). In addition, in other mitochondrial genomes such as those of *Podospora* (Koll et al. 1996), yeast (Cardazzo et al. 1997) and land plants regenerated from long-term somatic tissue cultures (Hartmann et al. 1994), SRSs have been shown to be associated with gene rearrangements. In Figure 22A I suggest two

hypothetical recombinatorial events that might have contributed to gene rearrangements during the evolution of green algal mitochondrial genomes.

5.4.2.3. Replication

The presence of sequences associated with the origin of mtDNA replication at the sites of gene rearrangement in yeast and vertebrate mitochondrial genomes suggested an involvement of these structures in the reorganization and thus the evolution of these genomes (Cardazzo et al. 1997, Macey et al. 1997). How the mitochondrial genome of *Chlamydomonas* replicates is not known. Both a recombination- and a reverse transcriptase-mediated model have been proposed (Vahrenholz et al. 1993) for the replication of the termini of the linear mtDNA of *C. reinhardtii*, but no suggestions have been made regarding the replication mechanisms of circular-mapping mtDNAs in other green algae such as *C. eugametos*. Rather surprisingly, I have found in both *C. eugametos* and *C. reinhardtii* mtDNA the conserved 3'-GGCCG-5' heavy strand sequence that has been found to be required for the *in vitro* replication of the light strand mtDNA in humans (Hixson et al. 1986); moreover, I have been able to fold the downstream sequences in O_L-like secondary structures similar to those described in vertebrate mitochondrial genomes (Macey et al. 1997). However, the stem element contains several non Watson-Crick basepairings and is more AT-rich than the human counterpart (Fig. 21). Could these structures represent non-functional origins of light strand replication due to accumulations of mutations that altered their secondary structures?

If so, it is tempting to assume that the GC-rich palindromic elements present in the intergenic spacers of *Chlamydomonas*-like mitochondrial genomes might act as surrogate origins of replication in a manner similar to that observed in yeast mtDNA (Goursot et al. 1982, Fangman and Dujon 1984). In addition, it is noteworthy that the heavy strand stem sequence 3'-GCC-5' that represents the initiation site for light-strand replication in mouse (Brennicke and Clayton 1981) is part of the stem in the palindromic elements of *C. reinhardtii* mtDNA. It is known that in vertebrate mtDNA alternative stem-loop structures can initiate replication of the light strand in a region that remains single-stranded for a long period during the asymmetrical replication process (Moritz and Brown 1987, Moritz 1991). In these periods of instability the chance is high that the 5'-end of the nascent light strand slips ahead to an alternative stem-loop structure (slipped-strand mispairing) and replicates a distant region. A model to explain the displacement of the O_L associated with a higher level of gene rearrangement among the vertebrate mitochondrial genomes has been proposed by Macey et al. (1997). It is therefore not unlikely that the extensive gene rearrangement observed between the two *Chlamydomonas* mitochondrial genomes (Fig. 20A) could be a consequence of the mtDNA replication occurring at multiple sites, first competing with and later probably displacing the conventional O_L.

5.4.2.4. Deletions

One of the most distinguishing features of mitochondrial genome organization of

Chlamydomonas relative to other green algal and land plant counterparts is its very reduced gene content. Although mostly speculative at this point, the "removal" of extensive coding regions during the evolution of the chlorophycean mitochondrial genome might have happened through deletion events mediated by short repeated sequences in a manner that appears to be rather common in fungal (e.g., Jamet-Vierny et al .1997 and the references therein) and human mitochondrial (Mita et al. 1990) as well as land plant plastid (e.g., Kawata et al. 1997 and the references therein) genomes.

The presence of single copies of SDRSs in the intergenic spacers of the *C. eugametos* mtDNA (Fig. 15) could represent signatures of excision events via intramolecular reciprocal cross-overs between two copies of a short direct repeated sequence (Fig. 22B). It has recently been shown that subsequent to the excision events associated with senescence in *Podospora anserina*, both the deleted and the senescent mtDNA each contain one copy of the two-copy short direct recombinogenic repeat (Jamet-Vierny et al. 1997). Moreover, the presence of tandemly repeated copies of the same or closely related repeated sequences in the *C. eugametos* mtDNA could be a consequence of short conversion tracts (Fig. 22B) shown to accompany the exchange of DNA strands during recombination events between slightly imperfect repeats (Jamet-Vierny et al. 1997).

It is rather surprising that the SRSs in the mtDNA of *C. eugametos* comply to all the features (discussed below) that have been found to characterize the repeated sequences associated with excision sites in *P. anserina* mtDNA: (i) the most frequent sequence

Figure 22. A. Recombinatorial events via short inverted repeats (solid and open flags) resulting in the interchange of the coding regions between the two sets of repeated sequences (upper row) or inversion of the regions between two copies of the same repeated sequence (lower row). B. Deletion events mediated by recombination via short perfect (upper row) or imperfect direct repeats (middle and lower row); the two latter events are accompanied by a short conversion tract and the duplication of a short direct repeat. C. Linearization of a circular genome consequent to the integration of a linear plasmid with small terminal inverted repeats (flags) that are homologous to a sequence on the circular chromosome; the model is consistent with the maize mitochondrial genome linearization event and suggests a way to explain the presence of the linear mitochondrial genome with long terminal inverted repeats (TIR) flanked by small terminal inverted repeats in *C. reinhardtii*.

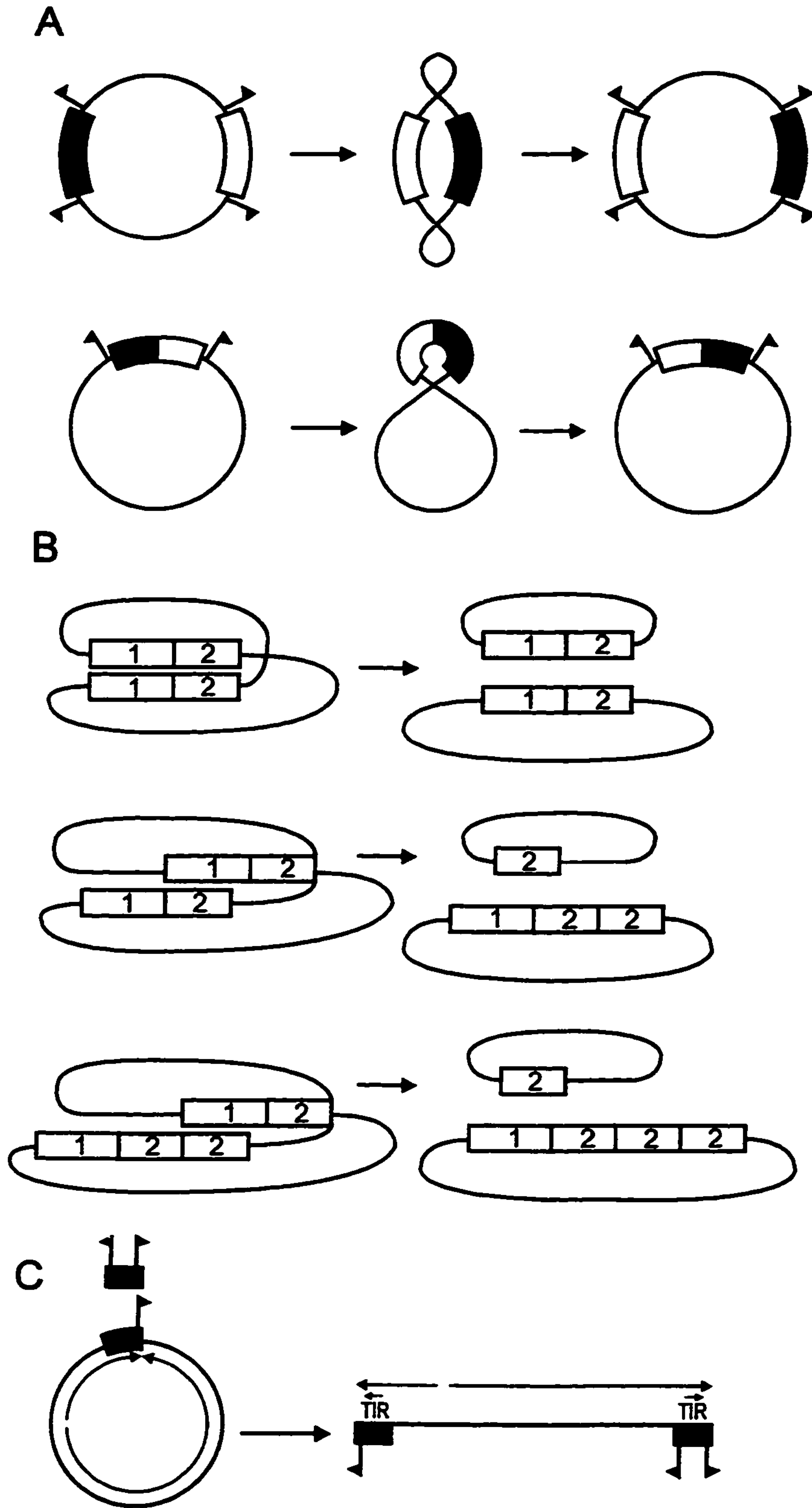


Figure 22

associated with the excision of several types of mitochondrial plasmids during the *P.* anserina senescence is the 11-bp sequence GGCGCAAGCTC (or its complementary sequence), which is similar to the central part of the consensus sequence of the *C. eugametos* mitochondrial SIRSs, GtCGCAAGCga (the nucleotide differences are in lower case) (Fig. 17A); (ii) the direct repeats are present at the excision sites; (iii) the SRSs have the potential to form stem-loop palindromes that include the excision site; and (iv) the inverted copies of short repeated sequences at the excision site allow the juxtaposition of distal (as far away as 18 kb) sites by complementary basepairing interactions (Turker et al. 1987).

Furthermore, it is intriguing that I have found in the *C. eugametos* mtDNA a recognition site of the *Drosophila* topo-II in a region close to the 5'-end of the L1-LSU rRNA coding module, which could represent an excision site given that the very 5'-end of the mitochondrial LSU rRNA gene is missing in *Chlamydomonas* as well as animal mtDNA. Topoisomerases, which have been shown to play a role in illegitimate recombination through cutting and joining the DNA sequences at specific sites, have been implicated in mitochondrial deletions (1.3 to 7.6 kb in size) via an homologous recombination mechanism mediated by flanking short (5-13 bp) repeats (Schon et al. 1989, Mita et al. 1990) in humans; in addition, several *Drosophila* topo-like recognition sequences have been found in the vicinity of various deletion breakpoints (Mita et al. 1990).

Lastly, the presence of the complex set of tandemly repeated sequences that contain palindromic elements and are flanked by AT- or CT-rich sequences in the mtDNA

of *C. eugametos* (Fig. 15) might not be fortuitous given that it has been previously suggested that both the length of the direct repeat and the secondary structure of the DNA region in the vicinity of the excision sites may contribute to the increased frequency of deletion at a particular spot on the mitochondrial genome (Mita et al. 1990). Sequences that could be involved in achieving a secondary structure that facilitates the excision process include: CT- and AT-rich regions that may favour the formation of bent DNA, palindromic sequences that could bring the breakpoints close together, as well as tRNA genes that can form secondary structures (Mita et al. 1990 and the references therein).

5.4.2.5. Genome conformation

Both green algal and fungal mitochondrial genomes appear as either circular- or linear-mapping DNA molecules within their group. Given that the mtDNA is circular-mapping in the primitive-like green flagellate *P. subcordiformis* as well as the more advanced *P. wickerhamii, C. eugametos* and land plants, it is likely that in the most recent common ancestor of green algae and land plants the mitochondrial genome was also a circular-mapping DNA molecule; consequently, the linear-mapping mitochondrial genomes observed in *C. reinhardtii* (Vahrenholz et al. 1993) and its colonial relative, *Pandorina morum* (Moore and Coleman 1989), and most likely several other chlorophycean lineages, are derived traits.

The actual in vivo conformation of the green algal mitochondrial genomes is debatable (Bendich 1993). If the circular-mapping mtDNAs are circular molecules *in*

vivo, as was shown for the liverwort *Marchantia polymorpha* (Oda et al. 1992), their linearization could have happened as a consequence of a recombination event between short repeated sequences present on a small linear molecule and their homologs on a larger circular chromosome (Fig. 22C). Similar events have been described during the linearization of the maize mitochondrial chromosome subsequent to the integration of linear episomes with terminal inverted repeats homologous to internal sequences on the circular chromosome (Schardl et al. 1984). Interestingly enough, in one of the long terminal inverted repeats (TIR) in the linear *C. reinhardtii* mtDNA the outermost 86-bp sequence is repeated in an inverted orientation in the region flanking the opposite end of the TIR (Vahrenholz et al. 1993) (Fig. 20A), arguing thus for a potential previous episomal existence of this TIR.

Large two-copy repeats have also been described in the mitochondrial genomes of *C. eugametos* (Denovan-Wright, Nedelcu and Lee in press) and *Platymonas subcordiformis* (Kessler and Zetsche 1995), but not of *Prototheca wickerhamii* (Wolff et al. 1994). Considering (i) the phylogenetic position of *Platymonas subcordiformis*, which is basal to *P. wickerhamii* and retains ancestral-like features relative to other green algae, and (ii) the presence of large repeats in the mitochondrial genome of both *P. subcordiformis* and the liverwort *M. polymorpha* (Oda et al. 1992), one can hypothesize that large direct repeats had already been present in the most recent common ancestor of the green algae and land plants. It is interesting, however, that while the large repeats are in a direct orientation in *C. eugametos*, they are in an inverted orientation in *P. subcordiformis* and *C. reinhardtii*; moreover, they have internal locations in the first two

taxa but terminal ones in the latter. It is worth mentioning that one of the direct repeats in *C. eugametos* is flanked by two $tRNA^{Met}$ coding regions, one of which might be a pseudogene (Denovan-Wright, Nedelcu and Lee in press). This situation could be related to an inversional event involving the sequence between the tRNA genes as seems to have happened during land plant chloroplast genome evolution (e.g., Hiratsuka et al. 1989). Such a hypothesis would thus imply that the current large two-copy direct repeat in the *C. eugametos* mtDNA was previously an inverted repeat, which is consistent with the presence of a two-copy inverted repeat in the more basal green algal lineage *P. subcordiformis*.

It is thought that two different processes occur simultaneously in plant mitochondrial genomes: a high-frequency reversible recombination involving large repeated sequences and a less reversible and less frequent recombination across smaller repeats (André et al. 1992). It was furthermore suggested that the shorter the repetitive sequences the lower the possibility that two consecutive recombination events happen at the same given site. Therefore, the reversible recombination across the same small repeated sequence to restore the initial arrangement has a low probability, favouring the maintenance of the new rearranged configuration (Hartmann et al. 1994). It is likely that both mechanisms coexisted during the evolution of the green algal mitochondrial genome but recombination involving small and short repeated sequences may have become predominant in the chlorophycean lineage due to accumulation of such sequences throughout the genome. Such a suggestion is also consistent with the lack of any detectable isomeric forms or multipartite structures of the green algal mitochondrial

genomes in which the lengths of the largest repeated sequences are much smaller than in their land plant counterparts.

5.4.3. Hypothetical scenarios

A hypothetical pathway to explain the fragmentation and scrambling of the rRNA coding regions in the chlorophycean green algal group, involving recombinatorial events mediated by short repeated sequences accumulated within the variable regions of the mitochondrial rRNA genes as well as intergenic spacers, has been presented in Chapter 4 (Nedelcu 1997). Figure 23 shows a hypothetical scenario via various recombination events between short and large repeated sequences to account for other evolutionary trends that can be defined among the green algal mitochondrial genomes investigated to date: (i) the evolution from a polycistronic genome organization such as in *P. subcordiformis* to a bicistronic one in *P. wickerhamii* and *C. reinhardtii* and possibly a monocistronic unit in *C. eugametos*, (ii) the breaking of the ancestral *nad5-nad4-nad2* gene cluster in *C. eugametos* and *C. reinhardtii* lineages, (iii) the "removal" of coding regions in the chlorophycean lineage, and (iv) the linearization of the *C. reinhardtii* mitochondrial genome.

However, the origin and accumulation of GC-rich SRSs in the *Chlamydomonas*-like but not *Prototheca*-like green algal mitochondrial lineage, as well as the factors that triggered the evolutionary events the SRSs might have been involved in, remain unknown.

Figure 23. Hypothetical recombination events to account for the evolutionary trends that can be defined among the green algal mitochondrial genomes investigated to date: (i) the evolution from a polycistronic genome organization such as in *P. subcordiformis* to a bicistronic one in *P. wickerhamii* and *C. reinhardtii* and possibly a monocistronic unit in *C. eugametos*, (ii) the breaking of the ancestral *nad5-nad4-nad2* gene cluster in *C. eugametos* and *C. reinhardtii* lineages, (iii) the "removal" of coding regions in the chlorophycean lineage, and (iv) the linearization of the *C. reinhardtii* mitochondrial genome. Thick curved arrows inside or outside the circles indicate the direction of transcription of that region; broken arrows denote interchange of the coding regions pointed by the arrows through recombinatorial events via two sets of two-copy short inverted repeats (see Fig. 22A); arrows pointing outside the circle indicate deletion events via recombinatorial events mediated by short direct repeats (see Fig. 22B); thin long curved arrows outside the circles indicate inversion events as a result of recombinatorial event between a two-copy short inverted repeat (see Fig. 22A); DR, IR, and TIR denote large direct, inverted, and terminal inverted repeats, respectively; flags denote small inverted repeats; open, gray, and cross-hatched blocks denote protein-coding regions, large repeated sequences, and deleted coding regions, respectively.

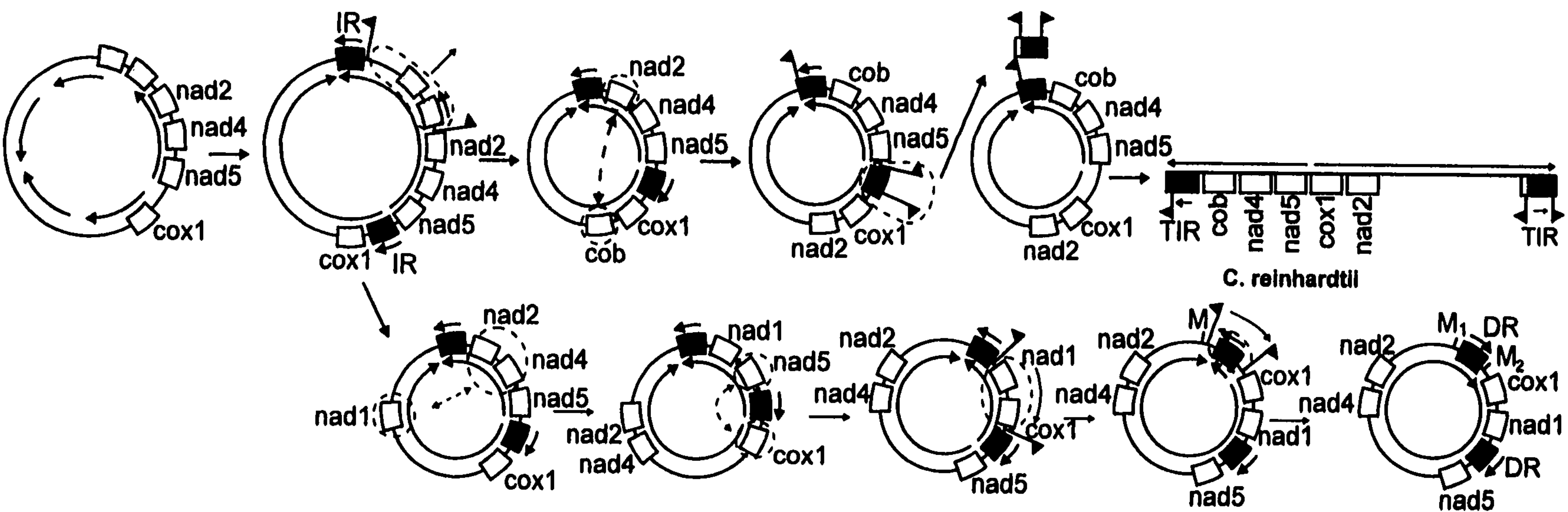


Figure 23

5.5. Conclusions

Current data on green algal mitochondrial genomes suggest an unexpected dichotomy within the group with respect to genome structure, organization, and sequence affiliations, thus indicating very distinct evolutionary changes undergone by its mitochondrial lineages. This chapter suggests that there is a correlation between this dichotomy and the differences in the abundance, base composition, and distribution of short repetitive sequences I have observed among green algal mitochondrial genomes. It is conceivable that the accumulation of GC-rich short repeated sequences in the *Chlamydomonas*- but not *Prototheca*-like mitochondrial genomes might have triggered evolutionary events responsible for the divergent evolutionary changes undergone by the two green algal mitochondrial lineages. The similarity in base composition, nucleotide sequence, abundance, and mode of organization I observed between the short repeated sequences present in *Chlamydomonas*-like and fungal mitochondrial genomes might extend to some of the roles that the short repeated sequences have been shown to have in the latter. Potential involvements I propose for the short repetitive sequences during the evolution of *Chlamydomonas*-like mitochondrial genomes include: fragmentation and scrambling of the ribosomal RNA coding regions, extensive gene rearrangements, coding region deletions, surrogate origins of replication, and chromosomal linearization.

Chapter 6

Introns and intronic *orfs* with potential involvement in the evolution of the chlorophycean mitochondrial genome

6.1. Introduction

It is increasingly suggested that mobile genetic elements such as group I and II introns as well as their intronic *orfs* may play an important role in the plasticity and evolution of mitochondrial genomes (Belcour et al. 1997). Mobile group I introns are considered to be comprised of two functionally and structurally distinct domains that evolved independently: a highly structured core sequence with self-splicing activity, and an open reading frame that encodes a DNA endonuclease (Bell-Pedersen et al. 1990, Lambowitz and Belfort 1993). The mobility of group I introns is dependent upon the production of a site-specific DNA endonuclease that promotes a double strand break repaired by gene conversion using the intron-containing gene as a template. Recently, it has been proposed that the *orfs* themselves are mobile elements that could colonize preexisting group I introns; such an hypothesis is supported by circumstantial data as well as experimental observations (Sellem and Belcour 1997 and the references therein). In addition, group II introns have also been shown to have the ability to home or move to a novel location (transposition), both in the same genome or between different genomes either in the same cell or in different lineages (horizontal transfer); such activity is facilitated by reverse transcriptase-like polypeptides encoded in their intronic *orfs* (Michel and Lang 1985).

Current data on mitochondrial genome organization indicate a rather large variation in intron number and insertion position among the green algal lineages investigated to date (discussed in Chapter 7). The intronic polymorphisms observed

among closely related taxa or interfertile strains suggest recent intron acquisition/loss events within the group; in addition, the mobility of some intronic sequences in interspecific crosses (Colleaux et al. 1990) suggests an ongoing role of these elements in the shaping of green algal mitochondrial genomes.

The limited information on intronic sequences in mitochondrial genes among green algal lineages makes it difficult to assess the presence/absence or type of intervening sequences in the mitochondrial genome of the most recent common ancestor of green algae. On the other hand, the presence of several introns as well as intronic *orfs* with similarity to introns and *orfs* found in non-homologous mitochondrial genes in green algae and fungi raises the possibility that horizontal transfer events occurred between the green algal and fungal ancestors. Given that introns and intronic *orfs* have been shown to be involved in fungal mitochondrial genome plasticity, it is of interest to find out if the green algal counterparts had any impact on the evolution of their mitochondrial genomes.

This chapter thus addresses the following question: Is there any correlation between the presence of particular types of intervening sequence and/or intronic *orfs* on the one hand, and the dichotomy observed in mitochondrial genome organization among green algae? This work (i) reports the presence of a degenerate group II intron in the intronless mitochondrial genome of *C. reinhardtii*; (ii) analyzes an intronic *orf* in the *C. eugametos* mitochondrial *nad5*; and (iii) suggests potential consequences of intron and *orf* acquisition to the evolution of green algal mitochondrial genomes.

6.2. Material and Methods

The mtDNA sequences analyzed were of *C. reinhardtii* (U03843) and *C. eugametos* (AF008237). Searches for amino acid sequence similarity have been done using the BLAST algorithm (Altschul et al. 1990) and nucleotide and amino acid sequence alignments have been performed using CLUSTAL V (Higgins et al. 1992).

