Molecular evidence for the antiquity of group I introns interrupting transfer RNA genes in cyanobacteria

Dissertation

zur Erlangung des Doktogrades

der Mathematisch-Naturwissenschaftlichen Fakultäten

der Georg – August – Universität zu Göttingen

vorgelegt von

David Fewer M.Sc.

aus Waterford/Irland

Göttingen 2001

D7

Referent: Prof. Dr. Thomas Friedl

Korreferent: Prof. Dr. Burkhard Büdel

Tag der mündlichen Prüfung:

Contents

Abstract

Genes interrupted by group I introns have been the perennial focus of evolutionary studies. Previous work has demonstrated the importance of lateral transfer in the evolutionary history of these auto-catalytic molecules. In this respect the group I intron interrupting the tRNA-Leu (UAA) gene in cyanobacteria and chloroplasts has attracted a great deal of scientific attention primarily because of its perceived age. Recent studies have concluded that the group I introns interrupting tRNA-fMet and tRNA-Arg (CCU) genes in cyanobacteria and proteobacteria have arisen through recent genetic exchange and suggest that the origin of the tRNA-Leu intron is also in doubt. However, direct phylogenetic evidence for these competing hypotheses has been lacking. In this study molecular systematic approaches were undertaken to examine the evolutionary history of the group I introns interrupting tRNA genes in chloroplasts, cyanobacteria, and α-proteobacteria. Highly congruent support was found for the co-evolution of the introns and the genomes in which they are inserted. The introns interrupting the tRNA-fMet and the tRNA-Leu (UAA) genes predate cyanobacteria and chloroplasts respectively while the tRNA-Arg (CCU) intron predates mitochondria. The scattered and sporadic distribution of the introns is best explained by pervasive parallel losses in the more derived lineages of cyanobacteria and α-proteobacteria (Sections 3.2-3.5). This study provides convincing phylogenetic evidence that the tRNA group I intron subfamily is ancient and this means that these introns are between 2.1 and 3.5 billion years old. This strengthens the argument for the antiquity of this class of RNA enzyme.

During phylogenetic analyses of cyanobacterial taxa containing group I introns it became apparent that the controversial sister taxa relationship between the non-heterocyst forming cyanobacteria *Chroococcidiopsis* PCC 7203 and the heterocyst forming cyanobacteria received highly congruent support with the inclusion of additional members of the genus and through independent and combined phylogenetic analyses of *rpoC1*, *tufA* and 16S rRNA gene datasets (Section 3.1). This is important because it means that the complex baeocyte differentiation process has arisen independently at least twice in the cyanobacterial radiation, that the morphological identical genus *Myxosarcina* is not closely related to *Chroococcidiopsis* and rejects *Chroococcidiopsis* as the most primitive living cyanobacterium.

1. Introduction

1.1. Cyanobacteria fossil record

The cyanobacteria have been tremendously important in shaping the course of evolution and instigating ecological change on early earth (Brocks et al. 1999, Summons et al. 1999). The oxygen atmosphere that we depend on was generated by numerous cyanobacteria photosynthesising during the Archaean and Proterozoic Era (Schopf 1993). Oyxgenic photosynthesis is the preferred physiology of cyanobacteria metabolic and it is proposed that oxygen excretion was occurring well before significant oxygen had accumulated in the atmosphere at about 2,000 million years ago (Holland and Buekes 1990, Knoll and Holland 1995).

In keeping with this, the cyanobacteria are widely held to have left a fossil record that extends far back into the Precambrian (Schopf 1993). Many fossil cyanobacteria show a striking resemblance to living genera of cyanobacteria and morphologies in the group are thus proposed to have remained much the same for billions of years (Schopf 1994, Knoll et al. 1986, Golubic and Hofmann 1976). The retention of morphological characters is proposed to be due to a low evolutionary rate in the cyanobacterial radiation (Castenholz 1992).

The earliest unicellular filamentous forms attributed to cyanobacteria were found in sedimentary rocks formed 3,465 million years ago (Schopf and Packer 1987, Schopf 1993). Possible corroboration for these early dates was provided in sedimentary carbon isotopes ratios that show a characteristic enrichment in ^{12}C (Schopf 2000). The presence of cyanobacteria in the Archean is also indicated by hopanoids found in 2,700 million year old sediments (Brocks et al. 1999, Summons et al. 1999). Stromatolites indicate that cyanobacteria inhabited the oceans more than 2,500 million years ago (Walter 1983, Golubic and Hoffmann 1976). There are heterocysts formed by nitrogen-fixing cyanobacteria in fossils aged 1,300-1,500 million years old (Golubic et al. 1995). Endolithic forms that reproduce by baeocytes were observed in rocks formed circa 1,400 million years ago (Schopf and Walter 1982). The oldest known lichen symbiosis dating to the late Devonian 400 million years ago, involves cyanobacteria of the *Chroococcidiopsis*, *Gloeocapsa* and *Cyanosarcina* type (Taylor et al. 1995). All of these studies point towards the antiquity of the cyanobacterial lineage.

1.2. Cyanobacterial systematics

Cyanobacteria comprise one of the largest, most ecologically diverse, successful and important group of bacteria on earth (Wilmotte 1994). The systematics of cyanobacteria has been unusually tumultuous. The earliest taxonomic monographs describe cyanobacteria as a kind of algal plant (Thuret 1875). Modern systematics is usually acknowledged as having begun with the system proposed by Geitler (1932). It has formed the basis of numerous revised systems (Elenkin 1949, Fritsch 1959, Golubic 1976). These systems share in common the view that the systematics of cyanobacteria should be based on traditional botanical criteria and their nomenclature ruled by the botanical code. A second system is based on the recognition that the blue-green algae are unquestionably bacteria and not simply a prokaryotic sister group to other bacteria (Stanier and van Niel 1962). On this basis investigators developed a provisional taxonomic scheme for cyanobacteria based on examination of strains in axenic culture using bacteriological rather than botanical criteria (Rippka et al. 1979, Rippka 1988, Rippka and Herdman 1992). Stanier and collaborators pioneered the use of physiological and ecotypic characters determined with axenic cultures. Characters employed included pigment composition, fatty acid analysis, heterotrophic growth, nitrogenase activity, DNA base composition and genome length (Kenyon et al. 1972, Herdman et al. 1979, Rippka et al. 1979). These physiological studies did not furnish many useful physiological characters and the basis of the bacteriological taxonomy largely rests on the use of morphological characters (Rippka et al. 1979).

A proposal to formally place the cyanobacteria under the bacteriological code rather than the botanical code met with immediate and vigorous opposition (Geitler 1979, Golubic 1979, Lewin 1979). Although the proposal was not adopted (Castenholz and Waterbury 1989, Lewin 1989, Waterbury and Rippka 1989) the Stanierian system with some modification now forms the basis of cyanobacterial taxonomy as described in the *Bergeys Manual of Systematic Bacteriology* (Boone and Castenholz 2001) a recognised authority in bacterial systematics. The possibility of having the same organism described under two different names in the Botanical and bacteriological codes would have created chaos. Mutual concessions and adaptations of the two codes have ensured that species described under one system were recognised under the other. In an effort to reconcile the differences between the botanical and bacteriological approaches investigators often use a compromise system (Anagnostidis and Komárek 1985, Komárek and Anagnostidis 1999). The strains of cyanobacteria used in this study are classified according to the *Bergeys Manual of Systematic Bacteriology* (Boone and Castenholz 2001).

1.3. Molecular evolution of cyanobacteria

The exact origin and diversification of cyanobacteria is one of the most interesting and controversial questions in microbial evolution. Pioneering phylogenetic studies established a systematic view of bacteria based on evolutionary relationships inferred by a direct comparison of homologous genes (Fox et al. 1977). Because of their low rates of substitution, rRNA sequences have proved to be useful for addressing questions concerning very ancient evolutionary divergence events (Li and Graur 1991).

The first molecular evolutionary studies on cyanobacteria confirmed the bacterial nature of cyanobacteria as well as links with photosynthetic plastids (Bonen and Doolittle 1976, Woese et al. 1975). The earliest proposal that photosynthetic plastids may be derived from micro-organisms of a bacterial nature was made more than a century ago (Schimper 1883). Mereshowsky (1905) proposed cyanobacteria as the progenitors of plastids. This was revived as the endosymbiotic theory in the late 1960's and early 1970's (Sagan 1967, Margulis 1970). It was proposed that multiple endosymbiotic events involving prokaryotes with distinct pigment complements gave rise to extant plastids (Raven 1970). The weight of evidence is such that it is now generally accepted that plastids are derived from cyanobacteria and that this happened on a single occasion (Delwiche et al. 1997, Gray 1988, Gray and Doolittle 1982) although the exact nature of the relationship was and continues to be a matter of debate (Turner et al. 1999, Lockhart et al. 1992).

Evidence for the presence of cyanobacteria in the Archaen has been claimed to be equivocal based on poorly preserved microfossils and indirect geochemical arguments (Brocks et al. 1999). Interpretations of microfossil evidence is frequently based upon the assumption that morphology is phylogenetically conserved. However, prokaryote and in particular cyanobacterial morphology correlates imperfectly with phylogeny (Giovannoni et al. 1988). Molecular phylogenetic studies have raised the possibility that the fossils are incorrectly attributed to modern cyanobacteria (Giovannoni 1988, Feng et al. 1997). Although the cyanobacteria are generally believed to be a particularly ancient group, the sequence similarities of their 16S rRNAs to one another and to those of other eubacteria show that other major eubacterial taxa diverged before modern cyanobacteria (Giovannoni et al. 1988). Among these other eubacterial taxa are the family Chloroflexaceae which diverged more deeply that cyanobacteria. Obligately anaerobic, phototrophic *Chloroflexus* species are known to form laminated microbial mats and are morphologically similar to microfossils in the earliest known stromatolites (Nubel et al. 2001). These considerations caution against the interpretation of the earliest microbial fossils as cyanobacterial in origin (Feng et al. 1997). In addition, divergence times based on a protein clock cast serious doubts on whether the 3,450 million year old microfossils truly represent modern cyanobacteria (Doolittle et al. 1996). This study itself has been heavily criticised for extrapolating too far back in

time based mostly on the vertebrate fossil record (Morell 1996). In addition the distance calculations did not take into account the rate of change at different amino acid positions (Hasegawa and Fitch 1996, Gogarten et al. 1996, Miyamoto and Fitch 1996) and potential problems with sequence data corruption by the presence of sequences imported during the endosymbiotic acquisition of organelles (Gogarten et al. 1996).

Phylogenetic studies to date on independent RNA and protein data sets show the position of cyanobacteria within the universal tree are equivocal but generally to be only a moderately deep branch within Bacteria (Brown et al. 2001, Gupta 1997, Giovannoni et al. 1988, Oyaizu et al. 1987). Nevertheless, the fossil and independent geological evidence is widely accepted and taken to imply that the cyanobacteria are indeed ancient and that the bacterial lines of descent which branch before cyanobacteria were around prior to the 3450 million year date (Doolittle 1997). While the precise position and divergence of cyanobacteria within the bacterial domain remains unresolved with the estimated age of modern cyanobacteria is 3,465 million years old. It is likely that the concatenation of sequence datasets and more in-depth phylogenetic will be require to resolve the exact phylogenetic position of cyanobacteria and perhaps reconcile the disparity in these studies.

1.4. Group I introns interrupt transfer RNA genes in cyanobacteria

Many eukaryotic genes have their coding regions interrupted by intervening sequences or introns. Group I introns represent a family of RNA molecules with a specific higher-order structure and the ability to catalyze their own excision by a common splicing mechanism (Cech 1990). Group I introns are divided into 11 subfamilies based on conserved primary- and secondary-structure elements (Michel and Westhof 1990). Group I introns interrupting transfer RNA genes form a rapidly expanding subfamily of group I introns. The group I introns which interrupt tRNA genes are found in a phylogenetically diverse set of bacteria including proteobacteria, cyanobacteria and their chloroplast derivatives (Paquin et al. 1997). All of these introns in tRNA genes are quite compact in structure and contain little more than the phylogenetically conserved core sequences required for splicing (Zuag et al. 1993).

The intron interrupting the gene encoding a tRNA-Leu (UAA) gene was first discovered in the chloroplast of *Zea mays* nearly two decades ago (Steinmetz et al. 1982). These introns are inserted between the wobble (first) and second bases of the UAA anticodon. In 1988 the tRNA-Leu (UAA) gene from the cyanelle of *Cyanophora paradoxa* was reported to be interrupted by a similar intron and it was predicted that the intron would also interrupt tRNA-Leu (UAA) genes in cyanobacteria (Evrard et al.

1988). While it is now known that the cyanelle of glaucocystophytes such as *Cyanophora paradoxa* are a part of primary lineage of plastids it the past it was generally believed that cyanelles arose recently and independently of all other plastid lineages (Lambert et al. 1985). It was predicted that cyanobacteria would also contain an intron in this gene (Evrard et al. 1988). Two years later two groups published papers simultaneously in *Science* describing tRNA-Leu (UAA) genes in cyanobacteria which were interrupted by an intron bearing remarkable similarity to the intron interrupting the tRNA-Leu (UAA) gene in land plant chloroplasts and that in the cyanelle of *Cyanophora paradoxa* (Xu et al. 1990, Kushel et al. 1990). This was the first report of an intron interrupting a bacterial gene and generated a great deal of scientific interest (Barinaga 1990, Belfort 1991, Shub 1991, Cavalier-Smith 1991, Liu 1991). It was predicted that perhaps other genes would be interrupted by such introns. In 1992 two more tRNA genes from proteobacteria were found to harbor introns (Reinhold-Hurek and Shub 1992). In 1994 an additional the tRNA-fMet gene which encodes the initiator tRNA in cyanobacteria was shown to be interrupted by an intron (Biniszkiewicz et al. 1994). In 1997 an intron more closely related to the tRNA-Arg (CCU) intron was described from tRNA-Leu (UAA) genes in cyanobacteria (Rudi and Jakobsen 1997).

The origins of group I introns is shrouded in mystery (Cech 1990). They are alternatively proposed to be recent invaders of the genome (Palmer and Lodgson 1991) or ancient relics of the putative precellular RNA world (Gilbert 1986). Group I introns are held to be the ancient, dating from the eubacterial cell 3,465 million years ago, and were originally restricted to tRNA (Cavalier-Smith 1991). It has been proposed that protein-spliced introns, which are usually confined to tRNA genes, evolved from these group I introns by a radical change in splicing mechanism in the common ancestor of eukaryotes and archaebacteria, perhaps only about 1700 million years ago (Cavalier-Smith 1991, Belfort and Weiner 1997). The strongest evidence that group introns tracing their ancestry to the RNA world comes from introns interrupting tRNA-Leu (UAA) gene in cyanobacteria and chloroplasts (Xu et al. 1990, Kushel et al. 1990). However, concerns over the high degree of sequence homology amongst these introns have been expressed and it is possible that the introns are in fact recent invaders of the genome (Barinaga 1990, Belfort 1991). Despite widespread acceptance of the antiquity of the introns little phylogenetic analyses have been undertaken to substantiate these claims. In fact, recent studies challenge the original conclusions of these studies (Rudi and Jakobsen 1997, Rudi and Jakobsen 1999).

1.5. Aims of thesis

In the light of the controversy now surrounding the origin and evolution of tRNA group I introns in cyanobacteria the preliminary aim of this thesis was to reconcile the evolutionary history of the introns

with that of the cyanobacteria harboring the intron. Molecular phylogenetic techniques were employed to compare the evolutionary history of the intron with that of the cyanobacteria. The small subunit rRNA gene was chosen to generate host-based phylogenies. During this study it because apparent that the controversial sister taxa relationship between *Chroococcidiopsis* and heterocysts forming cyanobacteria received highly congruent support upon the inclusion of additional taxa from the *Chroococcidiopsis* lineage. A multigene approach was developed to assess the role of taxon sampling in this controversial sister taxon relationship.

2. Materials and Methods

2.1 General Material and Methods

2.1.1 Origin of the strains of cyanobacteria and eukaryotes used in this study

The strains investigated in this study were obtained from the public culture collections the Culture Collection of Algae at the University of Texas (Starr and Zeikus 1993), the Provasoli-Guillard National Center for Culture Collection of Marine Phytoplankton (Andersen et al. 1997) and the Pasteur Culture Collection of Cyanobacteria (Rippka and Herdman 1992) and the Sammlung von Algenkulturen at University of Göttingen (Schlösser 1994, 1997, Friedl and Schlösser 2000). Additional strains of cyanobacteria were donated from the personal culture collection of Professors Burkhard Büdel (BB) and Dieter Mollenhauer (DM). The strains used in systematics surveys in this study are identified in the text by the following abbreviations: ATCC (American Type Culture Collection), CCMP (Provasoli-Guillard National Center for Culture of Marine Phytoplankton), IAM (Institute of Applied Microbiology Culture Collection),NIVA (Norwegian Institute for Water Research Culture Collection of Algae), PCC (Pasteur Culture Collection), SAG (Sammlung von Algenkulturen Göttingen), UGBG (University of Göttingen Botanical Garden) and UTEX (University of Texas Algal Collection). The strains were maintained on BG11 growth medium (Stanier et al. 1971) at 18/16 ºC under a light/dark regime of 12 h:12 h at a light intensity of about 20-50 µmol photons.m^{-2}.s^{-1} from white fluorescent bulbs. Glassware was autoclaved prior to use and cultures were changed every 2-3 months.

2.1.2 DNA extraction

Sample processing and DNA extraction were carried out in a laboratory dedicated exclusively to these purposes. Cells were mechanically broken using a cell homogenizer (Minibeadbeater, Biospec). DNA was extracted using the Invisorb Plant Spin Kit (Invitrogen) as recommended by the manufacturer and DNA solutions were stored at –20 ºC.

2.1.3 PCR conditions

Twenty to fifty ng of template DNA was added to a 0.5 ml Eppendorf containing 10 mM of each dNTP (Hybaid), 20 mM $MgCl_2$, 2 mM of each primer, 0.2 µl Taq polymerase (Eurogentec). Sterile distilled water was added to bring the volume up to 50 µl. All sample reactions were accompanied by negative controls (template-free reactions) to detect contamination and related problems. The PCR cycling parameters were as follows: 35 cycles of 94°C for 60 s, 50°C for 60 s, and 72°C for 60 s. All reactions were initiated with a 4-min denaturation at 94°C and ended with 7-min extension at 72°C. PCR primers used in this study (Table 2.1) were either taken from the literature or designed by

Table 2.1. PCR and cycle sequencing primers used in this study. All primers are given in the sense (5'-3') direction. Cycle sequencing primers were either labeled with IRD 700 or IRD 800 fluorescent dyes for use with the Licor L4200 automated sequencer or a CY5 fluorescent dye for use with the AL-Fexpress II automated sequencer.

Primer		Source
PCR primers		
16S rRNA gene		
PCR1	GAGTTTGATCCTGGCTCAG	Wilmotte et al. 1993
PCR 18	TCTGTGTGCCTAGGTATCC	Wilmotte et al. 1993
tRNA-Leu (UAA) gene		
LeuF	TGGYGRAATYGGTAGACGCWRCG	This study
LeuR	TGGGGRYRGAGRGACTYGAACYCTC	This study
tRNA-Leu (UAA) intron		
PorphF	GAACTGCTCAAATTCGGGG	This study
PorphR	GGTCTGGACTATCCCTTC	This study
tRNA-fMet gene		
MetF	GGGGTAGAGCAGYCTGGTAGCTCGT	This study
MetR	GAAGGTCRNRGGTTCAAATCCRSTCC	This study
tRNA-Arg (CCU) gene		
ArF	CACGTAGCTCAGCWGGATAGA	This study
ArR	GGRGGTTCGAATCCTCRTCCG	This study
tufA gene		
TF	CACGTDGAYTGYCCNGGNCACGCTG	This study
TR	ATNCGRTCNCCDGGCATAACCATTTC	This study
rpoC1 gene		
RF	TGGGGHGAAAGNACAYTNCCTAA	This study
RR	GCAAANCGTCCNCCATCYAAYTGBA	This study
Cycle sequencing primers		
16S rRNA gene		
PCR1	GAGTTTGATCCTGGCTCAG	Wilmotte et al. 1993
PCR18	TCTGTGTGCCTAGGTATCC	Wilmotte et al. 1993
460F	CCGTATTACCGCGGCTGCT	Wilmotte et al. 1993
460R	AGCAGCCGCGGTAATACGG	Wilmotte et al. 1993
847F	GTGAAACTCAAAGGAATTGA	Wilmotte et al. 1993
847R	TCAATTCCTTTGAGTTTCAC	Wilmotte et al. 1993
1027F	AGTCCCGCAACGAGCGCAAC	Wilmotte et al. 1993
1458R	AAGGAGGTGATCCARCCRCAC	Wilmotte et al. 1993
tRNA-Leu (UAA) gene		
LeuF	TGGYGRAATYGGTAGACGCWRCG	This study
LeuR	TGGGGRYRGAGRGACTYGAACYCTC	This study
tRNA-fMet gene		
MetF	GGGGTAGAGCAGYCTGGTAGCTCGT	This study
MetR	GAAGGTCRNRGGTTCAAATCCRSTCC	This study
tufA gene		
TF	CACGTDGAYTGYCCNGGNCACGCTG	This study
TR	ATNCGRTCNCCDGGCATAACCATTTC	
rpoC1 gene		
RF	TGGGGHGAAAGNACAYTNCCTAA	This study
RR	GCAAANCGTCCNCCATCYAAYTGBA	This study

comparative sequence alignments and synthesized by the MWG Biotech (Germany). PCR amplification was performed using either a Biometra Personal Cycler or MWG Biotech Primus 96 thermocycler. The products of the PCR were run on a 2% agarose gel and visually inspected for products. 16S rRNA, *rpoC1* and *tufA* genes used in this study were identified on the basis of their expected size from the annealing sites of the primers. Introns interrupting tRNA genes were identified on the basis of a 200-300 bp increase in the size expected for an uninterrupted tRNA gene. Slower migrating PCR products of approximately 280-380 bp were expected for interrupted tRNA genes while faster migrating PCR products of approximately 80 bp were expected for uninterrupted tRNA genes.

2.1.4 Cloning of PCR products

Occasionally double bands were amplified under the low stringency PCR conditions employed. The fragments of the expected sizes were cut from gels were ligated with pGEM-T easy vector (Promega) for 16 at 14°C. The ligated vectors were transformed using a standard heatshock method in which host *E. coli* cells were transferred from an icebucket to a waterbath at 42°C for 40 s and the returned to the icebucket (Sambrook et al. 1989). One-fifth of the transformation mixture was directly spread LB agar plate supplemented with 100 mg/ml of ampicillin. The plates were incubated overnight at 37°C. Recombinants were randomly picked and insert length was checked by colony PCR.

2.1.5 Automated sequencing

PCR products were cleaned with High Pure PCR Purification Kit (Roche), and cycle-sequenced directly with fluorescent labelled dyes (either CY5 or IRD labelled). The sequencing reactions were run out on ALFexpress II (Amersham Pharmacia) and L4200 (Licor) automated sequencers. Nucleotide sequences were determined for both strands.

2.1.6 Phylogenetic analyses

All analyses were performed using PAUP* V4.0b8 (Swofford 2000). In maximum parsimony (MP) analyses, the sites were either unweighted (uniform) or weighted (Bhattacharya and Medlin 1995) and then used as input for bootstrap analyses (1000 replications). In the weighted analyses the characters were rescaled according to the consistency index over an interval of 1-1000, an option implemented in PAUP* V4.0b8 (Swofford 2000). In parsimony searches, 10 random input orders and the tree bisection-reconnection branch-swapping algorithm (TBR) were employed to find the best tree. Best scoring trees were held at each step. MP analyses were performed with constant sites included and repeated with the constant sites excluded.

For distance analyses, two different approaches were used. In the first approach, a model of DNA substitution was chosen that fits the observed data best using likelihood ratio statistic as implemented in

the program MODELTEST 3.04 (Posada and Crandall 1998). This program evaluates the adequacy of 56 different models. Phylogenetic trees inferred from cyanobacterial and plastid ribosomal DNA sequences may be particularly sensitive to biases in base composition (Lockhart et al. 1992, 1994). Therefore, in the second approach distance analyses based on the LogDet transformation were also included (Steel et al. 1993). LogDet distances were calculated in two ways, firstly with all constant sites removed and secondly with the proportion of sites assumed to be invariable (*pinvar*) estimated from the MP and maximum likelihood trees and removed in proportion to base frequencies as implemented in PAUP* V4.0b8 (Swofford 2000). The models of DNA substitution and corresponding settings were selected in PAUP* 4.0b8 with DNA distances set to maximum likelihood. Distance trees were constructed using both the minimum evolution criterion (ME, Rzhetsky and Nei 1992), with the same heuristic search procedure as in the maximum parsimony analyses, and the neighbor-joining method (Saitou and Nei 1987). Bootstrap resampling (1000 replications) was performed on both ME and NJ trees.

Maximum likelihood (ML) searches were performed using the model estimated by MODELTEST as in distance analyses. The ML trees obtained was used as input tree for a second round of ML analyses to search for trees with smaller -ln likelihoods. A tree with a better likelihood score was not obtained. 100 bootstrap resamplings were performed under the maximum likelihood criterion where computation time and resources permitted.

2.1.7 Statistical tests

The likelihood and parsimony based Kishino-Hasegawa tests (Kishino Hasegawa 1989), the parsimony based Templeton test (Templeton 1983) and the parsimony based winning-sites test (Prager and Wilson 1988) were used to evaluate evolutionary hypotheses as implemented in PAUP* 4.0b8.

2.2. Materials and Methods for section 3.1

2.2.1 PCR Amplification of the 16S rRNA, rpoC1 *and* tufA *genes*

The PCR primer pair PCR1 and PCR18 were used to specifically amplify the cyanobacterial 16S rRNA gene (Table 2.1). These primers amplified the entire 16S rRNA gene, the internal transcribed spacer between the 16S rRNA and 23S rRNA and a portion of the 23S rRNA gene. The 16S rRNA gene was in this way amplified and directly sequenced from a number of strains of *Chroococcidiopsis*, additional members of the order Pleurocapsales and heterocyst forming cyanobacteria (see general materials and methods 2.1).

For the combined dataset of 19 cyanobacteria and chloroplasts portions of the genes encoding the *rpoC1* and *tufA* genes were amplified in order to construct a comparable dataset. These two genes have been used in phylogenetic studies independently (Delwiche et al. 1995, Palenik et al. 1992 and others???) and this study is the first to make use of concatenated alignments to address a molecular systematic problem, namely the effect of taxon sampling on the sister taxon relationship. The primer pair TF and TR (Table 2.1) were developed to amplify cyanobacterial *tufA*. The *tufA* gene encodes the elongation factor Tu, that has a central role in protein synthesis (Delwiche et al. 1995). Primer pair RF and RR were developed to amplify a fragment of the cyanobacterial *rpoC1* gene. The *rpoC1* gene encodes the gamma subunit of cyanobacterial RNA polymerases and is homologous to the chloroplast RNA polymerase C1 subunit (Bergsland and Haselkorn 1991).

2.2.2 Large 16S rRNA datasets

The 16S rRNA sequences determined in Section 3.1 were manually aligned with approximately 80 16S rRNAs representative of known cyanobacterial diversity using the multiple sequence editor BioEdit (Hall 1999). Since there is no gain in phylogenetic information by incorporating virtually identical sequences in tree inference, these were excluded from the phylogenetic analyses. The 16S rRNA secondary structure of *Synechococcus* PCC 6301 (Gutell 1993) was used to refine the alignment. Regions not clearly alignable for all sequences were excluded from the phylogenetic analyses. In total 1331 aligned positions of which 653 were variable and 464 parsimony informative were considered in the phylogenetic analysis. The final alignment comprised 13 rRNA coding regions newly determined in this study and 44 complete rRNA sequences (Table 2.2). These were representative of the cyanobacterial sequence groups previously identified (Wilmotte et al. 1994, Turner 1997, Honda et al. 1999). Taxa that had unstable positions in initial analyses were excluded. This meant that members of the LEPT group (Turner et al. 1999) were excluded from the final analyses. The monophyly of this sequence group could not be resolved in the present analyses. The bacteria *Bacillus subtilis* (AB016721), *Chlorobium tepidum* (M58468), and *Escherichia coli* (AE000129) were used as outgroup taxa in the final analyses. We performed analyses in which each of the outgroup taxa were in turn specified as the outgroup.

For this dataset the GTR+I+G model (Rodríguez et al. 1990) was found to be best with estimations of nucleotide frequencies (A = 0.2359, C = 0.2352, G = 0.3169, T = 0.2120), a rate matrix with 6 different substitution types, assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter α = 0.5173), and *pinvar* = 0.3882 as estimated by MODELTEST. Phylogenetic analyses were performed using these settings as outlined in the general materials and methods section 2.1.

Table 2.2. List of ingroup taxa and 16S rRNA sequence accession numbers used in this study. Taxonomic scheme according to Rippka and Herdman (1992) and Burger-Wiersma et al. (1989).

Taxon	16S rRNA	Taxon	16S rRNA
Cyanobacteria			
Order Chroococcales		Order Oscillatoriales	
Cyanobium gracile PCC 6307	AF001477	*Arthrospira* PCC 8005	X70769
Cyanothece PCC 7424	AF132932	*Leptolyngbya* PCC 7375	AF132786
Gloeobacter violaceus PCC 7421	AF132790	*Lyngbya aestuarii* PCC 7419	AJ000714
Gloeocapsa PCC 73106	AF132784	*"Oscillatoria" rosea* IAM M-220	AB003164
Gloeothece membranacea PCC 6501	X78680	*"Oscillatoria" neglecta* IAM-M82	AB003168
Microcystis aeruginosa PCC 7941	U40340	*Oscillatoria sancta* PCC 7515	AF132933
Synechoccoccus ATCC 700246	AF132775	*"Phormidium" mucicola* IAM M-221	AB003165
Synechoccoccus lividus ATCC 700243	AF132772	*Pseudanabaena* PCC 6903	AF132778
Synechococcus elongatus PCC 6301	X03538	*Spirulina major* PCC 6313	X75045
Synechococcus elongatus PCC 7002	D88289	*Spirulina subsalsa* IAM M-223	AB003166
Synechococcus "elongatus"	D83715	*Trichodesmium* NIBB 1067	X70767
Synechococcus WH 8101	AF001480		
Synechocystis PCC 6803	D64000	Order Nostocales	
		Anabaena cylindrica PCC 7122	AF091150
Order Pleurocapsales		*Anabaena* NIVA-CYA 281/1 [partial]	ALZ82797
"Chroococcidiopsis" PCC 6712	XXXX	*Calothrix desertica* PCC 7102	AF132779
Chroococcidiopsis SAG 2023	XXXX	*Cylindrospermum stagnale* PCC 7417	AF132789
Chroococcidiopsis SAG 2024	XXXX	*Nodularia spumigena* PCC 7804	AJ133181
Chroococcidiopsis SAG 2025	XXXX	*Nostoc microscopicum* DM 2028	XXXX
Chroococcidiopsis SAG 2026	XXXX	*Nostoc punctiforne* PCC 73102	AF027655
Chroococcidiopsis CCME 029 [partial]	AF279107	*Nostoc* PCC 7120	X59559
Chroococcidiopsis CCME 057 [partial]	AF279108	*Scytonema hofmanni* PCC 7110	AF132781
Chroococcidiopsis CCME 123 [partial]	AF279109		
Chroococcidiopsis CCME 171 [partial]	AF279110	Order Stigonematales	
"Chroococcidiopsis" CCMP 1489	XXXX	*Chlorogloeopsis fritschii* PCC 6718	AF132777
Chroococcidiopsis cubana SAG 39.79	XXXX	*Chlorogloeopsis* PCC 7518	X68780
Chroococcidiopsis thermalis PCC 7203	AB039005	*Fischerella muscicola* PCC 7414	AF132788
Dermocarpa violacea PCC 7301	AB039009	*Fischerella* SAG 2027	XXXX
Dermocarpella SAG 29.84	XXXX		
Myxosarcina PCC 7312	XXXX	Prochlorophyta	
Myxosarcina PCC 7325	XXXX	*Prochlorococcus* MIT 9313	AF053398
Pleurocapsa minor PCC 7327	AB039007		
Pleurocapsa minor SAG 4.99	XXXX	Chloroplasts	
Pleurocapsa PCC 7516	X78681	*Cyanophora paradoxa* UTEX 555	U30821
Stanieria PCC 7437	AF132931	*Marchantia polymorpha*	X04465
Xenococcus PCC7305	AF132783	*Porphyra purpureum*	U38804

Note: Organisms with their names in quotes are likely to have been misidentified.

2.2.3 Reduced Chroococcidiopsis *16S rRNA dataset*

A second 16S rRNA dataset was constructed with the number of ingroup taxa reduced to *Chroococcidiopsis*, representatives of heterocystous cyanobacteria, and members of the Pleurocapsales to allow the inclusion of additional partial sequences (Billi et al. 2001) from members of these groups. *Phormidium mucicola* IAM M-221 and *Gloeobacter violaceus* PCC 7421 were used as outgroup taxa. This

second data set was 474 aligned positions long of which 186 were variable and 121 parsimony informative. Genbank accession numbers for the newly determined 16S rRNA sequences are given in Table 2.2.

For this reduced data set (24 ingroup taxa), the TrN+I+G model of DNA substitution (Tamura and Nei 1993) was found best with estimations of nucleotide frequencies (A = 0.2645, C = 0.1962, G = 0.3240, T = 0.2153), a rate matrix with 6 different substitution types, gamma parameter $\alpha = 0.5943$ (number of rate categories=4), and *pinvar* = 0.3594 as estimated by MODELTEST. Phylogenetic analyses were performed using these settings as outlined in the general materials and methods section 2.1. The LogDet transformation was used to compare pairwise genetic distances within and among the *Chroococcidiopsis* and heterocystous cyanobacteria lineages.

2.2.4 Individual analyses of the genes comprising the combined dataset

The large alignment of 16S rRNA sequences was reduced to 19 taxa for which either *tufA* or *rpoC1* genes were publically available (Table 2.3). These taxa were chosen to minimize sequencing effort as well as allowing the role of taxon sampling in the sister taxon relationship to be examined. In total 1438 aligned positions of which 641 were variable and 455 parsimony informative were considered in the phylogenetic analysis. For this 16S data set the GTR+I+G model of DNA substitution was found best with estimations of nucleotide frequencies (A = 0.22431, C = 0.2323, G = 0.3072, T = 0.2174), a rate matrix with 6 different substitution types, gamma parameter $\alpha = 0.4134$ (number of rate categories=4), and *pinvar* = 0.2801 as estimated by MODELTEST.

Six novel sequences of *tufA* gene were generated in this study and aligned with 13 other publically available *tufA* gene sequences (Table 2.3). Gaps and third positions were excluded from the final analyses. In total 522 aligned positions of which 250 were variable and 173 parsimony informative were considered in the phylogenetic analysis. For this *tufA* data set the TvM+I+G model of DNA substitution was found best with estimations of nucleotide frequencies (A = 0.2943, C = 0.2268, G = 0.2569, T = 0.2220), a rate matrix with 6 different substitution types, gamma parameter $\alpha = 0.6606$ (number of rate categories=4), and *pinvar* = 0.2673 as estimated by MODELTEST.

Nine novel sequences of the *rpoC1* gene were generated in this study and aligned with 10 publically available *rpoC1* gene sequences (Table 2.3). Gaps and third positions were excluded from the final analyses. In total 429 aligned positions of which 270 were variable and 197 parsimony informative were considered in the phylogenetic analysis. For this *rpoC1* data set the GTR+G model of DNA substitution was found best with estimations of nucleotide frequencies (A = 0.3643, C = 0.1661, G = 0.2534, T = 0.2161), a rate matrix with 6 different substitution types and a gamma parameter $\alpha =$

Table 2.3. List of ingroup taxa and 16S, *tufA* and *rpoC1* sequence accession numbers used in the combined dataset. XXXX denotes sequenced obtained for this study. 'Genome' indicates sequences taken from unpublished microbial genomes.

Species	16S rRNA	*tufA*	*rpoC1*
Cyanobacteria			
Order Chroococcales			
Chamaesiphon PCC 7430	XXXX	XXXX	XXXX
Gloeobacter violaceus PCC 7421	AF132790	U09433	U52340
Gloeothece membranacea PCC 6501	X78680	U09434	XXXX
Synechococcus elongatus PCC 6301	X03538	X17442	XXXX
Synechococcus PCC 7002	AJ000716	AB025429	U52345
Synechocystis PCC 6803	AB001339	AB001339	AB001339
Order Pleurocapsales			
Chroococcidiopsis PCC 7203	AB039005	XXXX	XXXX
Myxosarcina PCC 7312	XXXX	XXXX	XXXX
Pleurocapsa minor SAG 4.99	XXXX	XXXX	XXXX
Order Oscillatoriales			
Leptolyngbya PCC 73110	X84810	U09444	XXXX
Leptolyngbya PCC 7375	AF132786	U09443	XXXX
Pseudanabaena PCC 7367	AF091108	XXXX	XXXX
Order Nostocales			
Nostoc PCC 7120	X59559	Genome	Genome
Nostoc punctiforme PCC 73102	AF027655	Genome	Genome
Order Stigonematales			
Fischerella muscicola PCC 7414	AF132788	XXXX	Z11153
Prochlorophyta			
Prochlorothrix hollandica	AF132792	U09445	Z11154
Chloroplasts			
Cyanophora paradoxa UTEX 555	U30821	U30821	U30821
Mesostigma viride NIES-296	AF166114	AF166114	AF166114
Porphyra purpurea	U38804	U38804	U38804
Outgroups			
Escherichia coli	U00096	U00096	U00096
Chlorobium tepidium	Genome	Genome	Genome
Pseudomonas aeruginosa	AE004091	AE004091	AE004091

0.4956 (number of rate categories = 4). Sequences obtained from the complete genomes of the bacteria *Pseudomonas aeruginosa*, *Chlorobium tepidum* and *Escherichia coli* were used as outgroup taxa in the individual analyses of all individual analyses (Table 2.3). Phylogenetic analyses for each individual dataset were performed using these individual settings as outlined in the general materials and methods section 2.1.

2.2.5 Analyses of the combined dataset

A combined dataset was prepared by concatenating all sequences from the individual 16S rRNA, *rpoC1* and *tufA* datasets. Sequences obtained from the complete genomes of the bacteria *Pseudomonas aeruginosa*, *Chlorobium tepidum* and *Escherichia coli* were used as outgroup taxa (Table 2.3). In order to assess the utility of combining data the partition homogeneity test (Farris et al. 1994, Swofford 2000) was applied to the of 16S rRNA, *tufA* and *rpoC1* datasets. The partition homogeneity tests suggested that third positions should be included from both *rpoC1* and *tufA*. Gaps, regions which were not clearly alignable the third positions in *rpoC1* and *tufA* were discarded from the combined dataset. In total 2389 aligned positions of which 1161 were variable and 825 parsimony informative were considered in the phylogenetic analysis.

For the combined data set the GTR+I+G model (Rodríguez et al. 1990) was found to be best with estimations of nucleotide frequencies (A = 0.2729, C = 0.2290, G = 0.2880, T = 0.2101), a rate matrix with 6 different substitution types, assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter α = 0.5891), and *pinvar* = 0.2841 as estimated by MODELTEST.

2.2.6 Statistical tests

Statistical tests outlined in the general materials and methods section 2.1 were performed on user-defined trees constructed to reflect three evolutionary scenarios previously either tacitly assumed or explicitly implied in the literature: (1) a monophyletic order Pleurocapsales, (2) *Myxosarcina* is a sister taxon to *Chroooccidiopsis* and (3) *Chroococcidiopsis* is the most primitive living cyanobacteria. User-defined trees were constructed for both the large 16s rRNA dataset (53 taxa) and the combined dataset (19 taxa) to reflect these three evolutionary scenarios. Due to the performance of multiple tests, the significance level of rejection of the null hypothesis should be adjusted via the Bonferoni correction to α = 0.0125.). In addition, the partition homogeneity test (Farris et al. 1994, Swofford 2000), was applied to the 16S rRNA, *tufA* and *rpoC1* gene datasets to assess under what conditions the three genes could be combined.

2.3. Materials and Methods for section 3.2

2.3.1 Systematic survey of cyanobacteria and chloroplasts

The primer pair MetF and MetR were constructed to amplify the anticodon stem, loop and anticodon of the cyanobacterial tRNA-fMet gene. Although bacterial tRNA-fMet genes have relatively few invariant nucleotides all have three consecutive G-C base pairs closing the anticodon stem (RajBhandary

1994). This run of Gs and Cs is exclusive to the initiator tRNA and are required to enter the P site of the ribosome and initiate translation (Lewin 1997). Primers were designed to end at the base of the anticodon stem so that uninterrupted tRNA-fMet genes could be identified by sequencing.

A systematic survey of the presence or absence of the tRNA-fMet intron was conducted based on the public availability of 16S rRNA gene sequences (Table 3.2.1). Additional strains of cyanobacteria for which 16S rRNA gene sequences had been newly determined (Section 3.1) were also included in this survey. Small subunit rRNA sequences were obtained from *Oscillatoria* PCC 6304, *Synechocystis* PCC 6906 and *Nostoc* DM 44 for which introns were detected were amplified and sequenced specifically for this study (see general materials and methods 2.1). In all cases direct one to one comparison of cyanobacteria 16S rRNA and tRNA-fMet intron was possible. An exception to this was the intron from *Scytonema hofmannii* UTEX 2349 (U10481) which was compared to the 16S rRNA gene from *Scytonema hofmannii* PCC 7110 (AF132781). It is expected that this will contribute a minor amount of incongruence to the comparisons of the two datasets.

2.3.2 Large 16S rRNA dataset

Ninety full-length 16S rRNA gene sequences were aligned manually using the 16S rRNA secondary structure of *Synechococcus* PCC 6301 (Gutell 1993) to refine the alignment. The express purpose of this alignment was to construct a tree to which the distribution of the tRNA-fMet intron could be mapped. Regions not clearly alignable for all sequences were excluded from the phylogenetic analyses. In total 1331 aligned positions of which 770 were variable and 591 parsimony informative were considered in the phylogenetic analysis.

The GTR+I+G model (Rodríguez et al. 1990) was found to be best with estimations of nucleotide frequencies (A = 0.2493, C = 0.2169, G = 0.2986, T = 0.2352), a rate matrix with 6 different substitution types, assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter α = 0.7779), and *pinvar* = 0.418 as estimated by MODELTEST.

A distance tree was constructed using the minimum evolution criterion (Rzhetsky and Nei 1992), with the same heuristic search procedure as in the maximum parsimony analyses using the GTR+I+G model of DNA substitution and settings as determined by MODELTEST. Bootstrap resampling (1000 replications) was performed on the ME tree.

2.3.3 The tRNA-fMet intron dataset

Seven previously characterised tRNA-fMet introns were downloaded from genbank and along with 9 novel sequences of the intron generated during this study were aligned manually with the computer program BioEdit (Hall 1999). The intron alignment was based on the secondary structure model for the intron interrupting the tRNA-fMet gene in *Synechocystis* PCC 6803 (Biniszkiewicz et al. 1994). Sites that were ambiguously aligned were not considered in the phylogenetic analysis. In total 243 aligned positions of which 182 were variable and 149 parsimony informative were considered in the phylogenetic analysis.

For this dataset the TIM+G model was found to be best with estimations of nucleotide frequencies (A = 0.2891, C = 0.2266, G = 0.3058, T = 0.1785), a rate matrix with 6 different substitution types and assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter α = 0.4606) as estimated by MODELTEST. Phylogenetic analyses for this dataset were performed using these settings as outlined in the materials and methods section 2.1.

2.3.4 Reduced 16S rRNA dataset for congruency tests

The taxon sampling in the large 16S rRNA sequence alignment was reduced to just those 16 cyanobacteria determined to contain a tRNA-fMet intron. Regions not clearly alignable for all sequences were excluded from the phylogenetic analyses. In total 1353 aligned positions of which 448 were variable and 391 parsimony informative were considered in the phylogenetic analysis.

The HKY+I+G model of DNA was found best with estimations of nucleotide frequencies (A = 0.2522, C = 0.2209, G = 0.3223, T = 0.2046), a rate matrix with 6 different substitution types, assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter α = 0.6603), and *pinvar* = 0.3665 as estimated by MODELTEST. The bacteria *Bacillus subtilis* (AB016721), *Chlorobium tepidum* (M58468), and *Escherichia coli* (AE000129) were used as outgroup taxa. Phylogenetic analyses for this dataset were performed using these settings as outlined in the general materials and methods section 2.1.

2.3.5 Host and intron divergences.

A generalised two-parameter model was used to calculate genetic distances using the maximum likelihood transformation using the TIM model of DNA substitution with a gamma distribution of rates as estimated by MODELTEST. This approach accounted for among-site variability (Swofford and Olsen 1990). Genetic distances were calculated for both the host and tRNA-fMet introns.

2.3.6 Statistical tests

The intron from *Stanieria* PCC 7437 is known to have discordant positions in intron and 16S rRNA generate trees (Biniszkiewicz et al. 1994, Paquin et al. 1997). User-defined trees were constructed to test the importance of the position of *Stanieria* PCC 7437. A user-defined tree based on host 16S rRNA gene sequences was compared to trees generated from the intron data with *Stanieria* PCC 7437 included and repeated with *Stanieria* PCC 7437 excluded. Statistical tests as outlined in the materials and methods section 2.1 were performed upon these user-defined trees. Due to the performance of multiple tests, the significance level of rejection of the null hypothesis should be adjusted via the Bonferoni correction to $\alpha = 0.0125$.). In addition, the partition homogeneity test (Farris et al. 1994, Swofford 2000), was applied to the tRNA-fMet intron and reduced 16S rRNA gene datasets to determine the level of congruence between the two datasets with and without *Stanieria* PCC 7437.

2.4 Materials and Methods for section 3.3

2.4.1 PCR amplification and sequencing of the divergent intron

During the systematic survey of cyanobacteria and plastids (Sections 3.4 and 3.5) by the primer pair LeuF and LeuR were employed to amplify the anticodon stem, loop and anticodon of the tRNA-Leu (UAA) gene. Suspicions were raised during the systematics survey the presence of either both uninterrupted and interrupted tRNA genes or a double band at the size expected for intron containing taxa. Cloning and sequencing (see general materials and methods 2.1) confirmed the suspicion that one of the bands was in each case attributable to the second atypical tRNA-Leu (UAA) intron.

2.4.2 Culture experiments

It was hypothesized that the eukaryotes and cyanobacteria which harbor the second type of tRNA-Leu (UAA) intron also harbor culture contaminants. The normal media (Schlösser 1994, Skulberg 1990) used to culture cyanobacteria, red algae and the glaucocystophyte suspected of being contaminated with heterotrophic bacteria was enriched. Five grams of tryptone, 1 gram of yeast extract, and 15 grams of agar were added to one litre with half strength seawater medium. Plates were poured under sterile conditions and allowed to polymerise. These plates were stored at +4°C until further use. Aliquots of the cultures of cyanobacteria, red algae and glaucocystophyte suspected to be contaminated with heterotrophic bacteria were used to innoculate these plates and incubated at +37°C and checked every 24 hours for signs of growth. Bacterial colonies were screened for the presence or absence of the intron using the LeuF and LeuR primer set with which there were originally identified (see general materials and methods 2.1). Positive colonies were screened with the primer set PorphF and PorphR to

ensure that the fragments generate were bona fide divergent tRNA-Leu (UAA) introns (see general materials and methods 2.1). These primers anneal to primary sequence regions unique to the divergent tRNA-Leu (UAA) intron and generate a DNA fragment 142-154 bp in length. Colonies positive for the presence of an intron sized tRNA gene were streaked and new colonies picked and examined microscopically until an pure culture was obtained. For the red algal culture *Bostrychia radicans* SAG 100.79, a contaminating bacterium which yielded a band of identical length to that of the intron was isolated into pure culture. A degenerate set of primers ArF and ArR were designed to amplify the anticodon loop and anticodon of tRNA-Arg (CCU) gene. These primers were used to screen culture contaminants for the presence of a tRNA-Arg (CCU) intron. PCR amplification was performed in the same manner as for the tRNA-Leu (UAA) gene. The *Bostrychia radicans* SAG 100.79 culture was treated with ampicillin (100 μg ml^{-1}) in an attempt to rid the culture of its contaminating bacterium.

2.4.3. DNA extraction

Liquid broths of the half-strength seawater medium were innoculated with a small amount of the contaminating bacterium from *Bostrychia radicans* SAG 100.79 and allowed to grow for 48 hours at +37°C on a shaker. Cells were pelleted and DNA was extracted as outlined in the general materials and method section 2.1. The PCR primer pair PCR1 and PCR18 were used to amplify and sequence the bacterial 16S rRNA gene despite the claim that these primers are specific for cyanobacteria (Wilmotte et al. 1993).

2.4.4 Alignment of introns interrupting tRNA genes

A subset of tRNA introns were downloaded from GenBank and other public databases. These sequences were combined with sequences obtained for this study in our lab. This was representative of all four tRNA genes known to be interrupted by group I introns in bacteria (Table 3.3.1). The sequences were aligned manually according to secondary structure models for the tRNA introns with the computer program BioEdit (Hall 1999). Regions not clearly alignable for all sequences were excluded from the phylogenetic analyses. In total 200 aligned positions of which 135 were variable and 118 parsimony informative were considered in the phylogenetic analysis.

The TIM+G model was found to be best with estimations of nucleotide frequencies (A = 0.2891, C = 0.2266, G = 0.3058, T = 0.1785), a rate matrix with 6 different substitution types and assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter α = 0.4606) as estimated by MODELTEST. Phylogenetic analyses for this dataset was unrooted and performed using these settings as outlined in the general materials and methods section 2.1.

2.4.5 Distribution of the tRNA-Arg intron and the phylogenetic position of contaminant

The 16S rRNA sequence determined from the bacterium contaminating the *Bostrychia radicans* SAG 100.79 culture was manually aligned against 51 16S rRNA gene sequences from other members of the α-proteobacteria. This taxon sampling corresponded to 3 proteobacteria for which introns had already being characterised and 47 proteobacteria for which complete genomes had been determined or for which genomes are currently being annotated (www.ncbi.nlm.nih.gov/Entrez). Regions not clearly alignable for all sequences were excluded from the phylogenetic analyses. In total 1383 aligned positions of which 681 were variable and 586 parsimony informative were considered in the phylogenetic analysis.

The GTR+I+G model (Rodríguez et al. 1990) was found to be best with estimations of nucleotide frequencies (A = 0.2493, C = 0.2169, G = 0.2986, T = 0.2352), a rate matrix with 6 different substitution types, assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter α = 0.7779), and *pinvar* = 0.418 as estimated by MODELTEST. The bacteria *Bacillus subtilis* (AB016721), *Clostridium botulinum* (X68317), and *Enterococcus faecalis* (AB036835) were used as outgroup taxa. Phylogenetic analyses for this dataset were performed using these settings as outlined in the general materials and methods section 2.1.

2.4.6. Congruency between 16S rRNA and tRNA-Arg intron

A second 16S rRNA dataset was constructed to address coevolution between the tRNA-Arg (CCU) intron identified in this study and previous studies and the α- proteobacteria from which they were identified. The taxon sampling in the original 16S rRNA sequence alignment was reduced to just those α- proteobacteria determined to contain a tRNA-Arg (CCU) intron. Regions not clearly alignable for all sequences were excluded from the phylogenetic analyses. In total 1454 aligned positions of which 461 were variable and 338 parsimony informative were considered in the phylogenetic analysis.

The HKY+I+G model of DNA was found best with estimations of nucleotide frequencies (A = 0.2522, C = 0.2209, G = 0.3223, T = 0.2046), a rate matrix with 6 different substitution types, assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter α = 0.6603), and *pinvar* = 0.3665 as estimated by MODELTEST. The bacteria *Bacillus subtilis* (AB016721), *Clostridium botulinum* (X68317), and *Enterococcus faecalis* (AB036835) were used as outgroup taxa. Phylogenetic analyses for this dataset were performed using these settings as outlined in the general materials and methods section 2.1.

4.4.7 Statistical tests

User-defined trees were constructed to test evolutionary hypotheses related to acceptance or rejection of intron horizontal transfer from a tRNA-Arg (CCU) to tRNA-Leu (UAA) gene (Figure 4.6). The Kishino-Hasegawa test (Kishino Hasegawa 1989), the Templeton test (Templeton 1983) and the winning-sites test (Prager and Wilson 1988) were used to evaluate these evolutionary hypotheses as implemented in PAUP* 4.0b8 and outlined in the general material and methods section 2.1. Due to the performance of multiple tests, the significance level of rejection of the null hypothesis should be adjusted via the Bonferoni correction to $\alpha = 0.0125$). In addition, the partition homogeneity test (Farris et al. 1994, Swofford 2000), was applied to the tRNA-Arg (CCU) intron and 16S rRNA gene datasets from α-proteobacteria to determine the level of congruence between the two datasets.

2.5 Materials and Methods for section 3.4

2.5.1. Systematic survey of cyanobacteria and plastids

A systematic survey of the presence or absence of the tRNA-Leu (UAA) intron was conducted based on the public availability of 16S rRNA gene sequences (Table 3.4.1). Additional strains of cyanobacteria for which 16S rRNA gene sequences had been newly determined (Section 3.1) were also included in this survey. The primer pair LeuF and LeuR were used to screen 57 cyanobacteria and 3 plastids for the presence or absence of the tRNA-Leu (UAA) intron.

2.5.2 Large 16S rRNA dataset

The 16S rRNA sequences for the 90 strains of chloroplast containing eukaryotes and cyanobacteria surveyed overall for the presence and absence of the intron were manually aligned using the multiple sequence editor BioEdit (Hall 1999). The 16S rRNA secondary structure of *Synechococcus* PCC 6301 (Gutell 1993) was used to refine the alignment. Regions not clearly alignable for all sequences were excluded from the phylogenetic analyses. In total 1331 aligned positions of which 770 were variable and 591 parsimony informative were considered in the phylogenetic analysis. The alignment was representative of the vast majority of cyanobacterial sequence groups identified to date (Wilmotte et al. 1994, Turner 1997, Honda et al. 1999). However, the THERM sequence group (Turner 1997) which consists of uncultured environmental samples of cyanobacteria could not be included here.

The GTR+I+G model was found to be best with estimations of nucleotide frequencies (A = 0.2475, C = 0.2269, G = 0.3092, T = 0.2164), a rate matrix with 6 different substitution types, assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories

= 4, shape parameter α = 0.5052), and *pinvar* = 0.3262 as estimated by MODELTEST. The bacteria *Bacillus subtilis* (AB016721), *Chlorobium tepidum* (M58468), *Chlorobium vibrioforme* (M62791) *Chlorobium limicola* (AB054671), *Agrobacterium tumefaciens* (AE007870) and *Escherichia coli* (AE000129) were used as outgroup taxa.

A distance tree was constructed using the minimum evolution criterion (Rzhetsky and Nei 1992), with the same heuristic search procedure as in the maximum parsimony analyses using the GTR+I+G model of DNA substitution and settings as determined by MODELTEST. Bootstrap resampling (1000 replications) was performed on the ME tree.

2.5.3. The tRNA-Leu (UAA) intron dataset

The tRNA-Leu (UAA) introns identified in this study and those identified in previous studies were manually aligned using the conserved secondary structure elements to identify homologous regions (Michel and Westhof 1990). Regions not clearly alignable for all sequences were excluded from the phylogenetic analyses. In total 211 aligned positions of which 144 were variable and 114 parsimony informative were considered in the phylogenetic analysis.

The GTR+G model was found to be best with estimations of nucleotide frequencies (A = 0.3956, C = 0.1377, G = 0.1898, T = 0.2770), a rate matrix with 6 different substitution types, and assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter α = 0.2258) as estimated by MODELTEST. Phylogenetic analyses for this dataset were performed using these settings as outlined in the general materials and methods section. A strict consensus of the trees generated in the approach outlined in this section was constructed to show the level of agreement between the individual trees.

2.5.4. Reduced 16S rRNA dataset for congruency tests

The taxon sampling in the large 16S rRNA sequence alignment was reduced to those 34 cyanobacteria determined to contain a tRNA-Leu (UAA) intron. Regions not clearly alignable for all sequences were excluded from the phylogenetic analyses. In total 1353 aligned positions of which 448 were variable and 391 parsimony informative were considered in the phylogenetic analysis.

The HKY+I+G model of DNA was found best with estimations of nucleotide frequencies (A = 0.2522, C = 0.2209, G = 0.3223, T = 0.2046), a rate matrix with 6 different substitution types, assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter α = 0.6603), and *pinvar* = 0.3665 as estimated by MODELTEST. The bacteria *Bacillus subtilis* (AB016721), *Chlorobium tepidum* (M58468), and

Escherichia coli (AE000129) were used as outgroup taxa. Phylogenetic analyses for this dataset were performed using these settings as outlined in the general materials and methods section.

2.5.5. Statistical tests

The partition homogeneity test (Farris et al. 1994, Swofford 2000), was applied to the a concatenated alignment comprised of 34 16S rRNA and tRNA-Leu (UAA) intron sequences.

2.6. Methods for section 3.5

2.6.1 Systematic survey of plastid containing eukaryotes

A systematic PCR survey using the primer pair LeuF and LeuR to determine the presence or absence of the intron interrupting tRNA-Leu (UAA) genes in diverse lineages of chloroplasts was undertaken (Table 3.5.1). BLAST searches of genbank and inspection of publically available complete genomes were used to supplement the survey.

2.6.2. Intron alignment

The set of 70 tRNA-Leu (UAA) intron sequences identified in this study was manually aligned using the conserved secondary structure elements to identify homologous regions (Michel and Westhof 1990, Bhattacharya et al. 1994). Regions not clearly alignable for all sequences were excluded from the phylogenetic analyses. In total 234 aligned positions of which 216 were variable and 206 parsimony informative were considered in the phylogenetic analysis.

Phylogenetic analyses of this large intron dataset analyses were limited to minimum evolution analyses. The TrN+G model was found to be best model of DNA substitution with estimations of nucleotide frequencies (A = 0.4063, C = 0.1362, G = 0.1942, T = 0.2633), a rate matrix with 6 different substitution types, and assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter $\alpha = 0.2420$). The TrN+G model of DNA substitution and corresponding settings were selected in PAUP* 4.0b8 with DNA distances set to maximum likelihood. Distance trees were constructed using the minimum evolution criterion (Rzhetsky and Nei 1992), with the same heuristic search procedure as in the maximum parsimony analyses. Bootstrap resampling (1000 replications) was performed on the resulting ME tree.

2.6.3. rbcL *and intron alignment*

Plastid encoded *rbcL* gene sequence publically available for 8 of the 10 strains of heterkonts identified in this study to contain tRNA-Leu (UAA) introns and were manually aligned (Table 3.5.1). Regions not clearly alignable for all sequences were excluded from the phylogenetic analyses. In total 1375 aligned positions of which 702 were variable and 328 parsimony informative were considered in the phylogenetic analysis.

For the *rbcL* data set, the GTR+G model was found to be best with estimations of nucleotide frequencies (A = 0.3051, C = 0.1561, G = 0.2097, T = 0.3291), a rate matrix with 6 different substitution types, and assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter α = 0.5128) as estimated by MODELTEST. The glaucocystophyte *Cyanophora paradoxa* was used as outgroup taxa (Table 3.5.1).

The taxon sampling in the original tRNA-Leu (UAA) intron sequence alignment was reduced to just those taxa for which *rbcL* sequences were available. These were rooted with the intron from *Cyanophora paradoxa* (Table 3.5.1). The TrN+G model was found to be best with estimations of nucleotide frequencies (A = 0.4355, C = 0.1063, G = 0.1563, T = 0.3020), a rate matrix with 6 different substitution types, and assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter α = 0.4025) as estimated by MODELTEST. Phylogenetic analyses for these datasets were performed using these settings as outlined in the general materials and methods section. The intron and *rbcL* alignments were also concatenated and used in the partition homogeneity test.

2.6.4. Nuclear encoded 18S rRNA alignment

An alignment of nuclear encoded 18S rRNA genes was created to examine the distribution of the tRNA-Leu (UAA) intron in the Chlorophyta (Table 3.5.1). Regions not clearly alignable for all sequences were excluded from the phylogenetic analyses. In total 1656 aligned positions of which 747 were variable and 505 parsimony informative were considered in the phylogenetic analysis. The glaucocystophytes *Cyanophora paradoxa* and *Glaucocystis nostochinearum* were used as an outgroup taxon (Table 3.5.1).

The "TrN+I+G" model was found to be best with estimations of nucleotide frequencies (A = 0.2381, C = 0.2303, G = 0.2724, T = 0.2592), a rate matrix with 6 different substitution types, assuming a heterogeneous rate of substitutions with a gamma distribution of variable sites (number of rate categories = 4, shape parameter α = 0.5547), and *pinvar* = 0.4245 as estimated by MODELTEST. Phylogenetic

analyses for these datasets were performed using these settings as outlined in the general materials and methods section.

2.6.5. Plastid encoded 16S rRNA and intron alignment

A plastid encoded 16S rRNA dataset was constructed to address coevolution between the tRNA-Leu (UAA) intron identified in this study and previous studies and those taxa for which chloroplast encoded 16S rRNA gene sequences were available. Plastid encoded markers were necessary to overcome the methodological problems introduced into such analyses by the secondary acquisition of plastid by non-photosynthetic eukaryotes (McFadden 1999). Regions not clearly alignable for all sequences were excluded from the phylogenetic analyses. In total 1320 aligned positions of which 553 were variable and 389 parsimony informative were considered in the phylogenetic analysis. The cyanobacterium *Pseudanabaena* PCC 7403 (AB039019) was used as an outgroup taxon.

The taxon sampling in the original tRNA-Leu (UAA) intron sequence alignment was reduced to just those taxa for which 16S rRNA was available. These were rooted with the intron from *Pseudanabaena* PCC 7403 (U83253). The intron and 16S rRNA alignments were concatenated used in the partition homogeneity test. The partition homogeneity test (Farris et al. 1994, Swofford 2001) was used to test whether the 16S rRNA and tRNA-Leu (UAA) intron datasets from all three primary photosynthetic lineages and the *rbcL* and tRNA-Leu (UAA) intron datasets from heterokonts are congruent. The taxa and genbank accession numbers for the genes used in these comparisons are listed in Table 3.5.1.