6.3. Results

6.3.1. A degenerate group II intron in the intronless mitochondrial genome of *Chlamydomonas reinhardtii*

The presence in the *C. reinhardtii* mitochondrial genome of a reverse transcriptase-like (*rtl*) coding region that is 3'-flanked by an LSU rRNA coding module whose sequence corresponds to the LSU rRNA coding region interrupted by a group II intron in *S. obliquus* mtDNA (see Fig. 11) raised the possibility that *rtl* in *C. reinhardtii* mtDNA is the remnant of an intronic *orf* previously harboured by a group II intron. To address this issue, I have analyzed the regions flanking the *rtl* nucleotide sequence in the *C. reinhardtii* mtDNA and searched for the consensus sequence motifs associated with the 5'- and 3'-splice sites characteristic for group II introns. Surprisingly, 24 nt upstream of the *rtl* initiation codon (Figure 24A), in the region corresponding to the 3'-end of the L_{3b} LSU rRNA fragment, I have found the sequence GUGUG that is similar to the very conserved GUGCG found at the 5'-splice site of most group II introns. Furthermore, the

region between the 23rd nucleotide downstream of the *rtl* termination codon and the 9th nucleotide upstream of the 5'-end of the L_8 LSU rRNA coding region shows sequence similarity to domain V and VI at the 3'-splice site of various group II introns (Figure 24A). There are, however, three short insertions in three of the five highly conserved blocks in the sequence corresponding to domain V in other group II introns; this is not unprecedented since a GC-rich cluster has also been found inserted in the domain IV of the *S. cerevisiae* mitochondrial COI-2 group II intron.

To find out if the sequence similarity observed between the *C. reinhardtii* mtDNA *rtl*/L_8 LSU rRNA intergenic spacer and sequences corresponding to domain V and VI of group II introns is biologically significant, I have modeled the *C. reinhardtii* nucleotide sequence and compared it to the secondary structure of domain V and VI in the *Saccharomyces cerevisiae* mitochondrial COI-2 group II intron (Fig. 24B). It is interesting that the *C. reinhardtii* mtDNA sequence (excluding the three short insertions mentioned above) that showed similarity to the 3'-end of group II introns, can be modelled into a structure strongly resembling that of domain V and VI of *S. cerevisiae* COI-2 intron. Moreover, it is noteworthy that the three short insertions in the *C. reinhardtii* sequence correspond to two loops in domain V of the *S. cerevisiae* COI-2 intron. In addition, although the domain VI in the two structures compared differ in size, the difference is limited to a region that varies in size among the known group II introns, from 2 nt in the *S. obliquus* mitochondrial LSU rRNA intron to 41 nt in a maize chloroplast tRNA intron (Fig. 24A).

Figure 24. *C. reinhardtii* L_{3b}/L_8 intergenic spacer showing A) nucleotide sequence and B) secondary structure similarity to highly conserved regions of group II introns. A. Alignment of the *C. reinhardtii* L_{3b}/L_8 intergenic spacer containing *rtl* and the nucleotide sequence at the 3'- and 5'-splice site of various group II introns; the alignment of the intronic sequences except that of *C. reinhardtii* is based on Kück et al. (1990). The highly conserved blocks of nucleotides are marked by continuous lines; bracketed numbers and vertical arrows pointing to the conserved regions indicate the size and location of three short insertions in the *C. reinhardtii* mtDNA sequence. Stars above the *C. reinhardtii* sequence denote nucleotides that are identical between at least two of the intronic sequences in the alignment. Cr, So, Pa, Sc, Zm, Eg, mt, and pt stand for *Chlamydomonas reinhardtii, Scenedesmus obliquus, Podospora anserina, Saccharomyces cerevisiae, Zea mays, Euglena gracilis,* mitochondria, and plastid, respectively. B. Comparison between the potential secondary structure of the *C. reinhardtii* *rtl*/L_8 LSU rRNA spacer and domain V and VI of the *S. cerevisiae* COI-2 group II intron. Continuous and dashed arrows indicate the position and the number of the nucleotide missing or inserted, respectively, in *C. reinhardtii* relative to the *S. cerevisiae* counterpart sequence; circled and bold nucleotides denote additional and identical nucleotides, respectively, relative to the *S. cerevisiae* counterpart. The boxed region in the *S. cerevisiae* domain VI is missing in *C. reinhardtii*; the location of the missing region in *C. reinhardtii* is marked by two short continuous arrows each indicating 9 nt.

A

group II intron

L3b ← 5'-GUGUG -(24 nt)— *rtl* —(23 nt)- UAG - - - - - - - - - - - - - - - UAU -3'-(9 nt)→ L8

1107 nt

(-10nt-) (-8nt-) (-7nt-)

		exon	(rtl)		
Cr mt	L3b/L8	5'-GUGUG	- 1154 nt -	UAG	GAGGUGUACACUAUUAAAGUCGGCAGUGCGGUUC
So pt	LSUrRNA	5'-GUGCG	- 527 nt -	CUU	GAGCCGUGUGCGGUGAAAGUCGCAUGCACGGUUC
So mt	petD-1	5'-GUGCG	- 3435 nt -	GAG	GAGCCGUAUGAGGUGAAAGUCUCACGUACGGUUC
Pa mt	COI-1(α)	5'-GUGCG	- 2438 nt -	GGA	GAGCCGUAUGCGGUGAAAGUCGCACGUACGGUUC
Sc mt	COI-1	5'-GUGCG	- 2361 nt -	GGA	AAGCCGUAUGAUGGGAAACUAUCACGAACGGUUU
Sc mt	COI-2	5'-GUGCG	- 2416 nt -	GGA	GAGCCGUAUGAUAUGAAAGUAUCACGUACGGUUC
Zm mt	COII	5'-GUGCG	- 727 nt -	GGA	GAGCUUUUUGCGGGGAAACUUGCAAGUCAACUUU
Zm pt	trnA	5'-UUGGG	- 686 nt -	GGA	GAGCACAGUACGAUGAAAGUUGUAAGCUGUGUUC
Cr pt	pslA1-1	5'-GUGCG	- 50 kb -	CUU	GAGCCGUAUGCGAAAAAACUCGCAUGUACGGUUC
Cr pt	pslA1-2	5'-GUGCG	- 90 kb -	UUU	GAGCCGUGUGCAGUGAAAAUUGCAUGCACGGCUC
Eg pt	psbA-4	5'-GUGCG	- 535 nt -	UUU	GAGCCUUGUGAUAUUAAAAUUUCAUGCAAGGUUC

domain V

						exon
Cr mt	L_3b/L_8	ACC	GAGCACCCAAGGUCU	AG	CUACGAUUUGCUAUAU	UAU -3'
So mt	LSUrRNA	UAC	UGGGGGGAAAGCCUG	-AG-	AGGGCCUACCUAUCCA	ACU -3'
So pt	petD-1	GGC	AGGGGUAAUAUGGUA	-19-	UAUCUUUCCCCUAGCC	UCU -3'
Pa mt	COI-1	GGA	GGAGAGUUUUUGGUU	-22-	GACUGGCUGCUUAUCC	UAC -3'
Sc mt	COI-1	GGG	AAAGGCUCUUUAACA	-12-	GGUUAAUUUGCUAUUU	CAU -3'
Sc mt	COI-2	GGA	GAGGGCUCUUUUAUA	-18-	GAUAGGUUUGCUACUC	UAC -3'
Zm mt	COII	GGG	GGGAGGCGG------------------GCGUCGACCC			AAC -3'
Zm pt	trnA	GGG	GGGGAGUUAUUGCCU	-41-	AGGCGGUGGUUUACCC	UGU -3'
Cr pt	pslA1-1	UUU	AGGAGGAUUUAAAAU	-13-	AAAACAAAUCCUACCU	GAC -3'
Cr pt	pslA1-2	UUA	AGGUUUAAAUAAAUU	- 8-	AGAAAAUUUAACUCCU	AAC -3'
Eg pt	psbA-4	UGU	GACGAUAUUUUU	- 6-	GAAAAUUUUAGUUU	AAC -3'

domain VI

Figure 24

B C. reinhardtii *rtl* / L_8 spacer

S. cervisiae COI - 2

Figure 24

6.3.2. An intronic *orf* with potential maturase/endonuclease/recombinase activity in the mitochondrial genome of *Chlamydomonas eugametos*

Out of the nine intervening sequences in the mitochondrial genes of *C. eugametos*, six contain intronic *orf*s (five introns are monoorfic, and one is biorfic) (Denovan-Wright, Nedelcu and Lee in press). A BLAST search for sequence similarity of the deduced amino acid sequence of the *orf* i5 located in the first intron of the *C. eugametos nad5* indicated high levels of similarity to intron-encoded maturases/endonucleases, mostly from fungal mitochondria. Comparisons of the *C. eugametos* i5 amino acid sequence with other intronic *orf*s revealed the presence of the two conserved dodecapeptide motifs characteristic for maturases/endonucleases of the LAGLI-DADG type (Fig. 25).

It is notable that the *C. eugametos* i5 amino acid sequence shows similarity to the intronic *orf* described in the *Saccharomyces cerevisiae* mitochondrial 21S rRNA gene (Fig. 25A) which is known to act as a transposase in the spreading of its host intron within populations of interfertile strains (Jacquier and Dujon 1985). Moreover, the *C. eugametos* i5 also shows similarity to the *orf* present in the single mobile intron of *Chlamydomonas smithii cob* and the second intron in the *Chlorogonium elongatum cob* sequence (Fig. 25B). It should be mentioned, however, that these two *cob* introns have distinct insertion sites that are different from those of the single and the first *cob* intron in *C. eugametos* and *Chlorogonium elongatum* (Denovan-Wright, Nedelcu and Lee in press, Kroymann and Zetsche 1997), respectively; in addition, the former two *cob* introns contain *orf*s encoding peptides of the LAGLI-DADG type, whereas, the *orf*s in the

Figure 25. Alignment of the *C. eugametos nad5* i5 amino acid sequence (CE_NAD) with intronic sequences from A) *Saccharomyces cerevisiae* mitochondrial 21S rRNA ω intron (SC_TRAN), B) *Chlorogonium elongatum* mitochondrial *cob* i2 intron (CE_COBi2), C) *Schizosaccharomyces pombe cox* I1 intron (SP_COX) and *Saccharomyces cerevisiae cob* i4 intron (SC_COB), D) *Mycobacterium leprae recA* intein (MYC_REC), and E) *Mycobacterium leprae recA* intein and *S. cerevisiae* mitochondrial 21S rRNA ω intron; a bold H denote a histidine residue shown to be required for the endonuclease activity of the *S. capensis* bi4 mitochondrial *orf*.

A

```
CE_NAD>   LYDL--NNFLDISGASTFWFFARKPAESRKGPGATLASSGFPHSFLFKQI
SC_TRAN>  MKNIKKNQVMNLGPNS-------KLLKEYKSQLIELNIEQFEAGI-----
          . ..  *. ....  * .      *      *.     *    *   ..

          ____P1_____
          LTGLLLGDGWLERHGKGTRLGVSCKHVYADVANWMQLMFYG---LGYHDK
          --GLILGDAYIRSRDEGKTY---CMQF-----EWKNKAYMDHVCLLYDQW
            **.***. .  .. *        * .       .* .  . .   * *..

          MYQVSPL-ECITRQGKISRYYQVRTFSFASLNKYYNLWYVNNIKIVPLD
          VLSPPHKKERVNHLGNLVITWGAQTFKHQAFNKLANLFIVNNKKTIPNNL
          .   .    *  .  . *..       .**    ..**  **  *** * .* .

                ______P2_____
          IDQYLTPLALAIWLMGDGS------GMRDGGFKISTHSFTKQENEFLVEL
          VENYLTPMSLAYWFMDDGGKWDYNKNSTNKSIVLNTQSFTFEEVEYLVKG
          ...****..** *.*.**.              . ..  ..*.***  .* *.**

          LLNKYDIKASIHRDGDEFNIYIWKQSVPKVKALVLPFFHVRVARHANPVF
          LRNKFQLNCYVKINKNKPIIYI--DSMS-----YLIFYN-----------
          * **....   .   . .    ***   .*..        * *..

          INGAMLRNKYITQCLSKKLSHICKKQITHNIASIKYRLIQRNALK
          -----LIKPYLIPQMMYKLPNT--------ISSETF-------LK
               * . *.    .  **..          *.*   .        **
```

B

```
                         ____P1_____
CE_NAD>       FKQILTGLLLGDGWLERHGKGTRLGVSCKHVYADVANWMQLMFYGLGY
CEL_COBi2>    FVQVAVGLLLSDAHAEVHGNGVRISFQQEKTFADYFSFVYGILERLGY
              * *.  ****.*.  * **.* *..         .**  . .   ..  ***

              HDKM---YQVSPLECITRQGKISRYYQVRTFSFASLNKYYNLWY-VNN
              VTRKGVSDGTGEVSFHTRNQGARTYFRLNTFTFSSLIWLRELFYDANG
                     . .    **.       *... **.*.**      .* *  *

                              ______P2______
              IKIVPLDIDQYLTPLALAIWLMGDGSGMRDGGFKISTHSFTKQENEFL
              VKVIRPELINYLTPISLRHAICGDGSST-DYGTSLSFNSFTYEECVLF
              .*..   .. .****..*    . ****.   * *   .* .*** .*   ..

              VELLLNKYDIKASIHRDG--DEFNIYIWKQSVPKVKALVLP
              TNMLKEKFGIIASVQSAGAPNQYRVYIQAASMNTLRAIVLP
               ..*..* *   *..  *  ... .**   *.   ..*.***
```

Figure 25

C

```
                                                   P1
CE_NAD>   FFARKPAESRKGPGATLASSGHPHSFLFKQILTGLLLGDGWLE-RHGKGTR-LGV-S
SP_COX>   FFNRHPFTRVKSCF----SSSSPSKFSFTQWLVGFTDGDGCFSISKQKIKNGKNKWS
SC_COB>   NIKRSSETTRKMLNN----S-MNKK--FNQWLAGLIDGDGYFGIVSKKYVS------
             *       *         *        *  *  *  *    ***       *

          -CKHV---YADVANWMQLMFY----GLG-------YHDKMYQVSPLECITRQGKI--
          LTKLTQNLY----NYRILYFIKRNLGIGSLYKESSTNTVIYRLRPREHLKKIIDIFD
          -LEITVALE-DE---MALKEIQNKFG-GSIKLRSGVKAIRYRLTNKTGMIKLINAVN
                          *       * *               *

          ------SRYYQVRTF--SF-----ASLNKYYNLWYVNNIKIVPLDIDQYLTPLALA
          QFFLLTKKYWDYYLFKKAFLILEDANLNSFEKNSKLEHIRIEKKSLKQY-SPVNLE
          GNIRNTKR------------------LVQFNKVCILLGIDFI------Y--PIKLT
             P2                      *             *          *   *  *
          -----IWLMG--DGSGMRDGGFKISTHSFTKQ---ENEFLVELL--LNKYDIKASIH
          KYLTKSWLIGFIFAEGSF-YLLQKSVPRIIHGFEITQNYEQPLLAQISEFLFNSQIS
          K--DNSWFMGFFDADGTINYSFKNNHPQLTKT--VTNKY----LQDVQEY--KNILG
                *         *                          *

          ----RDGDEFNIYIWKQSVPKVKALVLPFFHVRVARHANPV-FINGA-MLRNKYIT
          PKIKSKKNSLITNYSLSTSSKERMLFLSSYFENCFKGVKSLEFKIWSRSLR-KNY-
          GNIYFDK-SQNGYYKWSIQSKDMVL---NFINDYIKMNPSRTTKMNKLYLS-KEFY
                                *   *                         *  *

          QCLSKKLSHICKKQITHNIASIKYRLIQRN--ALK
          -----NFEQLRRARDLIRKLKNKYSRGSQHP-KDK
          -----NLKELKAYNKSSDSMQYKAWLNFENKWKNK
                                 *          *
```

D

```
                                          P1
CE_NAD>    GATLASSGFPHSF---LFKQILTGLLLGDGWL-----ERHGKGTRL-GVS
MYC_REC>   GDRVLAVE-PHMLSDQQFQVVL-GSLMGDGNLSPNLCDRNGVRFRLLGYG
           *  . .    ** .     *. .* * * *.*** *      .*.*   ** * .

           CKHVYADVANWMQLMFYGLGYHDKMYQVSPLECITRQGKISRYYQVRTFS
           CKQV--EYLQWKKAL------------MGNIRHTVRENSMGASF----ID
           **.*   .  .*  . .              . . .      *.  . .   .   .

                                                             P2
           FASLNKYYNL----WYVNNIKIVPLDIDQYLTPLALAIWLMGDGS----
           FTPLPELVELQRAVYLGDGKKFLSEEYLKALTPLVLAIWYMDDGSFTVG
           *..*     .*         .   *... .   . **** **** *.***

           -------GMRDGG-FKISTHSFTKQENEFLVELLLNKYDIKASIHR--D
           SKRVQERTAGGSGRIEICVDAMTEGTRVRLRDYLCDTHGLDVRLREVGS
                        .* . *   .. *        *    *    .*    *  .

           GDEFNIYIWKQSVPKVKALVLPFFHVRVARHANPVFINGAMLRNKYITQ
           AGKAVLVFSTAATAKFQSLIAPYVAPSMEYKLLPQFRGRGSVTPQFVEP
                 .  .    .*.   *   *   .            * *          . ..

           C----------LSKKLSHICKK----------------QITHNIASIKY
           TQQLVPARVLDVHVKLSTRSMNRFDIEVEGNHNYFVDGVMVHNSPETTT
                     .   ***                                **

           RLIQRNALK-
           GGK---ALKF
               ***
```

Figure 25

E

```
                                   _____P _____
CE_NAD>     GA------TLASSGFPHSFLFKQILTGLLLGDGWLERH---GKGTRL---GVS
MYC_REC>    GD---RVLAVEPHMLSDQQF--QVVLGSLMGDGNLSPNLCDRNGVRFRLLGYG
SC_TRAN>    KEYKSQLIEELNIEQFEAGI------GLILGDAYIRSR---DEGKTY------
                     .    .           *  ..**. .        *

            CKHVYADVANWMQLMFYG---LGYHDKMYQVSPL-ECITRQGKISRYYQV
            CKQV--EYLQWKK-ALMGNI---------------RHTVRENSMGASF--
            CMQF-----EWKNKAYMDHVCLLYDQWVLSPPHKKERVNHLGNLVITWGA
            * .      .* .    .                     .   .

                                                  _____P2_____
            RTFSFASLNKYYNL----WYVNNIKIVPLD-IDQYLTPLALAIWLMGDGS
            --IDFTPLPELVELQRAVYLGDGKKFLSEEYLKA-LTPLVLAIWYMDDGS
            QTFKHQAFNKLANL----FIVNNKKTIPNNLVENYLTPMSLAYWFMDDGG
                  ..    .*       .  * .. . .   ***. ** * *.**.

            -----------GMRDGG-FKISTHSFTKQENEFLVELLLNKYDIKASIHR
            FTVGSKRVQERTAGGSGRIEICVDAMTEGTRVRLRDYLCDTHGLDVRLRE
            -----KWDYNKNSTNKS-IVLNTQSFTFEEVEYLVKGLRNKFQLNCYVKI
                            . . .  .. *      *   * .   .     .

            --DGDEFNIYIWKQSVPKVKALVLPFFHVRVARHANPVFINGAMLRNKYI
            VGSAGKAVLVFSTAATAKFQSLIAPYVAPSMEYKLLPQFRGRGSVTPQFV
            --NKNKPIIYI--DSMS-----YLIFYN----------------LIKPYL
                      . .   . .        .                  . ..

            TQC-----------LSKKLSHICKKQITHNI-ASIKY---RLIQRN----
            EPTQQLVPARVLDVHVKLSTRSMNRFDIEVEGNHNYFVDGVMVHNSPETT
            IPQ-----------MMYKLPN--------TI-SSETF-------------
                         .  **.          . ..  .

            ----ALK-
            TGGKALKF
            -----LK-
                 **
```

Figure 25

two latter introns are of the GIY-YIG type.

Furthermore, the observation that the LAGLI-DADG maturase/endonucleases encoded in the fourth *cob* intron and first *cox*1 intron of *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, respectively, can stimulate intra-chromosomal recombination in *Escherichia coli* (Goguel et al. 1990, Manna et al. 1991), prompted amino acid sequence comparisons between these *orf*s and the *C. eugametos* i5. Figure 25C shows the alignment between the regions of similarity of the *C. eugametos* i5, *S. pombe* I1, and *S. cerevisiae* bi4. It is noteworthy that in addition to the regions corresponding to the two dodecapeptide motifs, there are several rather conserved blocks in which at least two of the intronic *orf* sequences are identical.

Rather intriguingly, analysis of *C. eugametos* i5 also revealed sequence similarity (Fig. 25D) to an intein sequence identified in the recombinase gene (*recA*) of *Mycobacterium leprae*. It should be noted that one region of similarity between the two sequences includes the hexapeptide motif present around the second splice site of protein introns. The alignment of partial sequences of the intronic *orf*s of *C. eugametos nad5*, *S. cerevisiae rnl*, and *M. leprae recA* (Figure 25E) reveals regions of similarity among the three sequences as well as regions in which the *C. eugametos* i5 sequence is more similar to either one or the other of the two sequences.

6.4. Discussion

6.4.1. Potential consequences of a group II intron acquisition to the evolution of the

chlorophycean mitochondrial genome

Several group II introns have been identified in mitochondrial genes of fungi, a liverwort, and vascular plants, but only one example of a group II intron has been reported to date in a green algal mitochondrial gene, i.e., the LSU rRNA gene of the chlorophycean taxon *Scenedesmus obliquus* (Kück et al. 1990). The *S. obliquus* mitochondrial *rnl* group II intron is 608 bp long and has a strong self-splicing activity in vitro; its unique occurrence, in an LSU rRNA gene, and autocatalytic properties (the intron represents the smallest self-splicing group II intron from an organelle gene), suggest that group II introns are remnants of transposable elements (Kück et al. 1990 and references therein).

The *S. obliquus rnl* group II intron is inserted into a highly conserved region of the gene, namely, that corresponding to the peptidyl-transferase center in the LSU rRNA mature molecule (Fig. 26A). It is interesting that the *C. reinhardtii* mtDNA region that has shown nucleotide sequence and secondary structure similarity to domain V and VI of many group II introns (Fig. 24A and B) is situated 100 nt upstream of the corresponding insertion site of the group II intron in the *S. obliquus* LSU rRNA coding region. This finding together with the presence of an *rtl* coding region in the *C. reinhardtii* intergenic spacer raises the possibility that this spacer is the remnant of a group II intron previously harboured by the *C. reinhardtii* LSU rRNA gene. If so, the insertion sites of the *S. obliquus* and *C. reinhardtii* mitochondrial *rnl* group introns appear 100 nucleotide apart. This is not unprecedented given that the group I intron found in the

Figure 26. A. The *C. reinhardtii* mitochondrial LSU rRNA peptidyl transferase center and the nucleotide sequence at the corresponding *S. obliquus rnl* group II insertion site (arrow) as well as around the potential integration site in the *C. reinhardtii* LSU rDNA. The bracketed U denotes the nucleotide difference between the *C. reinhardtii* and *S. obliquus* sequences downstream of the corresponding insertion site in *S. obliquus*; L_7 and L_8 represent the two LSU rRNA fragments accounting for the 3'-half of the *C. reinhardtii* LSU rRNA. B. Hypothetical recombination event between an LSU rRNA coding module (coding units HIJ, as defined in Chapter 3) that contains a group II intron (zigzag line) harbouring an *rtl* (solid circle) and a second LSU rRNA coding module (DE) situated on an oppositely oriented transcriptional unit. Flags denote short inverted repeats located in regions corresponding to variable regions in the LSU rRNA, one of which is situated in the vicinity of a group II intron; dotted arrow indicate a recombination event between a two-copy short inverted repeat resulting in the inversion of the region flanked by the repeats. Curved arrows outside the circle indicate the transcription orientation of that region, and grey and solid blocks denote protein-coding genes in the mtDNA region that becomes inverted. C. Strand-exchange event within a region of similarity between two LSU rRNA coding modules; the crossed arrows indicate the site where the two strands exchange. D, E, H, I, and J are LSU rRNA coding units as defined in Chapter 3; underlined sequences denote the conserved motifs at the 5'- and 3'- splice sites of the group II intron, and the nucleotides in brackets indicate short conversion tracts during the strand exchange event.