2.6.6. In vitro transcription and intron splicing

All plasticware, glassware and solutions were treated with diethy pyrocarbonate (DEPC) prior to use. The tRNA-Leu (UAA) intron from a selection of cyanobacteria and plastid containing eukaryotes (Table 3.5.2) were amplified using a modified LeuF and LeuR primer pair. A T7 promoter site was attached to the 5 ‘ end of the forward primer and used to generate RNA transcripts for in vitro splicing assays. The intron constructs contained 20 nucleotides of 5’ and 41 nt of 3’ flanking tRNA sequence, including the regions required for the P1 and P10 interactions (Cech et al. 1994). PCR products were incubated with Klenow enzyme, which polishes the sticky ends with a 3'->5' exonuclease activity (Liu and Schwartz, 1992).Run-off transcription were performed in a 20 µl reaction volume and contained 4 µl Transcription Buffer (Promega), 2 µl DTT (100 mM), 20 u RNAisin ribonuclease inhibitor, (0.5 µl), 4µl rNTPs (1 µl of each), 8.5 µl DNA/water and 1µl T7 RNA polymerase to give a final reaction volume of 20 µl and incubated at 37º C for one hour. The mixture was extracted with phenol/chloroform and the RNA was precipitated with ethanol.

Protocols for the assay of intron splicing were based upon the method of Zuag et al. (1993). The RNA was incubated in 2.5 µl of 250 mM HEPES, 0.75 µl of 500 mM $MgCl_2$ and 1.75 µl DEPC treated water. The reaction mixture was pre-incubated at 50º C for 15 minutes to renature the intron. The mixture was then incubated at 32º C for 2-3 minutes and splicing was initiated by the addition of 0.5 µl of rGTP was added and incubated for a further 15 minutes to promote splicing (Zuag et al. 1993). The reaction was mixed gently and incubated at 32º C for 60 minutes.

2.6.7. Northern analysis

The RNA splicing products were ethanol precipitated with 0.5 volumes of ammonium acetate and 2.5 volumes 100% ethanol and incubated at -20º C for 30 mins. The products were resuspended in 10 µl of RNA loading buffer and denatured for 10 minutes at 75 ºC. The RNA splicing products were separated on polyacrylamide (5%)/7 M urea gels and stained with ethidium bromide prior to photographic documentation. These gels were blotted onto positively charged nylon membranes (Amersham) through capillary blotting and cross-linked to the membrane. A sequences which included both intron and flanking exon regions was 3' end labelled. Hybridization was at 50º C overnight with final stringent washes of 0.1 x SSC and 0.1% (w/v) SDS at 60º C. DNA-RNA hybrids were visualized by exposure to x-ray plates.

3.1 *Chroococcidiopsis* and the heterocysts differentiating cyanobacteria are each others closest living relatives

3.1.1 Abstract

Many filamentous cyanobacteria reduce atmospheric nitrogen in specialized differentiated cells called heterocysts. In this study phylogenetic evidence is presented that shows members of the unicellular non-heterocyst differentiating genus *Chroococcidiopsis* and filamentous heterocyst differentiating cyanobacteria are each other's closest living relatives. Distance, maximum parsimony and maximum likelihood analyses of complete small subunit ribosomal RNA gene sequences yielded highly congruent support for the monophyly of *Chroococcidiopsis* and the heterocyst differentiating cyanobacteria. Additional support for the controversial sister taxon relationship was achieved through independent and combined phylogenetic analyses of *rpoC1*, *tufA* and 16S rRNA gene datasets. These results demonstrate that the order Pleurocapsales which traditionally contains *Chroococcidiopsis* is a polyphyletic assemblage with the ability to reproduce by multiple fission having arisen independently at least twice during the cyanobacterial radiation. This data also rejects *Myxosarcina* as a sister taxon to *Chroococcidiopsis* indicating that the numerous presumed shared derived characters thought to unite the two genera evolved independently. The sequence divergence within the *Chroococcidiopsis* lineage is comparable to and probably exceeds that in the entire heterocyst differentiating lineage. *Chroococcidiopsis* forms unique survival cells under nitrogen limiting conditions, and the sister-group relationship with the heterocystous cyanobacteria shown here suggests that differentiation of these cells and heterocysts may be related processes. Moderate bootstrap support for the monophyly of nearly all previously identified sequence groups was found. However, little support for deep level relationships between groups of modern cyanobacteria was recovered. The phylogenetic analyses presented here are consistent with a near simultaneous radiation of modern cyanobacteria.

3.1.2 Introduction

The origin and diversification of cyanobacteria are amongst the most interesting and controversial problems in microbial evolution. The use of small subunit rRNA gene to determine the evolutionary relationships among cyanobacteria has become more common in recent years (Wilmotte et al. 1993, Wilmotte et al. 1994, Nelissen et al. 1995, Turner 1997, Turner et al. 1999, Ishida et al. 2001). These studies have established a systematics of cyanobacteria based on evolutionary relationships inferred by a direct comparison of homologous genes and provided consistent support for the monophyly of just one of the orders of cyanobacteria (Turner 1997). Many aspects of cyanobacterial evolution continue

*Part of this section has been accepted for publication (Fewer et al. 2002).

to be a matter of debate and the branching order of modern cyanobacteria remain unresolved with the use of the small subunit rRNA (Turner et al. 1999). One such intractable problem is the controversial sister taxon relationship between *Chroococcidiopsis* and heterocyst differentiating cyanobacteria.

Members of the genus *Chroococcidiopsis* are widespread cosmopolitan unicellular cyanobacteria which grow in a near unparalleled range of extreme environments. They have been found in airspaces of porous rocks from Antarctic valleys and hot deserts worldwide where they often represent the sole photosynthetic organism (Friedmann and Ocampo-Friedmann 1976, Friedmann and Ocampo-Friedmann 1984, Büdel 1999). They are found in freshwater, marine and hypersaline environments (Dor et al. 1991) and have also been described from hot springs in Indonesia and nitrate caves worldwide (Geitler 1933, Friedmann 1962). They are also frequent partners in lichen symbioses in Savannah regions (Büdel and Henssen 1983, Büdel and Wessels 1991, Büdel et al. 2000). *Chroococcidiopsis* can survive prolonged desiccation and are amongst the most ionizing-radiation resistant bacteria known (Billi et al. 2000). Morphological characteristics of *Chroococcidiopsis* and resemblance to certain Proterozoic microfossils have been proposed to make it the most primitive living cyanobacterium (Friedmann et al. 1994). The recent focus on exobiology and the potential for life in extreme environments has generated a great deal of interest in microorganisms like *Chroococcidiopsis.* In keeping with this *Chroococcidiopsis* has been proposed as a candidate for terraforming Mars specifically because of its perceived antiquity and the ability to survive in such extreme environments (Friedmann and Ocampo-Friedmann 1995).

The genus *Chroococcidiopsis* has been traditionally assigned to the order Pleurocapsales which is comprised of unicellular cyanobacteria that reproduce by multiple fission (Waterbury and Stanier 1978, Waterbury 1989). This is a complex differentiation process involving either simultaneous division or the rapid successive division of single cells to yield a multitude of minute cells called baeocytes (Waterbury and Stanier 1978, Kunkel 1984, Waterbury 1989). Their common mode of reproduction is thought to distinguish members of the order Pleurocapsales from all other cyanobacteria (Rippka et al. 1979). No comparable division process is known in other bacteria and initial phylogenetic studies suggested that reproduction by multiple fission has evolved just once during the radiation of prokaryotes (Giovannoni et al. 1988, Wilmotte et al. 1994).

It has been suggested that the genus *Chroococcidiopsis* is polyphyletic and by implication that the order Pleurocapsales is also polyphyletic (Reeves 1996). Subsequently an association between *Chroococcidiopsis thermalis* PCC 7203 and heterocyst forming cyanobacteria was inferred with 16S rRNA sequence analyses (Turner 1997). Bootstrap analyses either directly contradicted (Ishida et al. 2001), found moderate support for the sister taxon relationship (Priscu et al. 1998, Bhattacharya et al. 1999),

or could not resolve the position of *Chrooococcidiopsis* (Rudi et al. 1997, Billi et al. 2001). This has raised doubts about the stability of the association between *Chroococcidiopsis* and the heterocyst forming cyanobacteria.

To resolve the phylogenetic position of *Chroococcidiopsis* and to investigate the monophyly of the Pleurocapsalean assemblage complete 16S rRNAs from eight additional strains assigned to the genus *Chroococcidiopsis*, representatives from remaining uncharacterized and characterized genera of the order Pleurocapsales, and two heterocyst forming cyanobacteria were amplified and sequenced. In tandem, sequences from the *rpoC1* and *tufA* genes were obtained for a number of cyanobacteria and *Chroococcidiopsis* PCC 7203. Distance, maximum parsimony and maximum likelihood analyses of individual and concatenated gene datasets consistently support the sister-group relationship between *Chroococcidiopsis* and the heterocyst forming cyanobacteria and the polyphyly of the order Pleurocapsales. Phylogenetic analyses presented here support the suggestion that there has been a near simultaneous ancient diversification of modern cyanobacteria.

3.1.3. Results and Discussion

3.1.3.1 16S rRNA dataset

Phylogenetic analyses here of the 16S rRNA dataset using a combination of distance, maximum parsimony and maximum likelihood methods yielded highly congruent support for the respective monophyly of two clades, one containing the bulk of the *Chroococcidiopsis* strains and the other containing the remaining genera of the order Pleurocapsales (Fig. 3.1.1). Two strains that have been assigned to *Chroococcidiopsis* sp., CCMP 1489 and PCC 6712 (Rippka and Herdman 1992, Andersen et al. 1997), did not cluster with other strains of that genus, but fell within a clade representing the bulk of the order Pleurocapsales (Fig. 3.1.1). A monophyletic Pleurocapsales inclusive of *Chroococcidiopsis* did not receive any bootstrap support and statistical tests rejected the forced monophyly of the Pleurocapsales (Table 3.1.1).

Phylogenetic analyses can yield results that are artifacts of outgroup choice, heterogeneous base composition, site-specific rate variation and taxon sampling (Swofford et al. 1996). Analyses in which each of the outgroup taxa were in turn specified as the sole outgroup taxon consistently resulted in strong

bootstrap support for the sister taxon association irrespective of which taxon was chosen as outgroup (98-99%). The association also received strong support using the LogDet transformation (99%), even with the removal of the constant sites (98%). LogDet transformation addresses potential concerns with heterogeneity of base composition (Lockhart et al. 1994). Finally, minimum evolution bootstrap analyses that accounted for site-specific rate variation with a gamma distribution of rates provided also high bootstrap support (99%) for the association (Fig. 3.1.1). A published 16S rRNA tree has recorded high (99%) neighbor-joining local bootstrap probabilities for *Chroococcidiopsis* and *Microcoleus* PCC 7420 to the exclusion of both heterocystous cyanobacteria and the remainder of the Pleurocapsales (Ishida et al. 2001). The inclusion of *Microcoleus* PCC 7420 in the dataset, resulted in reduced support for the sister taxon relationship (37%). However, no support for a sister taxon relationship between *Microcoleus* PCC 7420 and *Chroococcidiopsis* PCC 7203 was found in neighbor-joining analyses. Furthermore, the addition of the 4 new 16S rRNA sequences of *Chroococcidiopsis* yielded higher support of the sister taxon relationship between *Chroococcidiopsis* and the heterocyst-forming cyanobacteria (87%) to the exclusion of *Microcoleus* PCC 7420. The phylogenetic analyses presented here suggests that the inclusion of *Microcoleus* PCC 7420 is enough to lower support for the relationship but not to destablize it when more than one sequence from *Chroococcidiopsis* is included. This suggests that taxon sampling and long branch attraction may be important factors in the recovery of support for the sister taxon relationship.

The inclusion of the two strains PCC 6712 and CCMP 1489 with the remaining genera of the Pleurocapsales rather than the *Chroococcidiopsis* lineage was well supported in the 16S rRNA analyses (Fig. 3.1.1). However, the exact positions of these two strains in this clade were ambiguous (Fig. 3.1.1). A primary subdivision was found between strains of *Pleurocapsa*, *Myxosarcina, Dermocarpella*, and strain CCMP 1489, which were always allied in one clade with high bootstrap support, and *Xenococcus*, *Staniera* and strain PCC 6712 which were always basal to this clade (Fig. 3.1.1). Strain PCC 6712 was originally reported as *Chlorogloea* sp. (Kenyon 1972). Recent microscopic analysis of this strain revealed that in contrast to *Chroococcidiopsis* it does not form the characteristic cube-shaped colonies, that multiple fissions may occur immediately after the first binary division, and occasionally short pseudofilaments typical of *Pleurocapsa* have been observed (Büdel, unpubl. result). In addition physiological and genetic differences are known to set strain PCC 6712 aside from all other *Chroococcidiopsis* strains (Waterbury 1989, Rippka et al. 2001). Strain CCMP 1489 produces motile baeocytes (R.A. Andersen, unpubl. result) which does not warrant its assignment to the genus *Chroococcidiopsis* (Waterbury 1989, Rippka et al. 2001). The relationships of strains assigned to the genus *Chroococcidiopsis* found here parallels the results of comparisons of fatty acid content of these genera (Caudales et al. 2000). *C. thermalis* PCC 7203 was distinct from other members of the Pleurocapsales in its elevated percentages of polyunsaturated acids and was more similar to the heterocystous

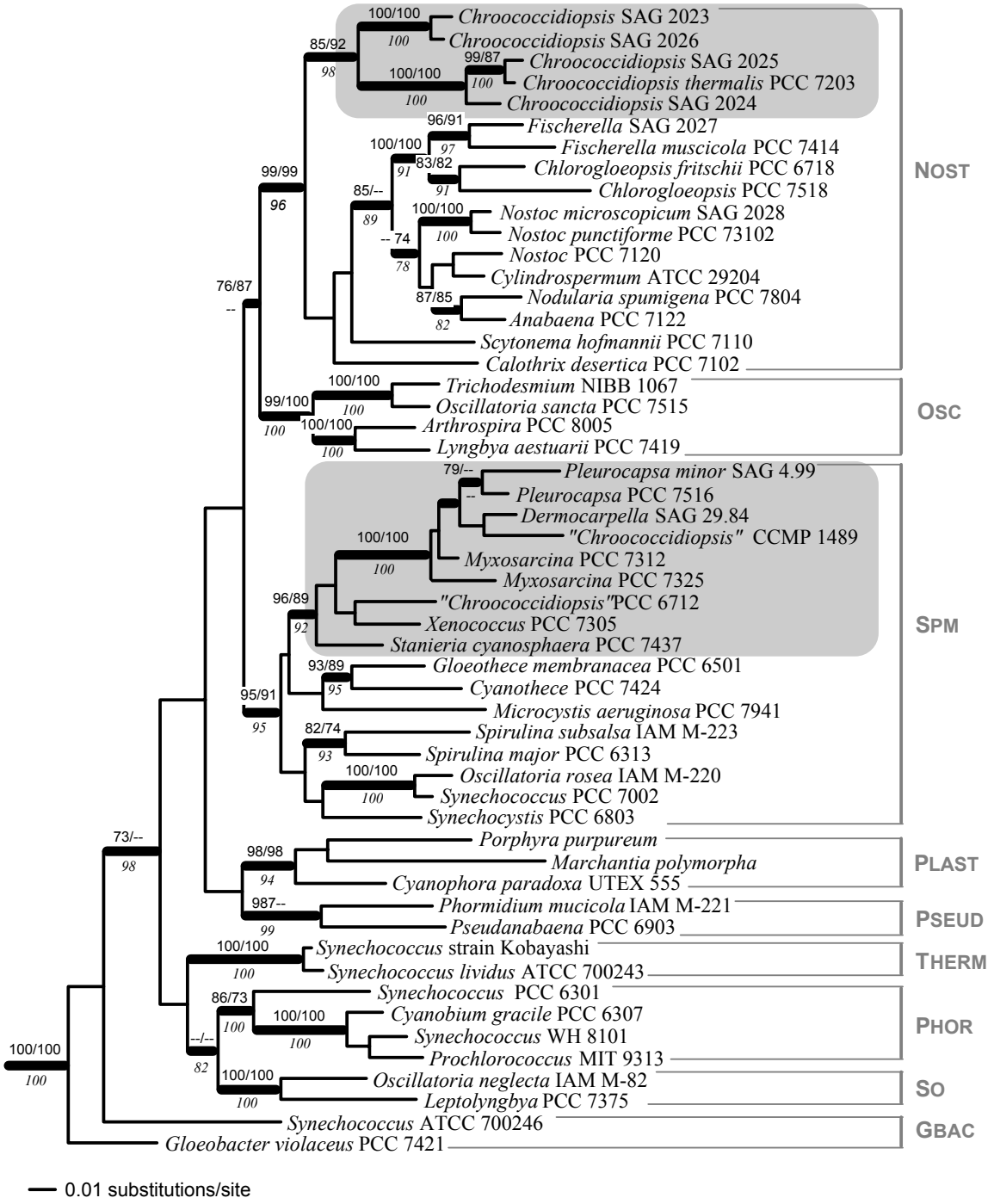

Figure. 3.1.1. Support for the sister taxon relationship between the genus *Chroococcidiopsis* and the heterocyst forming cyanobacteria. Boxed taxa form baeocytes and are part of the order Pleurocapsales. Sequences groups are marked on the right-hand side of the figure (Turner 1997). A maximum-likelihood tree showing based on the GTR+I+G model of DNA substitution (Ln-likelihood = -17,795.98). Thick lines mark internal nodes that were resolved in all distance, maximum parsimony and maximum likelihood trees (see text). Values above internal nodes are bootstrap values from 1000 replications as calculated from minimum evolution using the LogDet transformation (left) or the GTR+I+G model (right), values below branches were calculated from weighted parsimony. Only bootstrap values above 70% are shown.

Table 3.1.1. Results of the statistical tests performed on the monophyly of the Pleurocapsales, the *Myxosarcina-Chroococcidiopsis* sister taxon relationship and *Chroococcidiopsis* as the most primitive living cyanobacterium. The additional number of steps to the most parsimonious tree and the decrease in the ln likelihood ratio (second column) is indicated after the P value (first column).

User defined topology	Maximum Parsimony						Maximum Likelihood	
Kishino-Hasegawa			Templeton		Winning sites		Kishino-Hasegawa	
16S rRNA dataset (53 taxa)								
A monophyletic Pleurocapsales	0.0001*	+24	0.0001*	+38	0.0002*	+38	0.0011*	+51.3
Myxosarcina-Chroococcidiopsis alliance	<0.0001*	+66	<0.0001*	+79	<0.0001*	+79	<0.0001*	+219.5
Most primitive living cyanobacterium	<0.0001*	+33	<0.0001*	+65	0.0001*	+65	<0.0001*	+79.6
Combined dataset (19 taxa)								
A monophyletic Pleurocapsales	<0.0001*	+69	<0.0001*	+131	<0.0001*	+100	<0.0001*	+115.3
Myxosarcina-Chroococcidiopsis alliance	<0.0001*	+112	<0.0001*	+174	<0.0001*	+143	<0.0001*	+191.0
Most primitive living cyanobacterium	<0.0001*	+69	<0.0001*	+147	<0.0001*	+108	<0.0001*	+90.7

Note: The best tree length in maximum parsimony analysis of the 16S rRNA dataset is 3662. The best tree length in maximum parsimony analysis of the combined dataset is 3822. The ln Likelihood score in maximum likelihood analysis of the 16S rRNA dataset is -17,795.98. The ln Likelihood score in maximum likelihood analysis of the combined dataset is –18,386.43. Due to the performance of multiple tests, the significance level of rejection of the null hypothesis is adjusted via the Bonferroni correction to $\alpha = 0.01$.

cyanobacteria. However, *Chroococcidiopsis* PCC 6712 and *C. polansiana* CCMP 1489 both contained high amounts of saturated and unsaturated straight-chain fatty acids like other members of the Pleurocapsales. Strains of the *Chroococcidiopsis* lineage conform to the botanical type species *C. thermalis* and strain PCC 7203 has been proposed as reference strain of this species (Waterbury 1989). Thus it seems legitimate to consider all strains of the clade containing *Chroococcidiopsis* PCC 7203 as the true representatives of the genus *Chroococcidiopsis*.

Numerous systematic and physiological studies have tacitly assumed that multiple fission arose on just a single occasion (Waterbury and Stanier 1978, Rippka et al. 1979, Waterbury 1989). However, the results of this study reveal robust support for two monophyletic clades each comprised of cyanobacteria which reproduce by multiple fission and that are located in different parts of the tree (Fig. 3.1.1). This demonstrates that multiple fission has arisen on at least two occasions during the radiation of cyanobacteria. Previous studies have suggested, either by direct comment or pictorially in phylogenetic trees that the order Pleurocapsales might be polphyletic and that the number of occasions on which multiple fission has arisen is more numerous (Reeves 1996, Turner 1997, Rudi et al. 1997, Garcia-Pichel et al. 1998, Billi et al. 2001, Ishida et al. 2001). It may be that multiple fission in cyanobacteria, while unique in the prokaryotic world, represents an accelerated series of successive binary fission and is thus only a minor modification of the more common reproduction process (Komárek and Anagnostidis 1999).

The sister taxon relationship between the two genera *Chroococcidiopsis* and *Myxosarcina* was not supported in this study. Statistical tests also rejected the forced monophyly of *Myxosarcina* and *Chroococcidiopsis* (Table 3.1.1). Various morphological analyses have maintained that *Myxosarcina* and *Chroococcidiopsis* are closely related (Waterbury 1989, Komárek and Anagnostidis 1999). *Chroococcidiopsis* and *Myxosarcina* are almost morphologically identical and can often only be distinguished by the motility of baeocytes in *Myxosarcina* and the lack of motility in *Chroococcidiopsis* (Waterbury and Stanier 1978). Care needs to be taken when assigning rank to cyanobacteria that bear a morphological resemblance to *Chroococcidiopsis* given the overwhelming lack of support for the sister taxon relationship between the morphologically indistinguishable genera *Myxosarcina* and *Chroococcidiopsis* (Table 3.1.1). This represents perhaps one of the most spectacular cases of convergent evolution in morphology in cyanobacteria to date. The reduced dataset which allows the inclusion of additional partial sequences from *Chroococcidiopsis* (Billi et al. 2001) revealed a highly stable substructure (Fig. 3.1.1). The tree suggests that *Chroococcidiopsis* which inhabit deserts of Antarctica and Israel are closely related. Lichenized *Chroococcidiopsis* have arisen on more than one occasion (Fig. 3.1.1). Two strains from lichen associations (SAG 2023 and SAG 2026) cluster with strains from various desert locations and this was well supported in bootstrap tests (Fig. 3.1.1). The 16S rRNA coding regions from the *Chroococcidiopsis* strains SAG 2025 and SAG 39.79 were almost identical (except for 7 positions). Subsequently, there was no resolution among *Chroococcidiopsis* SAG 2025, *C. thermalis* and *C. cubana* because the sequence difference among these strains was reduced to a single position in the reduced data set. Interestingly, the branch lengths among the endolithic strains from deserts were longer than among the other strains. Given the limitations of that data set which was reduced in the number of sites and ingroup taxa, the monophyletic origin of the *Chroococcidiopsis* lineage was also resolved, but only with moderate bootstrap support (Fig. 3.1.1).

It has been claimed that *Chroococcidiopsis* is the most primitive living cyanobacterium based on the dominance of *Chroococcidiopsis* in extreme cold, arid, hot and saline environments around the globe and morphological resemblance to certain Proterozoic microfossils (Friedmann et al. 1994). It has been proposed that the preference for extreme habitats may be due to an inability to compete with more discriminating, specialized or aggressive species that are abundant in the more moderate environments (Friedmann and Ocampo-Friedmann 1995). However, phylogenetic analyses and statistical tests presented here failed to support the view that *Chroococcidiopsis* is the most primitive living cyanobacterium (Table 3.1.1). Instead, *Gloeobacter violaceus* PCC 7421 was shown to be the most primitive type of cyanobacteria in agreement with earlier findings (Nelissen et al. 1995). A

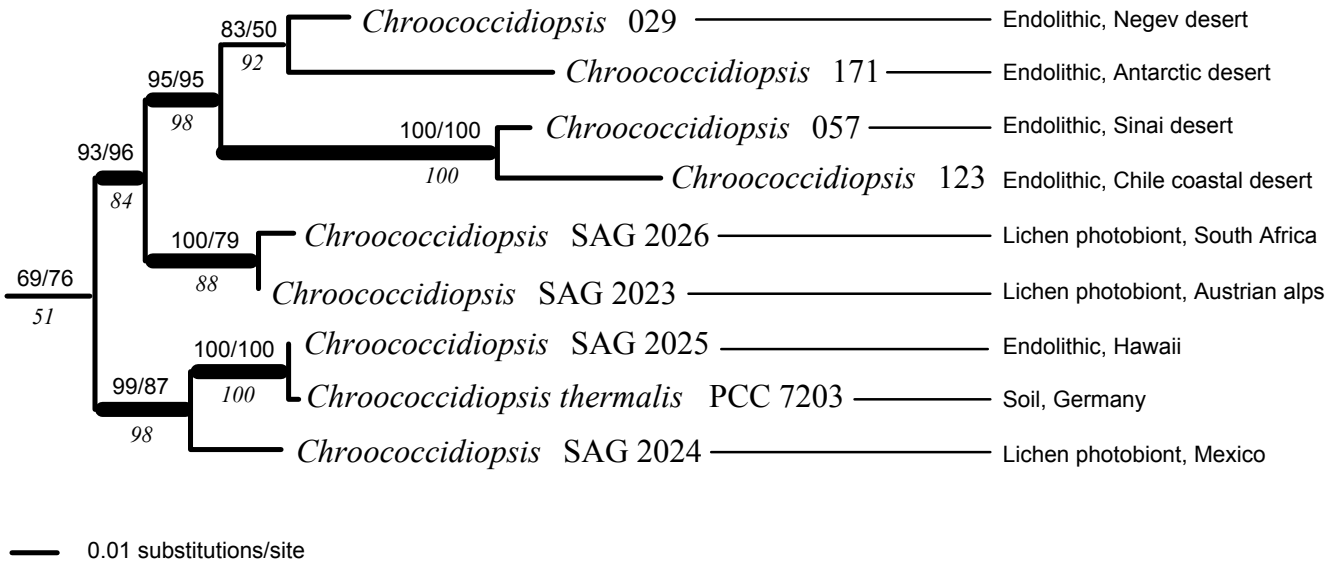

Figure 3.1.2. The phylogenetic structure within the *Chroococcidiopsi*s lineage based on partial sequences. A maximum-likelihood tree using the TrN+I+G model of DNA substitution (ln-likelihood = -3,108.47). Thick lines mark internal nodes that were resolved in all distance, maximum parsimony and maximum likelihood trees (see text). Values above the internal nodes are bootstrap values from 1000 replications as calculated from the LogDet transformation with a proportion of sites assumed to be invariable excluded (left), minimum evolution using the GTR+I+G model (right), and values below branches were calculated from weighted parsimony.

user-defined topology in which the *Chroococcidiopsis* lineage was forced to predate *Gloeobacter* PCC 7421 was rejected (Table 3.1.1). The genus *Gloeobacter* is characterized by the lack of thylakoid membranes (Rippka et al. 1979) and is generally accepted to be the earliest diverging cyanobacteria.

While not primitive *Chroococcidiopsis* may certainly be ancient. Cyanobacteria morphology is reputed to have remained unchanged for millions of years and fossils attributed to the genus *Chroococcidiopsis* may have a long geological history (Foster et al. 1989). The earliest unequivocal specimens of *Chroococcidiopsis* date from the Early Devonian Rhynie Chert 400 million years ago (Taylor et al. 1995). Genetic distances within the *Chroococcidiopsis* lineage (0.0054 to 0.1296) were comparable to that in the entire heterocystous cyanobacteria lineage (0.0208 to 0.2026). This roughly corresponded to the maximum genetic distances observed in the data set between different genera of the orders Nostocales and Stigonematales (0.1259). The largest distance among strains of *Chroococcidiopsis* was found between *C. cubana* SAG 39.79 and strain SAG 2023 (0.1296). The shortest distance observed was that between *C. thermalis* PCC 7203 and *C. cubana* SAG 39.79 (0.0054) which differed by only seven basepairs. The partial 16S rRNA *Chroococcidiopsis* sequences from hot and cold deserts world wide

(Billi et al. 2001) which clustered as sister taxon to the *Chroococcidiopsis* sequences presented here (Fig. 3.1.3) were not included in these estimates of genetic distance. Given that these partial 16S rRNA were even more divergent than the strains sequenced it is predicted here that the sequence divergence within the *Chroococcidiopsis* lineage actually exceeds that of the heterocystous cyanobacteria. The high sequence divergence is in keeping with the anticipation that genera like *Chroococcidiopsis* have been around for a long time and that the genus may represent a living fossil (Schopf 1994). Sequence divergence is expected to be large in taxa that have been around for a long time but which have not undergone parallel cladogeneses. This is supported by a number of studies which could find no rational basis for subdividing the genus *Chroococcidiopsis* into smaller genera or even species (Waterbury 1989, Friedmann and Ocampo-Friedmann 1995). Nevertheless, the strains of *Chroococcidiopsis* studied here may represent distinct species of cyanobacteria for which morphological or other characters have yet to be identified. Such cryptic speciation is not unprecedented (King and Hanner 1998) and the resolution of these conflicting hypotheses must await additional lines of evidence and characterization of additional strains of *Chroococcidiopsis*.

The heterocyst differentiating cyanobacteria, which are comprised of the orders Stigonematales and Nostocales, formed a single monophyletic lineage in most analyses (Fig. 3.1.1). While the Stigonematales were monophyletic in all analyses, members of the order Nostocales did not appear as an internally coherent clade (Fig. 3.1.1). Members of the order Stigonematales are united by the ability to form true branches. That monophyly of this group is well supported in the 16S rRNA analyses suggests that the ability to undergo division in more than one plane and to form true branches may be a good phylogenetic marker. However, lack of true branching does not define a monophyletic lineage. The order Nostocales, defined by the absence of true branching, appears as a paraphyletic assemblage (Fig. 3.1.1). *Calothrix* and *Scytonema,* two members of the Nostocales, are basal to a clade containing the order Stigonematales and the remainder of the order Nostocales in all analyses except for maximum parsimony (Fig. 3.1.1). In maximum parsimony trees, *Calothrix* and *Scytonema* were most closely associated with the *Chroococcidiopsis* lineage, but with insignificant bootstrap support. Maximum parsimony analyses did also not support the monophyly of the heterocystous cyanobacteria (Fig. 3.1.1). Though further work is needed, the results presented here are consistent with the idea that primitive heterocystous cyanobacteria divided in only one plane and did not form true branches (Turner 1997).

Chroococcidiopsis strains isolated from different environments are able to fix atmospheric nitrogen under anaerobic conditions (Rippka et al. 1979). *Chroococcidiopsis* is also known to differentiate specialized cells under nitrogen limiting conditions (Billi and Grilli-Caiola 1996). These cells in *Chroococcidiopsis* are currently interpreted as survival cells which allow survival through periods of

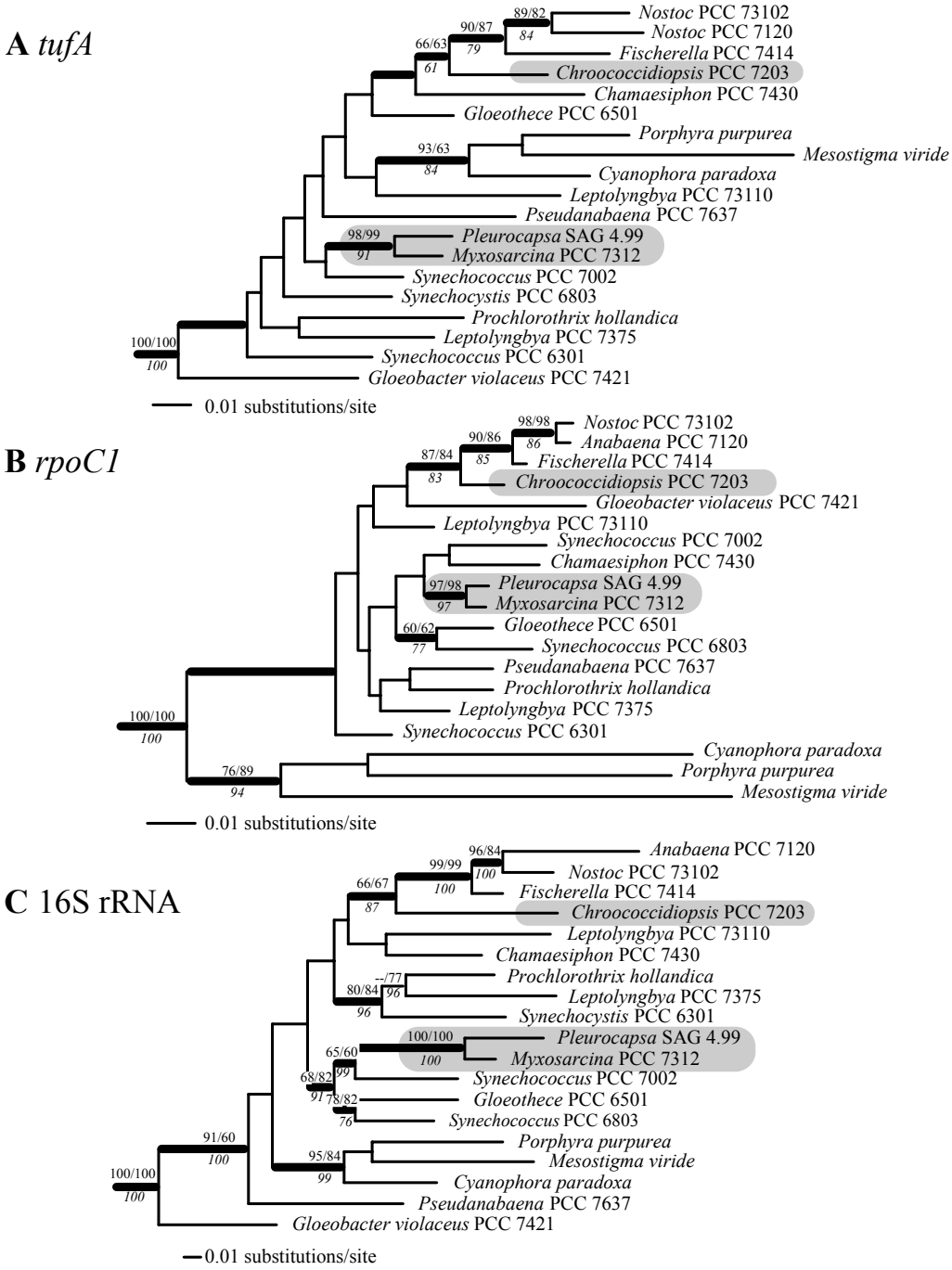

Figure 3.1.3. Phylogenetic support for the sister taxon relationship and the polyphyly of the order Pleurocapsales based on analyses of the individual datasets. Boxed taxa form baeocytes and are members of the order Pleurocapsales. A maximum likelihood trees of (A) *tufA* (lnL=-6174.43), (B) *rpoC1* (-lnL=5963.39) and (C) 16S rRNA (lnL=-10830.80). Thick lines mark internal nodes that were resolved in all distance, maximum parsimony and maximum likelihood trees (see text). Values above the nodes are bootstrap values from minimum evolution with the LogDet transformation and a proportion of sites assumed to be invariable calculated separately for each dataset and excluded (left), and with the maximum likelihood transformation (right) using the HKY+G, GTR+G and GTR+I+G models of DNA substitution for the *tufA*, *rpoC1* and 16S rRNA datasets respectively (right). Values below branches were calculated from weighted parsimony. 1000 bootstrap replications were used in each case.

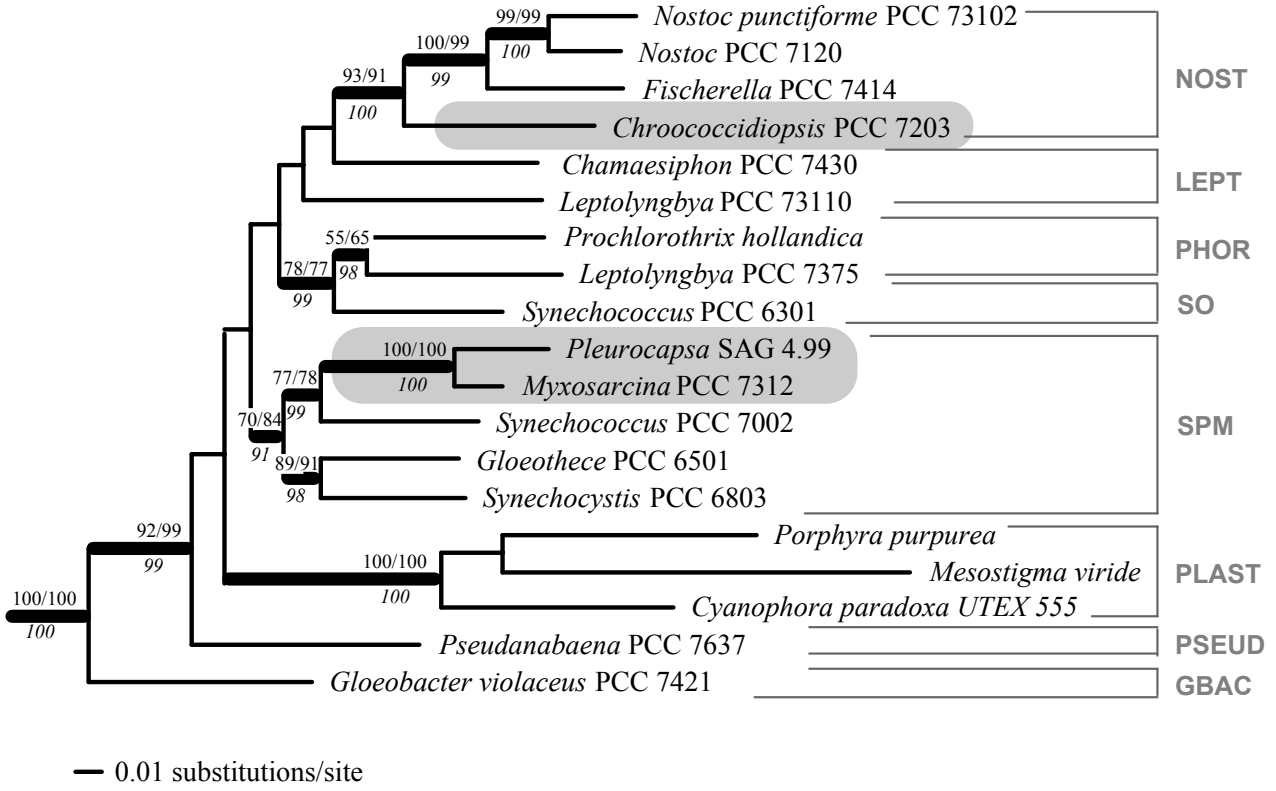

Figure 3.1.4. Phylogenetic support for the sister taxon relationship and the polyphyly of the order Pleurocapsales based on analyses of the combined dataset. Sequences groups are marked on the right-hand side of the figure (Turner 1997). Boxed taxa form baeocytes and are members of the order Pleurocapsales. A maximum likelihood based on the GTR+I+G model of DNA substitution (Ln-likelihood = -18,386.43). Thick lines mark internal nodes that were resolved in all distance, maximum parsimony and maximum likelihood trees (see text). Values above internal nodes are bootstrap values from 1000 replications as calculated from minimum evolution using the LogDet transformation (left) or the GTR+I+G model (right), values below branches were calculated from weighted parsimony. Only bootstrap values above 70% are shown.

nitrogen limitation and desiccation (Grilli-Caiola et al. 1993, Billi and Grilli-Caiola 1996). Heterocysts on the other hand allow prolonged survival in aerobic nitrogen limiting conditions (Wolk et al. 1994). Many heterocystous cyanobacteria also form akinetes and it has been proposed that the heterocyst differentiation mechanism is based on a pre-existing akinete differentiation mechanism (Wolk et al. 1994). Akinetes are restricted to the heterocyst forming cyanobacteria (Rippka et al. 1979). The survival cells of *Chroococcidiopsis*, and the heterocysts and akinetes of heterocyst differentiating cyanobacteria are formed under adverse environmental conditions. Little is known about the differentiation and exact function of the survival cells in *Chroococcidopsis*. However, the sister taxon

Table 3.1.2. Alignment length and bootstrap support for the monophyly of *Chroococcidiopsis* and heterocyst with individual and combined data sets.

Gene combination	Length (bp)	MP	ME	NJ	ML
All three genes combined	2,389	100	99	93	100
Protein genes combined	951	96	92	96	94
Small subunit ribosomal RNA (16S rRNA)	1,357	61	90	88	92
Gamma subunit of RNA polymerase (*rpoC1*)	429	54	29	68	71
Elongation Factor Tu (*tufA*)	522	67	31	37	62

Note: Maximum parsimony analyses (MP) weighted according to the rescaled consistency index, Neighbor joining analyses (NJ) and minimum evolution analyses (ME) performed with the LogDet transformation and invariable sites removed in proportion to base frequencies. Maximum likelihood (ML) according to the GTR model of DNA substitution with a gamma distribution of rates and invariable sites removed in proportion to base frequencies. The bootstrap values result from a 1000 replications for MP, ME and NJ and 100 replications for ML.

relationship with the heterocystous cyanobacteria shown here suggests that the developmental and physiological processes underlying the formation of survival cells in *Chroococcidiopsis* may be related with heterocyst differentiation. It is anticipated that identifying *Chroococcidiopsis* as the closest extant taxon to the heterocystous cyanobacteria will facilitate the illumination of the evolution of heterocyst differentiation.

3.1.3.2 Individual and combined analyses

Initial 16S rRNA analyses suggested that the support for the sister taxon relationship between *Chroococcidiopsis* PCC 7203 and the heterocysts forming cyanobacteria was sensitive to taxon choice. Inclusion of members taxa such as *Chamaesiphon* PCC 7430 and *Leptolygnbya* PCC 73110 reduced support for the relationship dramatically as well as other critical nodes in the tree. A multigene approach was developed to reaffirm support for the relationship and to resolve the relationship between these taxa and the *Chroococcidiopsis* lineage. The main principle behind combining data is that it allows for the amplification of phylogenetic signal and increased resolving power in those cases where signal is masked by homoplasy among individual datasets. The sequences from 16S rRNA, *rpoC1* and *tufA* genes were concatenated into a combined dataset. Individual loci did not provide consistent decisive support for the sister taxon association when *Chamaesiphon* PCC 7430 and *Leptolyngbya* PCC 73110 were included (Table 3.1.2). The partition homogeneity test revealed congruence (P = 0.156) between *tufA*, *rpoC1* and 16S rRNA data sets upon exclusion of the third positions from both the *rpoC1* and *tufA* datasets allowing the data to be combined. In keeping with this none of the strongly supported clades were mutually incompatible when third bases posi-

tions were excluded. Phylogenetic analyses based on the concatenated dataset provided much higher bootstrap support for the association in all analyses performed (90-100%). Even with the removal of constant sites from the dataset support for the sister taxon relationship always exceeded 90%. Support for the sister taxon relationship increased as a function of the number of basepairs resampled for both individual and concatenated datasets suggesting that much of the increased resolving power of the pooled data set relative to the individual data sets is because of the increased number of base pairs in the pooled data (Table 3.1.2).

There is little support in phylogenetic analyses of the 16S rRNA gene dataset for the relationship between the bulk of the groups of modern cyanobacteria (Fig. 3.1.1). The lack of resolution has been taken to imply the simultaneous divergence of modern groups of cyanobacteria (Giovannoni et al. 1988, Turner 1997). Although the events responsible for this apparent burst of evolution in the cyanobacterial line of descent are uncertain, as the first organisms to exploit water as the electron donor for photosynthesis, the common ancestor of the cyanobacteria and chloroplasts had available a novel and profoundly fertile physiological niche (Giovannoni et al. 1988). Molecular phylogenetic analyses in recent years have taken advantage of multigene concatenation to increase resolving power and provide solutions to previously intractable problems (Brown et al. 2001, Teeling et al. 2000, Springer et al. 1997). This current study makes use of an increase in phylogenetic signal but no concomitant resolution of the relationships was detected (Fig. 3.1.2). Deep level relationships between the sequence groups of cyanobacteria were unresolved in both individual and combined datasets this study (Figs. 3.1.3-3.1.2). Phylogenetic analyses of individual and combined datasets is consistent with a near simultaneous of cyanobacteria radiation in antiquity. If lineages diverge rapidly in time rapidly in time relative to the rate of character evolution there may be insufficient evidence present in short stretches of DNA sequences to reconstruct the exact order of splitting (Page and Holmes 1998). The exact branching order and timing of this radiation awaits the development of more reliable molecular phylogenies and accurate molecular clocks for the cyanobacteria.

3.2 Phylogenetic evidence for the antiquity of the intron interrupting the initiator tRNA gene in cyanobacteria

3.2.1 Abstract

Many phylogenetic studies on auto-catalytic group I introns have demonstrated the importance of horizontal transfer in the evolutionary history of these molecules. Horizontal transfer is proposed to be the dominating force in the distribution of the intron interrupting the gene encoding the initiator tRNA in cyanobacteria. In this study the intron is shown to have a scattered and sporadic distribution in modern cyanobacteria. Maximum likelihood, maximum parsimony and distance trees generated using sequences of both the introns and the 16S rRNA gene were highly congruent. The position of *Stanieria* PCC 7437 in the tRNA-fMet intron and 16S rRNA trees was discordant. However, analyses revealed support for just this single example of horizontal transfer. Two of the introns are shown to encode putative endonucleases, genetic elements known to promote horizontal transfer. Despite the presence of these endonucleases the phylogenetic placement of the intron cores was also congruent with the placement of the host 16S rRNA gene sequences. This demonstrates that it is the open reading frame which is the recent acquisitions and not the intron. Together this indicates a long-term immobility and vertical inheritance of the intron followed by subsequent loss in more derived lineages. The phylogenetic analyses strongly suggests the acquisition of the intron prior to the divergence of modern cyanobacteria making this intron the oldest known group I intron and strengthening the argument for the antiquity of this class of RNA enzyme.

3.2.2 Introduction

The evolutionary affinities of the auto-catalytic introns have long been of interest to evolutionary biologists and biochemists (Michel and Westhof 1990, Cech 1990). The existence of self-splicing and other RNA enzymes is interpreted as evidence for a primordial RNA world in which RNA functioned both in information storage and catalysis (Pace and Marshal 1985). Proponents of the RNA world hold that group I introns are ancient relics and trace their origin to the RNA world (Gilbert 1986, Cavalier-Smith 1991, Moore et al. 1993). Detractors point out that there is no evidence to support this and suggested instead that the distribution of the introns today is consistent with a recent origin of the group after transition from the putative RNA world to a largely DNA world (Diener 1989, Palmer and Lodgson 1991).

Tracing the historical origin of group I introns is made arduous by the ability of these introns to transpose themselves to novel locations. This is brought about by either the reversal of the auto-catalytic process leading to intron excision or the presence of site-specific endonuclease often encoded by the intron which facilitate horizontal transfer (Bell-Pedersen et al. 1990, Lambowitz and Belfort 1993). The emerging consensus is that the distribution of group I introns is overwhelmingly defined by horizontal transfer and that vertical inheritance has little or nothing to do with the distribution of the introns (Sogin et al. 1986, Hardy and Clark-Walker 1991, Nishida et al. 1993, Gargas et al. 1995, Turmel et al. 1995, Hibbett 1996, Cho et al. 1998, Nishida et al. 1998, Watanabe et al. 1998, Cho and Palmer 1999, Holst-Jensen et al. 1999, Perotto et al. 2000). It follows that disentangling the historical origins of group I introns is not a simple task.

A group I intron was shown to interrupt the tRNA-fMet gene which encodes the initiator transfer RNA in cyanobacteria seven years ago (Biniszkiewicz et al. 1994). To date 7 such introns have been characterised from cyanobacteria (Biniszkiewicz et al. 1994, Paquin et al. 1997). The group I intron in *Synechocystis* PCC 6803 can be divided into two functionally and structurally distinct domains. The intron core sequences form the characteristic secondary structure and tertiary elements necessary for catalyzing splicing (Cech et al. 1990). Emanating from the peripheral loops of the P1 core element in this structure is an endonuclease encoding open reading frames (Bonocora and Shub 2001). Many group I introns encode such site-specific endonucleases that impart mobility on the intron and catalyze their efficient spread from intron-containing to intronless alleles of the same gene in genetic crosses (Dujon 1980, Lambowitz and Belfort 1993, Belfort and Perlman 1995). The evolutionary importance of intron homing to the spread of group I introns across species barriers has been demonstrated repeatedly in the last decade (Hibbett 1996, Nishida et al. 1998, Bhattacharya et al. 1996, Turmel et al. 1995, Vaughn et al. 1995, Adams et al. 1998, Cho et al. 1998).

The sporadic distribution of the intron, the presence of a site-specific endonuclease in the intron of a single strain and supposed incongruence between the intron and the cyanobacterial host led workers to conclude that the distribution of the intron is explained by a series of recent horizontal transfers (Biniszkiewicz et al. 1994, Paquin et al. 1997, Paquin et al. 1999, Bonocora and Shub 2001). It has been pointed out that the phylogenetic evidence does not concur entirely with these workers conclusions (Rudi and Jakobsen 1999). Aspects of the tRNA-fMet intron topology are proposed to be congruent with the accepted host tree for cyanobacteria (Rudi and Jakobsen 1999, Turner et al. 1999). In addition, *Synechocystis* PCC 6803 contains the only known tRNA-fMet intron interrupted by a site-specific endonuclease (Bonocora and Shub 2001). It is generally accepted that group I introns predate such

such site-specific endonuclease (Gimble 2000). It is entirely possible that it is the endonuclease itself is the recent acquisition and not the intron.

Comparative studies of genome sequences have revealed that horizontal gene transfer has been very common over evolutionary time (Nelson et al. 1999). Horizontal gene transfer events have long been recognized as a principal force in the evolution of genomes (Doolittle 2000, Keeling and Palmer 2001). However, such events are often the subject of frequent and elaborate speculation and notoriously easy to invoke (Stanhope et al. 2001, Salzberg et al. 2001). Phylogenetic analyses arguably provide the most objective approach for determining the occurrence and directionality of horizontal gene transfer (Page 1994). While specific horizontal transfer events may be difficult to predict the genetic consequences of these events can be detected in sensitive molecular phylogenetic analyses. If the intron phylogenetic tree is discordant with the tree generated from 16S rRNA gene data then, the most probable explanation is that the sporadic distribution observed is indeed best explained by lateral transfer.

There has been no rigorous attempt to examine the evolutionary history of this intron in such a phylogenetic context. In order to test the two competing evolutionary hypotheses in this study a systematic survey to broaden the known distribution of the intron and to assess the frequency of the endonuclease encoding introns was undertaken. Phylogenetic analyses were employed to determine the level of congruence between the cyanobacteria host and intron. This allowed the resolution of the evolutionary history of the intron in cyanobacteria and shows that while horizontal transfer plays a role in the evolutionary history of this intron it is likely that the intron was present in the last common ancestor of modern cyanobacteria.

3.2.3 Results and Discussion

3.2.3.1 Scattered and sporadic distribution of the intron

In this study it is demonstrated that the intron interrupting the initiator tRNA is rare in cyanobacteria and has a scattered and sporadic distribution. A low-stringency PCR survey of 60 cyanobacteria and 3 chloroplasts identified just 9 taxa with tRNA-fMet genes interrupted by the group I intron (Table 3.2.1). The absence of the intron in the remaining taxa was confirmed by products corresponding in size to those expected for uninterrupted genes and automated sequencing across the insertion site for a subset of these products. The presence or absence of introns determined in 16 taxa in previous studies (Biniszkiewicz et al. 1994, Paquin et al. 1997) is here confirmed and these taxa were appended to the list (Table 3.2.1). The tRNA-fMet intron from *Stanieria* PCC 7437 was used to query the microbial genomes database on the NCBI BLAST page (www.ncbi.nlm.nih.gov/BLAST). The BLAST search results indicated that no introns were present in the genomes of the cyanobacteria *Prochlorococcus marinus* MED 4, *Prochlorococcus marinus* MIT 9313 or *Synechococcus* WH 8102. Sequence from the uninterrupted tRNA gene was used to query the database and revealed the presence of uninterrupted tRNA-fMet genes in each genome. The genome of *Synechocystis* PCC 6803 contains an intron as previously demonstrated (Biniszkiewicz et al. 1994, Paquin et al. 1997) while the genomes of *Nostoc* PCC 7120 and *Nostoc punctiforme* PCC 73102 both lack the intron as previously described (Paquin et al. 1997). Inspection of the complete plastid genomes of *Cyanophora paradoxa* (Stirewalt et al. 2000), *Porphyra purpurea* (Reith and Mullholland 1995), *Mesostigma* viride (Lemieux et al. 2000)*, Chlorella vulagaris* (Wakasugi et al. 1997), *Nephroselmis olivacea* (Turmel et al. 1999), *Marchantia polymorpha* (Shimada and Sugiura 1991) and *Pinus thunbergii* (Tsudzuki et al. 1994) revealed that all lack the intron. The sporadic nature of the intron is also supported the absence of the intron in *Synechococcus* PCC 6301 (Ecarot-Charrier and Cedergren 1976). Seven of the introns have been identified previously (Paquin et al. 1997) and 9 were novel to this study (Table 3.2.1). This more than doubles the number of tRNA-fMet introns known to science. The intron cores are similar in length (246-268 bp) with the exception of *Synechocystis* PCC 6803 and PCC 6906 which are 837 and 841 bp in length respectively. All introns share a high sequence homology (75% identity) and are all inserted between the universally conserved U of the anticodon loop and C of the CAU anticodon. Together this systematic survey revealed the presence of 16 introns in 90 taxa representing all the major cyanobacterial and plastids lineages at the ordinal level (Table 3.2.1).

Table 3.2.1. Strains of cyanobacteria and plastids used in this study as well as accession numbers for tRNA-fMet intron and 16S rRNA sequences (90 taxa). Taxonomic scheme according to Rippka and Herdman (1992) and Burger-Wiersma et al. (1989).

Strain	Source	16S rRNA	tRNA-fMet
Cyanobacteria			
Order Chroococcales			
Chamaesiphon polonicus SAG 32.87	This study	XXXX	———
Chamaesiphon subglobosus PCC 7430	Paquin et al. 1997	XXXX	———
Cyanobacterium stanieri PCC 7202	This study	AF132782	———
Cyanobium gracile PCC 6307	This study	AF001477	———
Cyanothece ATCC 51142	This study	AF132771	———
Cyanothece PCC 7418	This study	AF296872	———
Cyanothece PCC 7424	This study	AF132932	———
Dactylococcopsis salina PCC 8305	This study	AJ000711	———
Gloeobacter PCC 8105	This study	AF132791	XXXX
Gloeobacter violaceus PCC 7421	Paquin et al. 1997	AF132790	U83262
Gloeocapsa PCC 73106	This study	AF132784	———
Gloeothece membranacea PCC 6501	This study	X78680	———
Microcystis aeruginosa PCC 7806	This study	U03402	———
Microcystis aeruginosa PCC 7941	This study	U40340	———
Synechococcus elongatus PCC 6301	Ecarot-Charrier and Cedergren 1976	X03538	———
Synechococcus leopoliensis PCC 7942	This study	AF132930	———
Synechococcus PCC 7002	This study	AJ000716	———
Synechococcus PCC 7902	This study	AF216946	———
Synechococcus WH 8102	Genome	AF001480	———
Synechocystis PCC 6803	Paquin et al. 1997	D64000	U10482
Synechocystis PCC 6906	This study	XXXX	XXXX
Order Pleurocapsales			
Chroococcidiopsis BB 79.1	This study	XXXX	———
Chroococcidiopsis BB 80.1	This study	XXXX	———
Chroococcidiopsis BB 86.6	This study	XXXX	———
Chroococcidiopsis BB 90.5	This study	XXXX	———
Chroococcidiopsis CCMP1489	This study	XXXX	———
Chroococcidiopsis cubana SAG 33.74	This study	XXXX	———
Chroococcidiopsis PCC 6712	This study	XXXX	———
Chroococcidiopsis SAG 2023	This study	XXXX	XXXX
Chroococcidiopsis SAG 2025	This study	XXXX	XXXX
Chroococcidiopsis SAG 2026	This study	XXXX	XXXX
Chroococcidiopsis SAG 2027	This study	XXXX	XXXX
Chroococcidiopsis thermalis PCC 7203	This study	AB039005	———
Cyanosarcina BB 80.2	This study	XXXX	———
Dermocarpella SAG 29.84	This study	XXXX	———
Gloeocapsa BB 80.5	This study	XXXX	———
Myxosarcina PCC 7312	Paquin et al. 1997	XXXX	———
Myxosarcina PCC 7325	This study	XXXX	———
Pleurocapsa minor SAG 4.99	This study	XXXX	———
Pleurocapsa PCC 7516	This study	X78681	———
Stanieria cyanosphaera PCC 7437	Biniszkiewicz et al. 1994	AF132931	U10480
Xenococcus PCC 7305	This study	AF132783	———
Order Oscillatorales			
Arthrospira PCC 8005	This study	X70769	———
Geitlerinema PCC 7105	Paquin et al. 1997	AF132780	U83259

Leptolyngbya PCC 73110	This study	X84810	———
Leptolyngbya PCC 7375	Paquin et al. 1997	AF132786	———
Lygnbya aestuarii PCC 7419	This study	AJ000714	———
Oscillatoria acuminata PCC 6304	Paquin et al. 1997	XXXX	U83260
Oscillatoria neglecta IAM-M82	This study	AB003168	———
Oscillatoria rosea IAM-M220	This study	AB003164	———
Oscillatoria sancta PCC 7515	This study	AF132933	XXXX
Phormidium mucicola IAM-M221	This study	AB003165	———
Planktothrix prolifera NIVA 18	This study	X84811	———
Plectonema boryanum UTEX 485	This study	AF132793	———
Pseudanabaena catenata PCC 7403	Paquin et al. 1997	AB039019	———
Pseudanabaena catenata PCC 7408	This study	AB039017	———
Pseudanabaena PCC 6903	Paquin et al. 1997	AF132778	———
Pseudanabaena PCC 7367	This study	AF091108	———
Spirulina major PCC 6313	Paquin et al. 1997	X75045	———
Spirulina subsalsa IAM-M223	This study	AB003166	———
Trichodesmium NIBB 1067	This study	X70767	———
Order Stigonematales			
Chlorogloeopsis fritschii PCC 6718	This study	AF132777	———
Chlorogloeopsis PCC 7518	This study	X68780	———
Fischerella muscicola PCC 7414	This study	AF132788	———
Fischerella SAG 2027	This study	XXXX	———
Nostochopsis BB 92. 1	This study	XXXX	———
Order Nostocales			
Anabaena cylindrica PCC 7122	This study	AF091150	———
Aphanizomenon flos-aquae PCC 7905	This study	AJ133154	———
Calothrix desertica PCC 7102	Paquin et al. 1997	AF132779	———
Cylindrospermum stagnale PCC 7417	Biniszkiewicz et al. 1994	AF132789	U83261
Nodularia sphaeocarpa PCC 7804	This study	AJ133181	XXXX
Nostoc DM 44	This study	XXXX	XXXX
Nostoc muscorum PCC 7120	Paquin et al. 1997	X59559	———
Nostoc punctiforme PCC 73102	Paquin et al. 1997	AF027655	———
Nostoc SAG 2028	This study	XXXX	———
Scytonema hofmannii UTEX 2349	Biniszkiewicz et al. 1994	AF132781	U10481
Prochorophyta			
Prochlorothrix hollandica NIVA 8/90	This study	AF132792	———
Prochlorococcus marinus MIT 9313	Genome	AF053399	———
Prochlorococcus marinus MED 4	Genome	AF001466	———
Chloroplasts			
Glaucocystophyta			
Cyanophora paradoxa UTEX 555	Stirewalt et al. 2000	U30821	———
Glaucocystis nostochinearum SAG 45.88	This study	X82496	———
Rhodophyta			
Chondrus crispus	This study	Z29521	———
Palmaria palmata	This study	Z18289	———
Porphyra purpurea	Reith and Mullholland 1995	U38804	———
Viridiplantae			
Chlorella vulagaris C-27	Wakasugi et al. 1997	AB001684	———
Marchantia polymorpha	Shimada and Sugiura 1991	X04465	———
Mesostigma viride NIES-296	Lemieux et al. 2000	AF166114	———
Nephroselmis olivacea NIES 484	Turmel et al. 1999	AF110138	———
Pinus thunbergii	Tsudzuki et al. 1994	D17510	———

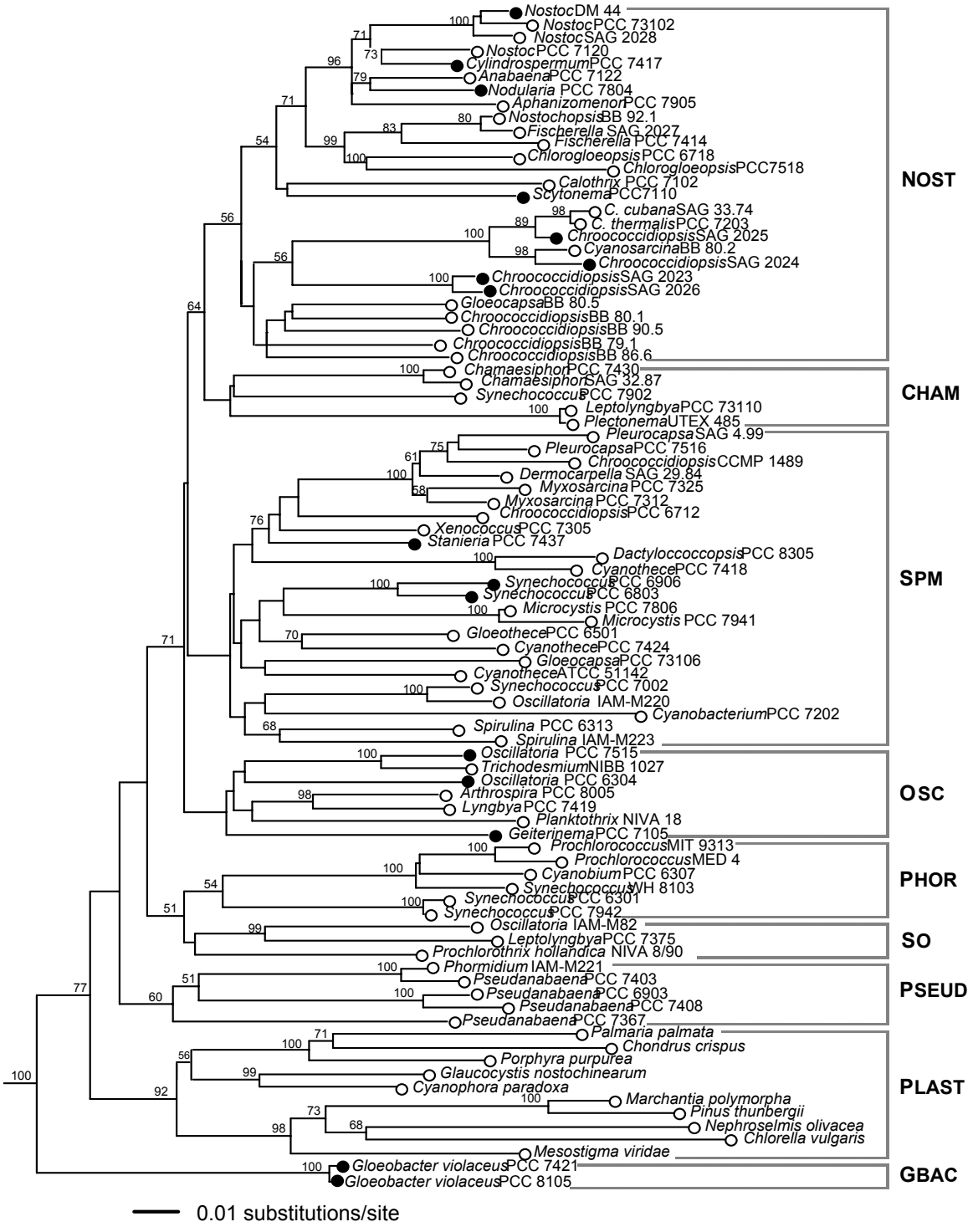

Figure 3.2.1. The sporadic and scattered distribution of the tRNA-fMet intron in cyanobacteria. The presence of an intron is denoted by filled circle and the absence of the intron by an empty circle. A minimum evolution tree based on the GTR+G+I model of DNA substitution. Branch lengths proportional to sequence change. Values above internal nodes are bootstrap values from 1000 replications as calculated from minimum evolution using the maximum likelihood transformation based on the GTR+G+I model of substitution. Only bootstrap values above 60% are shown.