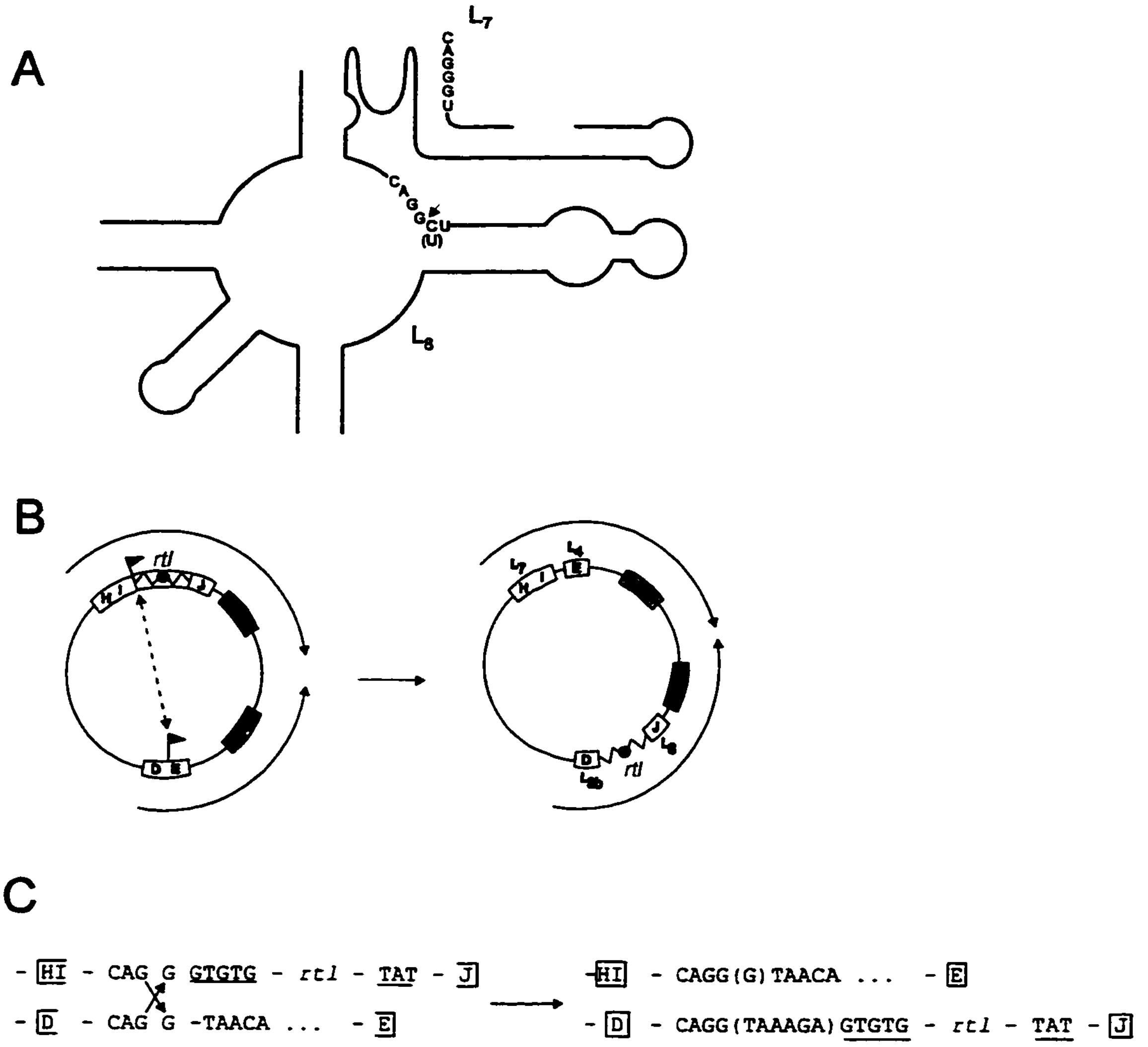


Figure 26

same peptidyl transferase center of *S. obliquus rnl* is also inserted 3 nt and 58 nt downstream of two group I introns in the LSU rRNA coding region of another *Chlamydomonas* taxa, *C. eugametos*. The presence of an insertion site that is 100 nt apart in *S. obliquus* relative to *C. reinhardtii* implies that the *rnl* group II introns were either (i) acquired independently in the two lineages or (ii) present in their most recent common ancestor at one or the other of the insertion sites, and subsequently transposed 100 nt in one of the lineages.

Given that introns are known to be always inserted in regions highly conserved among homologous genes, I have compared the nucleotide sequence of the region around the insertion site in *S. obliquus* with the counterpart in *C. reinhardtii*; surprisingly, I have noticed that although the level of conservation between the two sequences especially upstream of the insertion site is very high (i.e., GATAACAGG|TTTAATCAG versus GATAACAGG|CTTATCCGC), one of the few differences (in bold in the sequences above) refers to the first nucleotide downstream of the group II insertion site, that is T in *S. obliquus* but C in *C. reinhardtii* (Fig. 26A). Assuming that the integration or transposition of the *rnl* group II intron was dependent on a particular sequence similar to that around the current insertion site in *S. obliquus*, I have searched the *C. reinhardtii* LSU rRNA coding region upstream of the corresponding *S. obliquus* group II insertion site for the closest sequence bearing similarity to the *S. obliquus* CAGG|T. It should be noted that the integration sites of the yeast *cox1* α group II intron during transposition to new locations have also been found to be preceded by motifs similar to the 3'-end of the upstream *cox1* exon, i.e., TTGCAG versus TTTCAG (Sellem et al. 1993). Interestingly,

in the *C. reinhardtii* mitochondrial LSU rRNA sequence corresponding to a conserved region about 55 nt upstream of the *S. obliquus rnl* group II insertion site, I have found the sequence CAGGGT that is similar to the CAGGT at the insertion site in the *S. obliquus* LSU rRNA.

It is rather intriguing that the CAGGGT sequence in the *C. reinhardtii* LSU rDNA is situated in the close vicinity (11 nt upstream) of the breakpoint that accounts for the L_7 and L_8 LSU rRNA coding modules. It should be emphasized that L_7 and L_8 LSU rRNA coding modules are not adjacent in the *C. reinhardtii* mtDNA (Fig. 3) and L_8 is flanked on its 5'-end by the potential degenerate group II intron. Assuming that the LSU rRNA gene became gradually fragmented and scrambled through a series of recombinatorial events during chlorophycean mitochondrial genome evolution (discussed in Chapter 4), it is conceivable that the *C. reinhardtii* degenerate group II intron currently flanked on its 3'-side by the L_8 LSU rRNA coding module was in the green flagellate ancestor as part of a larger continuous coding region comprised of L_7 and L_8 LSU rRNA coding modules, similar to the situation present in *S. obliquus* mtDNA (discussed in Chapter 3). If so, one can envision (Fig. 26B) a recombination event between two LSU rRNA coding modules (coding units HIJ and DE, as defined in Chapter 3), mediated by short inverted repeats situated in variable regions of the LSU rRNA, one of which was in the proximity of a group II intron; this hypothetical event would fragment and scramble the LSU rRNA coding modules close to the group II boundary such that the group II intron becomes flanked by a new coding region, i.e., L_{3b}, on its 5'-end.

To support this hypothesis, I have searched for sequence signatures that will allow

the reconstitution of such an event. Interestingly, the 3'-end of L_7 (coding unit HI) contains the sequence CAGG**G**T**A**A**C**A which is similar to the sequence CAGGTA**A**A**G**A (differences are in bold) situated at the very 3'-end of L_{3b} (coding unit D) that precedes the 5'-end of the degenerate group II intron in *C. reinhardtii* mtDNA. Figure 26C suggests a recombination event (strand exchange) between two hypothetical LSU rRNA coding modules: an HIJ (as present in *S. obliquus*) that contains a group II intron inserted after the sequence CAGG (as in *S. obliquus* LSU rDNA) and a DE coding module (as present in *C. eugametos*). The strand exchange is initiated within a region of similarity including the CAGG sequence and is followed by a short conversion tract (Fig. 26C). Such an event results in the breaking of the HIJ coding module close to the 3'-splice site of the group II intron that consequently becomes adjacent to a new coding unit, namely, D; the intronic sequence is excised out from the chimeric LSU rRNA transcript but the two LSU rRNA fragments are not spliced together. To argue for the involvement of the group II intron in this particular gene rearrangement event could be the fact that the 3'-half LSU rRNA coding region in the *C. eugametos* mtDNA that lacks a group II intron is not broken in the corresponding *C. reinhardtii* region but, rather, 180 nt apart (Fig. 9).

It should be noted that the involvement of group II introns in genomic rearrangements has been previously suggested for fungal mitochondrial genomes. Among the 14 *Kluyveromyces lactis* strains investigated by Skelly et al. (1991), all of the six strains that contained a group II intron revealed rearranged mtDNA. Likewise, some of the observed rearrangements in *Podospora anserina* mitochondrial genome are due to the mobility of a particular group II intron (i.e., cox1-i1 or α) and are correlated with an

increased reverse transcriptase activity in senescent mycelia (Belcour et al. 1997). Although erroneous reverse splicing mechanisms are thought to contribute to mtDNA rearrangements in senescent fungi (Mueller et al. 1993), it has been suggested that group II splice-site-associated macrodeletions in fungal mtDNA might result from homologous recombination between the two copies of a group II intron duplicated during transposition events to new locations (Mueller et al. 1993 and the references therein). In addition, it is noteworthy that the breakpoints associated with the fragmentation and scrambling of the angiosperm mitochondrial *nad1*, *nad2* and *nad5* coding regions are situated within group II intronic sequences (Fauron et al. 1995 and the reference therein).

It is difficult at this point to distinguish between independent acquisitions of the *rnl* intron via horizontal transfer events in the *S. obliquus* and *C. reinhardtii* ancestors on the one hand, and a single acquisition in the green algal ancestor followed by a subsequent transposition event in one of the two lineages, on the other hand. It should be mentioned that yeast mitochondrial group II introns have been shown to both home and transpose *in vivo* (e.g., Skelly et al. 1991, Mueller et al. 1993, Sellem et al. 1993). Moreover, evidence for recent horizontal transfer of a group II intron between two yeast mitochondrial genomes has been provided (e.g., Hardy and Clark-Walker 1991).

More data on other chlorophycean mitochondrial genomes have to become available before one can assess the consequences, if any, of group II intron acquisition to the evolution of chlorophycean mitochondrial genomes. It should be emphasized, however, that the mitochondrial genome of the trebouxiophycean (sensu Friedl 1995) taxon *P. wickerhamii* as well as the sequenced mitochondrial genes of the prasinophycean

taxon *Platymonas subcordiformis* do not contain any intervening sequence of the group II type.

6.4.2. Potential roles of intron-encoded polypeptides in the evolution of the chlorophycean mitochondrial genome

A potential consequence of the acquisition of a group II intron harbouring an *orf* with reverse transcriptase activity to the evolution of chlorophycean mitochondrial genome is the loss of group I introns in the lineage leading to *C. reinhardtii*. It has been proposed that the reverse transcriptase activity encoded in some group II introns could cause the concomitant loss of group I and II introns (Wolff et al. 1993). Such an hypothesis is consistent with 1) the lack of any intronic sequences in *C. reinhardtii* mtDNA, 2) the presence of nine group I introns in *C. eugametos* whose mtDNA lacks an *rtl*, and 3) the presence of both group I and II introns in the *S. obliquus* LSU rDNA coding region whose group II intron does not harbour any intronic *orf* with potential reverse transcriptase activity. It should be mentioned, however, that the potential *C. reinhardtii* reverse transcriptase might not be functional because (i) the deduced amino acid sequence is missing motifs that are essential for the activity of reverse transcriptases from retroelements and (ii) no reverse transcriptase activity associated with the *rtl* expression was detected in the *C. reinhardtii* mitochondria (Faßbender et al. 1994).

The different codon usage bias in the *rtl* coding region relative to the standard protein-coding genes in the *C. reinhardtii* mtDNA led Boer and Gray (1988b) to suggest

that *rtl* has a separate, more recent origin. It remains to be learned whether *rtl* in the *C. reinhardtii* lineage was acquired as an independent mobile *orf* or if it was gained as the intronic *orf* of a mobile group II intron; assuming that the *C. reinhardtii* and *S. obliquus* group II introns have a common origin, the former scenario implies that the *S. obliquus rnl* group II intron never harboured such an *orf*, whereas the latter scenario would imply an independent loss of the intronic *orf* in the lineage leading to *S. obliquus*. It is noteworthy that sequences related to the vertebrate mtDNA origin of replication as well as to the *Chi* site involved in bacterial recombination have been found in the *C. reinhardtii rtl* (Nedelcu and Lee submitted) (discussed in Chapter 5). Although speculative at this point, the introduction of such sequences via the *rtl* in the chlorophycean mitochondrial genome might have had a contribution to its evolution. The characterization of the *S. obliquus* mitochondrial genome whose LSU rRNA gene contains a group II intron that lacks *rtl* might shed light on the potential consequences of an *rtl* acquisition to the chlamydomonadalean genome evolution provided that the *S. obliquus* mtDNA was completely devoid of *rtl* sequences.

The group I intron-encoded polypeptides could have also contributed to the plasticity of chlorophycean mitochondrial genomes. Although the *C. reinhardtii* mitochondrial genome is devoid of any group I intronic sequences, those of *C. eugametos*, *Chlamydomonas moewusii* and *Chlorogonium elongatum* contain a rather large and complex set of such sequences, many of which harbour *orfs* whose deduced amino acid sequences have similarities to maturase/endonucleases from other genetic systems (Kroymann and Zetsche 1997, Denovan-Wright, Nedelcu and Lee in press).

It is interesting that in addition to their role in RNA splicing and intron mobility, the LAGLI-DADG polypeptides encoded in many group I introns appear to be also involved in other processes such as DNA recombination and protein splicing. The bi4 maturase encoded by the fourth intron of the *S. cerevisiae cob* was shown to stimulate intra-chromosomal recombination in *E. coli* (Goguel et al. 1989). Although the exact mechanism of such activity is not known, the authors hypothesized that the *S. cerevisiae* bi4 maturase/endonuclease stimulates DNA strand-exchange, which is the first step in homologous recombination. In addition, it is known that this polypeptide binds double-stranded DNA, single-stranded DNA and RNA with similar affinities; binding to DNA might alter the DNA structure and enhance the binding of RecA protein (Delahodde et al. 1985).

Furthermore, it was shown that the accumulation of bi4 maturase in the yeast mitochondria of a splicing deficient mutant stimulates recombination between the mitochondrial genomes of *S. cerevisiae* and *Saccharomyces douglasii* (Kotylak et al. 1985). It is rather intriguing that the *C. eugametos* i5 is harboured by one of two introns found to have atypical nucleotides at one of the splice sites. In the first intron of the *C. eugametos nad5* the highly conserved U preceding the 5'-splice site of the intron is replaced by a C (Denovan-Wright, Nedelcu and Lee in press). It is noteworthy that in one of the only two other examples reported so far, i.e., an *Aspergillus cox1* intron, the C was shown not to be edited to restore the conserved U required for splicing (Hur and Waring 1995). Since it is not known yet whether such introns are capable of splicing or not, it is tempting to speculate that the *C. eugametos nad5* intron is not spliced (or

inefficiently spliced) and that, consequently, the i5 maturase accumulates in the mitochondria and stimulates recombination events. Such an hypothesis is consistent with the higher level of gene rearrangement in the *C. eugametos* mitochondrial genome relative to the *C. reinhardtii* counterpart (discussed in Chapter 5).

It is noteworthy that another intronic *orf*, the one in the first intron of *Schizosaccharomyces pombe cox1*, can stimulate intra-chromosomal recombination in *E. coli*; in addition, it was shown that a deletion in the intronic *orf*, which leads to a truncated polypeptide lacking the second dodecapeptide (P_2) region, abolishes both the RNA splicing and recombination stimulation in *E. coli* (Manna et al. 1991). Such an observation implies that P_2 and the downstream region are the ones most likely to be involved in the recombination processes. This is rather intriguing given that P_2 was shown to be involved in the maturase function of the bifunctional protein encoded by the group I intron aI4α of yeast mtDNA but to have no contribution to its endonuclease activity (Henke et al. 1995). However, it should be mentioned that there is evidence arguing for the involvement of the region C-terminal to the P_2 motif in both endonuclease and maturase function: the replacement of only two specific non-adjacent amino acids in the bi2 maturase encoded by the second intron in the *S. cerevisiae cob* is sufficient to gain homing-endonuclease activity (Szczepanek and Lazowska 1996). In this connection, it is interesting that one of the two amino acid differences associated with the presence of endonuclease activity in the *Saccharomyces capensis* bi2 *cob* maturase relative to the *S. cerevisiae* counterpart, i.e., histidine, is also present at the corresponding position in the *C. eugametos* i5 bi2 (see the bold H in Fig. 25E), arguing for a potential endonuclease

activity of this polypeptide.

It is also interesting that the *C. eugametos* i5 shows amino acid similarity with the *Micobacterium leprae recA* intein. Although the similarities between the two amino acid sequences might be only related to the potential endonuclease function of their polypeptides, a few aspects can be mentioned. First, the P_2 region, which is considered to be involved in the maturase activity of the LAGLI-DADG polypeptides (Henke et al. 1995), of the *C. eugametos* i5 showed the highest level of amino acid sequence identity (twelve out of 15 residues) to the *recA* intein sequence. Secondly, two adjacent residues are identical between the hexapeptide motif around the splice site of the *M. leprae recA* intein and the corresponding sequence in the *C. eugametos* i5. Lastly, it should be noted that the *C. eugametos nad5* intron that harbours i5 has an atypical nucleotide at its 5'-splice site (Denovan-Wright, Nedelcu and Lee in press); it is thus tempting to speculate that the intron is not excised at the RNA level but, rather, at the protein level or that this sequence represent an evolutionary intermediate between an RNA and a protein intron. It is also noteworthy that the observations regarding the splicing behaviour of the *M. leprae recA* intein led Davis et al. (1994) to suggests that protein introns may perform specific functions for their hosts, rather than being just selfish elements.

6.5. Conclusions

This chapter reports a degenerate group II intron in the intronless mitochondrial genome of *C. reinhardtii* as well as an intronic *orf* encoding a polypeptide of the LAGLI-

DADG type in the mitochondrial *nad5* of *C. eugametos*. Potential consequences of a group II acquisition as well as intronic group I and II *orfs* to the evolution of the chlorophycean mitochondrial genome include: fragmentation and scrambling of LSU rRNA coding regions,and facilitation and stimulation of recombination processes.

Chapter 7

Concerted modes and tempos of evolution of mitochondrial and chloroplast genomes in chlamydomonadalean green algae: A comparative analysis

7.1. Introduction

The increasing accumulation of information on organellar genome sequence, structure and organization in various lineages makes it possible to address questions such as: (i) how conserved are the organellar genomes in terms of DNA sequence, structure and organization among different evolutionary lineages; (ii) what are the mechanisms underlying the evolutionary processes in organellar genomes; (iii) are the mechanisms acting on mitochondrial and chloroplast genomes similar or different; (iv) do the mitochondrial and chloroplast genomes have the same tempo and mode of evolution in a given lineage; (v) what are the evolutionary forces shaping the organellar genomes?

Land plants and green algae are the only two groups that possess both mitochondria and chlorophyll *a/b*-containing chloroplasts. As the complete DNA sequences of several land plant chloroplast (see Shimada and Sugiura 1991 for references) and mitochondrial (Oda et al. 1992) genomes have become available, and as more and more genes have been mapped and sequenced, it has become obvious that the two organellar genomes exhibit different tempos and modes of evolution in the Chlorobionta: chloroplast genomes are more conserved in size and gene order but more variable in DNA sequence than the mitochondrial counterparts (Palmer 1990). No extensive analyses have been done, however, to assess the rates and patterns of evolutionary change of green algal organellar genomes, mainly due to incompleteness or disparity of the available data. Given that information on the structure, organization and DNA sequence of both mitochondrial and chloroplast genomes is now available, such issues can be addressed.

Nevertheless, the most data are from one green algal lineage, namely, the chlamydomonadalean lineage; thus the emphasis will be on this group. This chapter is not intended, however, to review the vast amount of information on green algal organellar genomes; rather, it will discuss only the information considered to have evolutionary significance from a comparative point of view. Extensive general reviews dealing with the structure, organization and evolution of both mitochondrial and chloroplast genomes have been published in the last decade (Palmer 1987, 1990, 1991, Palmer et al. 1985, Gray 1992, 1993, 1995, Gray and Spencer 1996, Gillham 1994, Rochaix 1995, Wolstenholme and Fauron 1995).

In the sections of this chapter I will (i) define evolutionary trends in the green algal mitochondrial and chloroplast genome lineages; (ii) compare the mitochondrial genome mode and tempo of evolution to that of the chloroplast counterpart within the *Chlamydomonas* lineage and among other green algal lineages; and (iii) contrast the rates and patterns of evolutionary change in the organellar genomes of the *Chlamydomonas* group and land plants.

7.2. Evolution of mitochondrial and chloroplast genome size in green algae

Mitochondrial and chloroplast genomes appear quite different in size both in the same lineage as well as among lineages. Mitochondrial genomes of land plants are large and extremely variable in size, ranging from 200 kb to 2500 kb (Palmer 1990). In contrast, land plant chloroplast genomes are smaller and seem to be rather conservative

in size, varying from 120 kb to 160 kb, in only few cases reaching 220 kb (see Palmer 1991 for a review). On the other hand, animal mitochondrial genomes are very small and extremely conserved in size, varying generally from 15.7 kb to 21 kb (see Wolstenholme and Fauron 1995 for a review), with the exception of scallop mitochondrial genomes that vary from 16.2 kb to 41 kb (see Gjetvaj et al. 1992 for references).

Surprisingly, the chlamydomonadalean mitochondrial genomes are 10-100-fold smaller than their angiosperm homologs, thus approximating the size of their metazoan counterparts, whereas the chlamydomonadalean chloroplast genomes are slightly larger than most of their angiosperm homologs. The analysis of the known mitochondrial and chloroplast genome sizes within the chlamydomonadalean group suggests a 1.5-fold variation in both organelles, from a 15.8-kb mtDNA in *C. reinhardtii* (Michaelis et al. 1990) to a 22.9-kb counterpart in *C. eugametos* (Denovan-Wright, Nedelcu and Lee in press), and from a 187-kb chloroplast DNA (cpDNA) in *Chlamydomonas pitschmannii* (Boudreau and Turmel. 1995) to a 292-kb homolog in *Chlamydomonas moewusii* (Boudreau et al. 1994). It is noteworthy that the same level of chloroplast and mitochondrial genome size variation is also observed in each of the two very divergent evolutionary lineages within the *Chlamydomonas* group, i.e., the "*C. reinhardtii*" and the "*C. eugametos*" lineages. Exceptional variation in genome size has, however, been described among the sexually incompatible members of the colonial chlamydomonadalean taxon, *Pandorina morum*, whose mitochondrial and chloroplast genome sizes vary from 20 kb to 38 kb, and from 150 kb to 450 kb, respectively (Moore and Coleman 1989, Moore 1990). It appears, therefore, that mitochondrial genome size is more conservative

among chlamydomonadalean taxa than among land plants (1.5-fold in the former, relative to 12.5-fold in the latter), whereas chloroplast genomes vary slightly more (1.5-fold) than do their land plant counterparts (1.3-fold). It is noteworthy that in contrast to land plants, in *Chlamydomonas* both organellar genomes exhibit the same level of size variation. Furthermore, it appears that the two organelle genomes followed parallel evolutionary pathways not only within the group but also within a given species: the mitochondrial and chloroplast genomes are both either small (e.g., in *C. reinhardtii* and *C. pitschmannii*) or large (e.g., in *C. eugametos* and *C. moewusii*) relative to the currently known size range in the *Chlamydomonas* group.

Among green algae, however, there is a 5-fold variation in mitochondrial genome size (excluding the 220-kb mitochondrial genome of *Bryopsis*), from 15.8 kb in *C. reinhardtii* (Michaelis et al. 1990) to 80 kb in *Chlorella pyrenoidosa* (Bayen and Rode 1973). Similarly, there is at least a 5-fold variation in chloroplast genome size among the green algal lineages investigated so far, from 89 kb in *Codium fragile* to 400 kb in a few members of three out of the five green algal classes, namely, the Charophyceae, Ulvophyceae, and Chlorophyceae (see Palmer 1991 for references). Although the degree of variation in genome size is overall higher among green algae, it is interesting that both organelle genomes seem to exhibit the same level of variation in genome size, a situation similar to that observed in *Chlamydomonas* but in contrast to that noted among land plants.