A

```
PCC 6906  MGSKLKGDIAEKAAILQALKHGWGVLRPIGDRLSYDMVLDVEGLLLKIGV
PCC 6803  -ST--------Q----R---L-----K-L-------L-F-----------

PCC 6906  KSSWFSEKTGNHVVDNRRTKTNRRNMLRQPYGETDFDFALAYIEELDIFY
PCC 6803  -----------Y-------R-----IV-S--RGN-----V--V---E---

PCC 6906  VFPVEVFIDYGSEIHLVEVGKRQRKPRSSEYRGAWHLILQKRAAYKEICT
PCC 6803  ----D---S---------TD--------FG--E--------G--Q--TSA
```

B

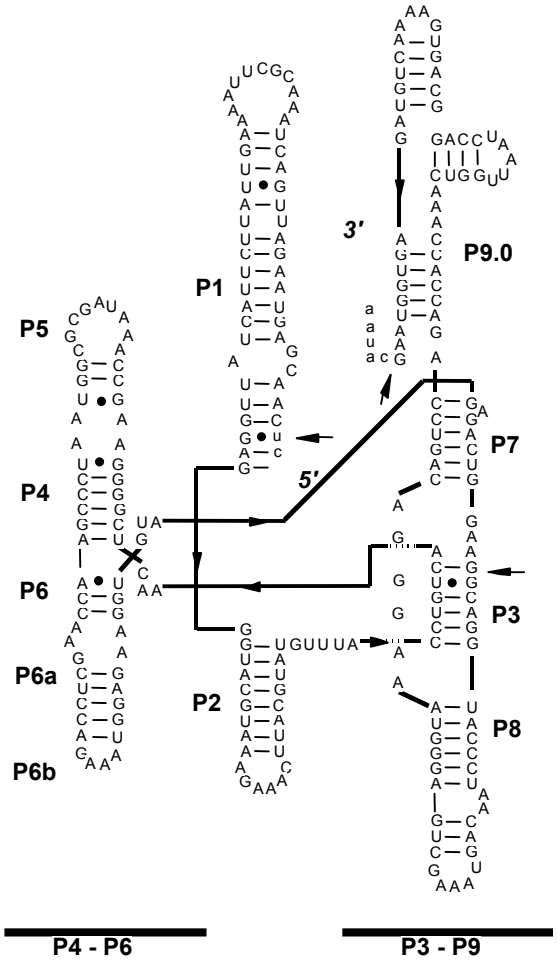

Figure 3.2.2. (A) Endonuclease amino acid composition in *Synechocystis* PCC 6803 and *Synechocystis* PCC 6906 which emanate from the P1 stem element of the intron catalytic core (B) A secondary structure models of the tRNA-fMet intron from *Scytonema hofmannii* UTEX 2349 showing the P1 stem element which the endonucleases in *Synechocystis* PCC 6906 and *Synechocystis* PCC 6803 interrupt.

The presence and absence of the intron in cyanobacteria was mapped to a minimum evolution tree based on the 16S rRNA gene (Fig. 3.2.1). This reaffirms the scattered and sporadic distribution of the

intron but demonstrates that it is present in many lineages of extant cyanobacteria (Fig. 3.2.1). Four genera of the heterocyst forming order Nostocales, *Nodularia*, *Scytonema*, *Nostoc*, and *Cylindrospermum*, contained tRNA-fMet genes interrupted by the intron. Interestingly, four members of the newly identified *Chroococcidiopsis* lineage (Section 3.1) contained tRNA-fMet introns. These two lineages of cyanobacteria are each others closest living relatives despite physiological and morphological dissimilarities (Section 3.1). A number of cyanobacterial lineages, as exemplified by the Turner sequence groups (Turner 1997) such as CHAM, PHOR, SO and PSEUD and plastids (PLAST) lack intron-containing representatives entirely (Fig. 3.2.1).

3.2.3.2 The open reading frames are recent acquisitions

Introns encoding open-reading frames were present in just two strains, *Synechocystis* PCC 6803 and PCC 6906 (Biniszkiewicz et al. 1994, Paquin et al. 1997). The open reading frames are both 453 base-pairs long and emanate from the P1 structural element (Fig. 3.2.2). They encode an 150 amino acid protein. Of the 150 residues. 117 were identical (78 % homologous). Even though there has been no attempt to assign a specific endonuclease function to the product of the ORF in *Synechocystis* PCC 6906 it is reasonable to assume that it also encodes a site-specific endonuclease given the high sequence homology between the two (Fig. 3.2.2). In addition, attempts at cloning of the PCR product containing the ORF showed that only those introns with ORFs containing point mutations or single base pair insertions that disrupted the open reading frame could be cloned. This suggests that the ORF encodes a functional endonuclease the expression of which is also toxic to *Escherichia coli* as has been shown for the ORF in *Synechocystis* PCC 6803 (Biniszkiewicz et al. 1994). The intron core of *Synechocystis* PCC 6803 and PCC 6906 were 85% identical while the 16S rRNA genes were 96% identical.

Many group I introns encode such site-specific endonucleases that catalyze their efficient spread from intron containing to intronless alleles of the same gene in genetic crosses (Dujon 1980, Lambowitz and Belfort 1993, Belfort and Perlman 1995). This process termed intron homing has been observed for introns located in a variety of mitochondrial and chloroplast genes (Delahodde et al. 1989, Wenzlau et al. 1989, Jacquier and Dujon 1985), in nuclear rRNA genes of the slime mold *Physarum* (Muscarella and Vogt 1989) and in protein genes of T-even phage (Quirk et al. 1989). Homing is initiated by the intron-encoded endonuclease, which makes a double stranded break at its target site within a recipient intronless allele and is then thought to proceed by the double-stranded-break repair pathway (Szotak et al. 1983). The evolutionary importance of intron homing to the spread of group I introns across species

barriers has been demonstrated repeatedly in the last decade (Hibbett 1996, Nishida et al. 1998, Bhattacharya et al. 1996, Turmel et al. 1995, Vaughn et al. 1995, Adams et al. 1998, Cho et al. 1998).

Most proteins encoded by group I introns ORFs exhibit two dodecapeptide motifs called LAGLIDADG (Hensgens et al. 1983). They probably belong to a large family of proteins whose members are present in several locations in the genomes. They could have evolved as independent mobile entities that colonized pre-existing group I introns providing them with a highly invasive property (Dujon 1989, Lambowitz and Belfort 1993, Sellem and Belcour 1997). Colonization of pre-existing group I introns by mobile ORFs is supported by circumstantial evidence that takes into account the various contexts similar ORFs can be found; examples include group I introns, an archael intron (Dalgaard et al. 1993), intergenic sequences (Sharma et al. 1992, Pel and Grivell 1993), and even coding sequences (Paquin et al. 1994). Moreover, in homologous group I introns the ORF sequence and position may vary (Mota and Collins 1988, Shub et al. 1988). The implied autonomous mobility of the ORFs has been experimentally demonstrated on a number of occasions (Sellem and Belcour 1997, Bechhoffer et al. 1994).

However, despite the presence of the endonuclease the phylogenetic position of the catalytic intron cores from *Synechocystis* PCC 6803 and PCC 6906 in phylogenetic trees derived from the intron cores is congruent with their phylogenetic position based on phylogenetic trees derived from their 16S rRNA genes (Figs. 3.2.3 and 3.2.4). The endonuclease encoding introns are restricted to one node on the intron generated tree (Fig. 3.2.4-3.2.4). In the systematic survey of over 90 cyanobacteria and chloroplasts just these two examples of endonuclease containing introns were identified (Fig. 3.2.1). Further inspection of publicly available databases revealed more that 30 uninterrupted tRNA-fMet genes. The open reading frame from *Synechocystis* PCC 6803 has been shown to express a endonuclease which recognizes a 20 basepair region of the initiator RNA gene (Bonocora and Shub 2001). The region recognized is conserved across the entire spectrum of cyanobacteria as well as in many eubacterial groups (Fig. 3.2.5). Despite conservation amongst innumerable bacterial and cyanobacterial tRNA-fMet genes this endonuclease has been found in just a single location in the 16S rRNA gene tree. It is proposed here that the endonuclease containing intron has arisen just once.

```
Species                          GenBank     -10  -5   0   +5   +10
```

	Endonuclease recognition	U10482	TCGTCGGGCTCATAACCCGA
Cyanobacteria	*Nostoc* PCC 7120	Genome	....
	Nostoc punctiforme PCC 73102	Genome	....
	Synechococcus WH 8102	Genome	....
	Prochlorococcus MIT 9313	Genome	....
	Prochlorococcus MED 4	Genome	....
	Synechococcus PCC 6301	K00311	....
	Synechocystis PCC 6803	U10482	....
	Gloeobacter PCC 7421	U83262	....
Chloroplasts	*Cyanophora paradoxa*	U30821	....
	Porphyra purpurea	U38804	....
	Cyanidium caldarium	AF022186	...................G
	Guillardia theta	AF041468	....
	Skeletonema costatum	AJ132263	....T...............
	Odontella sinensis	Z67753	....T...............
	Euglena gracilis	V00158	....A..........T—TG
	Astasia longa	NC_002652	....G........G......
	Toxoplasma gondii	NC_001799	............G......
	Mesostigma viride	AF166114	...TT...............
	Nephroselmis olivacea	AF137379	...CA.............TG
	Chlorella vulagaris	AB001684	....AA...........TT—
	Prototheca wickerhamii	AJ236874	....AA...........TT.
	Scenedesms obliquus	K00313	....G...........T—C
	Chlamydomonas rheinhardtii	X13879	....G..........T—TG.
	Marchantia polymorpha	11640	...CAA...........TTG
	Hordeum vulgare	AJ011807	..ACGA...........TTG
	Pinus thunbergii	D17510	..GCAA...........TTG
	Oenothera elata	AJ271079	..GCAA...........TTG
	Orobanche minor	AJ007722	..GCAA...........TTG
	Lotus japonicus	AP002983	..GCAA...........TTG
	Epifagus virginiana	M81884	..GCAA...........TTG
	Spinacia oleracea	AJ400848	..GCAA...........TTG
	Oryza sativa	X15901	..ACGA...........TTG
	Arabidopsis thaliana	AP000423	..GCAA...........TTG
	Zea mays	L02941	..ACGA...........TTG
	Nicotiana tabacum	Z00044	..GCAA...........TTG
Eubacteria	*Aquifex aeolicus*	AE000736	...C..............G
	Thermotoga maritima	AE001788	....T.............A.
	Deinococcus radiodurans	AE002006	...................G
	Thermus thermophilus	K00309	....
	Chlorobium tepidum	Genome	..................G.
	Neisseria meningitidis	AE002554	....
	Xylella fastidiosa	AE003877	....
	Vibrio cholerae	AE004123	....
	Treponema pallidum	AE001201	...................G
	Borrelia burgdorferi	AE001162	....
	Staphylococcus epidermidis	AF270116	....
	Mycobacterium smegmatis	AF330217	...CT.............AG
	Helicobacter pylori	AE000599	....TG.............A

Figure 3.2.3. Sequences of tRNA-fMet gene from bacteria and chloroplasts showing the proposed conservation of the endonuclease recognition sequence. Mutations in this region abolish endonuclease activity (Bonocora and Shub 2001). The accession genbank numbers for the tRNA-fMet genes are given to the right of the species name. 'Genome' indicates that the sequence for the tRNA-fMet gene was taken from the unpublished complete genome database at NCBI (www.ncbi.nlm.nih.gov). A dot indicates a basepair identical with the recognition sequence.

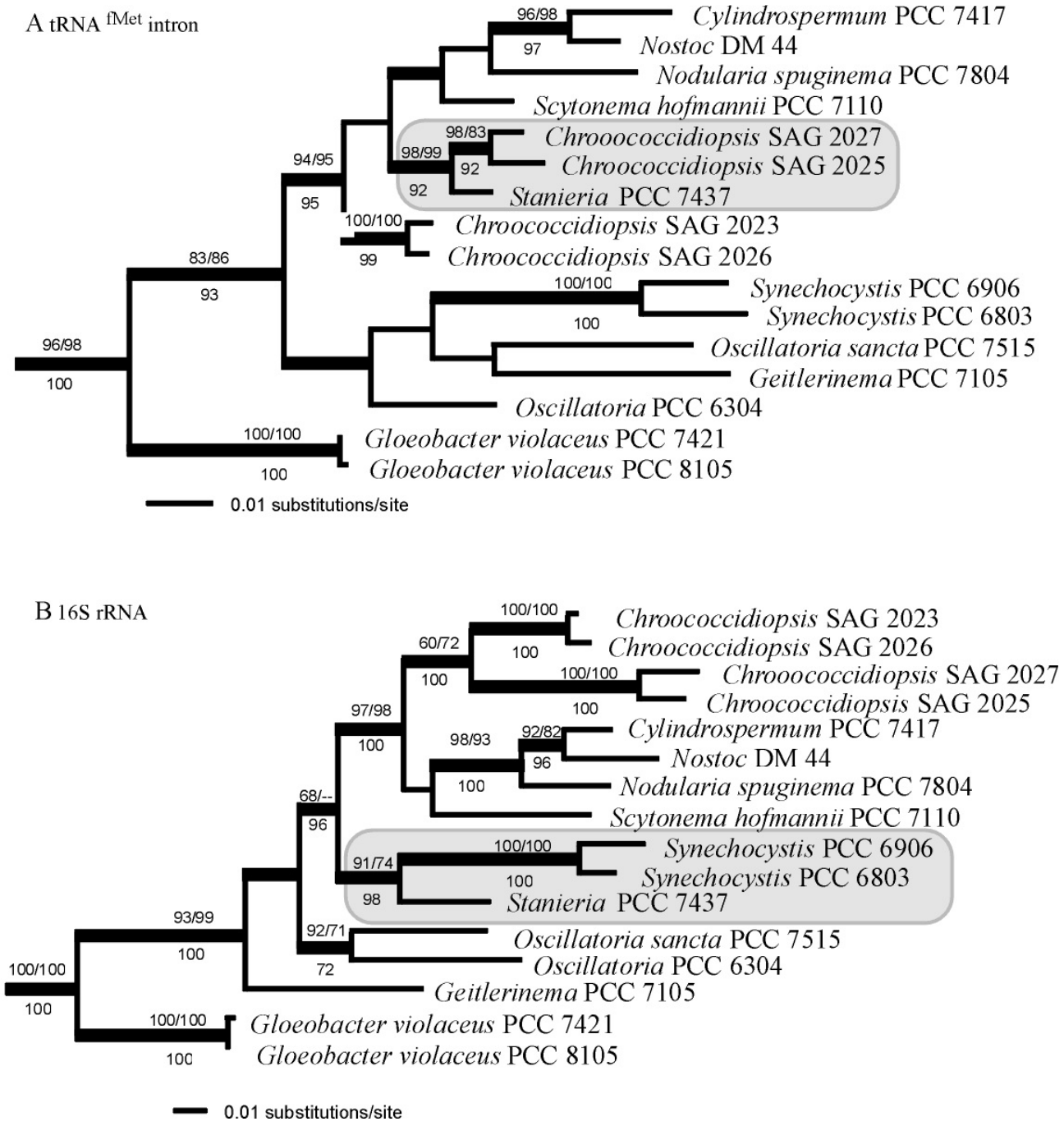

Figure 3.2.4. Phylogenetic trees showing the discordant position of *Staniera* PCC 7437 in the tRNA-fMet and 16S rRNA gene trees. The nearest relatives to the *Stanieria* PCC 7437 intron and 16S rRNA gene are boxed. (A) tRNA-fMet intron maximum likelihood tree based on the TIM+G model of DNA substitution (lnL=-4279.28). (B) 16S rRNA maximum likelihood tree based on the GTR+I+G model of DNA substitution (lnL=-8728.49). In both trees thick lines mark internal nodes that were resolved in all distance, maximum parsimony and maximum likelihood trees (see text). Values above internal nodes are bootstrap values from 1000 replications as calculated from minimum evolution using the LogDet transformation (left) or the TIM+G/GTR+I+G model (right), values below branches were calculated from weighted parsimony. Only bootstrap values above 60% are shown.

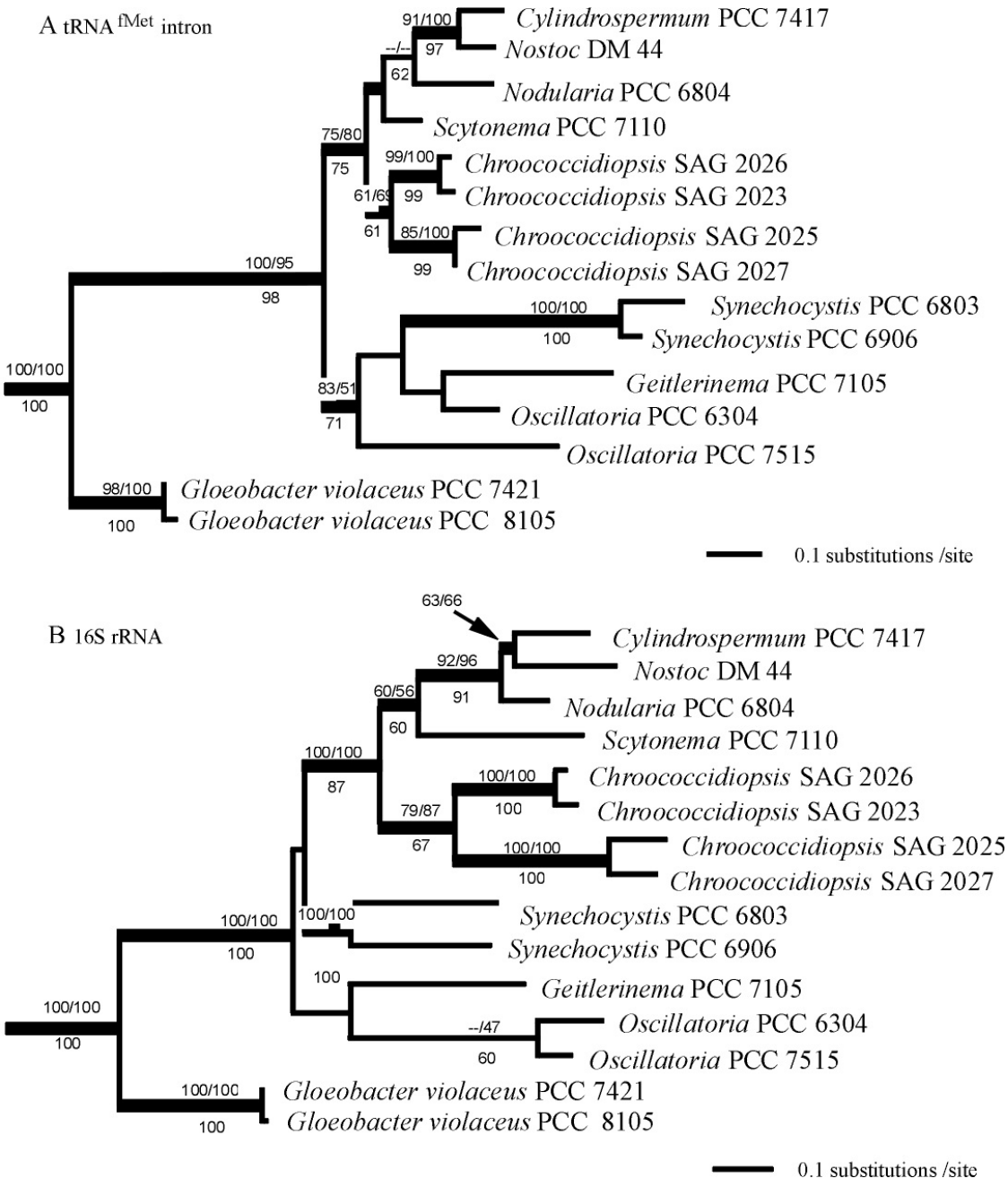

Figure 3.2.5. Phylogenetic trees showing the congruence between the remaining tRNA-fMet intron cores and 16S rRNA genes upon the exclusion of the *Staniera* PCC 7437 intron from analyses. (A) tRNA-fMet intron maximum likelihood tree based on the TIM+G model of DNA substitution (lnL=-4198.18). (B) 16S rRNA maximum likelihood tree based on the GTR+I+G model of DNA substitution (lnL=-8507.80). In both trees thick lines mark internal nodes that were resolved in all distance, maximum parsimony and maximum likelihood trees (see text). Values above internal nodes are bootstrap values from 1000 replications as calculated from minimum evolution using the LogDet transformation (left) or the TIM+G/GTR+I+G model (right), values below branches were calculated from weighted parsimony. Only bootstrap values above 60% are shown.

3.2.3.3 Instances of congruence between the host and intron

To assess the relative contributions of horizontal and genetic transmission to the introns phylogenetic history the 16S rRNA gene for strains that contained the intron was sequenced and compared to the intron phylogenies. Phylogenetic analyses employing maximum likelihood, maximum parsimony and distance methods revealed a highly stable substructure within the tRNA-fMet intron despite its short length (Fig. 3.2.4). Support for the sister taxon relationship between *Chroococcidiopsis* and heterocysts forming cyanobacteria was recovered (Figs. 3.2.4 and 3.2.5). All methods yielded congruent support for the monophyly of the newly identified *Chroococcidiopsis* lineage and the heterocyst forming cyanobacteria. Bootstrap support for a monophyletic order Nostocales was also recovered (60-82%). However, the support for a monophyletic *Chroococcidiopsis* was low (22-60%). The introns from *Gloeobacter* violaceus were resoundingly placed as the earliest branching introns taxa in accordance with the phylogenies based on 16S rRNA and a combination of *tufA*, *rpoC1* and 16S rRNA (Section 3.1). Support for the early placement of *Gloeobacter* intron ranges from 95-100%. Previous report that *Gloeobacter* grouped with *Cylindrospermum* in maximum likelihood analysis (Paquin et al. 1997) were not supported here. The two introns which encode putative endonucleases group together in all analyses with high bootstrap support. *Geitlerinema* PCC 7105, *Oscillatoria* PCC 7515 and *Oscillatoria* PCC 6304 grouped together (70-80%) but the relationship between the three intron cores was not fully resolved. In all analyses highly congruent support for the monophyly of the intron from *Stanieria* PCC 7437 and *Chroococcidiopsis* SAG 2025 was obtained (94-100%). This is in keeping with the previous studies which showed that *Stanieria* PCC 7437 occupied a position in introns trees that was incongruent with the position it occupies in trees based on the 16S rRNA gene (Paquin et al. 1997). However, in this study there is quite robust support for placing the intron from *Stanieria* PCC 7437 amongst the introns from the *Chroococcidiopsis* lineage (Fig. 3.2.4). In fact when *Stanieria* PCC 7437 is removed from and the analyses repeated support for a monophyletic *Chroococcidiopsis* is retrieved (Fig. 3.2.5). This is explained most parsimoniously by a horizontal transfer of the intron from the Chroococcidiopsis lineage into the tRNA-fMet gene of *Stanieria* PCC 7437.

A robust assay for horizontal transmission is to test whether has a different phylogenetic history than that of the host. An appropriate statistical test for differences in phylogenetic history is found in the partition homogeneity test (Farris et al. 1994, Swofford 2000), in which the sum of the lengths of the most parsimonious trees fitted to the two datasets independently is compared with the sum of the lengths of trees fitted to random partitions of the same data. The partition homogeneity test performed

Table 3.2.2. Results of the statistical tests performed on the effect of the presence of *Stanieria* PCC 7437. The additional number of steps to the most parsimonious tree and the decrease in the ln likelihood ratio (second column) is indicated after the P value (first column).

User defined topology	Maximum Parsimony						Maximum Likelihood	
Kishino-Hasegawa			Templeton		Winning sites		Kishino-Hasegawa	
Plus *Stanieria* PCC 7437	<0.0001*	+21	<0.0001*	+21	<0.0001*	+21	0.0026*	+32.99
Minus *Stanieria* PCC 7437	0.2014	+6	0.2008	+6	0.2863	+6	0.1461	+9.85

Note: The best tree length in maximum parsimony analysis with Stanieria PCC 7437 is 5567. The best tree length in maximum parsimony analysis without *Stanieria* PCC 7437 is 556. The ln Likelihood score in maximum likelihood analysis with *Stanieria* PCC 7437 is –2562.51 The ln Likelihood score in maximum likelihood analysis with *Stanieria* PCC 7437 is –2487.12. Due to the performance of multiple tests, the significance level of rejection of the null hypothesis is adjusted via the Bonferroni correction to $\alpha = 0.01$.

to test the *a priori* hypothesis that the tRNA-fMet intron is vertically inherited were always rejected with the inclusion of *Stanieria* PCC 7437 (P=0.01). The hypothesis of vertical inheritance with the exclusion of *Stanieria* PCC 7437 showed that the host and intron tree did not differ significantly from one another (P=0.25). Statistical tests performed with *Stanieria* PCC 7437 included always rejected the null hypothesis that the 16S rRNA and intron datasets are estimating the same tree (Table 3.2.2). When Stanieria PCC 7437 was excluded and the tests repeated the null hypothesis could not be rejected (Table 3.2.2). Such phylogenetic congruence is strong evidence for vertical transmission of all the other introns except for *Stanieria* PCC 7437.

While the results here demonstrate a primary role for vertical inheritance it does not preclude the occurrence of horizontal transfer in the evolutionary history of this intron. There is just a single horizontal transfer event supported by the phylogenetic analysis. The support for placing *Stanieria* PCC 7437 with *Chroococcidiopsis* SAG 2025 is very high (94-100%). The support for placing *Stanieria* PCC 7437 intron within the *Chroococcidiopsis* intron lineage is lower from (45 –90%). It seems likely that the intron interrupting the tRNA-fMet gene in *Stanieria* PCC 7437 has arisen through a reversal of the splicing process and cDNA recombination. It can be tentatively concluded that the direction of horizontal transfer is from the *Chroococcidiopsis* lineage to the single isolated *Stanieria* PCC 7437 individual in the order Pleurocapsales (Fig. 3.2.2).

3.2.3.4 The antiquity of the tRNA-fMet intron

These results presented here suggest that the tRNA-fMet intron was present in the ancestor of all extant cyanobacteria and has undergone pervasive loss resulting in the sporadic distribution which is observed today. Cyanobacteria are widely held to have a fossil record that extends back billions of years (Schopf 1994). The occurrence of microfossils have been undisputedly dated to 3.465 billion years ago and reported to resemble modern cyanobacteria (Schopf 1993). However, the exact age cyanobacteria has been contested recently (Doolittle et al. 1996). Protein clock comparisons suggest that gram-positive bacteria and cyanobacteria diverged from one another 2,100 to 2,500 million years ago casting doubts on whether the 3,465 million year old microfossils truly represent modern cyanobacteria (Feng et al. 1997). These studies themselves have attracted a lot of criticism (Morell 1996, Hasegawa and Fitch 1996, Gogarten et al. 1996, Miyamoto and Fitch 1996) and the overwhelming bulk of geological and paleontological studies to date point towards an ancient origin of cyanobacteria (Schopf and Packer 1987, Brocks et al. 1999, Summons et al. 1999, Schopf 1993, Holland and Buekes 1990, Knoll and Holland 1995, Schopf 1994, Knoll et al. 1986, Golubic and Hofmann 1976). Clearly conclusive molecular phylogenetic studies concentrating on the placement of the cyanobacteria amongst the other deeply branching bacteria (Boone and Castenholz 2001) is desirable. Conservative estimates would then age the intron at over 2 billion years old and potentially much older if a direct and literal reading of the fossil record is taken (Schopf and Packer 1987). The tRNA-fMet intron while having a scattered and sporadic distribution predates the cyanobacteria lineage and is thus the oldest known group I intron. This strengthens the case for the antiquity of group I introns interrupting tRNA genes in cyanobacteria and the antiquity of group I introns in general.

3.3 Phylogenetic analyses do not support the horizontal transfer of a group I intron from α-proteobacteria to cyanobacteria

3.3.1 Abstract

Many tRNA-Leu (UAA) genes from plastids and cyanobacteria contain a group I intron. Phylogenetic studies have revealed the presence of a second atypical divergent group I intron interrupting tRNA-Leu (UAA) genes in a handful of cyanobacteria. A horizontal transfer from the tRNA-Arg (CCU) introns of α-proteobacteria is proposed to have given rise to these introns in cyanobacteria. Evidence presented here strongly suggests that these divergent tRNA-Leu (UAA) introns are actually encoded by α-proteobacteria contaminating the cultures of cyanobacteria and eukaryotes identified in this study. The primary and secondary structural similarities between the atypical tRNA-Leu (UAA) introns and α-proteobacterial tRNA-Arg (CCU) introns are reflected in moderate support for the monophyly of these intron groups in all analyses. However, while the phylogenetic analyses and statistical tests presented here underscore the close relationship between the two groups of introns they do not support the long-distance horizontal transfer of a group I intron from a tRNA-Arg (CCU) gene in α-proteobacteria to a tRNA-Leu (UAA) gene in cyanobacteria. All phylogenetic analyses revealed congruence between the α-proteobacterial tRNA-Arg (CCU) introns and 16S rRNA despite a restricted and scattered distribution of the tRNA-Arg (CCU) intron among α-proteobacteria. This makes a transfer of an tRNA-Arg (CCU) intron to the tRNA-Leu (UAA) gene in antiquity implausible and suggests that the introns interrupting the tRNA-Arg (CCU) gene are over 2 billion years old.

3.3.2 Introduction

Until recently, it had been thought that RNA splicing occurs only in eukaryotes (Lewin 1997). It is now know that transfer RNA genes in proteobacteria and cyanobacteria contain group I introns that are removed at the RNA level (Paquin et al. 1997). The splicing event is auto-catalytic and does not require the participation of proteins (Cech 1990, Zuag et al. 1993). Unlike eukaryotes, transcription in bacteria is coupled with translation and movement of the ribosome into the intron may prevent formation of helices that stabilizes the catalytic core of the intron and potentially delay splicing of the intron (Ohman-Heden et al. 1993). Introns inserted in transfer RNA or ribosomal RNA genes do not face this problem of coupled transcription and translation, which may account for the prevalence of group I introns in bacterial tRNA genes (Woodson 1998).