Because the mitochondrial genome in *Platymonas subcordiformis* (Kessler and Zetsche 1995), a green flagellate that retains ancestral-like features, is larger (i.e., 42.8

kb) than in chlamydomonadalean taxa, it is most likely that the reduced genome size in the latter represents a derived condition among green algae. Moreover, one can hypothesize an evolutionary trend toward a smaller mitochondrial genome within the chlorophycean group, from a 45-kb mitochondrial genome as in, for example, *Scenedesmus obliquus* (Kück 1989), to a 15.8 kb homolog in *C. reinhardtii*. Although limited, the current data do not support a similar trend among green algae: at 42.8 kb, the mitochondrial genome of the ancestral-like green flagellate *Platymonas subcordiformis* is smaller than the 55.3-kb homolog in *Prototheca wickerhamii* (Wolff et al. 1994) and the 80-kb counterpart in *Chlorella pyrenoidosa*. There is no indication of a tendency toward a smaller chloroplast genome among chlorophycean taxa as suggested for the mitochondrial counterparts; however, chloroplast genome sizes much smaller than the average have been reported among ulvophycean (89 kb in *Codium fragile* [Manhart et al. 1989]) and charophycean (130 kb in *Spirogyra maxima* [Manhart et al. 1990]) taxa.

7.2.1. Factors contributing to variation in genome size

Generally, changes in genome size are the result of changes in sequence complexity and/or changes in the amount of repeated DNA (Palmer 1990). Changes in genome complexity occur through the deletion and insertion of unique sequences (intergenic regions, introns and open reading frames). The great variation in mtDNA size among land plants is mostly accounted for by changes in the complexity of spacer DNA. In contrast, cpDNAs in the same group vary relatively little in size, with less contribution

from changes in intergenic region size, intron number or gene content but more (i.e., nearly two-thirds of cpDNA size variation) from the expansion/contraction of the inverted repeat (see Palmer 1991 for a review).

7.2.1.1. Changes in intergenic spacer size

The intergenic spacers are very reduced in *Chlamydomonas* mitochondrial genomes (i.e., 16-17% of the genome). In *C. eugametos*, with the exception of two large intergenic regions of 902 nt and 1057 nt, respectively, the intergenic spacers range from 0 to 466 nt, with most of them being smaller than 72 nt. In the more compact mitochondrial genome of *C. reinhardtii*, with the exception of the terminal non-coding regions of about 530 nt each, most of the intergenic regions are smaller than 200 nt and missing whenever two rRNA gene pieces are adjacent. On the other hand, the intergenic regions in the mitochondrial genome of *P. wickerhamii* represent 29% of the genome, and in the majority of cases are 100-150 nt long with only two considerably longer regions (1118 nt and 1993 nt) (Wolff et al. 1994). However, only about 12% of the 7-kb difference in size between the two completely sequenced *Chlamydomonas* mitochondrial genomes can be accounted for by differences in intergenic region size between the two genomes, the rest being accounted for by differences in intron number (discussed later), whereas the 30-kb difference in mitochondrial genome size between the mitochondrial genomes of *Chlamydomonas* and *Prototheca* is a consequence of variation in both the intergenic region size as well as gene content (discussed later).

In contrast, the substantial difference in chloroplast genome size between closely related *Chlamydomonas* taxa (i.e., *C. moewusii* and *C. pitschmannii* as well as *C. reinhardtii* and *C. gelatinosa*) is mainly a consequence of multiple deletions/additions in the intergenic spacers (Boudreau and Turmel 1995, 1996). Similarly, most of the observed differences in cpDNA restriction patterns between the interfertile *C. reinhardtii* and *C. smithii* are correlated with insertions/deletions of short dispersed repeated sequences of 50 bp to 200 bp, which are ubiquitous in the intergenic regions of *C. reinhardtii* cpDNA (Rochaix 1978, Gelvin and Howell 1979, Palmer et al. 1985). On the other hand, 12% of the 50-kb difference in size between the interfertile taxa *C. moewusii* and *C. eugametos* is accounted for by a 6-kb insertion in one of the single copy-regions of the *C. moewusii* chloroplast genome (Lemieux et al. 1985); the rest is the consequence of a 21-kb insertion in its large inverted repeat (Turmel et al. 1987). Another type of insertion is represented by the two copies of a 2.4 kb DNA sequence, the Wendy element, that has many of the features of transposable elements (discussed later) and has only been described in the chloroplast genome of *C. reinhardtii* (Fan et al. 1995). The absence of a counterpart in any of the other chlamydomonadalean or land plant chloroplast genomes examined to date suggests that Wendy is a relatively recent acquisition in the *C. reinhardtii* lineage (Fan et al. 1995). As far as the difference in chloroplast genome size between *Chlamydomonas* and land plants, it is correlated with the presence of enlarged intergenic spacers in the former relative to the latter (Boudreau et al. 1994).

7.2.1.2. Changes in intron number

There is quite a variation among chlamydomonadalean taxa in terms of the number of introns present in their organellar genomes. The difference in size between the mitochondrial genomes, otherwise colinear (Boynton et al. 1987), of the two interfertile taxa, *C. reinhardtii* and *C. smithii*, is solely the result of a unique 1-kb intronic insertion in the *cob* sequence of *C. smithii* (Matagne et al. 1988, Colleaux et al. 1990). In addition, it appears that the difference in mitochondrial genome size between the members of the other pair of interfertile *Chlamydomonas* taxa, that is *C. eugametos* and *C. moewusii*, might also be correlated with the presence of optional introns (Denovan-Wright and Lee 1993). Moreover, 88 % of the difference in size between the *C. reinhardtii* and *C. eugametos* mitochondrial genomes can be accounted for by the presence of nine intervening sequences in the coding regions of the latter. There seems to be a variation also in the number of introns harboured by homologous genes among chlamydomonadalean taxa: two introns are present in the *cob* sequence of a chlamydomonadalean taxa more closely related to *C. eugametos* than to *C. reinhardtii* (Buchheim et al. 1996), namely *Chlorogonium elongatum* (Kroymann and Zetsche 1997), whereas only one intron is found in *C. eugametos* (Denovan-Wright, Nedelcu and Lee in press) and *C. smithii* (Colleaux et al. 1990) *cob*, but none in the homologous *C. reinhardtii* gene (Michaelis et al. 1990). More data must be available before one can suggest whether the *Chlamydomonas* ancestor had or did not have introns in its mitochondrial genome, but variation in intron content is clearly a significant factor responsible for the observed variation in mitochondrial genome size among *Chlamydomonas* taxa.

Similarly, there is quite a variation in intron content among *Chlamydomonas* chloroplast genomes (Turmel et al. 1991). Turmel et al. (1993) investigated the chloroplast large subunit (LSU) rRNA gene from 17 *Chlamydomonas* taxa and identified a total of 39 group I introns representing 12 insertion sites. Whereas *C. pitschmannii* and *C. reinhardtii* have no and only one intron, respectively, in their chloroplast LSU rRNA gene, *C. moewusii* and *C. eugametos* have five and six, respectively. On the other hand, four introns are present in the *C. reinhardtii psbA* gene, which is intronless in *C. eugametos* (Erickson et al. 1984, Lemieux et al. 1985); one of the four introns present in *C. reinhardtii* is missing in the interfertile *C. smithii* (Palmer et al. 1985). Additional optional introns are also present in the *C. reinhardtii rns*, *psaB* and *psbC* (see Turmel et al. 1993 for references).

Another case of optional insertions contributing to variation in size of the open reading frames in *Chlamydomonas* chloroplast genes is represented by the presence of one or two large insertion sequences that are not spliced out at the mRNA level in the genes coding for the catalytic subunit of the ATP-dependent Clp protease (*clpP*) of *C.reinhardtii* and *C. eugametos*, respectively (Huang et al. 1994), and in-frame long sequences of unknown identity juxtaposed within the gene coding for the RNA polymerase subunit C (*rpoC2*) of *C. reinhardtii* cpDNA (Fong and Surzycki 1992).

7.2.1.3. Changes in gene content

One of the most distinctive features of the mitochondrial genome among green

algae is the unexpected dichotomy in terms of gene content, with *Chlamydomonas*-like genomes having a much more reduced gene content relative to other green algal and land plant counterparts (Nedelcu in press) (discussed in Chapter 1).

The two *Chlamydomonas* mitochondrial genomes completely sequenced to date, i.e., those of *C. reinhardtii* (Michaelis et al. 1990 and references therein) and *C. eugametos* (Denovan-Wright, Nedelcu and Lee in press) have the same set of standard genes: seven respiratory protein-, three transfer RNA (tRNA)- as well as fragmented and scrambled SSU and LSU rRNA-coding regions (Fig. 3). However, a reverse transcriptase-like gene (Boer and Gray 1988b) and an additional tRNA coding region (Denovan-Wright, Nedelcu and Lee in press) have been identified in *C. reinhardtii* and *C. eugametos*, respectively. Nevertheless, different codon usage and deduced amino acid composition of the reverse transcriptase-like coding region led Boer and Gray (1988b) to suggest that *rtl* in *C. reinhardtii* had an independent more recent origin relative to the standard mitochondrial genes. Although in *Oenothera berteriana* mtDNA an independent open reading frame showing reverse transcriptase-like similarity has also been described (Schuster and Brennicke 1987), it is noteworthy that the *rtl* in *C. reinhardtii* mtDNA is flanked by intergenic regions that contain sequence motifs present at the splice sites of group II introns (discussed in Chapter 6). The additional $tRNA^{Met}$ coding region present in *C. eugametos* mtDNA flanks one of the two copies of a large direct repeat (Fig. 3) and may be the result of a duplication/inversion-related event since the other end of the large direct repeat is flanked by a conventional $tRNA^{Met}$ gene. Duplicated transfer RNA genes or tRNA pseudogenes have been previously reported at inversion ends in wheat and rice

chloroplast genomes (Howe et al. 1988, Shimada and Sugiura 1989).

The presence of virtually the same set of coding regions in the mitochondrial genomes of two *Chlamydomonas* taxa that belong to evolutionary lineages considered to have diverged very early in the evolution of the *Chlamydomonas* group allows us to speculate that (i) the feature of a very reduced gene content was already present in the most recent chlamydomonadalean common ancestor and (ii) gene content is not a significant contributor to the mitochondrial genome size variation within the *Chlamydomonas* group.

Surprisingly, the mitochondrial genomes of green algal taxa outside the chlamydomonadalean group do not seem to share the very reduced gene content of their *Chlamydomonas* counterparts. The *Platymonas subcordiformis* mtDNA encodes at least 12 respiratory proteins, seven tRNAs, two ribosomal proteins as well as LSU and SSU rRNAs (Kessler and Zetsche 1995). Moreover, the mitochondrial genome of *Prototheca wickerhamii* codes for 16 respiratory proteins (including three subunits of the ATPase complex which is entirely non-mitochondrial-encoded in *Chlamydomonas*), 26 tRNAs, ten ribosomal proteins as well as 5S, LSU and SSU rRNAs (Wolff et al. 1994). The presence (or absence) of a mitochondrial 5S rRNA gene and ribosomal protein genes is an additional distinctive feature among green algal lineages: *Prototheca wickerhamii* has a 5S rRNA gene as well as many ribosomal protein-coding genes (Wolff et al. 1994), *Platymonas subcordiformis* has at least a few ribosomal protein genes (Kessler and Zetsche 1995), whereas *C. reinhardtii* (Michaelis et al. 1990) and *C. eugametos* (Denovan-Wright, Nedelcu and Lee in press) have neither ribosomal protein nor 5S rRNA genes.

Given that the mitochondrial genome of the ancestral-like green algal lineage *Platymonas* as well as of the trebouxiophycean (sensu Friedl 1995) lineage, *Prototheca*, has a gene content more similar to their land plant counterparts, one can hypothesize that the most recent common ancestor of green algae and land plants contained quite a large number of genes in its mtDNA and that a reduction in gene content occurred in the chlorophycean lineage leading to *Chlamydomonas*. Given that *Chlamydomonas* and *Prototheca* may not be as closely related as previously thought (Friedl 1995) (i.e., they are in fact members of the chlorophycean and trebouxiophycean lineages [sensu Friedl 1995], respectively, whose divergence is probably very old) the differences in gene content between *Chlamydomonas* and *Prototheca* are less surprising.

The dichotomy in mitochondrial gene content among green algae is also reflected in different degrees of resemblance to their counterparts in land plants, the group considered their closest relatives at the nucleo-cytosolic level (Chapman and Ragan 1980; Chapman and Buchheim 1991). Whereas the mitochondrial genomes of *C. reinhardtii* and *C. eugametos* (and most likely of all the chlorophycean taxa) resemble more closely their ciliate/fungal/animal counterparts in terms of gene content, those of *Prototheca wickerhamii* and *Platymonas subcordiformis* (and most likely of all trebouxiophycean lineages) have a gene content similar to their land plant counterparts (Nedelcu in press) (discussed in Chapter 1).

No green algal chloroplast genome is fully described, but half of the *C. moewusii* and *C. eugametos* chloroplast genomes have been sequenced and the analysis of these genomes revealed half of the land plant cpDNA gene content (Boudreau et al. 1994).

Moreover, the two very divergent *Chlamydomonas* taxa, *C. reinhardtii* and *C. moewusii*, appear to share a similar gene complement (Fig. 4) (Boudreau et al. 1994); among the 75 genes mapped on the two *Chlamydomonas* chloroplast genomes, five coding regions have not been reported in any land plant cpDNA, and four of these have also not been identified in other green algal counterparts (Boudreau et al. 1994). On the other hand, the genes encoding the chlororespiratory NADH dehydrogenase subunits present in land plant cpDNA may be lacking in the chloroplast genome of *Chlamydomonas*. The information accumulated so far led Boudreau et al. (1994) to suggest that the presence of additional genes accounts only for a very small fraction of the increased size of the chlamydomonadalean chloroplast genome relative to the land plant homologs. It is noteworthy that some chloroplast genes display unusual structures and the corresponding transcripts are undetectable, suggesting that they might not be functional (Fong and Surzycki 1992).

Although there seems to be some variation in gene content among green algal chloroplast genomes (e.g., several additional genes involved in nucleotide metabolism reported in the cpDNA of the ulvophycean taxon *Acetabularia mediterranea* and a more reduced gene content in that of *Codium fragile*, see Palmer 1991 for discussion), the differences are not expected to be as striking as those observed among mitochondrial genomes.

7.2.1.4. Changes in the amount of repeated DNA

Large and short repeated sequences have been found in mitochondrial genomes of some green algae. The terminal non-coding regions of the linear mtDNA of *C. reinhardtii* each contain a copy of an inverted repeat (TIR) of about 530 nt (Fig. 3) (Vahrenholz et al. 1993). On the other hand, the two largest intergenic regions in the *C. eugametos* circular-mapping mitochondrial genome each contain a copy of a 260-nt direct repeat (DR) (Fig. 3) (Denovan-Wright, Nedelcu and Lee in press). TIRs were described also in the linear mitochondrial genome of the colonial chlamydomonadalean taxon *Pandorina morum* (Moore and Coleman 1989). Given that among sexually incompatible members of *P. morum* the TIRs range in size from 1.8 to 3.3 kb, it seems, at least in this taxon, that the variation in the amount of repeated DNA could be responsible in part for the observed intraspecific variation in mitochondrial genome size (from 20 kb to 38 kb). However, complete mitochondrial genome sequence data for more chlamydomonadalean taxa are needed in order to make any suggestion as to how significant if at all is the amount of repeated DNA in genome size evolution within the chlamydomonadalean group.

Among green algae there is even more variation than seen in *Chlamydomonas* in terms of the amount of repeated DNA in the organellar genomes. Although there are no large repeats in the circular-mapping mitochondrial genome of *Prototheca wickerhamii*, a two-copy inverted repeat of ca. 1.5 kb was reported in the *Platymonas subcordiformis* homolog (Kessler and Zetsche 1995).

Among the large repeated sequences in the chloroplast genome of *Chlamydomonas*, the most common is a two-copy inverted repeat (IR) found always as

two identical but oppositely oriented copies that divide the chloroplast genome into two rather equal single-copy regions (Rochaix 1978, Lemieux et al. 1985, Turmel et al. 1987, Boynton et al. 1992). As in land plants, the *Chlamydomonas* IR varies in size by spreading or shrinkage (Turmel et al. 1991). The 50-kb difference in chloroplast genome size between the very closely related *C. moewusii* and *C. eugametos* is mostly accounted for by an enlarged inverted repeat in *C. moewusii*, which was shown to be the consequence of a 21-kb insertion; the rest is accounted for by a 6-kb insertion in the single copy bordered by the LSU rRNA genes (Lemieux et al. 1985, Turmel et al. 1987). The expansion/contraction of the inverted repeat also accounts for the gain/loss of a 7-kb sequence containing the ATPase beta subunit gene (*atpB*) between the two closely related *C. reinhardtii* and *C. gelatinosa* (Boudreau and Turmel 1996). On the other hand, although there is a 53-kb difference in size between the chloroplast genomes of the distantly related *C. eugametos* and *C. reinhardtii*, their inverted repeats are about the same size in both taxa. Nevertheless, similar size does not necessarily imply identical sequence complexity, given that the inverted repeat in *C. reinhardtii* lacks the *rbcL*, which is located instead in a single-copy region (Malnoë et al. 1979, Dron et al. 1982). In contrast, although the inverted repeats of the interfertile *C. reinhardtii* and *C. smithii* are identical in gene content and similar in overall gene organization, they differ at the fine structure level by an extensive series of small deletions/additions: a minimum of 11 length mutations (20-1600 bp) are distributed throughout the IR of *C. smithii*, which makes it almost 1 kb larger that the 22-kb repeat of *C. reinhardtii* (Palmer et al. 1985). Comparisons between the inverted repeats of the members of each of the interfertile

Chlamydomonas pairs, *C. reinhardtii/C. smithii* and *C. eugametos/C. moewusii*, have revealed more deletion/addition differences than nucleotide substitutions (Palmer et al. 1985, Lemieux et al. 1985), which is opposite to the situation observed in higher plants (Zurawski et al. 1984). Among green algae, the chloroplast IR ranges in size between 20 and 41 kb and seems to be lacking in a few charophycean and ulvophycean lineages whose cpDNAs are also known to be smaller than the chloroplast genome size average (see Palmer 1991 for references).

7.2.2. Mechanisms possibly involved in the evolution of genome size

7.2.2.1. Length mutations

Changes in organellar genome sequence complexity occur primarily by length mutations, i.e., the addition of new sequences or the deletion of existing ones. Many of the small-length mutations in organelle genomes were found to be flanked by or close to short direct repeats, suggesting their occurrence during DNA replication or repair according to the "slippage-mispairing" model (Takaiwa and Sugiura 1982, Zurawski et al. 1984). The great majority of length mutations in land plant cpDNA are small, only 1-10 bp in size, and occur predominantly in noncoding DNA (intergenic spacers and introns). Length mutations of 10-1,200 bp in size occur less frequently than smaller ones and are more likely to occur by recombination than by replication; unequal crossing-over between misaligned tandem repeats could produce both deletions and additions, and

intramolecular recombination between short direct repeats could produce deletions (see Palmer 1991 for a review).

Repetitious DNA is very abundant in the mitochondrial genome of *Chlamydomonas*. Short dispersed repeats have been reported in five of the intergenic spacers of the *C. reinhardtii* mitochondrial genome (Boer and Gray 1991). A more abundant and complex set of repetitive sequences (in both direct and inverted orientation) has been found in the mitochondrial genome of *C. eugametos*; more than 80 repetitive elements, ranging in size from 6 nt to 17 nt, are dispersed throughout the intergenic regions as well as within several introns (Denovan-Wright, Nedelcu and Lee in press, Nedelcu and Lee submitted) (discussed in Chapter 5). Identical copies or closely related sequences of the same repeat are present tandemly repeated in various combinations; in only a few cases were individual copies of a repetitive sequence found apart from a cluster. The mitochondrial genome of *P. wickerhamii* is also rich in very complex repetitive motifs consisting of AT-rich tandem repeats in both intergenic spacers and introns (Wolff et al. 1994).

Similarly, the *C. reinhardtii* chloroplast genome possesses a family of about 40 short (100-300 bp) repeated sequences scattered throughout most of the genome (Rochaix 1978, Gelvin and Howell 1979), at least 13 of them being localized within the inverted repeat (Palmer et al. 1985). The short dispersed repeats account for ca. 22% of the total *C. reinhardtii* chloroplast genome and each is composed of shorter repeated motifs that occur in direct or inverted orientation and in various combinations (Gillham 1994). Given that a number of the observed small-length mutations mapped close to regions showing

differences in the number of these small repeats, Palmer et al. (1985) suggested that enhanced recombination within and between repeat elements may be related to the increased incidence of length mutations in *Chlamydomonas* relative to angiosperm cpDNA. Studies of experimentally induced mutants showed that the endpoints of most of the structural mutations, both deletions and inversions, mapped in the general vicinity of the 100-300-bp repeat elements scattered throughout the inverted repeat (Palmer et al. 1985). The high frequency of symmetrical alterations suggested the existence of a copy-correction mechanism for maintaining identity between the two copies of the inverted repeat. The mechanism for insertion of new sequences within the intergenic regions of *Chlamydomonas* cpDNA is not known. Because no dispersed repeats have been detected by Southern blot hybridization in the cpDNAs of *C. pitschmannii* and *C. eugametos,* Boudreau and Turmel (1995) did not favour the proliferation of existing sequences throughout the genome through unequal recombination as a mechanism responsible for the 56-kb difference in size between the intergenic regions in the cpDNAs of these two taxa. On the other hand, in the land plant cpDNA it was proposed (Tsai and Strauss 1989) that repeats were created and spread by duplicative transposition; however, no classical transposable element has been isolated from any land plant chloroplast. In contrast, the repeated sequences and the *orfs* (whose deduced amino acid sequence show some similarity with transposases and integrases of other mobile elements) associated with the Wendy element in the *C. reinhardtii* chloroplast genome (discussed later) argue for the existence, present or past, of a transposable element in this lineage (Fan et al. 1995).

Short repetitive sequences are also present in the chloroplast genomes of other

green algae. Insertions of repeated sequences as well as *orfs* with terminal repeated sequences account for larger intergenic spacers in the *Chlorella ellipsoidea* chloroplast IR relative to the *C. reinhardtii* counterparts (Yamada 1991). Because the smallest chloroplast IRs known to date contain at least the rRNA (*rrn*) operon, Yamada (1991) suggested that the IR might have been originally created from the duplication (in an inverted orientation) of the *rrn* operon, followed by its expansion to incorporate additional coding regions; the mechanism proposed for the expansion of the IR involves a double reciprocal recombination during the replication step. The model presented by Yamada (1991) requires the presence of repetitive sequences acting as recombination hot spots within and around the IR and is consistent with the location of the repetitive sequences in the inverted repeat of *Chlorella ellipsoidea* chloroplast genome.

7.2.2.2. Intron mobility

Although gain or loss of introns could be considered a special case of length mutation (Palmer 1991), the mechanisms underlying these processes are very different. It was proposed that the loss of introns is the result of two processes: (i) the reverse-transcription of an RNA whose intron sequences have been removed by splicing and (ii) homologous recombination between the intronless cDNA and the native gene (Dujon 1989). In mitochondria, the putative reverse transcriptases (RTs) encoded either by the group II intronic *orfs* or certain non-intronic *orfs* with reverse transcriptase similarity, as found in *C. reinhardtii* (Boer and Gray 1988b) and a few angiosperms (see Moenne et al.

1996 for references), could produce an intronless copy from a spliced RNA. The above mentioned processes seem more likely to account for the loss of group II introns; nevertheless, Wolff et al. (1993) considered that the reverse-transcriptases encoded in group II introns could have also been responsible for the loss of group I introns, culminating in their complete elimination from higher land plant mtDNA. It is noteworthy that although the overexpressed gene product of *C. reinhardtii rtl* does not appear to have a reverse transcriptase activity (Faßbender et al. 1994), such an activity has been recently detected in the mitochondria of potato (Moenne et al. 1996). On the other hand, in *Chlamydomonas* chloroplast, evidence for reverse transcriptase activity is still lacking, but recombination processes seem to be well developed (Dürrenberger et al. 1996).

To explain the transfer of group I introns to novel locations (intron transposition), two mechanisms have been proposed: one occurs at the DNA level and is promoted by an intron-encoded endonuclease (Dujon 1989), whereas the other starts at the RNA level and involves (i) the introduction of an intron RNA sequence into a foreign RNA by a reversal of a self-splicing reaction, (ii) the reverse-transcription of the recombined RNA and (iii) the integration of the cDNA sequence into the genomic DNA by homologous recombination (Woodson and Cech 1989). Turmel et al. (1993) presented evidence that reverse self-splicing reaction might have played the major role in the creation of novel intron insertion sites in the LSU rRNA genes as well as elsewhere in the chloroplast genome of *Chlamydomonas*. However, these authors also suggested that certain group I introns might have been introduced via lateral transfer facilitated by the site-specific

endonuclease encoded in these introns. Such a mechanism is considered to be responsible for the mobility (intron homing) of the chloroplast *rnl* intron of *C. eugametos* (Lemieux and Lee 1987, Gauthier et al. 1991, Bussieres et al. 1996) and *C. reinhardtii* (Dürrenberger and Rochaix 1991) as well as of the mitochondrial *cob* intron of *C. smithii* (Boynton et al. 1987, Matagne et al. 1988, Colleaux et al. 1990, Ma et al. 1992) to cognate positions within the corresponding intronless genes. Another case indicative of intron mobility is the apparent evolutionary transfer of a group I intron between the mitochondrial *rnl* of *Acanthamoeba castellanii* and the chloroplast counterpart of *Chlamydomonas*, either intracellularly in a remote photosynthetic common ancestor of the two lineages, or intercellularly, as a result of a recent lateral transfer event (Lonergan and Gray 1994, Turmel et al. 1995b).

7.2.2.3. Gene transfer

Among the processes of genetic flux (such as gene transfer, reverse gene transfer, gene substitution, gene sharing, gene recruitment, and gene loss - see Palmer 1991 for a review) that account for the differences in gene content among lineages, gene transfer seems to be the main contributor to changes in genome complexity in organellar genomes. However, Palmer (1991) considers that the incorporation of a foreign gene into an organelle genome (reverse gene transfer) is not very rare and includes events such the invasion of chloroplast genes into the higher plant mtDNA as well as the recent acquisition of a number of *orf*s within land plant and green algal chloroplast introns (see

Palmer 1991 for references).

To explain the gene transfer from one cellular compartment to another, Obar and Green (1985) proposed a stepwise model involving: (i) duplication of an organellar gene followed by the transfer of one copy to the nucleus; (ii) activation of the nuclear copy still keeping active the organellar counterpart; and (iii) inactivation and subsequent loss of the organellar gene copy. Gene transfer from the organelle to the nuclear genome seems to have been more important in the evolution of the mitochondrial than the chloroplast genome in *Chlamydomonas*. The presence of a higher number of genes in the mitochondrial genome of *Platymonas subcordiformis*, considered a descendant of the primitive green flagellates from which all the advanced green algal lineages have evolved, suggests that the feature of a very reduced gene content in *Chlamydomonas* is a derived trait among green algae. The mechanisms and causes responsible for such a massive reduction in the gene content of mtDNA in *Chlamydomonas* are not known, although a few suggestions have been made (discussed in Chapter 1 and 5). Recombination between short direct repeated sequences was proposed to have been responsible for the deletion of mitochondrial coding regions during the evolution of the *Chlamydomonas*-like genomes (Nedelcu 1997, Nedelcu and Lee submitted) (discussed in Chapter 5). It is conceivable that the accumulation of GC-rich short direct repeated sequences with recombinogenic properties in the lineage leading to *Chlamydomonas* could have promoted recombination events responsible for the deletion not only of protein and tRNA genes but also of rRNA-coding regions (discussed in Chapter 4 and 5).

7.3. Evolution of mitochondrial and chloroplast genome organization in green algae

7.3.1. Mitochondrial and chloroplast genome structure

An unexpected dichotomy in mitochondrial genome conformation has been observed among green algal taxa: linear mtDNA molecules have been isolated from *C. reinhardtii* (Boer et al. 1985), *C. smithii* (Boynton et al. 1987) and the colonial chlamydomonadalean alga *Pandorina morum* (Moore and Coleman 1989) but circular-mapping mtDNAs have been reported for other chlamydomonadalean taxa, *C. eugametos* (Denovan-Wright and Lee 1992), *C. moewusii* (Lee et al. 1991) and *C. pitschmannii* (Boudreau and Turmel 1995), chlorophycean taxa such as *Chlorella* and *Scenedesmus obliquus*, as well as trebouxiophycean and prasinophycean taxa like *Prototheca wickerhamii* and *Platymonas subcordiformis*, respectively (Kück 1989, Moore and Coleman 1989, Waddle et al. 1990, Kessler and Zetsche 1995).

The actual in vivo conformation of green algal mitochondrial genomes is debatable (Bendich 1993). Electrophoretic migration patterns of the circular-mapping mitochondrial genomes of *C. moewusii* and *C. eugametos* (Boer et al. 1985, Lee et al. 1991) leave open the possibility of their existing in vivo as linear, larger-than-unit-size genomes (Bendich 1993). It is noteworthy that although circular-mapping mitochondrial genomes have also been reported in land plants, they may exist in vivo predominantly as larger-than-unit-genome-size linear structures (Bendich 1993, 1996). If the circular-mapping green algal

mtDNAs are circular molecules in vivo, their linearization in some chlamydomonadalean lineages could have been the consequence of a recombination event between short repeated sequences on a small linear episome and their homologs on the circular chromosome, as described during the linearization of the maize mitochondrial genome (Schardl et al. 1984). It is intriguing that one of the long terminal inverted repeats in the linear *C. reinhardtii* mtDNA is flanked by small inverted repeats (Fig. 3) thus arguing for a potential previous episomal existence of this TIR (Nedelcu and Lee submitted) (discussed in Chapter 5).

In contrast, chloroplast genomes appear to be circular in conformation in all *Chlamydomonas* investigated to date; however, they exist as a 50:50 mixture of two genetically identical but physically distinct molecules that differ only in the relative orientation of their single-copy regions as a result of high-frequency intramolecular recombination events between the two copies of the IR (Aldrich et al. 1985, Palmer et al. 1985). In addition, intermolecular recombination events involving the short repeated sequences in the *C. reinhardtii* cpDNA have been suggested to yield dimer and multimer cpDNA molecules (Boynton et al. 1992, Boudreau and Turmel 1996).

7.3.2. Mitochondrial and chloroplast gene order

The gene order among the green algal mitochondrial genomes investigated to date appears to be as variable as observed among the vascular plant mitochondrial genomes. Protein-coding genes are highly interspersed with tRNA genes as well as rRNA gene

pieces in the two *Chlamydomonas* mitochondrial genomes fully sequenced to date, namely of *C. reinhardtii* and *C. eugametos*. Moreover, none of the coding regions is flanked by homologous counterparts, i.e., there is no gene cluster common to the two *Chlamydomonas* genomes (Boer and Gray 1988a, Denovan-Wright, Nedelcu and Lee in press) (Fig. 3). Given that in *C. eugametos* the protein-coding genes are more interspersed with rRNA-coding regions than they are in *C. reinhardtii*, it seems likely that the mitochondrial genome of *C. eugametos* has undergone additional gene rearrangements relative to its *C. reinhardtii* counterpart.

The level of mitochondrial gene rearrangement among green algae is difficult to assess due to the lack of complete mitochondrial genome sequences for green algal lineages other than *Chlamydomonas* and *Prototheca*, as well as the very reduced gene content of the *Chlamydomonas* mitochondrial genome relative to the *Prototheca* counterpart. Nevertheless, among the 12 genes common to both *Chalmydomonas* and *Prptotheca* lineages, there is only one gene cluster that is common to *Prototheca wickerhamii* and *C. reinhardtii*, i.e., the *nad5-nad4* cluster, and none between *P. wickerhamii* and *C. eugametos*. It is noteworthy that among the genes that have been mapped on the mitochondrial genome of *Platymonas subcordiformis*, the only gene cluster that is shared with *Prototheca* is the same *nad5-nad4* cluster. Probably the most unexpected variation in gene order among the green algal mitochondrial genomes has to do with one of the most conserved gene clusters in the land plant counterparts, i.e, the one comprising the rRNA genes: in *Platymonas*, as in most other mitochondrial systems, *rnl* and *rns* are located on the same DNA strand, whereas in *Prototheca* and land plant

mitochondrial genomes the *rns* and *rnl* are encoded on opposite DNA strands. It is also interesting that there might be at least five polycistronic units in the mtDNA of *Platymonas*, but as few as two in *Prototheca* and *C. reinhardtii* mtDNA, and only one in *C. eugametos*.

In contrast to the situation in land plants, the chloroplast gene order is quite variable among *Chlamydomonas* taxa. Although the chloroplast genomes of the interfertile members of each pair of taxa, *C. eugametos/C. moewusii* and *C. reinhardtii/C. smithii* are colinear (Turmel et al. 1987, Boynton et al. 1992), the cpDNAs are so extensively rearranged between the two pairs (i.e., *C. eugametos/C. moewusii* vis-a-vis *C. reinhardtii/C. smithii*) that rearrangements cannot be described in terms of simple individual events (Fig. 4). It has been suggested that the great evolutionary distance separating these taxa might be responsible for such a high level of rearrangements although the possibility that cpDNA rearranges at a fast rate in *Chlamydomonas* cannot be disregarded (Lemieux and Lemieux 1985). Interestingly, a comparative analysis of chloroplast gene order in the two closely related *Chlamydomonas* taxa, *C. moewusii* and *C. pitschmannii*, has revealed a level of rearrangement close to that observed among all land plants: one or two inversions and possibly one or three events of expansion/contraction of the inverted repeat (Boudreau and Turmel 1995). Unexpectedly however, the level of rearrangements appears much more extensive between *C. reinhardtii* and *C. gelatinosa* (i.e., at least nine inversions and one expansion/contraction event of the IR), although chloroplast LSU rDNA sequence-based phylogenies suggest they are as closely related as are *C. moewusii* and *C. pitschmannii* (Boudreau and Turmel 1996).

Most changes in gene order in the two pairs of closely related *Chlamydomonas* specieas are located in the single-copy region bordering the *rns* gene (Boudreau and Turmel 1995, 1996). In the course of these rearrangements, a few cpDNA sequences have moved from one single copy region to the other, a phenomenon not observed in higher plant cpDNA.

Chlamydomonas cpDNAs lack the extensive operon structure of their land plant counterparts; the best example is represented by the six ATPase genes (*atp*), which are organized into two operons in all other chloroplasts but are scattered singly around the genome in a species-specific manner in *Chlamydomonas* (Palmer 1991). It has been suggested (Turmel et al. 1988) that the numerous rearrangements during *Chlamydomonas* cpDNA evolution resulted in the disruption of the polycistronic transcription units inherited from the prokaryotic ancestor; in contrast, different evolutionary pressures during land plant evolution determined an increased number of polycistronic transcription units and a more compact genome organization. Most of the ancestral operons still present in land plant cpDNAs have been lost in the *Chlamydomonas* lineage; of 76 genes mapped on five *Chlamydomonas* cpDNAs, 40 represent 15 conserved clusters, only five of which are similar to the primitive operons present in land plant chloroplast genomes, and only one having exactly the same gene content as the land plant equivalent (Boudreau et al. 1994, Boudreau and Turmel 1996).

Similarly, although limited, studies on cpDNA in green algae outside the chlamydomonadalean group have disclosed a highly variable gene order with only few conserved gene clusters, suggesting that the evolutionary pattern of green algal cpDNA is less conservative than that of their land plant counterparts (Palmer 1991). The few

green algal chloroplast genomes mapped to date do not share similar gene orders with either one another or with *Chlamydomonas* counterparts. Some gene clusters, however, are present in more than one lineage: for instance, the cytochrome f-cytochrome b/f complex subunit 4 gene cluster (*petA*-*petD*) present in *Chlamydomonas* is also found in *Scenedesmus obliquus* (Kück 1989, Kück et al. 1990) and it was suggested that a more extended gene cluster, including *petA*-*petD*-*trnR*^{UCU}, might have been present in the most recent common ancestor of *Chlamydomonas* and *Scenedesmus* (Boudreau et al. 1994). Moreover, two clusters, PSI P700 apoprotein A1-PSI P700 apoprotein A2 (*psaA*-*psaB*) and PSII 44kd protein-PSII D2 protein (*psbC-psbD*), are shared by *Spirogyra* and *Codium* (Manhart et al. 1990). DNA rearrangements involving ancestral polycistronic units also occurred during the evolution of green algal chloroplast genomes; *rns* and *rnl* that are co-transcribed in all bacteria, as well as *Chlamydomonas* and land plant chloroplasts, are separately transcribed in *Spirogyra* (Manhart et al. 1990), *Codium* (Manhart et al. 1989) and *Chlorella ellipsoidea* (Yamada and Shimaji 1987). Moreover, the organization of the *rrn* genes is very different between species of the same genus: in *Chlorella ellipsoidea* the *rrn* operon present in duplicate in the IR is split into two back to back operons (Yamada and Shimaji 19987) whereas in *Chlorella vulgaris* the *rrn* operon is represented by a single copy, due to the absence of an IR, and is coded on the same DNA strand (Kapoor et al. 1997).

7.3.2.1. Factors contributing to gene rearrangement

Palmer (1990) suggested several factors that are likely to promote more inversional recombination in land plant mtDNA relative to cpDNA: (i) more short dispersed repeats that could serve as points for homologous recombination; (ii) larger intergenic regions that could tolerate inversions; (iii) more monocistronic rather than multicistronic mitochondrial operons.

It is interesting that short repeated sequences have been found in the intergenic spacers of both *C. reinhardtii* (Boer and Gray 1991) and *C. eugametos* (Denovan-Wright, Nedelcu and Lee in press, Nedelcu and Lee submitted) mtDNAs. Furthermore, the presence of fewer and more reduced gene clusters in *C. eugametos* mitochondrial genome relative to the *C. reinhardtii* homolog, as well the disruption of the conserved *nad5-nad4* gene cluster in the former, are correlated with a more abundant and complex set of short GC-rich repetitive sequences in the intergenic regions of the former relative to the latter (Nedelcu and Lee submitted). It is noteworthy that short repeated sequences are also present in the mitochondrial genome of *P. wickerhamii*, but they are highly AT-rich (Wolff et al. 1994).

Not surprisingly, short dispersed repeats are also particularly abundant in the highly rearranged cpDNAs of *Chlamydomonas*. The dispersed repeats in the *C. reinhardtii* and *C. gelatinosa* cpDNA are composed of short repeated units occurring in different combinations at different loci on the genome (Boynton et al. 1992, Boudreau and Turmel 1996). The insertion of repeated sequences in intergenic spacers has been suggested to have been a significant factor in promoting the disruption of the ancestral polycistronic operons in the lineage leading to *Chlamydomonas* (Boudreau et al. 1994).

The very different number of repeated sequences between the cpDNAs of the members of two apparently equally distant pairs of closely related *Chlamydomonas* taxa, *C. reinhardtii*/*C. gelatinosa* and *C. moewusii*/*C. pitschmannii*, correlates directly with the very different level of gene rearrangement between the cpDNAs of the two lineages (Boudreau and Turmel 1996). It is noteworthy that the tobacco and *Marchantia* chloroplast genomes that differ in gene order by a single 30 kb inversion, despite 400 million years of evolutionary separation, contain no dispersed repeats larger than 50 bp (Palmer 1990). However, short dispersed repeats of 50-1000 bp are unusually abundant in most of the land plant genomes that are highly rearranged. Short terminal degenerate inverted repeats and four-nucleotide directly repeated sequences as well as additional degenerate copies of these sequences in direct or inverted orientation have been also found within the two copies of the Wendy element suggested to be involved in rearrangement events in the chloroplast genome of *C. reinhardtii* (discussed later) (Fan et al. 1995).

7.3.2.2. Mechanisms possibly involved in gene rearrangement

In higher plant mitochondria, recombination has been invoked to explain the genomic rearrangements accounting for their complex mitochondrial genome structures. It has been proposed that the frequency of homologous recombination is related to the presence of recombinogenic repeated sequences (Palmer and Shields 1984). A model invoking large direct and inverted repeat-mediated inter- and intramolecular recombination

events involving various isomeric forms of the mitochondrial genome has been proposed to illustrate the multipartite structure and the evolution of the maize mitochondrial genome (Fauron et al. 1995).

Because there is no evidence for a multipartite structure of the mitochondrial chromosome in *Chlamydomonas*, recombination events similar to those proposed for the land plant mitochondrial genome do not seem very likely to have happened in the *Chlamydomonas* counterparts. Rather, I suggest that intramolecular recombination between one and two sets of inverted repeats, which results in the inversion and interchange of the regions flanked by the repeats, respectively, might have played an important role in the evolution of mitochondrial genomes in *Chlamydomonas* (discussed in Chapter 5).

Moreover, I think that mtDNA replication occurring at multiple sites, first competing with, and later replacing the conventional origin of light strand replication could have been involved in the extensive gene rearrangement observed in *Chlamydomonas*, in a manner similar to that proposed for the vertebrate mitochondrial genome (Macey et al. 1997). The GC-rich palindromic sequences present in the intergenic spacers of *C. reinhardtii* and *C. eugametos* could act as surrogate origins of replication of the mitochondrial DNA, similar to the situation observed in yeast mtDNA (Nedelcu and Lee submitted) (discussed in Chapter 5).

In land plant chloroplast genomes, on the other hand, homologous or illegitimate recombination between short repeats (as small as 11-16 bp) and/or tRNA genes (*trn*), both functional and pseudogenes, is considered the major cause of gene rearrangements (see

Boudreau and Turmel 1995, 1996 for references). The relative abundance of appropriately oriented and located (between, rather than within transcription units) short dispersed repeats seems to be a major factor in determining the prevalence of cpDNA inversions in land plants (Palmer 1991).

Similarly, recombination events between short dispersed repeats have also been proposed to account for the various rearrangements described in the *Chlamydomonas* chloroplast genome (discussed by Boudreau and Turmel 1996). Moreover, the finding that *trn*-specific oligonucleotide probes hybridized near the endpoints of an inversion in *C. pitschmannii* cpDNA led Boudreau and Turmel (1995) to raise the possibility of intra- or intermolecular recombination events between duplicated tRNA genes being responsible for the observed inversion.

In addition, in *C. reinhardtii*, the Wendy element is considered to have played a major role in the shuffling of chloroplast gene clusters in this lineage relative to other *Chlamydomonas* lineages; the fact that both copies of Wendy were found to be flanked by gene clusters that are contiguous in *C. moewusii* but are separated and inverted relative to each other in *C. reinhardtii* argues for such an involvement. The mechanisms involved in such rearrangements might have involved Wendy-dependent illegitimate homologous or site-specific recombination events, or both (Fan et al. 1995).

7.4. Evolution of mitochondrial and chloroplast gene structure and organization in green algae

7.4.1. Intron-containing coding regions

Both mitochondrial and chloroplast coding regions in green algae are interrupted by introns. Only introns of the group I family have been found in mitochondrial rRNA- and protein- but not tRNA-coding regions of *Chlamydomonas*. It is interesting, however, that the only other available mtDNA sequence of a chlorophycean taxon outside the chlamydomonadalean group, i.e., a partial sequence of mitochondrial *rnl* from *Scenedesmus obliquus*, revealed in addition to a group I intron, a group II intronic sequence (Kück et al. 1990). Because the insertion site of the *rnl* group I intron in *S. obliquus* is four nucleotides downstream of the insertion site of one of the *C. eugametos* mitochondrial *rnl* group I introns, it is difficult to make any suggestions as whether the most recent common ancestor of these two lineages that belong to the two main evolutionary lineages within the chlorophycean group, namely the DO and CW lineages, had or did not have introns in its LSU rRNA gene. Nevertheless, two group I introns have been found in *Prototheca wickerhamii* mitochondrial *rnl*: one intron is present at the same position as the mitochondrial *rnl* group I intron in *S. obliquus* (Wolff et al. 1993), whereas the other intron shares an identical insertion site with one of the three mitochondrial *rnl* group I introns in *C. eugametos*. This finding could argue for the presence of at least two introns in the mitochondrial LSU rRNA gene in the most recent common ancestor of the trebouxiophycean and chlorophycean lineages, followed by the loss of one of the introns as well as the acquisition of new introns at new insertion sites in each of the two lineages. It seems likely, therefore, that the most recent common

ancestor of green algae had introns in its mitochondrial coding regions, and subsequently introns were independently lost or acquired in distinct evolutionary lineages.

Interestingly, two of the three group I introns present in *cox1* of *Prototheca wickerhamii* are located at positions identical to the sites of insertion of liverwort mitochondrial *cox1* introns, and it has been suggested that they were already present in the common chlorophyte/embryophyte ancestor (Wolff et al. 1993). Moreover, both *Scenedesmus obliquus* and liverwort mtDNAs contain introns of the group II family, which represent the only type of intron found in the mitochondrial genes of angiosperms. It is interesting to note that the reverse transcriptase-like coding region present in the *C. reinhardtii* mtDNA seems to be in fact the intronic *orf* of a degenerate group II intron (discussed in Chapter 6). These observations may suggest that the most recent common ancestor of the green algal/land plant group contained both group I and II introns in its mitochondrial coding regions and that a massive loss of all of the former, and most of the latter, has occurred in the tracheophyte (vascular plants) as well as *C. reinhardtii* lineage.

In contrast, *Chlamydomonas* chloroplast genes contain both group I and II introns. The distribution of 39 group I introns representing 12 insertion sites is highly variable among 17 *Chlamydomonas* species and does not suggest the same phylogenetic relationships among *Chlamydomonas* lineages as do the chloroplast rDNA sequences (Turmel at al. 1993). Because the *rnl* of cyanobacteria and of the *Chlorella* and land plant chloroplast lineages lack introns, it was suggested that all of the intron insertion positions in *Chlamydomonas* are of recent origin and that some of them might have arisen after the divergence of the two main *Chlamydomonas* lineages (Turmel et al. 1993).

However, it was recently reported that the *Chlorella vulgaris* chloroplast *rnl* is interrupted by a group I intron inserted at the same position as the single group I intron in the *C. reinhardtii rnl*; intronic sequence comparisons as well as amino acid sequence similarities of their intronic *orfs* suggested that the two closely related self-splicing *rnl* group I introns in *C. vulgaris* and *C. reinhardtii* descended from the same group I intron present in the most recent common ancestor of these two lineages (Kapoor et al. 1997). Moreover, because the chloroplasts of both land plants and their closest relatives, the charophycean green algae, display overall a very small number of group I introns (Palmer 1991), it has been suggested that the proliferation of group I introns occurred in the *Chlamydomonas* lineage after the charophycean divergence (Turmel et al. 1993).

Whereas *Chlamydomonas* chloroplast genes have multiple introns in most of the genes possessing introns, higher plant chloroplast genes contain single introns. Also, although land plant chloroplast tRNA genes harbour long single introns, no split tRNA genes have been found in algal chloroplasts. The only introns with features of the group II family identified to date in *Chlamydomonas* chloroplast genes are those found in *psaA* of several *Chlamydomonas* taxa (see Turmel et al. 1995a).

It is interesting that some of the land plant mitochondrial and chloroplast protein-coding genes consist of scattered exons flanked by 5'- or 3'- segments of group II introns; the exons are separately transcribed and *trans*-spliced (see Turmel et al. 1995a for references). In green algae, such an organization has only been described among chloroplast genes. *Chlamydomonas psaA* coding regions are made up of three exons scattered around the genome and *trans*-spliced (Kück et al. 1987, Turmel et. al. 1991,

Choquet et al. 1988, Goldschmidt-Clermont et al. 1991, Turmel et al. 1995a). Although the location of the three exons in the genome is different between two very divergent taxa, *C. moewusii* and *C. reinhardtii*, the information contained is similar, indicating that the most recent common ancestor of these two lineages possessed a *psaA* coding region interrupted in a similar manner (Boudreau et al. 1994). In contrast, *psaA* is not *trans*-spliced in land plants or other genera of algae (with the exception of *Euglena gracilis*), but *rps12* (which is uninterrupted in *Chlamydomonas*, *Euglena*, and *Cyanophora*) is *trans*-spliced in all examined land plants (Sugiura 1989).

An additional feature contributing to variation in chloroplast gene structure in *Chlamydomonas* is the presence of translated large insertion sequences in the *clpP* gene of *C. reinhardtii* and *C. eugametos* (Huang et al. 1994) as well as chimeric RNA polymerase coding regions juxtaposed in-frame with long sequences of unknown origin in *C. reinhardtii* (Fong and Surzycki 1992).

7.4. Fragmented coding regions

In some green algal lineages, both mitochondrial and chloroplast rRNA coding genes are fragmented into coding modules whose transcripts are not spliced together into covalently continuous rRNA molecules (Nedelcu et al. 1996, Nedelcu 1997). However, the rRNA gene organization and expression as well as evolutionary origins are very different between the two organellar genomes.

Mitochondrial rRNA genes in *Chlamydomonas* are not only highly fragmented but

also scrambled (i.e., the gene pieces are interspersed with other coding regions and do not follow the 5'-3' transcriptional order of their counterparts in conventional continuous genes) (Fig. 3). The SSU and LSU rRNA-coding regions are fragmented into four and eight gene pieces, respectively, in *C. reinhardtii* (Boer and Gray 1988a) and into three and six gene pieces, respectively, in *C. eugametos* (Denovan-Wright and Lee 1994). The rRNA-coding modules are extensively interspersed with each other as well as with protein-coding and tRNA genes, and the rRNA fragments are most likely excised from longer multicistronic transcripts following precise endonucleolytic scissions (Boer and Gray 1988a). Although the rRNA pieces are not spliced together, they have the ability to interact through intermolecular base pairing, to restore the conserved core of the rRNA secondary structure (Boer and Gray 1988a, Denovan-Wright and Lee 1994), and are able to assemble into mitochondrial ribosomes (Denovan-Wright and Lee 1995).