Many tRNA-Leu (UAA) genes from chloroplasts contain a group I intron. An intron is also inserted in the same gene at the same position in cyanobacteria, the bacterial progenitors of plastids (Xu et al. 1990, Kushel et al. 1990). This self-splicing intron has received widespread attention because evolutionary analyses suggest that it was present in the common ancestor of cyanobacteria and chloroplasts and is at least 1 billion years old (Xu et al. 1990, Kushel et al. 1990). It was anticipated that tRNA-Leu (UAA) introns would be found in other bacteria (Kushel et al. 1990, Xu et al. 1990). However, the preliminary reports of tRNA-Leu (UAA) introns in bacteria other than cyanobacteria proved erroneous (Kushel et al. 1990, Shub 1991).

Introns interrupting isoacceptors of a tRNA-Arg (CCU) and a tRNA-Ile (CAT) gene were subsequently reported from proteobacteria (Rheinhold-Hurek and Shub 1992). To make the paralogy of the tRNA group I intron subfamily even more complicated an additional study has identified a subset of the tRNA-Leu (UAA) introns in cyanobacteria which are more similar to the introns interrupting tRNA-Arg (CCU) genes in α-proteobacteria than the typical tRNA-Leu (UAA) introns which are widespread in cyanobacteria and plastids (Rudi and Jakobsen 1997). These divergent introns are easily recognized from their sequence homology to the tRNA-Arg (CCU) introns of α-proteobacteria. The presence of similar introns in diverse bacteria such as α-proteobacteria and cyanobacteria is held to be explained most parsimoniously by a horizontal transfer of an intron from α-proteobacteria to cyanobacteria in antiquity (Rudi and Jakobsen 1999, Paquin et al. 1999). This horizontal transfer event would account for the sporadic distribution of the intron in cyanobacteria and the sequence similarity (Paquin et al. 1999).

Further examples of tRNA-Arg (CCU) introns in widely divergent α-proteobacteria were recently identified amongst members of the Rickettsiaceae Rhodospirillaceae and Rhizobiaceae subdivisions of α-proteobacteria (Paquin et al. 1999). The sporadic distribution of the intron and in extant α-proteobacteria was taken to imply that the intron had arisen recently and that its scattered distribution could be explained through horizontal transfer and exchange of genetic material between distantly related individuals (Paquin et al. 1999).

To date no phylogenetic analysis has been undertaken to substantiate either claims of horizontal transfer. Additional examples of the second type of divergent tRNA-Leu (UAA) intron were found in a systematic survey of introns interrupting tRNA-Leu (UAA) genes in cyanobacteria and plastids (Sections 3.4 and 3.5). This prompted the reappraisal of these divergent introns evolutionary history. The proposed recent origin of the tRNA-Arg (CCU) introns in α-proteobacteria is not substantiated in phylogenetic analyses presented here. This study demonstrates that the concerns over potential culture con-

tamination originally dismissed are duly founded and that the atypical divergent tRNA-Leu (UAA) introns interrupt genes in α-proteobacteria.

3.3.3 Results and discussion

3.3.3.1 Distribution of divergent tRNA-Leu (UAA) intron in cyanobacteria and chloroplasts

The adventitious find of the second divergent type of intron in red algae during a systematic survey of the more typical intron in cyanobacteria and plastids prompted the reappraisal of the origins of this intron (Section 3.4 and 3.5). During this systematic survey over 116 cyanobacteria and chloroplasts were screened for the presence or absence of the typical tRNA-Leu (UAA) intron (Table 3.3.1). These PCR analyses revealed the presence of the divergent intron in three members of the Rhodophyta (*Porphyridium sordidum* SAG 44.94, *Rhodella violacea* SAG 30.97, and *Bostrychia radicans* SAG 100.79) one glaucocystophyte (*Cyanoptyche gloeocystis* SAG 4.97) and one cyanobacteria (*Nodularia* SAG 50.79). Within the Rhodophyta the typical introns are found in both orders of red algae, the monophyletic order Florideophyceae and more ancient paraphyletic order Bangiophyceae (Oliveira and Bhattacharya 2000, Freshwater et al. 1994). The glaucocystophytes and rhodophytes are two of the three primary lineages of chloroplasts (Douglas 1998) showing that these introns are found in widely divergent plastid lineages and that their distribution spans vast phylogenetic distance. All five introns are inserted between the wobble and second bases of the UAA anticodon and are similar in length (203-240 bp). This more than doubles the number of known divergent tRNA-Leu (UAA) introns and extends the distribution of these introns to include plastids.

3.3.3.2 Lack of congruence between tRNA-Leu (UAA) intron and "host"

The evolutionary history of the second type of tRNA-Leu (UAA) intron was reconstructed using a combination of distance and maximum parsimony analyses (Fig. 3.3.1). The maximum likelihood and parsimony/distance analyses differed slightly from the parsimony and distance and this analyses was presented separately (Fig. 3.3.2). There are several instances of incongruence between the phylogeny of the divergent tRNA-Leu (UAA) group I introns (Figs. 3.3.1 and 3.3.2) and what is already known of the cyanobacteria and plastid phylogeny from published studies on 16S rRNA, *atpB* and *rpoC1* and *tufA* genes (Morden et al. 1992, Wilmotte et al. 1994, Delwiche et al. 1995, Palenik and Haselkorn 1992, Turner 1997, Oliveira and Bhattacharya 2000, Turner et al. 1999). The plastid introns are non-monophyletic with the intron from the glaucocystophyte *Cyanoptyche gloeocystis* SAG 4.97 grouping with the intron from the cyanobacterium *Aphanizomenon flos-aquae* NIVA 142 separate from the remainder of the plastids introns (Figs. 3.3.1 and 3.3.2). While the three introns from the rhodophyte

Table 3.3.1. Strains of cyanobacteria, bacteria, and chloroplasts for which the atypical intron was identified. The typical tRNA-Leu (UAA) introns which are widespread in cyanobacteria and plastids are marked (T) while the divergent tRNA-Leu (UAA) introns suspected of interrupting tRNA genes in proteobacteria are marked (D).

Species	tRNA intron	16S rRNA
Cyanobacteria		
tRNA-fMet introns		
Chroococcidiopsis SAG 2023	XXXX	XXXX
Chroococcidiopsis SAG 2026	XXXX	XXXX
Cylindrospermum stagnales PCC 7417	U83261	AF132789
Gloeobacter violaceus PCC 7421	U83262	AF132790
Nostoc DM 44	XXXX	XXXX
tRNA-Leu (UAA) introns		
Cyanobacteria		
Anabaena PCC 7122 (T)	XXXX	AF091150
Aphanizomenon flos-aquae NIVA 142 (D)	AJ228697	none
Microcystis aeruginosa NIVA 228/1 (D)	AJ228695	none
Nodularia PCC 7804 (T)	XXXX	AJ133181
Nodularia sphaerocarpa SAG 50.79 (D)	XXXX	none
Nostoc sp. NIVA 308 (D)	AJ228696	none
Pseudanabaena PCC 7403 (T)	XXXX	AB039019
Chloroplasts		
Bostrychia radicans SAG 100.79 (D)	XXXX	none
Porphyridium sordidum SAG 44.94 (D)	XXXX	none
Rhodella violacea SAG 30.97 (D)	XXXX	none
Cyanoptyche gloeocystis SAG 4.97 (D)	XXXX	none
Proteobacteria		
tRNA-Arg (CCU) introns		
Agrobacterium tumefaciens A136	X66220	AF388030
Agrobacterium tumefaciens C 58	AE007870	AE007870
Anaplasma marginale	AF081791	AF311303
Azospirillum halopraeferens Au5	AF081792	X79731
Magnetospirillum magnetotacticum ATCC 31632	Genome	Genome
tRNA-Ile (CAT) intron		
Azoarcus sp. BH 72	X66221	AF011344

lineage group together with high bootstrap support (98-99%) the order Bangiophyceae (*Rhodella* and *Porphyridium*) are paraphyletic (98-100%) with respect to the order Florideophyceae (*Bostrychia*) in direct contradiction to previous studies which demonstrated the monophyly of the order Florideophyceae (e.g. Freshwater et al. 1994). The heterocystous cyanobacteria do not form a monophyletic group in direct contraction of the vast majority of phylogenetic analyses presented to date (e.g. Wilmotte et al. 1994, Nelissen et al. 1995 Turner et al. 1999, Ishida et al. 2001, Section 3.2). Given this incongruence between the divergent group I introns and the organisms which harbor them a single horizontal

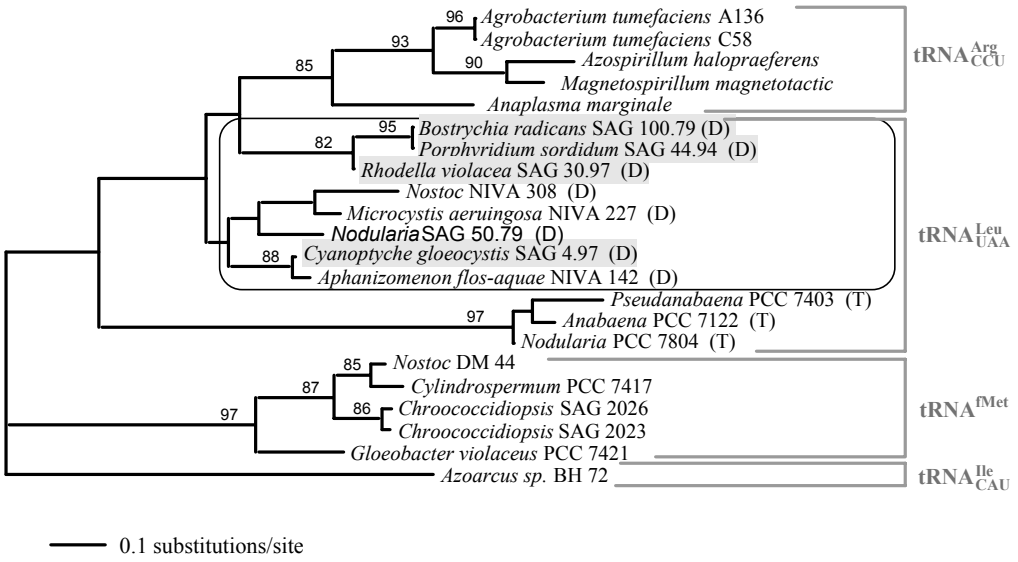

Figure 3.3.1. Phylogenetic relationship between the group I introns interrupting transfer RNA genes in cyanobacteria and proteobacteria. A maximum likelihood tree based on TIM+G model of DNA substitution (lnL=-2232.30). Maximum likelihood bootstrap values using the TIM+G model of DNA are given at the node. Only bootstrap values above 60% are shown. The typical tRNA-Leu (UAA) introns which are widespread in cyanobacteria and plastids are marked (T) while the divergent tRNA-Leu (UAA) introns suspected of interrupting tRNA genes in proteobacteria are marked (D) and boxed. The divergent introns isolated from plastid containing eukaryotes are shaded.

transfer in antiquity cannot explain their current distribution. A recent origin of the introns through horizontal transfer necessitates at least four independent insertions events whereas a single ancient horizontal transfer invokes innumerable losses of the divergent intron in cyanobacteria and plastids to explain the current distribution.

3.3.3.3 Evidence suggesting divergent introns interrupt genes in culture contaminants

To account for the presence of both the typical and divergent tRNA-Leu (UAA) (*Nodularia* SAG 50.79, *Nostoc* NIVA 308, *Aphanizomenon flos-aquae* NIVA 142, *Cyanoptyche gloeocystis* SAG 4.97) as well as the presence uninterrupted typical tRNA-Leu (UAA) genes and atypical tRNA-Leu (UAA) introns in plastids and cyanobacteria (*Porphyridium sordidum* SAG 44.94, *Rhodella violacea* SAG 30.97, *Bostrychia radicans* SAG 100.79 and *Microcystis* NIVA 228) introns in the genomes of some cyanobacteria and the presence of a second tRNA-Leu (UAA) gene was invoked (Rudi and Jakobsen 1999). The presence of duplicated tRNA genes in the same genome is not unprecedented (Bult et al. 1996). However, inspection of the tRNA content in the draft analyses of four unpublished genomes and one complete genome of cyanobacteria failed to reveal the presence of a second divergent type of

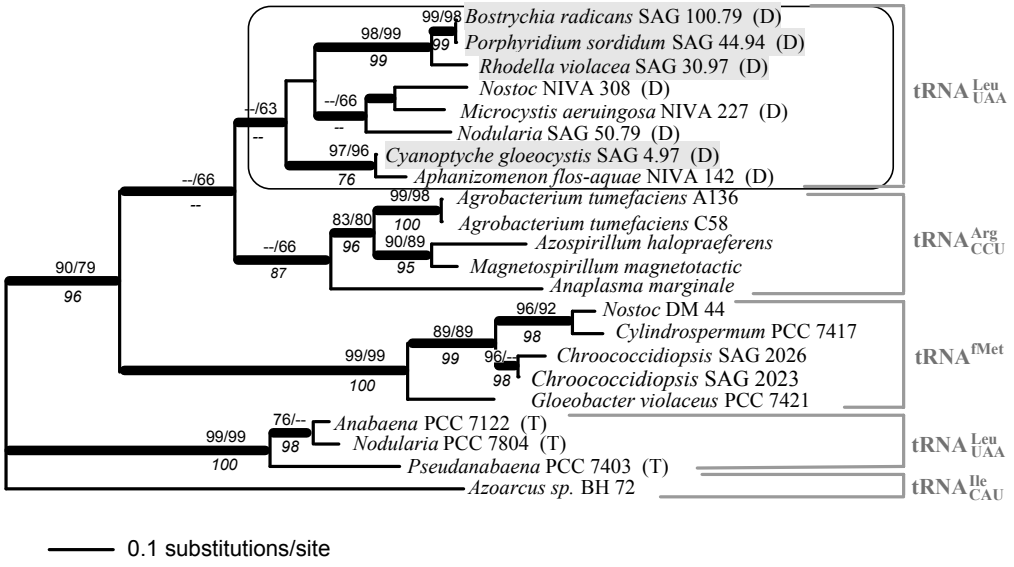

Figure 3.3.2. Phylogenetic relationship between the group I introns interrupting transfer RNA genes in cyanobacteria and proteobacteria. The tree is a maximum parsimony tree and values above the node neighbor-joining (TIM+G) left minimum evolution (TIM+G) right and weighted maximum parsimony below the node. Thick lines mark internal nodes that were resolved in all distance, maximum parsimony trees (see text). Only bootstrap values above 60% are shown. The typical tRNA-Leu (UAA) introns which are widespread in cyanobacteria and plastids are marked (T) while the divergent tRNA-Leu (UAA) introns suspected of interrupting tRNA genes in proteobacteria are marked (D) and boxed. The divergent introns isolated from plastid containing eukaryotes are shaded.

tRNA-Leu (UAA) gene. The genomes of *Prochlorococcus marinus* MED4, MIT9313, *Synechococcus* WH 8102, *Nostoc* PCC 7120, and *Synechocystis* PCC 6803. Likewise complete plastid genomes from the glaucocystophyte *Cyanophora paradoxa* (Stirewalt et al. 2000) and the red algae *Porphyra purpurea* (Reith and Mullholland 1995) and *Cyanidium caldarium* (Glockner et al. 2000) also contained just a single tRNA-Leu (UAA) gene. Thus to account for the phylogeny and distribution of the divergent intron it is necessary to invoke the independent duplication of tRNA-Leu (UAA) genes in widely divergent cyanobacteria and plastid genomes followed by multiple independent horizontal transfer. Alternatively, an ancient duplication of the tRNA-Leu (UAA) gene followed by widespread loss of this duplicated tRNA gene could be envisaged. The strains used in this study and those from which the divergent intron were identified in previous studies are all heavily contaminated with heterotrophic bacteria. Bacteria may be closely associated with or attached to the cyanobacteria and many cyanobacteria are difficult to maintain or to even introduce into pure culture (Paerl 1996, Schlösser 1994). Indeed the possibility of such a potential culture contamination is acknowledged in one of these studies

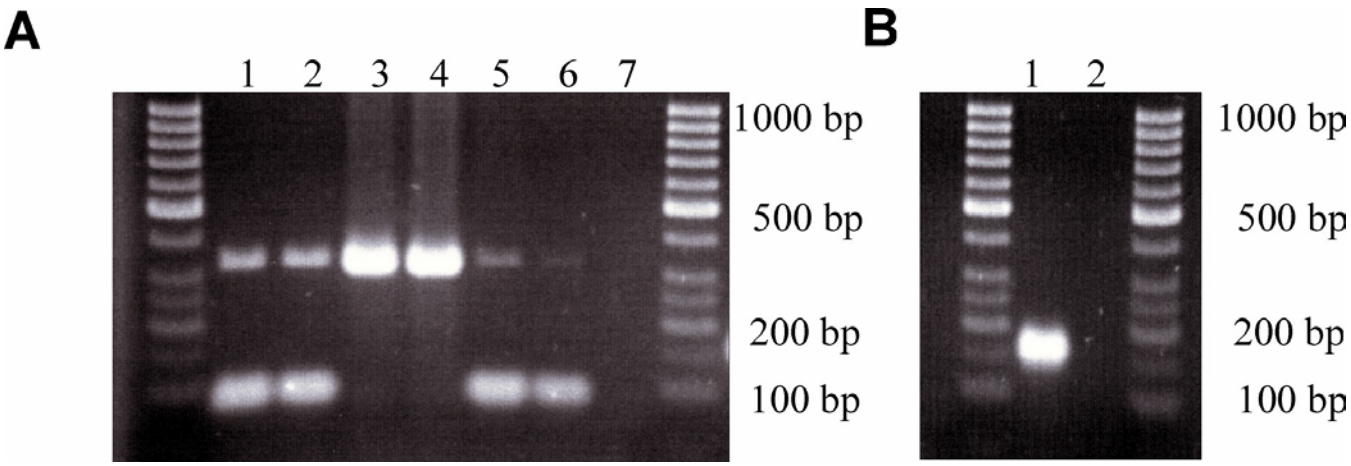

Figure 3.3.3. PCR amplification of tRNA-Leu (UAA) genes in a PCR experiment with the red alga *Bostrychia radicans* SAG 100.79 and its culture contaminant. (A) The 2% agarose gel shows the migration of PCR amplified fragments of the tRNA-Leu (UAA) gene including the entire intron, anticodon, anticodon loop using the primers LeuF and LeuR. 1-2: Amplification of intron and uninterrupted tRNA-Leu (UAA) gene from *Bostrychia radicans* SAG 100.79. 3-4: An intron amplified from a bacterium isolated from the *Bostrychia radicans* SAG culture. 5-6: Amplification of the intron and uninterrupted genes from a *Bostrychia radicans* SAG 100.79 culture treated with ampicillin. 7: A template free PCR reaction. (B) The first lane shows the amplification of a fragment of the group I intron using the primer combination PorphF and PorphR. The second lane is a template free PCR reaction Both gel images contain a 50 bp ladder (MBI Fermentas) the sizes of which are indicated on the right hand side of the gel.

(Rudi and Jakobsen 1999). A culture contaminant of *Bostrychia radicans* SAG 100.79 was isolated using an enriched culture media based on the half strength seawater medium (Schlösser 1994). This bacterium was grown free of the red algae. PCR experiments with *Bostrychia radicans* SAG 100.79 culture contaminating bacterium, in which the tRNA-Leu (UAA) gene was amplified, yielded a single PCR product of identical length to the intron containing tRNA-Leu (UAA) gene (Fig. 3.3.3 A). PCR with primers designed to amplify a fragment of the group I intron strongly suggested that this was a bona fide group I intron (Fig. 3.3.3 B). The putative intron containing tRNA was sequenced directly and confirmed to be identical to the intron sequenced from the *Bostrychia radicans* SAG 100.79 culture.

The 16S rRNA gene of the contaminating bacterium was amplified and sequenced and used to query the BLAST server (www.ncbi.nlm.nih.gov/BLAST). The top BLAST hit results strongly suggested that the unidentified bacterium cultured from *Bostrychia radicans* SAG 100.79 was closely related to members of the α-proteobacterial genus *Sphingomonas*. The sequence was properly aligned against other

16S rRNA gene sequences from proteobacteria and the subsequent phylogenetic analyses reaffirmed this initial conclusion (Fig. 3.3.4). Members of the genus *Sphingomonas* are characterized as gram-negative, non-spore forming rods that display a single polar flagellum when they are motile (Yabuuchi et al. 1990). *Sphingomonas* strains have been isolated from a wide variety of sources including soil, both marine and fresh waters, marine life, and from plants (Takeuchi et al. 2001). The red algae *Bostrychia radicans* SAG 100.79 was originally isolated from Chesapeake Bay in the USA (Schlösser 1994). The marine nature of the genus *Spingomonas* and its association with eukaryotes makes it plausible that this contaminating bacterium also originated from a marine habitat.

Ampicillin was added to the *Bostrychia radicans* SAG 100.79 culture media as it acts against many gram-negative and gram-positive bacteria by inhibiting the production of new cell wall (Lewin 1997). This bacteriostatic antibiotic binds to and inhibit enzymes involved in the transpeptidation (cross-linking) of peptidoglycan preventing the bacterium from dividing. Ampicillin does not kill the micro-organism outright and PCR experiments show that the band corresponding in size to the intron is much fainter consistent with the ampicillin action in preventing the bacterium from dividing (Fig. 3.3.3).

The presence of the divergent type of tRNA-Leu (UAA) intron in proteobacteria would explain why the intron is found in such a wide diversity of cyanobacteria and chloroplasts and has such a complex phylogeny an observation difficult to rationalize otherwise without having to invoke widespread loss of the intron in all other chloroplasts or multiple independent horizontal transfers. This calls into question the original conclusions (Rudi and Jakobsen 1997) and those of subsequent workers (Rudi and Jakobsen 1999, Paquin et al. 1999).

3.3.3.4 Distribution of tRNA-Arg (CCU) intron in proteobacteria

Sequence data from the intron interrupting the tRNA-Arg (CCU) in *Anaplasma marginale* was used to query the microbial genomes database on the NCBI BLAST page (www.ncbi.nlm.nih.gov/BLAST). An additional unannotated intron was found in the linear genome of *Agrobacterium tumafaciens* C58 (AE008320.1) and an intron from the unfinished genome *Magnetospirillum magnetotacticum* were found from database hits in the homology search at NCBI. Both the *Agrobacterium tumafaciens* and *Magnetospirillum magnetotacticum* introns are similar in primary sequence to the other known tRNA-Arg (CCU) group I introns all of which are restricted to the α-proteobacteria group. All five introns are inserted after the T of the CCT anticodon and are similar in length (217-237 bp). Draft sequence analyses of incomplete genomes was checked for the presence of an uninterrupted tRNA-Arg (CCU) gene. The distribution of the intron was mapped to a maximum likelihood tree based on 16S rRNA for those proteobacteria from which the intron has been characterised and for which either unpublished or complete genome is publically available (Fig. 3.3.4). This confirmed the sporadic and scattered distribution

of the tRNA-Arg (CCU) intron in α-proteobacteria.

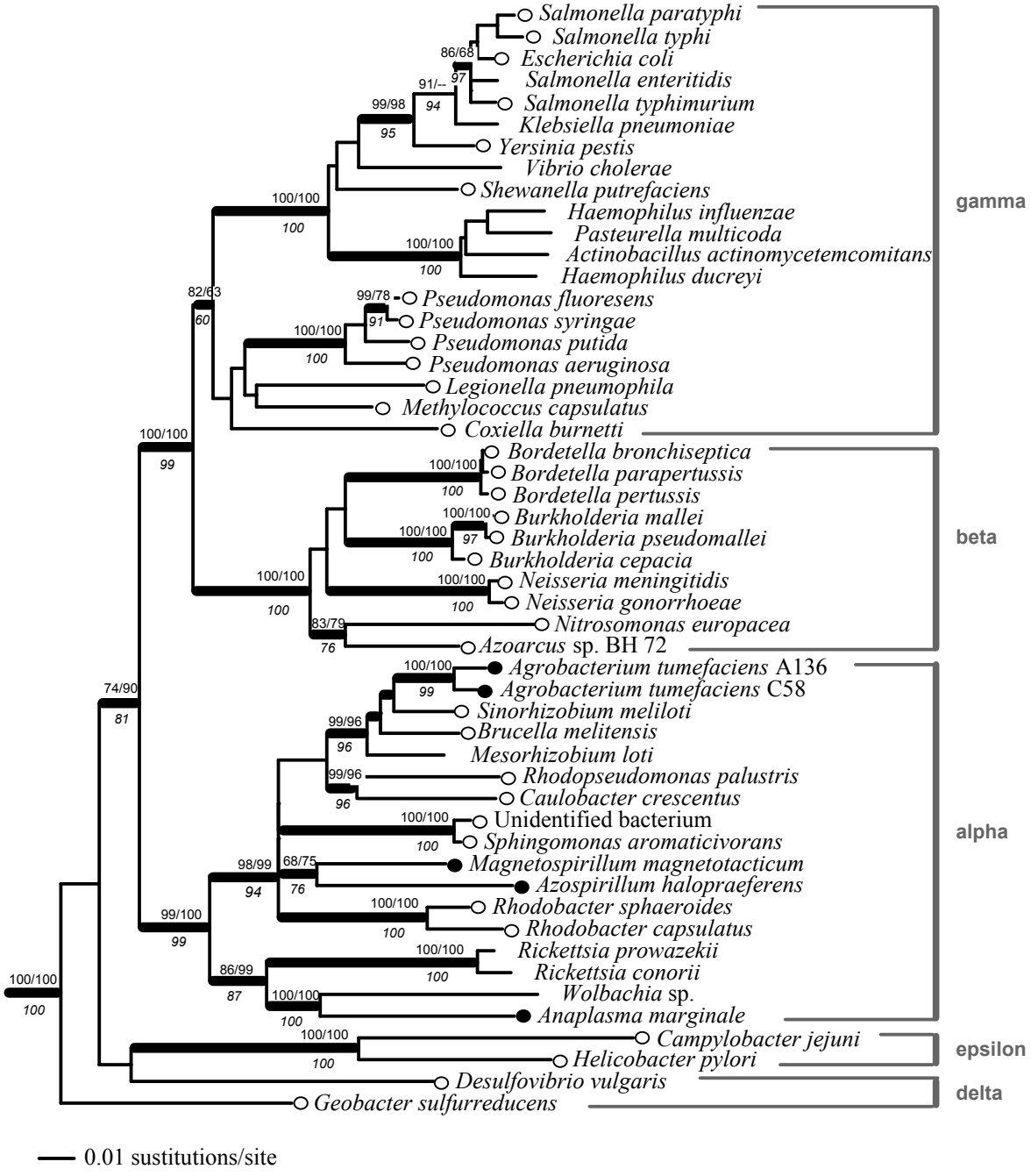

Figure 3.3.4 A phylogeny of the proteobacteria based on 16S rRNA showing the distribution of the intron interrupting the tRNA-Arg (CCU) in proteobacteria. Maximum likelihood tree (lnL=-4,887,654.41) based on the GTR+G+I model of DNA substitution Open circles at the end of nodes denotes an uninterrupted tRNA gene and a filled circle denotes the presence of an intron. The absence of a circle denotes the absence of the tRNA-Arg (CCU) gene. Thick lines mark internal nodes that were resolved in all distance, maximum parsimony and maximum likelihood trees (see text). Values above internal nodes are bootstrap values from 1000 replications as calculated from minimum evolution using the LogDet transformation (left) or the

GTR+I+G model (right), values below branches were calculated from weighted parsimony. Only bootstrap values above 60% are shown.

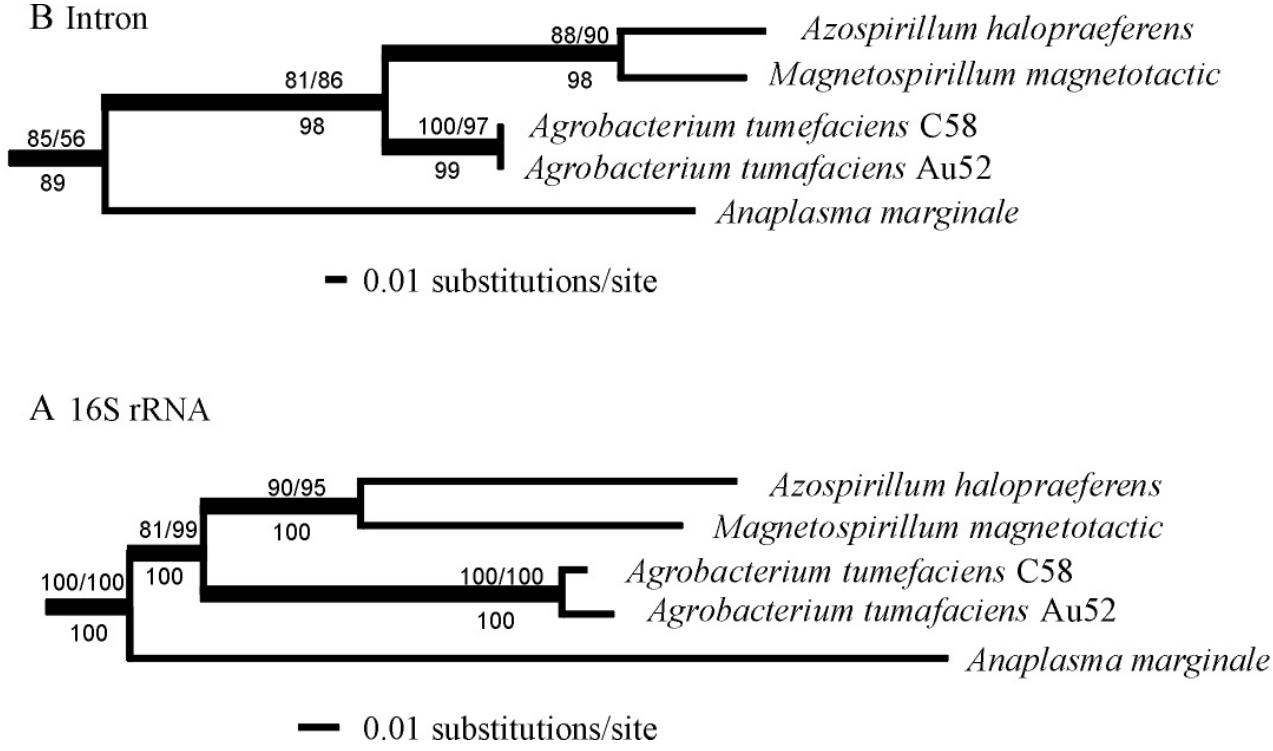

Figure 3.3.5. Congruence between the five introns interrupting tRNA-Arg (CCU) genes and the 16S rRNA from α-proteobacteria. (A) 16S rRNA maximum likelihood tree based on the GTR+G+I model of DNA substitution (lnL=-2,866,069.11). (B) tRNA-Arg (CCU) intron maximum likelihood tree based on the TIM+G model of DNA substitution (lnL=-1489.80). In both trees thick lines mark internal nodes that were resolved in all distance, maximum parsimony and maximum likelihood trees (see text). Values above internal nodes are bootstrap values from 1000 replications as calculated from minimum evolution using the LogDet transformation (left) or the TIM+G model (right), values below branches were calculated from weighted parsimony. Only bootstrap values above 60% are shown. The typical tRNA-Leu (UAA) introns which are widespread in cyanobacteria and plastids are marked (T) while the divergent tRNA-Leu (UAA) introns suspected of interrupting tRNA genes in proteobacteria are marked (D).