Discontinuous mitochondrial rRNAs have been reported not only among *Chlamydomonas* taxa (Denovan-Wright et al. 1996), but also in chlorophycean taxa with a CW or DO flagellar configuration as well as chlorococcalean taxa phylogenetically related to them (Nedelcu et al. 1996). A trend in the evolution of this trait, that is a tendency toward an increase in the degree of discontinuity, from continuous mitochondrial rRNAs to the highly fragmented mitochondrial rRNAs in *C. eugametos* and *C. reinhardtii,* was suggested (Nedelcu et al. 1996, Nedelcu 1997) (discussed in Chapter 3).

Although mitochondrial rRNA genes are highly fragmented and scrambled in both *C. reinhardtii* and *C. eugametos*, the distribution of the coding information among their coding modules, as well the order of these modules within the genome, is different

between the two species (Denovan-Wright and Lee 1994). Calculations of the minimal number of transpositions required to convert hypothetical ancestral rRNA gene organizations to the arrangements present in the two *Chlamydomonas* taxa, as well as a limited survey of the size of mitochondrial LSU rRNAs in other *Chlamydomonas* species, led Denovan-Wright et al. (1996) to propose that the last common ancestor of *Chlamydomonas* algae possessed fragmented mitochondrial rRNA genes whose coding modules were nearly colinear with their counterparts in conventional continuous rRNA genes. The model presented by the authors predicted that in taxa basal to the *Chlamydomonas* group, mitochondrial rRNA genes would be fragmented but not scrambled. The presence of scrambled but not highly fragmented mitochondrial LSU rRNA coding regions in *Scenedesmus obliquus*, however, suggested that scrambling may have developed at an early stage in the evolution of discontinuous and scrambled rRNA genes within the chlorophycean green algal group, probably in parallel with the fragmentation events (Nedelcu 1997) (discussed in Chapter 4).

The mechanisms responsible for either the fragmentation or the scrambling of the mitochondrial rRNA coding regions in *Chlamydomonas* are not known yet, although several suggestions have been made. Nedelcu (1997), however, proposed a model that could disrupt and scramble a coding region in a single step through an intramolecular homologous recombination event between two sets of two-copy inverted repeats (discussed in Chapter 4). It was suggested, furthermore, that the fragmented and scrambled mitochondrial rRNA coding regions in the chlorophycean green algal group may have been generated through multiple recombination events triggered by the

accumulation of short repeated sequences within the variable regions of the rRNA genes and the intergenic spacers of these mitochondrial genomes (Nedelcu 1997). Comparisons among the locations of the short repeated elements within the mitochondrial genome of *C. reinhardtii*, *C. eugametos* and the available sequence of other chlorophycean taxa revealed similarities regarding the positions of these repeats relative to the rRNA-coding units within the respective genomes (Nedelcu and Lee submitted) (discussed in Chapter 5).

It is interesting that chloroplast LSU but not SSU rRNA-coding regions are also fragmented in some green algal lineages. Three internal transcribed spacers (ITSs) located at the same position in the chloroplast *rnl* of all 17 *Chlamydomonas* taxa investigated by Turmel et al. (1993) interrupt this gene into four gene pieces whose transcripts are not covalently linked after the removal of the ITSs from the primary transcript. Unlike the introns, but like the break points in the mitochondrial rRNA-coding regions, the ITSs are located within highly variable regions of primary and/or secondary structure. The ITSs in the *Chlamydomonas* chloroplast *rnl* are usually less than 300 nt long and differ substantially in size and base composition. Although they are always excised post-transcriptionally from a precursor RNA to yield four mature rRNA species, no common sequence motif to account for a similar processing recognition signal has been identified, suggesting that either different recognition signals or specific three-dimensional topology of the ribosome might be involved in the processing of ITSs (Turmel et al. 1993). It is noteworthy that the size and base composition differences among corresponding ITSs in different *Chlamydomonas* taxa are not consistent with the

phylogenetic relationships suggested by the LSU rRNA-coding sequences (Turmel et al. 1993).

This feature of fragmented chloroplast LSU rRNA-coding regions is not confined to *Chlamydomonas* taxa. In *Chlorella ellipsoidea*, an insert that does not have the characteristics of an intron has been reported at the same position as ITS3 of *Chlamydomonas* (Yamada and Shimaji 1987). Moreover, Nedelcu et al. (1996) showed that the chloroplast LSU rRNAs in green algal lineages from three green algal classes (sensu Mattox and Stewart 1984), the Chlorophyceae, Pleurastrophyceae, and Micromonadophyceae, have fragmented chloroplast LSU rRNAs, in most cases the fragmentation pattern being similar to that described in *Chlamydomonas*. The distinct patterns observed in some lineages are most likely due to the absence or inability to process one of the ITSs. On the other hand, although three ITSs, one of which accounts for the 4.5S rRNA species, have been identified in the maize chloroplast LSU rRNA gene (Kössel et al. 1985), they are situated at different positions than in *Chlamydomonas*. It is interesting that the chloroplast *rnl* of *Chlamydomonas* as well as the mitochondrial *rnl* of higher plants are both missing the variable region in which the ITS that accounts for the 4.5S rRNA species is situated (see Turmel 1993 for references).

It was proposed that there is a direct evolutionary connection between variable regions and ITSs, in the sense that variable regions might have in fact evolved from the ITSs separating the rRNA coding modules in the progenote (Gray and Schnare 1995). Although most of the ITSs in contemporary rRNA genes represent most likely derived rather than primitive traits, the acquisition of the processing sites responsible for the

excision of the contemporary ITSs could be considered a "reversion to a primitive state" (Gray and Schnare 1995).

Another unusual gene organization has been reported for the chloroplast RNA polymerase genes of *C. reinhardtii*. The RNA polymerase beta subunit coding region (*rpoB*) is divided into two *orfs* separated by a 616 bp spacer. Although evidence for the transcription of these *orfs* is missing, if expressed, they most likely encode separate polypeptides (Fong and Surzycki 1992).

7.5. Evolution of mitochondrial and chloroplast DNA sequences in green algae

In land plants, estimated rates of synonymous (silent) nucleotide substitution per site in mitochondrial and chloroplast protein-coding genes are lower than in nuclear genes (reviewed by Palmer, 1991; Bousquet et al., 1992; Laroche et al., 1997). No extensive studies on point mutation levels in green algal mitochondrial, chloroplast and nuclear genes have yet been published. However, K.J. Prendergast and R.W. Lee (personal communication) have noted that in *Chlamydomonas*, the number of synonymous substitutions per site, in contrast to land plants, is higher in mitochondrial than chloroplast protein-coding genes and both are higher than the only one nuclear protein-coding gene (*rbcS*) for which such a value can be calculated at present.

The number of substitutions in *Chlamydomonas* mitochondrial SSU and LSU rRNA genes is severalfold higher than the accumulated substitutions in land plant mitochondrial counterparts (Denovan-Wright et al. 1996). Mitochondrial rRNA sequences

of *P. wickerhamii* also seem to have a high rate of nucleotide substitution and, together with the *Chlamydomonas*, ciliate, fungal and yeast counterparts, constitute a rapidly evolving group (associated with long branches in phylogenetic analyses), in marked contrast to the slowly evolving land plant mitochondrial rRNA sequences. Although the number of transitional substitutions is probably saturated in the rapidly evolving mitochondrial rRNA sequences, Denovan-Wright et al. (1996) showed that the apparent affiliation of the *Chlamydomonas* sequences with ciliate/fungal/yeast counterparts, and therefore their separation from the land plant sequences, is not due to a "long-branch length attract" artifact (i.e., the grouping of rapidly evolving sequences together, in spite of their true phylogenetic relatedness).

Surprisingly, green algal chloroplast *rnl* sequences also display extensive sequence divergence. The various *Chlamydomonas* lineages studied by Turmel et al. (1993) revealed at least twice the range of sequence variation seen in all land plants. Moreover, within some *Chlamydomonas* lineages (including those leading to *C. reinhardtii* and *C. eugametos*) the level is greater than that found between the bryophyte *Marchantia* and the monocot *Oryza* (Turmel et al. 1993). Although the chloroplast genomes of the closely related *C. pitschmannii* and *C. eugametos/C. moewusii* are extremely similar in gene order they appear to be very divergent in DNA sequence, as deduced from differences in their cpDNA restriction patterns (Boudreau and Turmel 1995). The ratio of point mutations to length mutations in *Chlamydomonas* cpDNA is, however, substantially lower in *Chlamydomonas* than in the angiosperm chloroplast DNAs, due to the increased level of length mutations in the former (Palmer et al. 1985).

The lack of knowledge as to the exact time of divergence of different green algal lineages makes it difficult to assess absolute rates of nucleotide substitutions in *Chlamydomonas* organellar genomes and to compare their tempo of DNA sequence evolution with that of other counterparts. Although the chlamydomonadalean group appears to be at least 350 million years old, the first chlorophycean green algal fossils are around 900 million years old (Tappan 1980). It seems reasonable to assume, therefore, that the *C. reinhardtii*/*C. eugametos* divergence could be as old as 900 million years or as recent as 400 million years. On the other hand, the bryophyte/tracheophyte (vascular plants) divergence is believed to have occurred about 400 million years ago (Schopf 1970). If nucleotide substitution levels in organelle DNAs of *Chlamydomonas* were roughly equal to or up to twice the level observed in their land plant counterparts, comparable rates of nucleotide substitutions in *Chlamydomonas* and land plants could be hypothesized. In contrast, levels of nucleotide substitution exceeding twice the level noted among land plant counterparts would suggest a higher rate of nucleotide substitution within *Chlamydomonas*. The observed levels of sequence divergence in organellar rRNA genes in *Chlamydomonas*, i.e., severalfold and at least two-fold higher in mitochondrial and chloroplast rRNA genes, respectively, indicate higher and slightly higher rates of nucleotide substitution in *Chlamydomonas* mitochondrial and chloroplast rRNA genes, respectively, relative to their land plant counterparts. However, substitution rates of protein-coding genes from all three genetic compartments from various lineages within the chlamydomonadalean group have to be available before the tempo of DNA sequence evolution in *Chlamydomonas* organellar genomes can be assessed with confidence.

It is not fully understood why the DNA sequences in different genomes of a given lineage or among different groups have different evolutionary rates. Palmer (1990) suggested that error-free replication mechanisms, better postreplication repair systems or copy-correction mechanisms might explain the overall low substitution rates in plant organelle genomes. It is noteworthy that in contrast to land plants, the rDNA nucleotide substitution levels in both *Chlamydomonas* organellar genomes appear relatively high. It is possible that this opposite trend could be determined by the same factors proposed for land plant organellar genomes but acting in an opposite direction: error-prone replication mechanisms, inefficient postreplication repair systems or copy-correction mechanisms. To explain the observation that mitochondrial and chloroplast genomes in land plants have both low rates of DNA sequence evolution, it was suggested that they might be under common nuclear control (Palmer 1990). Such a control can also be hypothesized for *Chlamydomonas* given that both organellar DNA sequences seem to have evolved at high rates.

7.6. Conclusions

Although land plant organellar genomes revealed "contrasting modes and tempos of genome evolution" (Palmer 1990), the organellar genomes within the chlamydomonadalean group seem to exhibit concerted modes and tempos of evolution (Table 9). The 1.5-fold variation currently observed in the size of both organelle genomes is mostly accounted for by changes in the spacer DNA and intron number, with

less contribution from changes in gene content and amount of repeated DNA. Gene order is highly variable in both mitochondrial and chloroplast genomes, the level of gene rearrangement being correlated with the abundance of short dispersed repeated sequences throughout the genome. Intron-containing-, fragmented-, and fragmented and scrambled coding regions are common features of mitochondrial and chloroplast gene structure and organization within the group. The level of rDNA sequence divergence in both mitochondrial and chloroplast genomes is higher in the *Chlamydomonas* lineage than in land plants and is most likely due to higher nucleotide substitution rates in *Chlamydomonas* organellar DNAs. The mechanisms and the selective pressures responsible for both the overall high rate of evolution in *Chlamydomonas* organellar genomes relative to the land plant counterparts, as well as their apparent concerted evolution are not fully understood.

Table 9. Factors and their contribution to organellar genome evolution in land plants and *Chlamydomonas*.

	Land plant		***Chlamydomonas***	
	mtDNA	**cpDNA**	**mtDNA**	**cpDNA**
Genome size evolution				
Changes in spacer DNA	significant	insignificant	moderate	moderate
Changes in intron number	insignificant	insignificant	significant	significant
Changes in gene content	insignificant	insignificant	insignificant	insignificant
Changes in amount of repeated DNA	insignificant	significant	significant	significant
Genome organization evolution				
Genomic rearrangements	significant	insignificant	significant	significant
Gene structure and organization evolution				
Intronic insertions	significant	insignificant	significant	significant
Gene fragmentation	insignificant	significant	significant	significant
DNA sequence evolution				
Nucleotide substitutions	insignificant	moderate	significant	significant

General Conclusions

We are probably still far from understanding the evolutionary forces that acted on the primitive organellar genomes and determined such a great diversity in the contemporary organelles. As much as we would like to define (i) prototypes for ancestral mitochondrial and chloroplast genomes, respectively, from which all the present organelle lineages diverged, as well as (ii) similar modes of organelle genome evolution in related eukaryotic lineages, it may not be possible. Although the vertebrate and higher plant organelle genome types seem to be quite easy to define, difficulties are encountered when trying to define an organelle genome type for the lower eukaryotic groups. The generally accepted view is that these groups are ancient and their divergence is older, therefore, they have evolved for a longer time, so that the present forms are more diverse. However, we should not disregard the fact that at the time of vertebrate or land plant divergence, the endosymbiotic associations were well established and the interactions between the cellular compartments quite stable, whereas at the time of protist and algal divergence, the genetic and environmental diversity and instability were probably still significant evolutionary forces. In this context, adaptative pressures represented by changes of habitat correlated with changes in the life history of the primitive eukaryotic lineages, acting on less stable genetic potentials may have been important factors in determining different evolutionary paths reflected in the great diversity we notice among the present lower eukaryotic organelle genomes.

Appendix

Polytomella was considered by Brown et al. (1976) to be one of the most primitive Volvocales, in a direct line of evolution with *Chlamydomonas*. The beta-tubulin nucleotide sequences (Conner et al. 1989) and the COXI amino acid sequences (Antamarian et al. 1996) of *Polytomella* and *Chlamydomonas reinhardtii* indicate a 98.6% and 75% similarity, respectively, suggesting a close phylogenetic relationship between these two flagellates.

Based on the cell shape and ultrastructural features of the pyrenoid and flagellar apparatus, Lembi (1975) recognized two distinct groups within the quadriflagellate *Carteria* genus: group I (e.g., *Carteria radiosa*) lacking an anterior papilla and retaining a flagellar apparatus similar to that of *Chlamydomonas reinhardtii*, and group II (e.g., *Carteria olivieri, Carteria crucifera*) having an anterior papilla and a flagellar apparatus far more elaborated and complex than that of group I. O'Kelly and Floyd (1984) have introduced the terms "clockwise absolute orientation" (CW) and "counterclock absolute orientation" (CCW) to define the orientation of the flagellar basal bodies of chlorophycean and ulvophycean green algae, respectively; they also suggested that the CW orientation gradually developed from the ancestral CCW configuration, and that evolutionary intermediates could be found among chlorophycean with almost directly opposed basal bodies. In this view, *Carteria* group II is the basal sister lineage to the clade containing all the green algae with a CW flagellar configuration (e.g, *Chlamydomonas, Polytomella,*

Carteria group I), as also indicated by cladistic analyses based on molecular, organismal and combined data sets (Buchheim and Chapman 1992).

Regarding *Scenedesmus obliquus* phylogenetic affiliations within Chlorophyceae, cladistic analyses of nuclear SSU rRNA sequences, organismal data and combined data sets (Buchheim and Chapman 1992), assigned this autosporic species as a lineage basal to the clade comprised of chlorophycean green algae with a CW configuration in their flagellar apparatus (e.g., *Chlamydomonas, Polytomella, Carteria*). Moreover, phylogenetic analyses based on nuclear 18S rRNA sequences of autosporic and zoosporic chlorococcalean green algae (Wilcox et al. 1992) grouped the autosporic *Scenedesmus obliquus* with zoosporic taxa with a directly opposed (DO) flagellar apparatus configuration, in a clade sister to the one containing CW taxa.

Hormotilopsis gelatinosa and *Planophila terrestris* have been classified by O'Kelly et al. (1994) into a new order, Chaetopeltidales, created to accommodate *Chaetopeltis*-like taxa, species whose quadriflagellate zoospores display almost directly opposed basal bodies in their flagellar apparatus. They also suggested that the group containing green algae with clockwise flagellar apparatus configurations (e.g., *Chlamydomonas, Polytomella, Carteria)* derived from a *Chaetopeltis*-like ancestor.

Based on distinct ultrastructural features, *Uronema belkae* was removed from the Ulotrichales and assigned to the Chaetophorales (Stewart et al. 1973); the species belonging to this order have quadriflagellate zoospores whose flagellar apparatus includes directly opposed upper basal bodies and lower basal bodies in the clockwise absolute orientation. The chaetophoralean and the chaetopeltidalean zoospores share similar

features in their rhizoplast attachment sites and position, and in their flagellar apparatuses configuration, some of them being different from the one described for other algae with directly opposed basal bodies. However, some zoospore and vegetative cell features of the chaetophoralean algae are different enough from those of chaetopeltidalean algae to determine O'Kelly et al. (1994) to consider the Chaetophorales as a separate lineage within Chlorophyceae, a lineage derived, however, from a *Chaetopeltis*-like ancestor. Cladistic analyses based on nuclear SSU and LSU rRNA sequences, organismal data, and combined data sets (Buchheim and Chapman 1992) also indicated the Chaetophorales as the basal chlorophycean clade, diverging prior to *Scenedesmus obliquus*.

Hafniomonas montana, previously known as *Pyramimonas montana* was removed from Prasinophyceae and included within Volvocales, Chlorophyceae (Ettl and Moestrup 1980). This naked flagellate has a flagellar pit whose structure is considered an intermediate between pits like those of *Tetraselmis* (Prasinophyceae) and large papillas like those described for some species of *Carteria* (Mattox and Stewart 1984). Although *Hafniomonas montana* has a flagellar apparatus whose major components display a counterclockwise configuration (but only a slight basal body offset) and other distinct features, O'Kelly et al. (1994) consider that *Hafniomonas* and the Chaetopeltidales are sister taxa and that *Hafniomonas* is ancestral to other Chlorophyceae. Melkonian (1990a) also pointed out that *Hafniomonas* bears relationships to the subgenus *Vestigifera* of *Pyramimonas* and to species of *Carteria*, and therefore, may be related to chlorophycean algae producing quadriflagellate motile cells.

As autosporic taxa, *Prototheca wickerhamii* and *Chlorella vulgaris* are classified

within the Chlorococcales. Surprisingly, on nuclear SSU rRNA sequence phylogenies (Wilcox et al. 1992, Steinkötter et al. 1994), autosporic taxa fall into two clade: one, including *Scenedesmus obliquus* and *Chlorella fusca* groups with zoosporic forms having a DO type of flagellar apparatus, and the other, including *Prototheca wickerhamii, Chlorella vulgaris, Chlorella minutissima* and *Nanochlorum eucariotum* appears distinct enough from other chlorococcalean taxa, both autosporic and zoosporic, to branch outside the higher level clade containing the rest of the chlorophycean algae (DO and CW). Moreover, this second clade seems to have the greatest affinity to pleurastrophycean taxa (sensu Mattox and Stewart 1984), organisms possessing a counterclockwise configuration in their flagellar apparatus.

Although the monophyly of the Pleurastrophyceae class (sensu Mattox and Stewart 1984) is not resolved, the Microthamniales (sensu Melkonian 1982, 1990b) including *Pleurastrum terrestre* and other taxa sharing a distinctive set of ultrastructural characters and a unique type of zoospore (flattened, biflagellate cells with a counterclockwise orientation of their basal bodies) does certainly represent a monophyletic assemblage of taxa. Moreover, the higher level clade containing the Microthamniales and the autosporic chlorococcalean sister group, is a sister clade to the one including all the remaining chlorophycean, both DO and CW lineages (Steinkötter et al. 1994).

Pyramimonas parkae is classified within Micromonadophyceae because exhibits the primitive characters of a persistent interzonal mitotic spindle and a counterclockwise configuration in its flagellar apparatus (Mattox and Stewart 1984). This class is considered a natural group in the sense that its members had a common ancestor, but that

ancestor apparently gave rise to all the other green algal groups as well.

References

Aldrich, J., B. W. Cherney, E. Merlin, C. Williams, and L. Mets. 1985. Recombination within the inverted repeat sequences of the *Chlamydomonas reinhardtii* chloroplast genome produces two orientation isomers. Curr. Genet. 9:233-238.

Altschul, S. F., W. Gish, E. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

André, C. A., A. Levy, and V. Walbot. 1992. Small repeated sequences and the structure of plant mitochondrial genomes. TIG. 8:128-132.

Antamarian, A., R. Coria, J. Ramirez, and D. Gonzalez-Halphen. 1996. The deduced primary structure of subunit I from cytochrome *c* oxidase suggests that the genus *Polytomella* shares a common mitochondrial origin with *Chlamydomonas*. Biochim. Biophys. Acta. 1273:198-202.

Bayen, M., and A. Rode. 1973. The 1.700 DNA of *Chlorella pyrenoidosa*: Heterogeneity and complexity. Plant Sci. Lett. 1:385-389.

Belcour, L., M. Rossignol, F. Koll, C. H. Sellem, and C. Oldani. 1997. Plasticity of the mitochondrial genome in *Podospora*. Polymorphism for 15 optional sequences: group-I, group-II introns, intronic ORFs and an intergenic region. Curr. Genet. 31:308-317.

Bell-Pedersen, D. , S. Quirk, J. Clyman, and M. Belfort. 1990. Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. Nucleic Acids Res. 18:3763-3770.

Bendich, A. J. 1996. Structural analysis of mitochondrial DNA molecules from fungi and plants using moving pictures and pulse-field gel electrophoresis. J. Mol. Biol. 255: 564-588.

Bendich, A. J. 1993. Reaching for the ring: the study of mitochondrial genome structure. Curr. Genet. 24:279-290.

Benslimane, A. A., C. Hartmann, B. Ouenzar, and A. Rode. 1996. Intramolecular recombination of a mitochondrial minicircular plasmid-like DNA of date-palm mediated by a set of short direct-repeat sequences. Curr. Genet. 29:591-594.

Boer, P. H., and M. W. Gray. 1991. Short dispersed repeats localized in spacer regions of *Chlamydomonas reinhardtii* mitochondrial DNA. Curr. Genet. 19:309-312.

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

Boer, P. H., and M. W. Gray. 1988b. Genes encoding a subunit of respiratory NADH dehydrogenase (ND1) and a reverse transcriptase-like protein (RTL) are linked to the ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. EMBO J. 7:3501-3508.

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

Boer, P. H., L. Bonen, R. W. Lee, and M. W. Gray. 1985.
Proc. Natl. Acad. Sci. USA 82:3340-3344.

Boudreau, E., and M. Turmel. 1996. Extensive rearrangements in the chloroplast DNAs of *Chlamydomonas* species featuring multiple dispersed repeats. Mol. Biol. Evol. 13:233-243.

Boudreau, E., and M. Turmel. 1995. Gene rearrangements in *Chlamydomonas* chloroplast DNAs are accounted for by inversions and by the expansion/contraction of the inverted repeat. Plant Mol. Biol. 27:351-364.

Boudreau E., C. Otis, and M. Turmel. 1994. Conserved gene clusters in the highly rearranged chloroplast genomes of *Chlamydomonas moewusii* and *Chlamydomonas reinhardtii*. Plant Mol. Biol. 24:585-602.

Boynton, J. E., N. W. Gillham, S. M. Newman, and E. H. Harris. 1992. Organelle genetics and transformation of *Chlamydomonas*. Pp. 3-64 *in* R. G. Herrmann, ed. Cell Organelles. Springer-Verlag, New York.

Boynton, J. E., E. H. Harris, B. D. Burkhart, and P. M. Lamerson. 1987. Transmission of mitochondrial and chloroplast genomes in crosses of *Chlamydomonas*. Proc. Natl. Acad. Sci. USA 84:2391-2395.