3.3.3.5 Phylogenetic congruence between tRNA-Arg intron and host 16S rRNA

16S rRNA gene sequences were used to reconstruct the evolutionary history of the five α-proteobacteria known to harbor tRNA-Arg (CCU) introns (Fig. 3.3.5). Simultaneously the evolutionary history of the tRNA-Arg (CCU) intron was reconstructed (Fig. 3.3.5). Maximum parsimony, distance and maximum likelihood analyses all yielded support for congruence between the intron and the α-proteobacteria harboring it (Fig. 3.3.5). *Anaplasma marginale* was placed basal to the remainder of the α-proteobacteria 16S rRNA gene sequences and tRNA-Arg (CCU) introns in all analyses (Fig. 3.3.5). In the distribution analyses based on 16S rRNA gene sequences *Anaplasma marginale* has a more sta-

ble position and groups with the *Drosophila melanogaster* fruit fly endosymbiont *Wolbachia* (Fig 3.3.4). A partition homogeneity test was undertaken and the null hypothesis that both the 16S rRNA and intron datasets are estimating the same tree could not be rejected at (P= 1.00). The absence of the tRNA-Arg (CCU) intron in the vast bulk of α-proteobacteria examined in this study can be explained through eight losses of the intron.

The distribution of the tRNA-Arg (CCU) intron is consistent with selective pressure to streamline the genome of rapidly dividing bacteria (Darnell and Doolittle 1986). While the streamlining theory was originally proposed to account for the absence of spliceosomal introns in bacteria it nevertheless may be applied to group I introns interrupting genes in bacteria. The vast bulk of α-proteobacteria lack an intron encoding tRNA-Arg (CCU) gene (Fig. 3.3.4). It seems likely that the intron, given that the tRNA-Arg (CCU) intron is vertically inherited and has a sporadic distribution amongst α-proteobacteria, fits the predictions of the streamlining theory. A sporadic distributions of insertion elements such as group I introns is often the sole criterion for inferring horizontal transfer (Healy et al. 1999, Pronk and Sanderson 2001). An important finding of this study is that evolutionary relationships among group I introns cannot be concluded solely on the basis sporadic distribution alone. Numerous studies have shown the importance of reconstructing evolutionary history of gene families before reaching conclusions on their origin and evolution an this appears to have been overlooked in this case (Logsdon and Faguy 1999, Kyrpides and Olsen 1999, Eisen 1998, Stanhope et al. 2001). A more detailed study on the distribution and evolutionary history of the tRNA-Arg (CCU) intron in α-proteobacteria and other proteobacterial subdivisions is called for.

Gene sequence data strongly support a monophyletic origin of the mitochondrion from the rickettsial subgroup of the α-proteobacteria (Yang et al. 1985). Members of the rickettsial subdivision of the α-proteobacteria, a group of obligate intracellular parasites that includes genera such as *Rickettsia*, *Anaplasma*, and *Ehrlichia*, are considered to be among the closest known eubacterial relatives of mitochondria (Gray and Spencer 1996). Comparison of complete genomes from mitochondria and α-proteobacteria has yielded many sequence similarities between the two and affirmed this relationship (Gray et al. 1999). It is estimated that at least 2 billion years have passed since free-living α-proteobacteria developed a beneficial symbiotic relationship with their host cells and became permanent passengers (Doolittle 1997). The distribution of the tRNA-Arg (CCU) intron in this respect is interesting because it is found in Rickettsiaceae and outside this group in the Rhodospirillaceae and Rhizobiaceae subdivisions of α-proteobacteria (Fig. 3.3.1). The evidence for the vertical inheritance of the tRNA-Arg (CCU) introns in the α-proteobacteria presented here suggests that the intron may have been present in last common ancestor of α-proteobacteria. This would imply that the tRNA-Arg (CCU) intron is over 2 billion years old.

Table 3.3.2. Results of the statistical tests performed to assess the support for the four scenarios presented in figure 3.3.6. The four topologies do not differ significantly from one another in any of the tests. Due to the performance of multiple tests, the significance level of rejection of the null hypothesis is adjusted via the Bonferroni correction to $\alpha = 0.01$

User defined topology	Maximum Parsimony						Maximum Likelihood	
Kishino-Hasegawa			Templeton		Winning sites		Kishino-Hasegawa	
No support	0	(best)	0	(best)	0	(best)	+4.12	0.4681
Arginine to leucine	+2	0.3185	+2	0.3173	+2	0.6250	+8.48	0.2508
Leucine to arginine	+1	0.6559	+1	0.6547	+1	1.0000	+0.95	0.7950
Leucine to arginine	+14	0.2016	+14	0.2008	+14	0.2863	0	(best)

Note: The best tree length in maximum parsimony analysis is 508. The ln Likelihood score in maximum likelihood analysis is –2,232.30.

3.3.3.6 Lack of support for horizontal transfer between cyanobacteria and α-proteobacteria

In the analyses presented here the evolutionary relationship between the tRNA-Arg (CCU) and tRNA-Leu (UAA) introns is unclear (Figs. 3.3.1 and 3.3.2). The primary and secondary structural similarities between the divergent cyanobacterial tRNA-Leu (UAA) introns and α-proteobacterial tRNA-Arg (CCU) introns are reflected in moderate support for the monophyly of these intron groups in all analyses (46-66%). Maximum parsimony and distance analyses yielded support for the reciprocal monophyly of the tRNA-Arg (CCU) introns (60-87%) and tRNA-Leu (UAA) introns (39-63%). Maximum likelihood analyses favored a topology which suggested a horizontal transfer from the tRNA-Leu (UAA) gene to the tRNA-Arg (CCU) gene but with negligible bootstrap support (0-3%). In addition, the monophyly of the tRNA-Arg (CCU) clade of introns argues strongly against the lateral transfer of introns from the tRNA-Arg (CCU) gene to the tRNA-Leu (UAA) gene (Figs. 3.3.1, 3.3.2 and 3.3.5). For this to be true we would expect a non-monophyletic clade of tRNA-Arg (CCU) introns within which the atypical tRNA-Leu (UAA) introns would nest. This is a compelling argument against the α-proteobacteria to cyanobacteria horizontal transfer. Statistical tests undertaken to examine the support for user-defined alternative topologies (Fig. 3.3.6) show that there is no support for a horizontal transfer between tRNA-Arg (CCU) and tRNA-Leu (UAA) genes (Table 3.3.2). Tests which force evolutionary scenarios which can be accounted for by a horizontal transfers are not statistically worse that a scenario in which there is no support for horizontal transfer (Fig. 3.3.6, Table 3.3.2). However, the maximum likelihood test favors horizontal transfer from the tRNA-Leu (UAA) gene to the tRNA-Arg (CCU) gene while the maximum parsimony tests favor the no-support scenario (Fig. 3.3.6, Table 3.3.2).

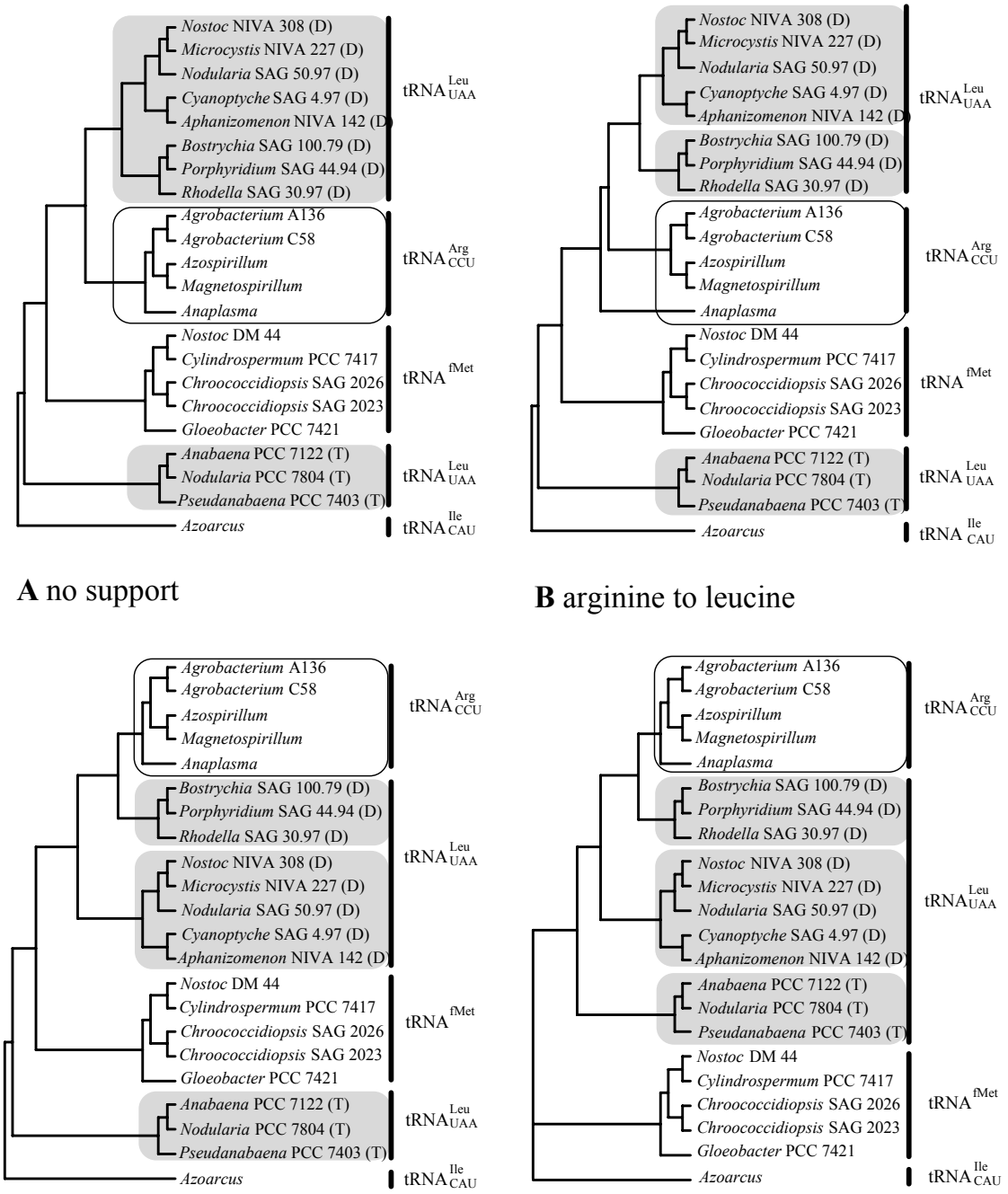

Figure 3.3.6. Hypothetical phylogenetic trees showing various evolutionary models related to acceptance or rejection of intron horizontal transfer from tRNA-Arg (CCU) to tRNA-Leu (UAA) genes. The results of statistical tests based on these trees are given in Table 3.3.2. (A) Phylogenetic history with no support for horizontal transfer. (B) Phylogenetic history in support of horizontal transfer from the tRNA-Arg (CCU) to the tRNA-Leu (UAA) gene. (C-D) Phylogenetic history which supports lateral transfer from the tRNA-Leu (UAA) to the tRNA-Arg (CCU) gene. tRNA-Leu (UAA) introns are shown in shaded boxes and tRNA-Arg (CCU) introns are shown in empty boxes. The typical tRNA-Leu (UAA) introns which are widespread in cyanobacteria and plastids are marked (T) while the divergent tRNA-Leu (UAA) introns suspected of interrupting tRNA genes in proteobacteria are marked (D).

The phylogenetic analyses and statistical tests presented here do not support the long-distance horizontal transfer of a group I intron from the tRNA-Arg (CCU) gene in α-proteobacteria to the tRNA-Leu (UAA) gene in cyanobacteria. It seems highly likely that the divergent tRNA-Leu (UAA) introns currently attributed to cyanobacteria are actually all proteobacterial. This will mean that self-splicing introns in eubacteria are much more common than previously suspected. Phylogenetic analyses must be a central component of any evolutionary study of these self-splicing introns. Importantly, phylogenetic reconstruction is critical to synthesizing, from the rapidly accumulating wealth of sequence data from the bacterial domain, a more comprehensive view of group I intron evolution. It is anticipated here that the innumerable microbial genome sequencing projects currently underway will yield further examples of group I introns from bacteria and shed light on this problem and perhaps provide proof for the long-standing speculation that the typical tRNA-Leu (UAA) introns predate the cyanobacterial radiation.

3.4 Multiple independent losses of the intron interrupting a leucine transfer RNA in cyanobacteria

3.4.1 Abstract

The tRNA-Leu (UAA) intron is widely held to be the most ancient naturally occurring group I intron. High levels of sequence homology amongst the introns has led to suggestions that they may in fact have a more recent origin. However, direct evidence for this has been lacking. To readdress the evolutionary history of this intron a systematic survey was undertaken here to determine the phylogenetic distribution of the intron in cyanobacteria. The systematic survey, which included 90 of cyanobacteria and plastids, identified 34 introns. The distribution of the intron is much more sporadic than previously suspected with over half of all modern cyanobacteria lacking the intron. Statistical tests revealed incongruence between the intron and 16S rRNA datasets. However, a strict consensus maximum parsimony, distance and maximum likelihood analyses based on the conserved catalytic cores of the intron revealed 7 instances of congruence with the 16S rRNA dataset. No instances of horizontal transfer received phylogenetic support. Moderate support for a single origin of the plastid and cyanobacteria introns reaffirms the monophyly of these introns and suggests that the introns are at least 1000 million years old. Concerns over the possible horizontal transfer within the cyanobacterial lineage are not substantiated at the ordinal level. However, no deep level support for either an ancient or recent origin of the intron was recovered. The lack of stable structure within the tRNA-Leu (UAA) intron phylogenetic trees and the lack of congruence between the intron and 16S rRNA datasets is surprising but in the absence of any further evidence to the contrary the distribution of the intron is consistent with an ancient origin followed by multiple independent losses.

3.4.2 Introduction

It is now known that widely divergent plastids contain a group I intron in the tRNA-Leu (UAA) gene (Evrard et al. 1988). An intron is also inserted in the same gene at the same position in cyanobacteria, the bacterial progenitors of plastids, suggesting an ancient bacterial origin for this intron (Xu et al. 1990, Kushel et al. 1990). Recent attempts to map the distribution of this intron in plastids underscore the current interest in determining the evolutionary history and origins of this self-splicing intron (Besendahl et al. 2000). The homology groupings that were based upon the intron sequence data do not agree with the relationships inferred from the small subunit rRNA sequence data.

To date there have been a number of evolutionary studies on group I introns from a wide array of organisms. In most of these reports, the group I introns were observed more or less sporadically and did not reflect host phylogeny, which suggests horizontal transmissions between distinct lineages (e.g., Sogin et al. 1986, Nishida et al. 1993, Gargas et al. 1995, Turmel et al. 1995, Hibbett 1996, Cho et al. 1998, Nishida et al. 1998, Watanabe et al. 1998, Cho and Palmer 1999, Holst-Jensen et al. 1999, Perotto et al. 2000). Only a few studies have provided convincing evidence that supports stable maintenance of particular group I introns over long periods of evolutionary time (Xu et al. 1990, Kuhsel et al. 1990, Bhattacharya et al. 1994, Nikoh and Fukatsu 2001). Of these studies those on the tRNA-Leu (UAA) intron (Xu et al. 1990, Kuhsel et al. 1990) have generated the most scientific interest primarily because of the perceived age of the intron.

The primary sequence of the introns from cyanobacteria and plastids show remarkable conservation. The striking similarity between the cyanobacterial and plastid introns is consistent with a recent divergence and this sequence homology has led to suggestions that the intron may in fact have a more recent origin than originally proposed (Barinaga 1990, Rudi and Jakobsen 1999). Group I introns without open reading frames (ORF's) are capable of transposing themselves from one insertion site to another through reverse transcription and splicing (Lambowitz and Belfort 1993, Mohr and Lambowitz 1991, Woodson and Cech 1989, Yang et al. 1996). The process by which the intron removes itself from the tRNA-Leu (UAA) gene is chemically reversible (Cech 1990). The intron can in a reversal of the splicing process reinsert itself back into a site in any RNA fragment which is homologous to its internal guide sequence which in the tRNA-Leu (UAA) intron is particularly short (Zuag et al. 1993). A concomitant reverse transcription and conversion event will result in the intron being immortalized in the cell line (Roman and Woodson 1998). By this process it is suggested that the distribution of the tRNA-Leu (UAA) introns observed today could have arisen through a series of horizontal transfers following the establishment the modern lines of cyanobacterial and plastid descent (Barinaga 1990, Daros and Flores 1996, Rudi and Jakobsen 1999).

Despite these concerns it has been proposed that the evolutionary history of the tRNA-Leu (UAA) intron in cyanobacteria is consistent with vertically inheritance (Paquin et al. 1997). However, to date little evidence has been forthcoming to substantiate this proposal. In the light of the overwhelming systematic importance of the tRNA-Leu (UAA) intron a survey to characterize tRNA-Leu (UAA) introns from diverse cyanobacteria was undertaken to readdress the evolutionary history of this intron.

3.4.3 Results and discussion

3.4.3.1 The distribution of the intron in cyanobacteria and chloroplasts

Sixty taxa of cyanobacteria and chloroplast containing eukaryotes were screened for the presence of the intron by low stringency PCR (Table 3.4.1). A total of 17 introns were identified and sequenced in this study (Table 3.4.1). The presence or absence of introns determined in 29 taxa in previous studies (Paquin et al. 1997, Rudi et al. 1999, Strehl et al. 1999) is here confirmed and these taxa were appended to the list (Table 3.4.1). The tRNA-Leu (UAA) intron from *Synechococcus* PCC 6301 was used to query the microbial genomes database on the NCBI BLAST page (www.ncbi.nlm.nih.gov/BLAST). The BLAST results indicated that no introns were present in the genomes of the cyanobacteria *Prochlorococcus marinus* MED 4, *Prochlorococcus marinus* MIT 9313 or *Synechococcus* WH 8102. Sequence from the uninterrupted tRNA gene was used to query the database and revealed the presence of uninterrupted tRNA-Leu (UAA) genes in each genome. The genomes of *Nostoc punctiforme* PCC 73102 and *Nostoc* PCC 7120 both contain introns as previously demonstrated (Kushel et al. 1990, Paquin et al. 1997). Inspection of the complete plastid genomes of *Marchantia polymorpha* (Shimada and Sugiura 1991), *Pinus thunbergii* (Tsudzuki et al. 1994), *Chlorella vulagaris* (Wakasugi et al. 1997), *Prototheca wickerhamii* (Knauf 2001) and *Cyanophora paradoxa* (Stirewalt et al. 2000) revealed the presence of split tRNA-Leu (UAA) genes and introns while inspection of the complete genomes of *Porphyra purpurea* (Reith and Mullholland 1995) and *Cyanidium caldarium* (Glöckner et al. 2000) revealed the presence of unsplit tRNA-Leu (UAA) genes (Table 3.4.1). The insertion site of the intron is conserved with all 33 introns of the introns are inserted between the first wobble and second base of the UAA anticodon. The introns are highly variable in length ranging from 211 to 320 bp. This variation in length was due primarily to the extension of the P5 element in members of the heterocyst forming cyanobacteria as previously described (Paulsrud and Lindblad 1998). Together this systematic survey revealed the presence of 34 introns in 89 taxa representing all the major cyanobacterial lineages and plastids at the ordinal level (Table 3.4.1).

The distribution of the intron was mapped to a minimum evolution tree based on LogDet distances for all 89 taxa used in the systematic survey (Fig. 3.4.1). The intron is present in all heterocyst forming cyanobacteria (NOST) examined in this study (Fig. 3.4.1). Heterocyst forming cyanobacteria constitute a natural monophyletic assemblage in all analyses presented to date (Wilmotte et al. 1994, Turner 1997, Turner et al. 1999, Ishida et al. 2001). However, the retention of the intron in this lineage is not universal and there are some reports of intronless genes in this group (Rudi and Jakobsen 1999, Wright et al. 2001). Outside of the heterocysts forming cyanobacteria the distribution of the intron is much more sporadic and not as clear-cut as previously supposed (Paquin et al. 1997). The sequence groups CHAM and GBAC (Turner 1997) did not contain any representatives with tRNA-Leu (UAA) genes

Table 3.4.1 Strains of cyanobacteria and plastids used in this study and the accession numbers for tRNA-Leu (UAA) intron and 16S rRNA sequences. Taxonomic scheme according to Rippka and Herdman (1992) and Burger-Wiersma et al. (1989). Organisms with their names in quotes are likely to have been misidentified.

Strain	Source	16S rRNA	tRNA-Leu (UAA)
Cyanobacteria			
Order Chroococcales			
Chamaesiphon polonicus SAG 32.87	This study	XXXX	———
Chamaesiphon subglobosus PCC 7430	Paquin et al. 1997	XXXX	———
Cyanobacterium stanieri PCC 7202	This study	AF132782	———
Cyanothece ATCC 51142	This study	AF132771	———
Cyanothece PCC 7418	This study	AF296872	———
Cyanothece PCC 7424	This study	AF132932	———
Dactylococcopsis salina PCC 8305	This study	AJ000711	———
Gloeobacter PCC 8105	This study	AF132791	———
Gloeobacter violaceus PCC 7421	Paquin et al. 1997	AF132790	———
Gloeocapsa PCC 73106	This study	AF132784	XXXX
Gloeothece membranacea PCC 6501	This study	X78680	———
Microcystis aeruginosa PCC 7806	This study	U03402	———
Microcystis aeruginosa PCC 7941	This study	U40340	———
Synechococcus elongatus PCC 6301	Sugita et al. 1995	X03538	D42186
Synechococcus gracile PCC 6307	This study	AF001477	———
Synechococcus leopoliensis PCC 7942	This study	AF132930	XXXX
Synechococcus PCC 7002	This study	AJ000716	———
Synechococcus PCC 7902	This study	AF216946	———
Synechococcus WH 8102	Genome	AF001480	———
Synechocystis PCC 6803	Paquin et al. 1997	D64000	———
Synechocystis PCC 6906	Paquin et al. 1997	XXXX	———
Order Oscillatorales			
Arthrospira PCC 8005	This study	X70769	XXXX
Geitlerinema PCC 7105	Paquin et al. 1997	AF132780	———
Leptolyngbya PCC 73110	This study	X84810	———
Leptolyngbya PCC 7375	Paquin et al. 1997	AF132786	U83256
Lygnbya aestuarii PCC 7419	This study	AJ000714	XXXX
Oscillatoria neglecta IAM-M82	This study	AB003168	XXXX
Oscillatoria PCC 6304	Paquin et al. 1997	XXXX	U83255
Oscillatoria rosea IAM-M220	This study	AB003164	———
Oscillatoria sancta PCC 7515	This study	AF132933	———
Phormidium mucicola IAM-M221	This study	AB003165	———
Planktothrix prolifera NIVA-CYA 18	This study	X84811	XXXX
Plectonema boryanum UTEX 485	This study	AF132793	———
Pseudanabaena PCC 6903	Paquin et al. 1997	AF132778	———
Pseudanabaena PCC 7367	This study	AF091108	XXXX
Pseudanabaena PCC 7403	Paquin et al. 1997	AB039019	U83253
Pseudanabaena PCC 7409	This study	AF132787	———
Spirulina major PCC 6313	This study	X75045	———
Spirulina subsalsa IAM-M223	This study	AB003166	———
Trichodesmium NIBB 1067	This study	X70767	———
Order Pleurocapsales			
"Chroococcidiopsis" CCMP1489	This study	XXXX	———
"Chroococcidiopsis" PCC 6712	This study	XXXX	———
Chroococcidiopsis BB 79. 1	This study	XXXX	———

Chroococcidiopsis BB 80. 1	This study	XXXX	———
Chroococcidiopsis BB 86.6	This study	XXXX	———
Chroococcidiopsis BB 90.5	This study	XXXX	———
Chroococcidiopsis cubana SAG 33.74	This study	XXXX	———
Chroococcidiopsis SAG 2023	This study	XXXX	XXXX
Chroococcidiopsis SAG 2025	This study	XXXX	———
Chroococcidiopsis SAG 2026	This study	XXXX	XXXX
Chroococcidiopsis SAG 2027	This study	XXXX	———
Chroococcidiopsis thermalis PCC 7203	This study	AB039005	———
Cyanosarcina BB 80.2	This study	XXXX	
Dermocarpella SAG 29.84	This study	XXXX	———
Gloeocapsa BB 80.5	This study	XXXX	
Myxosarcina PCC 7312	Paquin et al. 1997	XXXX	———
Myxosarcina PCC 7325	This study	XXXX	———
Pleurocapsa minor PCC 7327	Rudi and Jakobsen 1999	Z82810	AJ228713
Pleurocapsa minor SAG 4.99	This study	XXXX	———
Pleurocapsa PCC 7516	This study	X78681	———
Stanieria cyanosphaera PCC 7437	Paquin et al. 1997	AF132931	———
Xenococcus PCC 7305	This study	AF132783	———
Order Stigonematales			
Chlorogloeopsis fritschii PCC 6718	This study	AF132777	XXXX
Ficherella SAG 2027	This study	XXXX	XXXX
Fischerella muscicola PCC 7414	Paquin et al. 1997	AF132788	U83258
Nostochopsis BB 92. 1	This study	XXXX	XXXX
Order Nostocales			
Anabaena PCC 7122	This study	AF091150	XXXX
Aphanizomenon PCC 7905	Rudi and Jakobsen 1999	AJ133154	AJ228706
Calothrix desertica PCC 7102	Paquin et al. 1997	AF132779	U83252
Cylindrospermum PCC 7417	Paquin et al. 1997	AF132789	U83250
Nodularia spumigena PCC 7804	This study	AJ133181	XXXX
Nostoc PCC 7120	Xu et al. 1990	X59559	M38692
Nostoc punctiforme PCC 73102	Paquin et al. 1997	AF027655	U83254
Nostoc SAG 2028	This study	XXXX	XXXX
Scytonema hofmannii PCC 7110	Kuhsel et al. 1990	AF132781	M61164
Prochorophyta			
Prochlorothrix hollandica	Rudi and Jakobsen 1999	AF132792	AJ228699
Prochlorococcus marinus MIT 9313	Genome	AF053399	———
Prochlorococcus marinus MED4	Genome	AF001466	———
Chloroplasts			
Glaucocystophyta			
Cyanophora paradoxa UTEX 555	Stirewalt et al. 2000	U30821	M22563
Glaucocystis nostochinearum SAG 45.88	This study	X82496	XXXX
Rhodophyta			
Chondrus crispus	This study	Z29521	———
Palmaria palmata	This study	Z18289	———
Porphyra purpurea	Reith and Mullholland 1995	U38804	———
Virideplantae			
Chlorella vulagaris C-27	Wakasugi et al. 1997	AB001684	AB001684
Marchantia polymorpha	Shimada and Sugiura 1991	X04465	X04465
Mesostigma viride NIES-296	Lemieux et al. 2000	AF166114	———
Pinus thunbergii	Tsudzuki et al. 1994	D17510	D17510
Prototheca wickerhamii	Knauf 2001	AJ245645	AJ245645

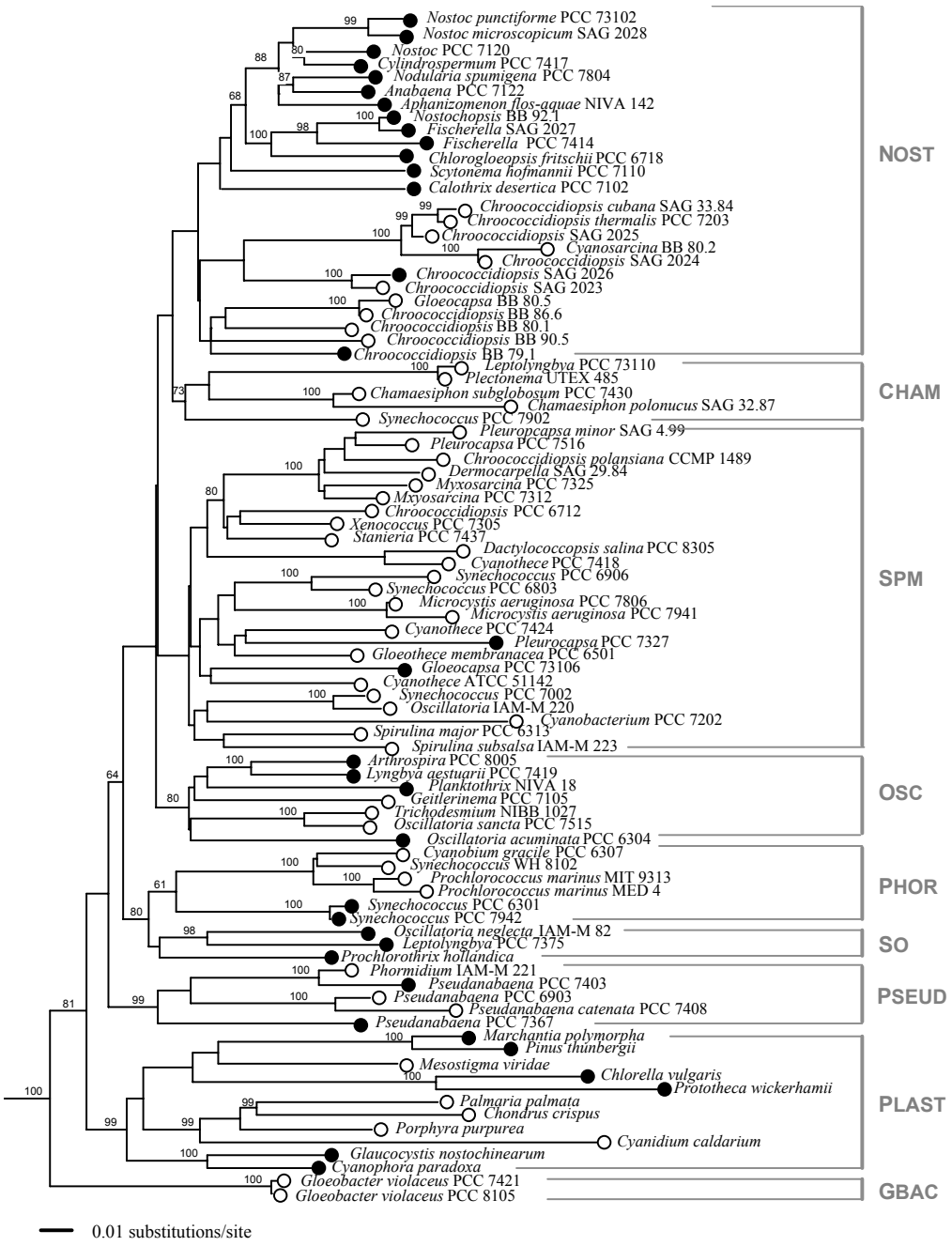

Figure 3.4.1. The sporadic distribution of the tRNA-Leu (UAA) intron in cyanobacteria based on a 16S rRNA tree. The minimum evolution tree is based on LogDet distances with constant sites excluded in proportion to base frequencies. Open circles denote the absence tRNA-Leu (UAA) intron while filled circles denoted the presence of the intron. Numbers above the nodes are bootstrap values (1000 replicates) from a minimum evolution analyses using the LogDet transformation. Only values above 60% are shown. The taxa presented in the tree are grouped according to Turner 1997.

interrupted by introns. The remainder of the sequence groups contained members that either had uninterrupted or interrupted tRNA-Leu (UAA) genes with no real discernible phylogenetic pattern (Fig. 3.4.1). Members of the SPM group are notable for the absence of the intron (Fig. 3.4.1). However, it is important to note that the lack of intron containing tRNA-Leu (UAA) genes is not universal in this lineage as previously proposed (Paquin et al. 1997). Both *Pleurocapsa* PCC 7327 and *Gloeocapsa* PCC 73106 are here demonstrated to contain tRNA-Leu (UAA) introns. In general, the intron distribution is sporadic. This is important because sporadic distribution of intron containing taxa is one criterion used to infer the horizontal transfer of group I introns, often in the absence of any supporting phylogenetic evidence (Paquin et al. 1997).

3.4.3.2 Phylogenetic analyses

Intron and host 16S rRNA datasets were constructed and concatenated to compare the phylogenetic history of the two. An appropriate statistical test for differences in phylogenetic history is found in the partition homogeneity test (Farris et al. 1997, Swofford 2001), in which the sum of the lengths of the most parsimonious trees fitted to the two datasets independently is compared with the sum of the lengths of trees fitted to random partitions of the same data. We compared the 16S rRNA gene and the intron and the datasets differ significantly by the partition homogeneity test ($P = 0.001$). However, empirical studies suggest that the partition homogeneity test can over-estimate incongruence. Recent studies have corroborated this observation (Cunningham 1997, Gatesy et al. 1999, Smith 2000). Templetons test, the winning sites test, and the Kishino-Hasegawa test were undertaken to compare trees generated from the intron and 16S rRNA dataset also demonstrated that the two datasets estimate trees which are statistical different from one another. Each of the maximum likelihood and maximum parsimony tests revealed incongruence between the intron and 16S rRNA datasets ($P = <0.001$). Such phylogenetic incongruence is strong evidence for horizontal transmission. To examine the potential phylogenetic incongruence more rigorously we performed more sensitive phylogenetic analyses.

Maximum parsimony based analyses yielded topologies which conflicted with the 16S rRNA host topology (not shown). The accuracy of different types of data analyses is controversial (Swofford and Olsen 1990). Rejection of the null hypothesis in the statistical tests performed was either maximum parsimony or maximum likelihood based. The tree topologies themselves, however, exhibited considerable variation. Depending on the tree reconstruction method different resultant topologies were generated. Individual maximum parsimony, distance and maximum likelihood analyses of the intron dataset gave conflicting trees with poor bootstrap support for the majority of the nodes (not shown). These individual analyses were used to construct a single strict consensus tree in which nodes common to all analyses were retained and the conflicting nodes reduced to polytomies. This was compared to a

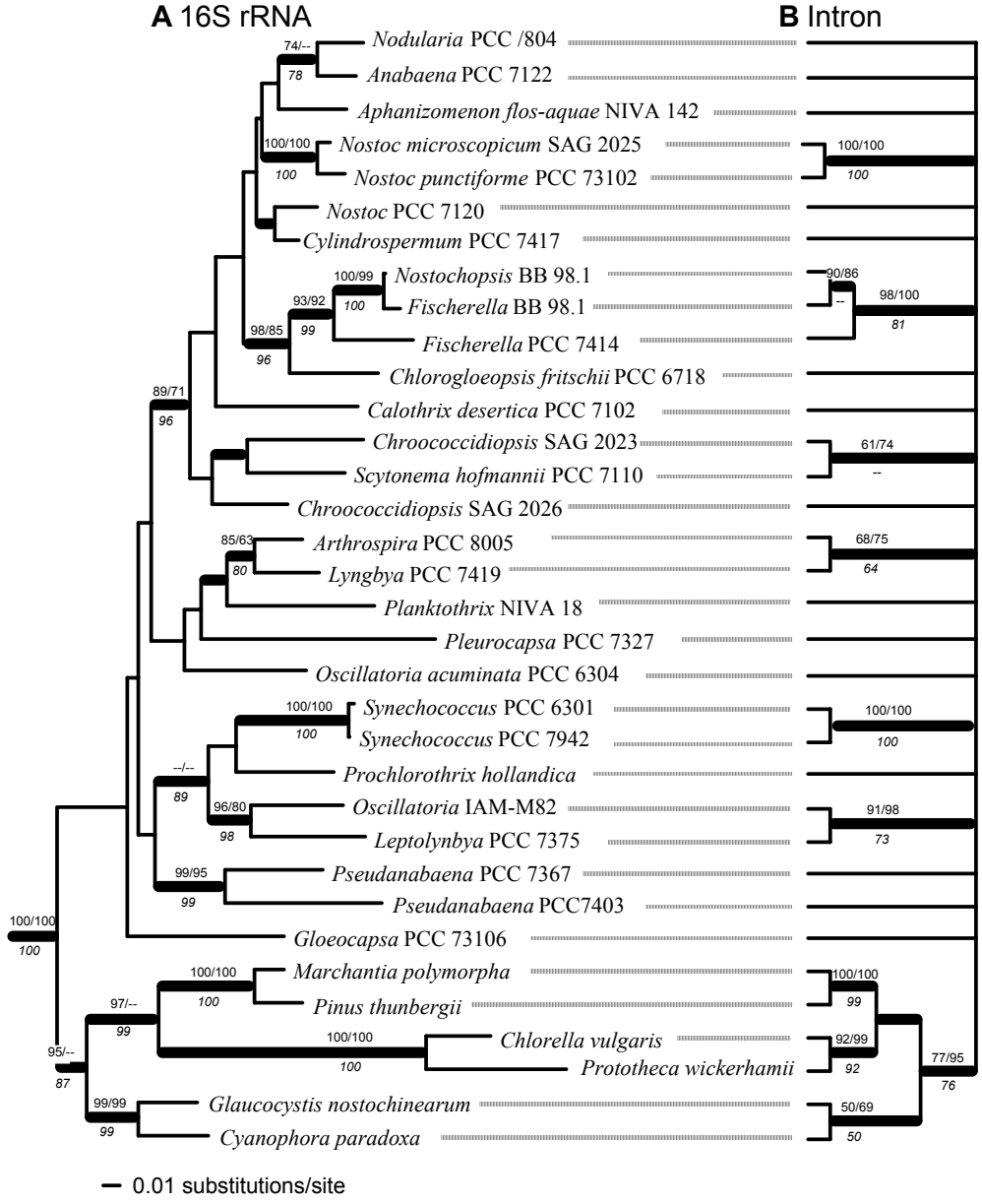

Figure 3.4.2. Congruence between the tRNA-Leu (UAA) intron and the 16S rRNA gene. (A) The maximum likelihood tree on the left is generated from 16S rRNA data using the GTR+I+G model of DNA substitution (Ln=-664484). Thick lines mark internal nodes that were resolved in all distance, maximum parsimony and maximum likelihood trees (see text). Values above internal nodes are bootstrap values from 1000 replications as calculated from minimum evolution using the LogDet transformation (left) or the GTR+I+G model (right), values below branches were calculated from weighted parsimony. Only bootstrap values above 70% are shown. (B) A strict consensus of minimum evolution and maximum likelihood trees based on the TIM+G model of DNA substitution and a maximum parsimony tree. Only nodes resolved in all three topologies remain. Values above internal nodes are bootstrap values from 1000 replications as calculated from minimum evolution using the LogDet transformation (left) or the GTR+I+G model (right), values below branches were calculated from weighted parsimony.

maximum likelihood tree generated from the host 16S rRNA dataset (Fig. 3.4.2). The bootstrap support for the nodes common in all analyses was mapped to the internal nodes common to all analyses (Fig. 3.4.2).

The strict consensus of the maximum parsimony, maximum likelihood and distance analyses yielded support for just 7 instances of congruence between host generated and intron generated phylogenies (Fig. 3.4.2). No instances of horizontal transfer which could be associated with the phylogenetic incongruence suggested in the statistical tests could be found in either the strict consensus tree or the bootstrapping results (Fig. 3.4.2). The strict consensus does not necessarily indicate that the introns diverged simultaneously but that there is insufficient evidence to determine the exact order in which they diverged. It is here assumed that the phylogenetic incongruence is artifactual and stems from conflicts between the individual phylogenetic analyses.

In both the strict consensus tree and the bootstrap analyses relatively robust support (76-95%) for the monophyly of plastid introns was recovered in keeping with the findings of previous studies (Besendahl et al. 2000). Plastid intron monophyly also received strong support using the LogDet transformation (77%) which addresses potential concerns with heterogeneity of base composition (Lockhart et al. 1994). The remaining instances of congruence were confined to terminal nodes which received high bootstrap support in the 16S rRNA trees (Fig. 3.4.2). One notable exception was a sister taxon relationship between *Scytonema hofmannii* PCC 7110 and *Chroococcidiopsis* SAG 2023. This did not receive high bootstrap support in analyses of the host 16S rRNA but received moderate support in distance analyses of the intron dataset. Phylogenetic analyses of the *Chroococcidiopsis* lineage and heterocyst forming cyanobacteria show them to be sister taxa (Section 3.1). The results here are taken to indicate that taxon sampling has an effect on the bootstrap support for reciprocal monophyly of the two lineages. It should be noted however that support for the monophyly of the two lineages (Fig. 3.4.2) remains high here (71-96%).

It is generally acknowledged that the length of the sequence used to reconstruct the evolutionary history of a gene has a profound effect on phylogenetic reconstruction (Poe and Swofford 1999). An increase in the amount of homologous sequence examined often leads to an increase in the resolving power of systematic studies (Teeling et al. 2000, Brown et al. 2001). Small sequence size of the group I introns leads to insufficient resolution and low confidence in phylogenetic analysis (Goddard and Burt 1999). It may be that the lack of resolution with the intron dataset will confound systematics studies which aim to resolve the evolution of this intron.

A recent study proposed an rapid division of a cyanobacterium during the life cycle to explain the loss of the intron (Strehl et al. 1999). The absence of an intron in the tRNA-Leu (UAA) gene and the additional finding that the ribosomal operon exists as a single copy suggests that the deletion of non-essential sequences played a major role in the evolution of *Prochlorococcus marinus* (Strehl et al. 1999). It is hypothesized that a small genome may have been advantageous in the adaptation to very oligotrophic marine conditions (Strehl et al. 1999). Thus the absence of the tRNA-Leu (UAA) intron in some cyanobacteria reflects the selection against retention of the intron in fast growing cells because their replication transcription and excision involves a bioenergetic expense. However, not all of the cyanobacteria which lack the intron are rapidly dividing cyanobacteria (Boone and Castenholz 2001). It is therefore not inconceivable that there have been multiple independent losses of the intron.

The sporadic distribution of the intron revealed in this study was unanticipated. The sporadic distribution of group I introns can be explained by two models: intron insertion or intron deletion (Burke 1988). The first model, intron insertion, is hypothesized to have begun with a gene devoid of introns followed by subsequent insertion of one or more introns (Burke 1988). The deletion model proposes a gene initially containing one or more introns, after which precise deletion of these introns occurs; that is, non-mobile introns are destined to be lost over time if they cannot reinfect homologous sites. In the light of the phylogenetic evidence and the absence of any support for lateral transfer it is likely that the intron is indeed ancient as originally hypothesized. Any model which proposes horizontal transfers to account for the high sequence homology amongst the introns has to explain the monophyly of the plastid introns in this study and others (Besendahl et al. 2000, Paquin et al. 1997). While the high sequence homology of the introns is consistent with a recent origin it should be noted that there is no phylogenetic evidence to support this at present. While the analyses presented here yields support for vertical inheritance this does not preclude a role for horizontal transfer in the evolutionary history of this intron.

3.5 Universal retention of the tRNA-Leu (UAA) intron in the chloroplasts of landplants is coupled with the near pervasive loss of the intron in all other chloroplasts

3.5.1 Abstract

In this study the distribution of the group I intron interrupting the tRNA-Leu (UAA) gene in chloroplasts was surveyed. The tRNA-Leu (UAA) intron is common in members of the land plant lineage and widely used in micro-evolutionary studies. It is shown here that the distribution of this intron is much more sporadic in all other eukaryotic lineages containing chloroplasts. The present-day distribution of the intron in plastids is consistent with an evolutionary history characterised by vertical transmission, with no losses in land plants, several losses among green algae and their secondary derivatives, and nearly pervasive loss in the red algae and their secondary derivatives. The long-standing hypothesis that the plastid introns have lost their ability to catalyze their own excision is confirmed. Early branching plastids can complete the first step of the two-step reaction wheras all land plants can complete neither step in vitro. Cyanobacteria that predate the endosymbiotic event retain the ability to self-splice. This pattern is consistent with the argument that plastids introns have lost the ability to catalyze their own excision. All leucine introns retain the catalytic core necessary for splicing. Although speculative, the heavily biased distribution of this intron in plastids suggests that the processing of the tRNA intron in the land plant lineage may differ form all other lineages.

6.2 Introduction

The enslavement of a photosynthetic prokaryote by a primitive eukaryote and its conversion to energy-producing organelle was one of the key evolutionary events contributing to the biodiversity of present day organisms. The three primary lineages of photosynthetic organelles and their derivatives have been shown to constitute a monophyletic group suggesting that endosymbiosis which gave rise to these organelles occurred on just a single occasion (Martin et al. 1998, Morden et al. 1992, Delwiche et al. 1995, Turner et al. 1999). These three lineages are thought to have diverged nearly simultaneously from each other (Delwiche and Palmer 1997), with the cyanelles most of the glaucocystophytes most likely being the first divergence (Martin et al. 1998). Many eukaryotes have been shown to have recieved their chloroplasts second through a symbiotic relationship with another photosythetic eukrayote (Douglas 1998). A consensus on the timing and mechanisms that gave rise to such evolutionary chimaeras is emerging through the phylogenetic analyses of a combination of nuclear and plastid encoded genes.

A group I intron interrupting tRNA-Leu (UAA) gene in *Zea mays* was first reported over two decades ago (Steinmetz *et al*. 1982). This intron is now known to interrupt the tRNA-Leu (UAA) gene of both cyanobacteria and chloroplasts and is thought that this intron is at least a billion years old predating the endosymbiotic event that gave rise to chloroplasts (Xu et al. 1990, Kushel et al. 1990). To date the plastid encoded tRNA-Leu (UAA) introns have been identified members of the Chlorophyta (Kushel et al. 1990, Wakasugi et al. 1997), the heterokonts secondary derivatives of the Rhodophyta (Kushel et al. 1990, Besendahl et al. 2000) and a single member of the Glaucocystophyta (Evrard et al. 1988).

The tRNA-Leu (UAA) intron is widely used in micro-evolutionary studies in higher plants (e.g. Gielly and Taberlet 1994, Kajita et al. 1998, Wittzell 1999, Bayer et al. 2000, Fukuda et al. 2001). A tacit assumption behind these studies is that the intron is vertically inherited. While this seems likely to be true it has never been explicitly demonstrated. This is important because it has been suggested on a number of occasions that the distribution of the intron in cyanobacteria and chloroplasts could be affected by reverse transcription of the intron into homologous positions in unrelated organisms (Daros and Flores 1996, Rudi and Jakobsen 1999). The first step of reverse splicing is mechanistically the same as intron circulation a phenomenon known to occur in vitro and in vivo in cyanobacteria and plants (Zuag et al. 1993, Daros and Flores 1996).

Group I introns have a distinct and conserved RNA secondary structure essential for splicing and often undergo auto-catalytic excision from coding regions (Kruger et al. 1982). The chloroplast tRNA-Leu (UAA) introns are missing structural elements which are essential for completing auto-excision in other group I introns and it was predicted that they would splice inefficiently in the absence of protein (Cech 1990). Consistent with such an explanation, the chloroplast tRNA-Leu (UAA) intron cannot splice in vitro (Evrard et al. 1988, Xu *et al*. 1990). In contrast it has been proposed that all tRNA-Leu (UAA) introns from cyanobacteria are self-splicing (Xu *et al*. 1990). Thus it is hypothesised that the intron evolved from a self-splicing intron into an intron that is completely dependent upon protein factors for excision following the endosymbiotic event that gave rise to chloroplasts. However, despite widespread acceptance of the hypothesis that self-splicing capabilities of the tRNA-Leu (UAA) intron from chloroplasts and retention of self-splicing abilities of the intron in cyanobacteria there has been no attempt to verify this on a broader scale. To date this hypothesis has not been tested rigorously and is based on just a few isolated attempts to induce splicing (Evrard et al. 1988, Xu et al. 1990). It is not clear if all chloroplasts lack the ability to catalyze their own excision.

This study aims to complete the distribution of the tRNA-Leu (UAA) intron in chloroplasts, assess if the intron has been vertically inherited over 1 billion years, to fill out the gaps in our knowledge of the phylogenetic distribution of the intron in early branching chloroplasts and to test the auto-catalytic

properties of the introns from these different lineages of chloroplasts. Twenty-four new intron sequences were determined and phylogenetic methods were used to elucidate the evolutionary history of these and the published tRNA-Leu (UAA) intron sequences.

3.5.3 Results and discussion

3.5.3.1 Distribution of the intron

The intron was PCR amplified from a diverse selection of cyanobacteria and chloroplasts and identified 24 new tRNA-Leu (UAA) genes interrupted by introns (Table 1). All putative intron containing products were sequenced and shown to contain a intron positioned at the identical site between the U (wobble) and the A (second base) of the UAA anticodon loop.

Green algae and land plants trace their evolutionary history to a unique common ancestor. This lineage is phylogenetically subdivided into two distinct assemblages, the Chlorophyta and the Streptophyta (Friedl 1997). The Chlorophyta includes the Chlorophyceae, Trebouxiophyceae, Ulvophyceae, and Prasinopohyceae, whereas the Streptophyta includes the Charophyceae, bryophytes, ferns and all other multicellular land plants collective termed the Embryophyta. All streptophytes examined here were shown to contain interrupted tRNA-Leu (UAA) genes (Table 3.5.1). Within the streptophyta, the intron was present at least in the common ancestor of the charophytes and appears to be universally present there after (Table 3.5.1). The tRNA-Leu intron from *Zea mays* (X86563) was used to query the BLAST server (http://www.ncbi.nlm.nih.gov/BLAST). From the top BLAST hits a minimum of 2500 group I introns interrupt tRNA-Leu (UAA) genes in the chloroplasts of land plants were identified n the Genbank database. This suggests that the tRNA-Leu intron has been present for the 470-million-year history of plants on land the bryophytes, pteridophytes and gymnosperms and ecological dominant angiosperms. Within the green algae at the base of the land-plant lineage while the majority of the taxa examined contained an intron there are several instances of uninterrupted tRNA-Leu genes (Table 3.5.1). Three losses are apparent in zygnematales, *Chara hispida* (Kuhsel et al. 1990), *Klesormidium flaccidium* and *C. brebissonii* (Besendahl et al. 2000).

In contrast, just 7 of the 28 green algal members of the Chlorophyta examined contained introns (Table 3.5.1). The ulvophytes *Derbesia marina* and *Bryopsis plumosa*, the chlorophytes *Bracteacoccus minor,* and *Hormotilopsis tetravacuolaris*, and the Trebouxiophytes *Leptosira obovata*, *Prototheca wickerhamii* and *Chlorella vulgaris* contain tRNA-Leu (UAA) genes interrupted by the intron. None of the

eight prasinophytes examined contained the intron. Of these the introns from *Leptosira obovata*, *Hormotilopsis tetravacuolaris* and *Bracteacoccus minor* are novel to this study. The absence of the intron in *Mestostigma viridae* (Lemieux et al. 2000), *Nephoselmis olivacea* (Turmel et al. 1999) supports the sporadic distribution of the intron in the Chlorophyta.

Two cryptophytes, four haptophytes, twenty heterokonts, and eighteen red algae were screened for the presence or absence of the intron. No intron was found in any of the 10 bangiophytes or 6 florideophytes that were examined. These results are consistent with sequence data from the bangiophytes *Cyanidioschyzon merolae*, *Cyanidium caldarium* (Ohta 1997) and *Porphyra purpurea* (Reith and Munholland 1995) that show intron lacking tRNA-Leu (UAA) genes. Kuhsel et al. (1990) reported that *Smithora naiadum* also lacks a plastid tRNA-Leu (UAA) intron. Initially reports which suggested that the tRNA-Leu (UAA) gene of *Gracilaria lemaneiformis* contained and intron was subsequently shown to be erroneous (Besendahl et al. 2000). Together, these data suggest that over 40 red algae and probably most extant Rhodophyta lack this intron in their tRNA-Leu (UAA) genes.

The plastids of the cryptophytes, heterokonts, and haptophytes are most likely derived from independent secondary endosymbioses of red algae by three disparate eukaryotic hosts (Douglas et al. 1991, Medlin et al. 1995, 1997, Daugbjerg and Andersen 1997, Douglas and Penny 1999, Douglas et al. 2001). While it remains debatable, it is generally accepted that these three lineages of plastid are monophyletic and derive from basal red algae (Besendahl et al. 2000). The PCR analyses of the cryptophyte and haptophyte lineges also failed to reveal the presence of the tRNA-Leu (UAA) intron. This result is supported by sequence data of the uninterrupted tRNA-Leu gene from the cryptophyte *Guillardia theta* (Douglas and Penny 1999). However, it is important to note just a small number of taxa were examined from these two lineages and the possibility that at some point in the future taxa with intron containing tRNA-Leu genes will be discovered cannot be ruled out. Within the heterokonts, we found taxa with and without the tRNA-Leu (UAA) intron. The chrysophyte *Ochromonas danica* and the diatom *Odontella sinensis* both lack the intron (Kowallik et al. 1995). The six diatoms (Bacillariophyceae) examined here all lacked the intron (Table 3.5.1). All of the 10 strains of xanthophytes and phaeophytes that were examined contained the intron. The phaeophytes and xanthophytes form a monophyletic group within the Heterokontophyta in both nuclear small subunit ribosomal DNA and chloroplast rbcL phylogenies (Daugbjerg and Andersen 1997, Potter et al. 1997).

The two members of the Glaucocystophytes examined in this study, *Glaucocystis nostochinearum* and *Cyanoptyche gloeocystis* contained the tRNA-Leu (UAA) intron. This is supported by the presence of the introns in the cyanelle of *Cyanophora paradoxa* (Evrard et al. 1988). Glaucocystophytes are currently believed to contain the earliest branching plastids (Helmchen et al. 1995, Turner et al. 1999).

Table 3.5.1. The presence or absence of the intron as determined by PCR for representatives of all the major linages of chloroplasts. A dash is used to indicate the absence of an intron in the third column. Accession numbers for the plastid encoded 16S rRNA and *rbc*Lgene sequences used in the partition homogeneity test and the nuclear encoded 18S rRNA gene sequences are given in columns 4-7. A dash is used to indicate taxa which were not used in these studies. In all cases four XXXX is used to denote unpublished data obtained for the purpose of this study.

Species	Family	Intron	*rbc*L	16S rRNA	18S rRNA
Streptophyta					
Charophytes*					
Chara sp. UGBG	This study	XXXX	———	X75519	U18493
Chlorokybus atmophyticus SAG 48.80	This study	XXXX	———	XXXX	M95612
Closterium cornu SAG 132.80	Besendahl et al. 2000	AF182366	———	———	———
Closterium ehrenbergii SAG 134.80	Besendahl et al. 2000	AF182367	———	———	AF115437
Closterium littorale SAG 611-7	Besendahl et al. 2000	AF182363	———	———	AF115438
Coleochaete orbicularis UTEX 2651	Besendahl et al. 2000	AF182368	———	U24579	M95611
Cosmarium botrytis SAG 136.80	Besendahl et al. 2000	AF182362	———	———	———
Cylindrocystis brebissonii SAG 615-1	Besendahl et al. 2000	———	———	———	AF115439
Genicularia spirotaenia SAG B 54.86	Besendahl et al. 2000	AF182365	———	———	X74753
Klebsormidium flaccidum SAG 335-2b	This study	———	———	———	X75520
Mesotaenium caldariorum SAG 230-1	This study	XXXX	———	———	X75763
Penium margaritaceum SAG B 22.82	Besendahl et al. 2000	AF182364	———	———	AF115440
Roya obtusa SAG B 168.80	Besendahl et al. 2000	AF182361	———	———	———
Spirogyra sp. SAG 170.80	This study	XXXX	———	U24596	U18523
Bryophyta					
Funaria hygrometrica	La Farge et al. 2000	AF231175	———	———	———
Marchantia polymorpha	Ohyama et al. 1986	X04465	———	X04465	———
Mnium hornum	Besendahl et al. 2000	AF182360	———	———	———
Pellia epiphylla	Meißner et al. 1998	AF071842	———	AF152609	———
Phaeoceros	This study	XXXX	———	AF244561	———
Polytrichum commune	Newton et al. 2000	AF231907	———	AF244563	———
Sphagnum palustre	Newton et al. 2000	AF231902	———	U24592	———
Tortula obtusissima	La Farge et al. 2000	AF231180	———	———	———
Embryophyta					
Adiantum raddianum UGBG	This study	XXXX	———	AF244549	———
Alnus viridis	Gielly and Taberlet 1995	X76810	———	U03555	———
Angiopteris evecta UGBG	This study	XXXX	———	U24580	———
Antirrhinum kelloggii	Freeman unpubl.	AF034886	———	———	———
Arabidopsis thaliana	Fangan et al. 1994	X74573	———	AP000423	———
Betula pubescens	Gielly and Taberlet 1994	X75698	———	———	———
Botrychium lunaria	Meißner et al. 1998	AF071833	———	U24581	———
Cycas circinalis UGBG	This study	XXXX	———	———	D38245
Ephedra UGBG	This study	XXXX	———	U24584	———
Equisetum scirpoides	Besendahl et al. 2000	AF182359	———	U24593	———
Gnetum gnemon UGBG	This study	XXXX	———	AJ007508	———
Hordeum vulgare	Gielly and Taberlet 1994	X75705	———	———	———
Huperzia lucidula	Wikström et al. 1999	AJ224591	———	AF244556	———
Huperzia reflexa	Wikström et al. 1999	AJ224581	———	———	———
Isoetes UGBG	This study	XXXX	———	U24585	———
Juniperus rigida	Kusumi et al. 2000	AB030064	———	U24586	———
Lycopodiella inundata	Wikström et al. 1999	AJ224605	———	———	———
Lycopodiella lateralis	Wikström et al. 1999	AJ224609	———	———	———
Lycopodium fastigiatum	Wikström et al. 1999	AJ224595	———	U24587	———
Nicotiana tabacum	Shinozaki et al. 1986	Z00044	———	Z00044	———
Nymphaea mexicana UGBG	This study	XXXX	———	AF244560	L24404

Ophioglossum petiolatum	Meißner et al. 1998	AF071849	———	U24589	———
Picea mariana	Clouser et al. unpubl.	AF156801	———	———	———
Pinus thunbergii	Tsudzuki et al. 1994	D17510	———	D17510	———
Psilotum nudum	Besendahl et al. 2000	AF182358	———	U24590	———
Spartina townsendii	Ferris et al. 1997	Z69913	———	———	———
Triticum aestivum	Gielly and Taberlet 1994	X75709	———	AB042240	———
Zea mays	Maier et al. 1995	X86563	———	X86563	———
Chlorophyta					
Parsinophyceae					
Mantoniella squamata	This study	———	———	———	X73999
Mesostigma viride	Lemieux et al. 2000	———	———	———	AJ250108
Nephroselmis olivacea	Turmel et al. 1999	———	———	———	X74754
Pseudoscourfieldia marina	This study	———	———	———	AF122888
Pycnococcus provasolii	This study	———	———	———	AF122889
Pyramimonas amylifera SAG 22.86	This study	———	———	———	AB017122
Scherffelia dubia	This study	———	———	———	X68484
Tetraselmis striata SAG 41.85	This study	———	———	———	U05039
Ulvophyceae					
Gloeotilopsis paucicellulare SAG 463.1	This study	———	———	———	Z28970
Byropsis plumosa	Kuhsel et al. 1990	M61159	———	———	———
Cladophora albida UTEX LB 1477	This study	———	———	———	Z35317
Derbesia marina	Besendahl et al. 2000	Genbank?	———	———	———
Enteromorpha intestinalis SAG 320-1a	This study	———	———	———	AJ000040
Neomeris dumentosa UTEX LB 2691	This study	———	———	———	Z33469
Ulva fasciata UTEX LB1422	This study	———	———	———	
Chlorophyceae					
Botryococcus braunii SAG 30.81	This study	———	———	———	X78276
Bracteacoccus minor UTEX 66	This study	XXXX	———	———	U63097
Carteria eugametos UTEX 233	This study	———	———	———	AF182817
Chlamydomonas reinhardtii	This study	———	———	———	U57697
Chloromonas sp.		———	———	———	U70795
Hormotilopsis tetravacuolaris UTEX 946	This study	XXXX	———	———	U83124
Trentepholia UTEX 1227	This study	———	———	———	XXXX
Trebouxiophyceae					
Chlorella vulgaris C-27	Ohta et al. 1997	AB001684	———	AB001684	AJ242757
Coccomyxa sp. SAG 49.81	This study	———	———	———	XXXX
Leptosira obovata UTEX 319	This study	XXXX	———	———	Z68695
Microthamnion UTEX 1914	This study	———	———	———	Z28974
Prototheca wickerhamii	Knauf unpubl.	AJ245645	———	AJ245645	X56099
Trebouxia arboricola SAG 219.1a	This study	———	———	———	Z68705
Trebouxia jamesii UTEX 2233	This study	———	———	———	Z68700
Glaucocystophyta					
Cyanophora paradoxa UTEX555	Evrard et al. 1988	M22563	M22563	M22563	X68483
Glaucocystis nostochinearum SAG 45.88	This study	XXXX	———	X82496	X70803
Cyanoptyche gloeocystis SAG 4.99	This study	XXXX	———	———	———
Rhodophyta					
Bangiophyceae					
Cyanidioschyzon merolae	Ohta 1997	———	———	———	———
Cyanidium caldarium RK1	Glockner et al. 2000	———	D63676	———	———
Porphyra purpurea	Reith and Mullholland 1995	———	U38804	———	———
Porphyridium aerugineum SAG 110.79	This study	———	———	———	———
Porphyridium purpureum SAG 112.79	This study	———	———	———	———
Porphyridium purpureum SAG 1380-1e	This study	———	———	———	———
Porphyridium purpureum SAG 44.94	This study	———	———	———	———
Rhodella violacea SAG 30.97	This study	———	———	———	———

Rhodochaete parvula SAG 8.99	This study	———	———	———	———
Florideophyceae					
Acrochaetium virgatulum SAG 1.81	This study	———	———	———	———
Antithamnion plumula SAG 3.86	This study	———	———	———	———
Audouinella hermannii SAG 206.80	This study	———	———	———	———
Bonnemaisonia hamifera SAG 118.79	This study	———	———	———	———
Bostrychia radicans SAG 100.79	This study	———	———	———	———
Callithamnion byssoides SAG 6.86	This study	———	———	———	———
Hildenbrandia rivularis SAG 17.96	This study	———	———	———	———
Heterokontophyta					
Bacillariophyceae					
Asterionella formosa SAG 8.95	This study	———	———	———	———
Cyclotella meneghiniana SAG 1020-1a	This study	———	———	———	———
Gomphonema parvulum SAG 1032-1	This study	———	———	———	———
Nitzschia palea SAG 1052-3b	This study	———	———	———	———
Odontella sinensis	Kowallik et al. 1995	———	———	———	———
Phaeodactylum SAG 1090-1a	This study	———	———	———	———
Pinnularia neomaior SAG 80.94	This study	———	———	———	———
Xanthophyceae					
Botrydiopsis alpina SAG 806-1	This study	XXXX	AF015587	———	———
Botrydium granulatum SAG 805-4	This study	XXXX	AF064743	———	———
Heterococcus caespitosus SAG 835-2a	This study	XXXX	AF084610	———	———
Vaucheria bursata	Kuhsel et al. 1990	M61165	AF015589	———	———