Brennicke, A., and D. A. Clayton. 1981. Nucleotide assignment of alkali-sensitive sites in mouse mitochondrial DNA. J. Biol. Chem. 256:163-165.

Brennicke, A., L. Grohmann, R. Hiesel, V. Knoop, and W. Schuster. 1993. The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. FEBS Lett. 325:140-145.

Buchheim, M. A., and R. L. Chapman. 1992. Phylogeny of *Carteria* (Chlorophyceae) inferred from molecular and organismal data. J. Phycol. 28:362-374.

Buchheim, M. A., J. A. Buchheim, and R. L. Chapman. 1997. Phylogeny of *Chloromonas* (Chlorophyceae): A study of 18S ribosomal RNA gene sequences. *J.* Phycol. 33:286-93.

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

Buchheim, M. A., M. Turmel, E. A. Zimmer, and R. L. Chapman. 1990. Phylogeny of *Chlamydomonas* (Chlorophyta) based on cladistic analysis of nuclear 18S rRNA sequence data. J. Phycol. 26:689-699.

Burger, G., I. Plante, K. M. Lonergan, and M. W. Gray. 1995. The mitochondrial DNA of the amoeboid protozoon, *Acanthamoeba castellanii*: Complete sequence, gene content and genome organization. J. Mol. Biol. 245:522-537.

Burton, M. D., and J. Moore. 1974. The mitochondria of the flagellate, *Polytomella agilis*. J. Ultrastr. Res. 48:414-419.

Bussieres, J., C. Lemieux, R. W. Lee, and M. Turmel. 1996. Optional elements in the chloroplast DNAs of *Chlamydomonas eugametos* and *C. moewusii*: unidirectional gene conversion and co-conversion of adjacent markers in high-viability crosses. Curr. Genet. 30:356-365.

Cardazzo, B., T. Rinaldi, L. Frontali, G. Carignani, and C. Palleschi. 1997. Evolution of mitochondrial genomes in yeast: A study of mitochondrial divergence in two closely related species, *Saccharomyces douglasii* and *Saccharomyces cerevisiae*. Mol. Biol. Evol. 14:200-203.

Chang, D. D., and D. A. Clayton. 1989. Mouse RNAse MRP RNA is encoded by a nuclear gene and contains a decamer sequence complementary to a conserved region of mitochondrial RNA substrate. Cell 56:131-139.

Chapman, D.J., and M. Ragan. 1980. Evolution of biochemical pathways: evidence from comparative biochemistry. Annu. Rev. Plant Physiol. 31:639-678.

Chapman, R. L., and M. A. Buchheim. 1992. Green algae and the evolution of land plants: inferences from nuclear-encoded rRNA gene sequences. BioSystems 28:127-137.

Chapman, R. L., and M. A. Buchheim. 1991. Ribosomal RNA gene sequences: Analysis and significance in the phylogeny and taxonomy of green algae. Crit. Rev. Plant Sci. 10:343-368.

Choquet, Y., M. Goldschmidt-Clermont, J. Girard-Bascou, U. Kück, P. Bennoun, and J-D Rochaix. 1988. Mutant phenotypes support a *trans*-splicing mechanism for the expression of the tripartite *psaA* gene in the *C. reinhardtii* chloroplast. Cell 52:903-913.

Coleman, A. W., and L. J. Goff. 1991. DNA analysis of eukaryotic algal species. J. Phycol.. 27:463-473.

Colleaux, L., M-R. Michel-Wolwertz, R. F. Matagne, and B. Dujon. 1990. The apocytochrome *b* gene of *Chlamydomonas smithii* contains a mobile intron related to both *Saccharomyces* and *Neurospora* introns. Mol. Gen. Genet. 233:288-296.

Conner, W. T., M. D. Thompson, and C. D. Silflow. 1989. Structure of the three beta-tubulin encoding genes of the unicellular alga, *Polytomella agilis*. Gene 84:345-358.

Corliss, J. O. 1990. Endosymbionts of Protozoa. Zoological Science. 7(Suppl.):167-177.

Darlix, J-L., and J-D. Rochaix. 1981. Nucleotide sequence and structure of cytoplasmic 5S RNA and 5.8S RNA of *Chlamydomonas reinhardii*. Nucl. Acids Res. 9:1291-1299.

Davis, E. O., H. S. Thangaraj, P. C. Brooks, and M. J. Colston. 1994. Evidence of selection for protein introns in the RecAs of pathogenic mycobacteria. EMBO J. 13:699-703.

Dayhoff, M. O., and R. M. Schwartz. 1981. Evidence on the origin of eukaryotic mitochondria from protein and nucleic acid sequences. *Ann. NY Acad. Sci.* 361:92-104.

Delahodde A., J. Banroques, A. M. Becam, V. Goguel, J. Perea, R. Schroeder, C. Jacq. 1985. Purification of yeast bi4 mRNA maturase from *Escherichia coli* and from yeast. Nucleic acids binding properties. Pp. 79-88 *in* E. Quagliariello, E. C. Slater, F. Palmieri, C. Saccone, and A. M. Kroon, eds. Achievements and perspectives in mitochondrial research, vol. 2. Elsevier, Amsterdam.

Delwiche, C. F., M. Kuhsel, and J. D. Palmer. 1995. Phylogenetic analysis of *tuf*A sequences indicates a cyanobacterial origin of all plastids. Mol. Phylogenet. Evol. 4:110-128.

Denovan-Wright, E. M., and R. W. Lee. 1995. Evidence that the fragmented ribosomal RNAs of *Chlamydomonas* mitochondria are associated with ribosomes. FEBS Lett. 370:222-226.

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

Denovan-Wright, E. M., and R. W. Lee. 1993. *Chlamydomonas eugametos* mitochondrial genome. Pp. 2.170-2.171 *in* O'Brien SJ, ed. Genetic Maps 6th edit. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Denovan-Wright, E. M., and R. W. Lee. 1992. Comparative analysis of the mitochondrial genomes of *Chlamydomonas eugametos* and *Chlamydomonas moewusii*. Curr. Genet. 21:197-202.

Denovan-Wright, E. M., D. Sankoff, D. F. Spencer, and R. W. Lee. 1996. Evolution of fragmented mitochondrial ribosomal RNA genes in *Chlamydomonas*. J. Mol. Evol. 42:382-391.

Denovan-Wright, E. M., A. M. Nedelcu, and R. W. Lee. 1997. Complete sequence of the mitochondrial DNA of *Chlamydomonas eugametos*. Plant Mol. Biol. (in press)

Devereux, R., A. R. Loeblich III, and G. E. Fox. 1990. Higher plant origins and the phylogeny of green algae. J. Mol. Evol. 31:18-24.

de Zamaroczy, M., G. Faugeron-Fonty, and G. Bernardi. 1983. Excision sequences in the mitochondrial genome of yeast. Gene 21:193-202.

Dieckmann, C. L., and B. Gandy. 1987. Preferential recombination between GC clusters in yeast mitochondrial DNA. EMBO J. 6:4197-4203.

Douglas, S. E. 1994. Chloroplast origins and evolution. Pp. 91-118 *in* D. A. Bryant, ed. The Molecular Biology of Cyanobacteria. Kluwer Academic Publishers, The Netherlands.

Douglas, S. E. 1992. Eukaryote-eukaryote endosymbioses: insights from studies of a cryptomonad alga. BioSystems. 28:52-68.

Dron, M., M. Rahire, and J-D. Rochaix. 1982. Sequence of the chloroplast DNA region of *Chlamydomonas reinhardtii* containing the gene of the large subunit of ribulose bisphosphate carboxylase and parts of its flanking genes. J. Mol. Biol. 162:775-793.

Dujon, B. 1989. Group I introns as mobile genetic elements: facts and mechanistic speculations -a review. Gene 82:91-113.

Dürrenberger, F., and J-D. Rochaix. 1991. Chloroplast ribosomal intron of *Chlamydomonas reinhardtii*: *in vitro* self-splicing, DNA endonuclease activity and *in vivo* mobility. EMBO J. 10:3495-3501.

Dürrenberger, F., A. J. Thompson, D. L. Herrin, and J-D. Rochaix. 1996. Double strand break-induced recombination in *Chlamydomonas reinhardtii* chloroplasts. Nucl. Acids Res. 24:3323-3231.

Erickson, J. M., M. Rahire, and J-D. Rochaix. 1984. *Chlamydomonas reinhardtii* gene for the 32,000 mol. wt. protein of photosystem II contains four large introns and is located entirely within the chloroplast inverted repeat. EMBO J. 3:2753-2762.

Ettl, H. 1976. Die Gattung *Chlamydomonas* Ehrenberg *Beih.* Nova Hedwigia 49:1-1122.

Ettl, H., and O. Moestrup. 1980. Light and electron microscopical studies on *Hafniomonas* gen. nov. (Chlorophyceae, Volvocales), a genus resembling *Pyramimonas* (Prasinophyceae). Pl. Syst. Evol. 135:177-210.

Fauron, C., M. Gasper, Y. Gao, and B. Moore. 1995. The maize mitochondrial genome: dynamic, yet functional. TIG 11:228-235.

Faßbender, S., K-H. Brühl, M. Ciriacy, and U. Kück. 1994. Reverse transcriptase activity of an intron encoded polypeptide. EMBO J. 13:2075-2083.

Fan, W-H., M. A. Woelfle, and G. Mosig. 1995. Two copies of a DNA element, 'Wendy', in the chloroplast chromosome of *Chlamydomonas reinhardtii* between rearranged gene clusters. Plant Mol. Biol. 29:63-80.

Fangman, W., and B. Dujon. 1984. Yeast mitochondrial genomes consisting of only A·T base pairs replicate and exhibit suppressiveness. Proc. Natl. Acad. Sci. USA. 81:7156-7160.

Feagin, J., E. Werner, M. J. Gardner, D. H. Williamson, and R. J .M. Wilson. 1992. Homologies between the contiguous and fragmented rRNAs of the two *Plasmodium falciparum* extrachromosomal DNAs are limited to core sequences. Nucleic Acids Res. 20:879-887.

Floyd, G. L., H. J. Hoops and J. A. Swanson. 1980. Fine structure of the zoospore of *Ulothrix belkae* with emphasis on the flagellar apparatus. Protoplasma. 104:17-31.

Fong, S. E., and S. J. Surzycki. 1992. Chloroplast RNA polymerase genes of *Chlamydomonas reinhardtii* exhibit an unusual structure and arrangement. Curr. Genet. 21:485-497.

Friedl, T. 1995. Inferring taxonomic positions and testing genus level assignments in coccoid green lichen algae: A phylogenetic analysis of 18S ribosomal RNA sequences from *Dictyochloropsis reticulata* and from members of the genus *Myrmecia* (Chlorophyta, Trebouxiophyceae cl. nov.). J. Phycol. 31:632-639.

Friedl, T., and C. Zeltner. 1994. Assessing relationships of some coccoid green lichen algae and the Microtamniales (Chlorophyta) with 18S ribosomal RNA gene sequence

comparisons. J. Phycol. 30:500-506.

Gaillard, C., F. Strauss, and G. Bernardi. 1980. Excision sequences in the mitochondrial genome of yeast. Nature 283:218-220.

Gauthier, A., M. Turmel, and C. Lemieux. 1991. A group I intron in the chloroplast large subunit rRNA gene of *Chlamydomonas eugametos* encodes a double-strand endonuclease that cleaves the homing site of this intron. Curr. Genet. 19:43-47.

Gelvin, S. B., and S. H. Howell. 1979. Small repeated sequences in the chloroplast genome of *Chlamydomonas reinhardtii*. Mol. Gen. Genet. 173:315-322.

Gibbs, S. P. 1981. Chloroplasts of some groups may have evolved from endosymbiotic eukaryotic green algae. Ann. NY Acad. Sci. 361:193-207.

Gibbs, S. P. 1978. The chloroplasts of *Euglena gracilis* may have evolved from symbiotic green algae. Can. J. Bot. 56:2883-28839.

Gillham, N. W. 1994. Organelle Genes and Genomes. Oxford University Press, New York, Oxford.

Gjetvaj, B., D. I. Cook, and E. Zouros. 1992. Repeated sequences and large-scale variation of mitochondrial DNA: A common feature among scallops (Bivalvia:Pectinidae). Mol. Biol. Evol. 9:106-24.

Goguel, V., A. Bailone, R. Devoret, and C. Jacq. 1989. The bi4 RNA mitochondrial maturase of *Saccharomyces cerevisiae* can stimulate intra-chromosomal recombination in *Escherichia coli*. Mol. Gen. Genet. 216:70-74.

Goldschmidt-Clermont, M., Y. Choquet, J. Girard-Bascou, F. Michel, M. Schirmer-Rahire, and J-D. Rochaix. 1991. A small chloroplast RNA may be required for trans-splicing in *Chlamydomonas reinhardtii*. Cell 65:135-143.

Goursot, R., M. Mangin., and G. Bernardi. 1982. Surrogate origins of replication in the mitochondrial genome of *ori°* petite mutants of yeast. EMBO J. 1: 705-711.

Graham, L. E. 1982. Cytology, ultrastructure, taxonomy, and phylogenetic relationships of Great Lakes filamentous algae. J. Great Lakes 8:3-9.

Grant, D., and K-S. Chiang. 1980. Physical mapping and characterization of *Chlamydomonas* mitochondrial molecules: their unique ends, sequence homogeneity, and conservation. Plasmid. 4:82-96.

Gray, M. W. 1995. Mitochondrial evolution. Pp. 635-659 *in* C. S. Levings III, and I.

K. Vasil, eds. The molecular Biology of Plant Mitochondria. Kluwer Academic Publishers, The Netherlands.

Gray, M. W. 1993. Origin and evolution of organelle genomes. Curr. Opin. Genet. Dev. 3:884-890.

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

Gray, M. W. 1989. Origin and Evolution of Mitochondrial DNA. Annu. Rev. Cell Biol. 5:25-50.

Gray, M. W., and D. F. Spencer. 1996. Organellar evolution. Pp.109-126 *in* D. McL. Roberts, P. Sharp, G. Alderson, and M. Collins, eds. Evolution of Microbial Life. Society for General Microbiology Symposium 54. Cambridge University Press.

Gray, M. W., and M. N. Schnare. 1995. Evolution of rRNA gene organization. Pp. 49-69 *in* R. A. Zimmermann, A. E. Dalhberg, eds. Ribosomal RNA: structure, evolution, processing and function in protein biosynthesis. CRC Press, Boca Raton, Florida.

Gray, M. W., and P. H. Boer. 1988. Organization and expression of algal (*Chlamydomonas reinhardtii*) mitochondrial DNA. Phil. Trans. Royal Soc. London B. 319:135-47.

Gray, M. W., R. Cedergren, Y. Abel, and D. Sankoff. 1989. On the evolutionary origin of the plant mitochondrion and its genome. Proc. Natl. Acad. Sci. USA 86:2267-2271.

Guillard, R. L. 1973. Methods for microflagellates and nanoplankton. Pp. 69-85 *in* J. R. Stein, Ed. Handbook of Phycological Methods. Culture Methods and Growth Measurements. Cambridge University Press, New York.

Gutell, R. R. 1992. Evolutionary characteristics of 16S and 23S rRNA structures. Pp. 243-309 *in* H. Hartman and K. Matsuno, eds. The Origin and Evolution of the Cell. World Scientific Publishing Co. Pte. Ltd., Singapore.

Gutell, R. R., M. N. Schnare, and M. W. Gray. 1992. A compilation of large subunit (23S and 23S-like) ribosomal RNA structures. Nucl. Acids Res. 20(Suppl.):2095-2109.

Hardy, C. M., and G. D. Clark-Walker. 1991. Nucleotide sequence of the *COX1* gene in *Kluyveromyces lactis* mitochondrial DNA: evidence for recent horizontal transfer of a group II intron. Curr. Genet. 20:99-114.

Hartman, C., H. Recipon, M-F. Jubier, C. Valon, E. Delcher-Besin, Y. Henry, J. de

Buyser, B. Lejeune, and A. Rode. 1994. Mitochondrial DNA variability detected in a single wheat regenerant involves a rare recombination event across a short repeat. Curr. Genet. 25:456-464.

Heinonen, T. Y. K., M. N. Schnare, P. S. Young, and M. W. Gray. 1987. Rearranged coding segments, separated by a transfer RNA gene, specify the two parts of a discontinuous large subunit ribosomal RNA in *Tetrahymena pyriformis* mitochondria. J. Biol. Chem. 262:2879-2887.

Henke, R. M., R. A. Butow, and P. S. Perlman. 1995. Maturase and endonuclease functions depend on separate domains of the bifunctional protein encoded by the group I intron aI4α of yeast mitochondrial DNA. EMBO J. 14:5094-5099.

Herrin, D. L., and G. W. Schmidt. 1988. Rapid, reversible staining of Northern blots prior to hybridization. BioTechniques 6:196.

Higgins, D. G., A. J. Bleasby, and R. Fuchs. 1992. CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. 8:189-191.

Hiratsuka, J., H. Shimada, R. Whittier et al. (16 co-authors). 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. 217:185-194.

Hixson, J. E., T. W. Wong, and D. A. Clayton. 1986. Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. J. Biol. Chem. 261:2384-2390.

Hori, H., and S. Osawa. 1987. Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. Mol. Biol. Evol. 4:445-472.

Hori, H., B-L. Lim, and S. Osawa. 1985. Evolution of green plants as deduced from 5S rRNA sequences. Proc. Natl. Acad. Sci. *USA*. 82:820-823.

Howe, C. J., R. F. Barker, C. M. Bowman, and T. A. Dyer. 1988. Common features of three inversions in wheat chloroplast DNA. Curr. Genet. 13:343-349.

Huang, C., S. Wang, L. Chen, C. Lemieux, C. Otis, M. Turmel, and X-Q Liu. 1994. The *Chlamydomonas* chloroplast *clpP* gene contains translated large insertion sequences and is essential for cell growth. Mol. Gen. Genet. 244:151-159.

Hudspeth, M. E. S., R. D. Vincent, P. Perlman, D. S. Shumar, L. O. Treisman, and L. Grossman. 1984. Expandable *var1* gene of yeast mitochondrial DNA: In-frame insertions can explain the strain-specific protein size polymorphisms. Proc. Natl. Acad. Sci. USA.

81:3148-3152.

Hunt, D. H., and B. C. Hyman. 1997. Animal mitochondrial DNA recombination. Nature 387:247.

Hur, M., and R. B. Waring. 1995. Two group I introns with a C•C basepair at the 5'-splice-site instead of the very highly conserved U•G basepair: is selection post-transcriptional? Nucleic Acids Res. 23:4466-4470.

Hüttenhofer, A., H. Sakai, and B. Weiss-Brummer. 1988. Site-specific AT-clusters insertions in the mitochondrial 25S rRNA genes of the yeast *S. cerevisiae*. Nucleic Acids Res. 16:8665-8674.

Jacquier, A., and B. Dujon. 1985. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell 41:383-394.

Jamiet-Vierny, C., J. Boulay, and J.-F. Briand. 1997. Intramolecular cross-overs generate deleted mitochondrial DNA molecules in *Podospora anserina*. Curr. Genet. 31:162-170.

Jupe, E. R., R. L. Chapman, and E. A. Zimmer. 1988. Nuclear ribosomal RNA genes and algal phylogeny - the *Chlamydomonas* example. BioSystems. 21:223-230.

Kairo, A., A. H. Failamb, E. Gobright, and V. Nene. 1994. A 7.1 kb linear DNA molecule of *Theileria parva* has scrambled rRNA sequences and open reading frames for mitochondrial encoded proteins. EMBO J. 21:223-230.

Kantz, T. S., E. C. Theriot, E. A. Zimmer, and R. L. Chapman. 1990. The Pleurastrophyceae and Micromonadophyceae: A cladistic analysis of nuclear rRNA sequence data. J. Phycol. 26:711-721.

Kapoor, M., T. Nagai, T. Wakasugi, K. Yoshinaga, and M. Sugiura. 1997. Organization of chloroplast ribosomal RN genes and in vitro self-splicing activity of the large subunit rRNA intron from the green alga *Chlorella vulgaris* C-27. Curr. Genet. 31:503-510.

Kawata, M., T. Harada, Y. Shimamoto, K. Oono, and F. Takaiwa. 1997. Short inverted repeats function as hotspots of intermolecular recombination giving rise to oligomers of deleted plastids DNAs (ptDNAs). Curr. Genet. 31:179-184.

Kessler, U., and K. Zetsche. 1995. Physical map and gene organization of the mitochondrial genome from the unicellular green alga *Platymonas* (*Tetraselmis*) *subcordiformis* (*Prasynophyceae*). Plant Mol. Biol. 29:1081-1086.

Koll, F., J. Boulay, L. Belcour, and Y. d'Aubenton-Carafa. 1996. Contribution of ultra-short invasive elements to the evolution of the mitochondrial genome in the genus

Podospora. Nucleic Acids Res. 24:1734-141.

Kössel, H., E. Natt, G. Strittmatter, E. Fritzsche, A. Gozdzicka-Jozefiak, and D. Przybyl. 1985. Structure and expression of rRNA operons from plastids of higher plants. Pp. 183-199 *in* L. van Vloten-Doting, G. S. P. Groot, and T. C. Hall, eds. Molecular form and function of the plant genome. Plenum Publishing Corp, New York.

Kotylak, L., J. Lazowska, D. C. Hawthorne, and P. P. Slonimski. 1985. Intron encoded proteins of mitochondria: key elements of gene expression and genomic evolution. Pp. 1-20 *in* E. Quagliariello, E. C. Slater, F. Palmieri, C. Saccone, A. M. Kroon, eds. Vol. 2. elsevier, Amsterdam.

Kroymann, J., and K. Zetsche. 1997. The apocytochrome-b gene in *Chlorogonium elongatum* (*Chlamydomonadaceae*): an intronic GIY-YIG ORF in green algal mitochondria. Curr. Genet. 31:414-418.

Kück, U. 1989. The intron of a plastid gene from a green alga contains an open reading frame for a reverse transcriptase-like enzyme. Mol. Gen. Genet. 218:257-265.

Kück, U., I. Godenhardt, and U. Schmidt. 1990. A self-splicing group II intron in the mitochondrial large subunit rRNA (LSU rRNA) gene of the eukaryotic alga *Scenedesmus obliquus*. Nucl. Acids Res. 18:2691-2697.

Kück, U., Y. Choquet, M. Schneider, M. Dron, and P. Bennoun. 1987. Structural and transcription analysis of two homologous genes for the P700 chlorophyll *a*-apoproteins in *Chlamydomonas reinhardtii*: Evidence for *in vivo trans*-splicing. EMBO J. 6:257-265.

Lambowitz, A. M., and M. Belfort. 1993. Introns as mobile genetic elements. Annu. Rev. Biochem. 62:587-622.

Lang, B. F., G. Burger, C. J. O'Kelly, R. Cedergren, G. B. Golding, C. Lemieux, D. Sankoff, M. Turmel, and M. W. Gray. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature 387:493-497.

Lang, B. F., L. G. Goff, and M. W. Gray. 1996. A 5S rRNA gene is present in the mitochondrial genome of the protist *Reclinomonas americana* but is absent from the red algal mitochondrial DNA. J. Mol. Biol. 61:607-613.

Lee, R. W., C. Dumas, C. Lemieux, and M. Turmel. 1991. Cloning and characterization of the *Chlamydomonas moewusii* mitochondrial genome. Mol. Gen. Genet. 231:53-58.

Lembi, C. A. 1975. The fine structure of the flagellar apparatus of *Carteria*. J. Phycol. 11:1-9.

Lemieux, C., and R. W. Lee. 1987. Non-reciprocal recombination between alleles of the chloroplast 23S rRNA gene in interspecific *Chlamydomonas* crosses. Proc. Natl. Acad. Sci. USA 84:4166-4170.

Lemieux, B., and C. Lemieux. 1985. Extensive sequence rearrangements in the chloroplast genomes of the green algae *Chlamydomonas eugametos* and *Chlamydomonas reinhardtii*. Curr. Genet. 10:213-219.

Lemieux, B., M. Turmel, and C. Lemieux. 1985. Chloroplast DNA variation in *Chlamydomonas* and its potential application to the systematics of this genus. BioSystems 18:293-298.

Lemieux, C., M. Turmel, and R. W. Lee. 1980. Characterization of chloroplast DNA in *Chlamydomonas eugametos* and *Chlamydomonas moewusii* and its inheritance in hybrid progeny. Curr. Genet. 2:139-147.

Lockhart, P. J., D. Penny, M. D. Hendy, C. J. Howe, T. J. Beanland, and A. W. Larkum. 1992. Controversy on chloroplast origins. FEBS Lett. 301:127-131.

Lonergan, K. M. 1993. The ribosomal RNA gene region of *Acanthamoeba castellanii* mitochondrial DNA: Organization, mode of expression and evolution. PhD thesis, Dalhousie University, Halifax, Nova Scotia, Canada, 227 pp.

Lonergan, K. M., and M. W. Gray. 1994. The ribosomal RNA gene region in *Acanthamoeba castellanii* mitochondrial DNA. A case of evolutionary transfer of introns between mitochondrial and plastids? J. Mol. Biol. 239:476-499.

Ma, D-P., Y-T. King, Y. Kim, and W. S. Luckett Jr. 1992. The group I intron of apocytochrome *b* from *Chlamydomonas smithii* encodes a site-specific endonuclease. Plant Mol. Biol. 18:1001-1004.

Macey, J. R., A. Larson, N. B. Ananjeva, Z. Fang, and T. J. Papenfuss. 1997. Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. Mol. Biol. Evol. 14:91-104.

Maddison, W.P., and D. R. Maddison. 1992. Tracing characters evolution. Pp. 237-263 *in* W. P. Maddison, and D. R. Maddison, eds. MacClade. Analysis of Phylogeny and Character Evolution. Sinauer Associates, Inc., Sunderland, Massachusetts.

Malnoë, P., J-D. Rochaix, N. H. Chua, and P. F. Spahr. 1979. Characterization of the gene and messenger RNA of the large subunit of ribulose-1,5-disphosphate carboxylase in *Chlamydomonas*. J. Mol. Biol. 133:417-434 .

Manhart, J. R., R. W. Hoshaw, and J-D. Palmer. 1990. Unique chloroplast genome in

Spirogyra maxima (Chlorophyta) revealed by physical and gene mapping. J. Phycol. 26:490-494.

Manhart, J.R., K. Kelly, B. S. Dudock, and J. D. Palmer. 1989. Unusual characteristic of *Codium fragile* chloroplast DNA revealed by physical and gene mapping. Mol. Gen. Genet. 216:417-421.

Manna, F., D. R. Massardo, L. Del Giudice, A. Buonocore, A. G. Nappo, P. Alifano, B. Schäfer, and K. Wolf. 1991. The mitochondrial genome of *Schizosaccharomyces pombe*. Stimulation of intra-chromosomal recombination in *Escherichia coli* by the gene product of the first *cox1* intron. Curr. Genet. 19:295-299.

Margulis, L. 1981. Symbiosis in Cell Evolution. W. H. Freeman. San Francisco.

Martin, W., C. C. Somerville, and S. Loiseaux-De-Goer. 1992. Molecular phylogenies of plastid origin and algal evolution. J. Mol. Evol. 35:385-404.

Matagne, R. F., D. Rongvaux, and R. Loppes. 1988. Transmission of mitochondrial DNA in crosses involving diploid gametes homozygous or heterozygous for mating type locus in *Chlamydomonas*. Mol. Gen. Genet. 214:257-262.

Mattox, K. R., and K. D. Stewart. 1984. Classification of the green algae: A concept based on comparative cytology. Pp. 29-72 *in* D. E. G. Irvine and D. M. John, eds. Systematics of the Green Algae. Academic Press, London and Orlando.

McCourt, R. M. 1995. Green algal phylogeny. Trends Ecol. Evol. 10:159-163.

McCracken, D., M. J. Nadakavukaren, and J. R. Cain. 1980. A biochemical and ultrastructural evaluation of the taxonomic position of *Glaucosphaera vacuolata* Korsch. New Phytol. 86:39-44.

McFadden, G., and P. Gilson. 1995. Something borrowed, something green: lateral transfer of chloroplasts by secondary endosymbiosis. Tree 10:12-18.

Michaelis, G., C. Vahrenholz, and E. Pratje. 1990. Mitochondrial DNA of *Chlamydomonas reinhardtii*: The gene for apocytochrome *b* and the complete functional map of the 15.8 kb DNA. Mol. Gen. Genet. 233:211-216.

Michel, F. , and B. F. Lang. 1985. Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. Nature 316:641-643.

Mita, S., R. Rizzuto, C. T. Moraes, S. Shanske, E. Arnaudo, G. M. Fabrizi, Y. Koga, S. DiMauro, and E. A. Schon. 1990. Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA. Nucleic Acids Res. 18:561-

567.

Moenne, A., D. Begu, and X. Jordana. 1996. A reverse transcriptase activity in potato mitochondria. Plant Mol. Biol. 31:365-372.

Moore, L. J. 1990. The nature and extent of intraspecific variation in chloroplast DNAs of sexually isolated populations of *Pandorina morum* Bory. PhD Thesis, Brown University, Providence, Rhode Island.

Moore, L. J., and A. W. Coleman. 1989. The linear 20 kb mitochondrial genome of *Pandorina morum* (Volvocaceae, Chlorophyta). Plant Mol. Biol. 13:459-465.

Morden, C. W, C. F. Delwiche, M. Kushel, and J. D. Palmer. 1992. Gene phylogenies and the endosymbiotic origin of plastids. ByoSystems 29:75-90.

Moritz, C. 1991. Evolutionary dynamics of mitochondrial DNA duplications in partenogenetic geckos, *Heteronotia binoei*. Genetics 129:221-230.

Moritz, C., and W. M. Brown. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizard mitochondrial DNA. Science 233:1425-1427.

Mueller, M. W., M. Allmaler, R. Eskes, and J. Schweyen. 1993. Transposition of group II intron *aI1* in yeast and invasion of mitochondrial genes at new locations. Nature 366:174-176.

Nedelcu, A. M. 1997. Contrasting mitochondrial genome organizations and sequence affiliations among green algae: Potential factors, mechanisms, and evolutionary scenarios. J. Phycol. (in press).

Nedelcu, A. M. 1997. Fragmented and scrambled mitochondrial ribosomal RNA coding regions among green algae: A model for their origin and evolution. Mol. Biol. Evol. 14:506-517.

Nedelcu, A. M., and R. W. Lee. 1997. Modes and tempos of mitochondrial and chloroplast genomes evolution in *Chlamydomonas*: A comparative analysis. (in press) *in* J.-D. Rochaix, ed. Molecular Biology of *Chlamydomonas*: Chloroplasts and Mitochondria. Kluwer Publishers, The Netherlands.

Nedelcu, A. M, and R. W. Lee. 1997. Short repetitive sequences in green algal mitochondrial genomes: Potential roles in mitochondrial genome evolution. Mol. Biol. Evol. (submitted).

Nedelcu, A. M., D. F. Spencer, E. M. Denovan-Wright, and R. W. Lee. 1996.

Discontinuous mitochondrial and chloroplast large subunit ribosomal RNAs among green algae: Phylogenetic implications. J. Phycol. 32:103-111.

Obar, R., and J. Green. 1985. Molecular archaeology of the mitochondrial genome. J. Mol. Evol. 22:243-251.

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

Oh-Hama, T., and E. Hase. 1980. Formation of protochlorophyll(ide) in wild type mutant C-2A' of *Scenedesmus obliquus*. Plant Cell Physiol. 21:1263-1272.

O'Kelly, C. J. 1992. Flagellar apparatus architecture and the phylogeny of "green" algae: chlorophytes, euglenoids, glaucophytes. Pp. 315-345 *in* D. Menzel, ed. Cytoskeleton of the Algae. CRC Press, Boca Raton, Florida.

O'Kelly, C. J., and G. L. Floyd. 1984. Flagellar apparatus absolute orientations and the phylogeny of the green algae. BioSystems 16:227-251.

O'Kelly, C. J., S. Watanabe, and G. L. Floyd. 1994. Ultrastructure and phylogenetic relationships of Chaetopeltidales ord. nov. (Chlorophyta, Chlorophyceae). J. Phycol. 30:118-128.

Orr-Weaver, T., J. Szostak, and R. Rothstein. 1981. Yeast transformation: A model system for the study of recombination. Proc. Natl. Acad. Sci. USA. 78:6354-6358.

Palmer, J. D. 1991. Plastid chromosomes: Structure and evolution. vol. 7A. Pp 5-33. *in* L. Bogorad, and I. K. Vasil, eds. The Molecular Biology of Plastids. Cell Culture and Somatic Cell Genetics of Plants. Academic Press, San Diego.

Palmer, J. D. 1990. Contrasting modes and tempos of genome evolution in land plant organelles. Trends Genet. 6:115-120.

Palmer, J.D. 1987. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. Am. Nat. 130:S6-S29.

Palmer, J. D., and C. R. Shields. 1984. Tripartite structure of the *Brassica campestris* mitochondrial genome. Nature 307:437-440.

Palmer, J. D., J. E. Boynton, N. W. Gillham, and E. H. Harri. 1985. Evolution and recombination of the large inverted repeat in *Chlamydomonas* chloroplast DNA. Pp. 269-278 *in* K. E. Steinback, S. Bonitz, C. J. Arntzen, and L. Bogorad, eds. Molecular Biology

of the Photosynthetic Apparatus. Cold Spring Harbor Laboratory, New York.

Ragan, M. A., and D. J. Chapman. 1978. A Biochemical Phylogeny of the Protists. Academic Press Inc. New York.

Ragnini, A., and H. Fukuhara. 1988. Mitochondrial DNA of the yeast *Kluyveromyces*: guanine-cytosine rich sequence clusters. Nucleic Acids Res. 16:8433-8442.

Rochaix, J-D. 1995. *Chlamydomonas reinhardtii* as the photosynthetic yeast. Annu Rev. Genet. 29:209-230.

Rochaix, J-D. 1978. Restriction endonuclease map of the chloroplast DNA of *Chlamydomonas reinhardtii*. J. Mol. Biol. 126:597-617.

Rochaix, J-D., and P. Malnoë. 1982. Use of DNA-RNA hybridizations for locating chloroplast genes and for estimating the size and abundance of chloroplast DNA transcripts. Pp.377-490 *in* M. Edelman, R.B. Hallick, and N-S Chua, eds. Methods in Chloroplast Molecular Biology. Elsevier Biomedical Press, Amsterdam.

Rochaix, J-D., M. Rahire, and F. Michel. 1985. The chloroplast ribosomal intron of *Chlamydomonas reinhardtii* codes for a polypeptide related to mitochondrial maturases. Nucl. Acids Res. 13:975-984.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sankoff, D., G. Leduc, N. Antoine, B. Paquin, B. F. Lang, and R. Cedergren. 1992. Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. Proc. Natl. Acad. Sci. USA 89:6575-6579.

Schardl, C. L., D. M. Lonsdale, D. R. Pring, and K. R. Rose. 1984. Linearization of maize mitochondrial chromosomes by recombination with linear episomes. Nature 310:292-296.

Schnare, M., N., and M. W. Gray. 1990. Sixteen discrete RNA components in the cytoplasmic ribosome of *Euglena gracilis*. J. Mol. Biol. 251:73-83.

Schon, E. A., R. Rizzuto, C. T. Moraes, H. Nakase, M. Zevian, and S. DiMauro. 1989. A direct repeat is a hotspot for large-scale deletion of human mitochondrial DNA. Science 244:346-349.

Schopf, J. W. 1970. Precambrian micro-organisms and evolutionary events prior to the origin of vascular plants. Biol. Rev. Camb. Philos. Soc. 45:319-352.

Schuster, W., and A. Brennicke. 1994. The plant mitochondrial genome: physical structure, information content, RNA editing, and gene migration to the nucleus. Annu. Rev. Plant. Physiol. Plant. Mol. Biol. 45:61-78.

Schuster, W., and A. Brennicke. 1987. Plastid, nuclear and reverse transcriptase sequences in the mitochondrial genome of *Oenothera*: is genetic information transfer between organelles via RNA? EMBO J. 6:2857-2863.

Sellem, C. H., and L. Belcour. 1996. Intron open reading frames as mobile elements and evolution of a group I intron. Mol. Biol. Evol. 14:518-526.

Sellem, C. H., G. Leceller, and L. Belcour. 1993. Transposition of a group II intron. Nature 366:176-178.

Shimada, H., and M. Sugiura. 1991. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucl. Acids Res. 19:983-995.

Shimada, H., and M. Sugiura. 1989. Pseudogenes and short repeated sequences in the rice chloroplast genome. Curr. Genet. 16:293-301.

Skelly, P. J., C. M. Hardy, and G. D. Clark-Walker. 1991. A mobile group II intron of a naturally occurring rearranged mitochondrial genome in *Kluyveromyces lactis*. Curr. Genet. 20:115-120.

Sor, F., and H. Fukuhara. 1983. Unequal excision of complementary strands is involved in the generation of palindromic repetitions of rho$^-$ mitochondrial DNA in yeast. Cell 32:391-396.

Sor, F., and H. Fukuhara. 1982. Nature of an inserted sequence in the mitochondrial gene coding for the 15S ribosomal RNA of yeast. Nucleic Acids Res. 10:1626-1633.

Sprinzle, M., C. Steegborn, F. Hübel, and S. Steinberg. 1996. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. 24:68-72.

Starr, R.C., and J. A. Zeikus. 1993. UTEX- The Culture Collection of Algae at the University of Texas at Austin. J. Phycol. (Suppl). 29:1-106.

Steinkötter, J., D. Bhattacharya, I. Semmelroth, C. Bibeau, and M. Melkonian. 1994. Prasinophytes form independent lineages within the Chlorophyta: Evidence from ribosomal RNA sequence comparisons. J. Phycol. 30:340-345.

Stewart, K. D., and K. R. Mattox. 1984. The case for a polyphyletic origin of mitochondria: Morphological and molecular comparisons. J. Mol. Evol. 21:54-57.

Stewart, K. D., and K. R. Mattox. 1980. Phylogeny of Phytoflagellates. Pp. 433-462 *in* E. R. Cox, ed. Phytophlagellates. Vol 2. Elsevier/North-Holland, New York .

Sugiura, M. 1989. The chloroplast chromosomes in land plants. Annu. Rev. Cell Biol. 5:51-70.

Szczepanek T., and J. Lazowska. 1996. Replacement of two non-adjacent aminoacids in the *S. cerevisiae* bi2 intron-encoded maturase is sufficient to gain a homing endonuclease activity. EMBO J. 15:3758-3767.

Takaiwa, F., and M. Sugiura. 1982. Nucleotide sequence of the 16S--23S spacer region in an rRNA gene cluster from tobacco chloroplast DNA. Nucl. Acids Res. 10:2665-2676.

Tappan, H. 1980. The Paleology of Plant Protists. Freeman WH and Co, San Francisco.

Tsai, C-H., and S. H. Strauss. 1989. Dispersed repetitive sequences in the chloroplast genome of Douglas-fir. Curr. Genet. 11:543-552.

Turker, M. S., J. M. Domenico, and D. J. Cummings. 1987. Excision-amplification of mitochondrial DNA during senescence in *Podospora anserina*. A potential role for a 11 base-pair consensus sequence in the excision process. J. Mol. Biol. 198:171-185.

Turmel. M., Y. Choquet, M. Goldschmidt-Clermont, J-D. Rochaix, C. Otis, and C. Lemieux. 1995a. The *trans*-spliced intron 1 in the *psa*A gene of the *Chlamydomonas* chloroplast: a comparative analysis. Curr. Genet. 27:270-279.

Turmel, M., V. Côté, C. Otis, J-P. Mercier, M. W. Gray, K. M. Lonergan, and C. Lemieux. 1995b. Evolutionary transfer of ORF-containing group I introns between different subcellular compartments (chloroplast and mitochondrion). Mol. Biol. Evol. 12:533-545.

Turmel, M., R. R. Gutell, J-P. Mercier, C. Otis, and C. Lemieux. 1993. Analysis of the chloroplast large subunit ribosomal RNA gene from 17 *Chlamydomonas* taxa. Three internal transcribed spacers and 12 group I intron insertion sites. J. Mol. Biol. 232:446-467.

Turmel, M., E. Boudreau, J. Boulanger, J-P. Mercier, C. Otis, and C. Lemieux. 1991. Chloroplast DNA evolution and phylogenetic relationships in *Chlamydomonas*. Pp. 816-827 *in* E. C. Dudley, ed. The Unity of Evolutionary Biology, Proc. ICSEB IV. Dioscorides Press, Portland, Oregon.

Turmel, M., J. Boulanger., M. N. Schnare, M. W. Gray, and C. Lemieux. 1991. Six groups I introns and three internal transcribed spacers in the chloroplast large subunit ribosomal RNA gene of the green alga *Chlamydomonas eugametos*. J. Mol. Biol.

218:293-311.

Turmel, M., B. Lemieux, and C. Lemieux. 1988. The chloroplast genome of the green alga *Chlamydomonas moewusii*: Localization of protein-coding genes and transcriptionally active regions. Mol. Gen. Genet. 214:412-419.

Turmel, M., G. Bellemare, and C. Lemieux. 1987. Physical mapping of differences between the chloroplast DNAs of the interfertile algae *Chlamydomonas eugametos* and *Chlamydomonas moewusii*. Curr. Genet. 11:543-552.

Vahrenholz, C., G. Riemen, E. Pratje, B. Dujon, and G. Michaelis. 1993. Mitochondrial DNA of *Chlamydomonas reinhardtii*: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication. Curr. Genet. 24:241-247.

Vaidya, A. B., R. Akella, and K. Suplick. 1989. Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6 kilobase-pair DNA of a malarial parasite. Mol. Biochem. Parasitol. 35:97-108.

Van de Peer Y, J-M. Neefs, and R. De Watcher. 1990. Small ribosomal subunit RNA sequences, evolutionary relationships among different life forms, and mitochondrial origins. J. Mol. Evol. 30:463-476.

Waddle, J. A., A. M. Schuster, K. W. Lee, and R. H. Meints. 1990. The mitochondrial genome of an exosymbiotic *Chlorella*-like green alga. Plant. Mol. Biol. 14:187-195.

Watanabe, S., and G. L. Floyd. 1989. Ultrastructure of the quadriflagellate zoospores of the filamentous green algae *Chaetophora incrassata* and *Pseudoschizomeris caudata* (Chaetophorales, Chlorophyceae) with emphasis on the flagellar apparatus. Bot. Mag. Tokyo. 102:533-546.

Weiller, G., C. M. E. Schueller, and R. J. Schweyen. 1989. Putative target sites for mobile G+C rich clusters in yeast mitochondrial DNA: Single elements and tandem arrays. Mol. Gen. Genet. 218:272-283.

Wilcox, L. W., L. A. Lewis, P. A. Fuerst, and G. L. Floyd. 1992. Assessing the relationships of autosporic and zoosporic chlorococcalean green algae with 18S rDNA sequence data. J. Phycol. 28:381-386.

Wolff, G., and U. Kück. 1993. Organization and coding capacity of mitochondrial genomes of algae. Pp. 101-113 *in* A. Brennicke, and U. Kück, eds. Plant Mitochondria with Emphasis on RNA Editing and Cytoplasmic Male Sterility. VCH Verlagsgesellschaft, Weinheim.

Wolff, G., and U. Kück. 1990. The structural analysis of the mitochondrial SSU rRNA

implies a close phylogenetic relationship between mitochondria from plants and from the heterotrophic alga *Prototheca wickerhamii*. Curr. Genet. 17:347-351.

Wolff, G., I. Plante, B. F. Lang, U. Kück, and G. Burger. 1994. Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii*. J. Mol. Biol. 237:75-86.

Wolff, G., G. Burger, B. F. Lang, and U. Kück. 1993. Mitochondrial genes in the colourless alga *Prototheca wickerhamii* resemble plant genes in their exons but fungal genes in their introns. Nucl. Acids. Res. 21:719-726.

Wolstenholme, D. R., and C. M-R. Fauron. 1995. Mitochondrial genome organization. Pp 1-59 *in* C. S. Levings III, and I. K. Vasil, eds. The molecular biology of plant mitochondria. Kluwer Academic Publishers, The Netherlands.

Wong, T. W., and D. A. Clayton. 1985. In vitro replication of human mitochondrial DNA: Accurate initiation at the origin of light-strand synthesis. Cell 42:951-958.

Woodson, S. A., and T. R. Cech. 1989. Reverse self-splicing of the *Tetrahymena* group I intron: implication for the directionality of splicing and for intron transposition. Cell 57:335-345.

Yamada, T. 1991. Repetitive sequence-mediated rearrangements in *Chlorella ellipsoidea* chloroplast DNA: completion of nucleotide sequence of the large inverted repeat. Curr. Genet. 19:139-147.

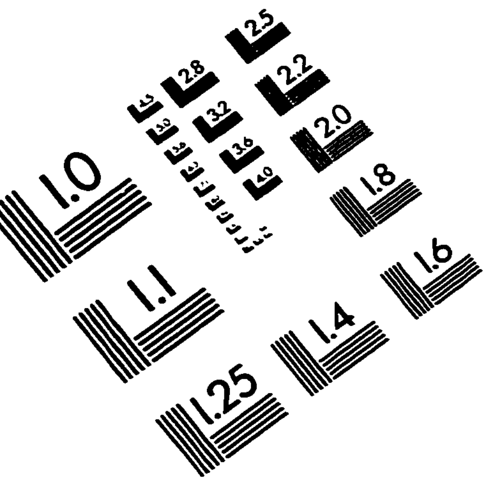
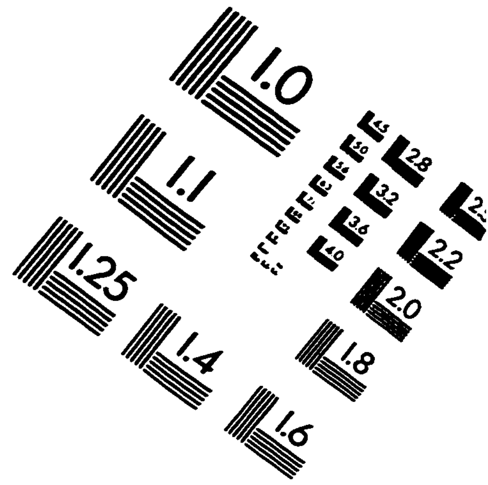
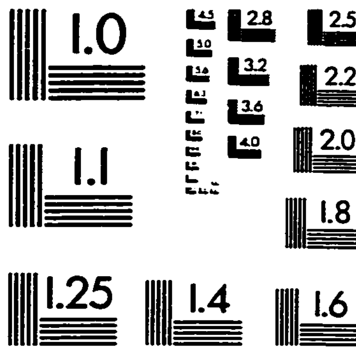
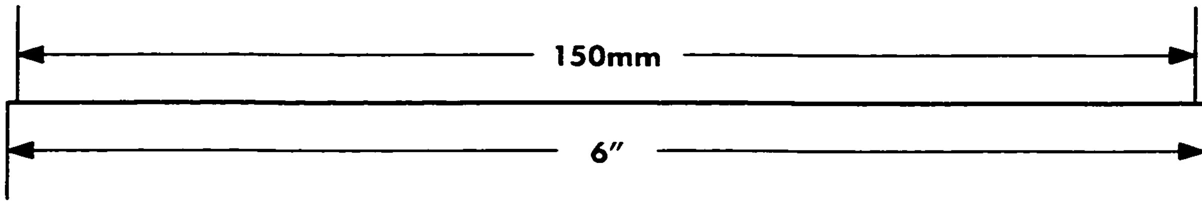
Yamada, T., and M. Shimaji. 1987. An intron in the 23 S rRNA gene of the *Chlorella* chloroplasts: complete nucleotide sequence of the 23 S rRNA gene. Curr. Genet. 11:347-352.

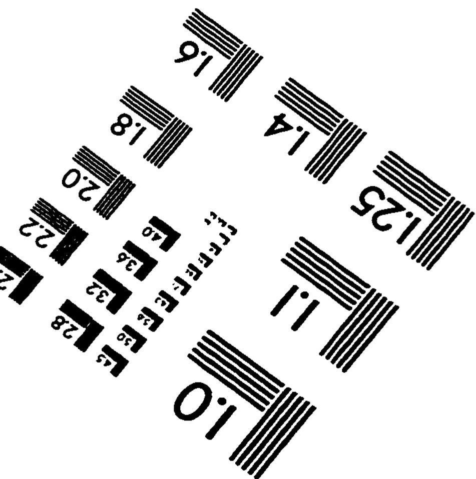
Yin, S., J. Heckman, and U. L. RajBhandary. 1981. Highly conserved GC-rich palindromic DNA sequences flank tRNA genes in *Neurospora crassa* mitochondria. Cell 26:325-332.

Zinn, A. R., J. K. Pohlman, P. S. Perlman, and R. A. Butow. 1988. *In vivo* double-strand breaks occur at recombinogenic G+C-rich sequences in the yeast mitochondrial genome. Proc. Natl. Acad. Sci. USA. 85:2686-2690.

Zurawski, G., M. T. Clegg, and A. H. D. Brown. 1984. The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics 106:735-749.

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved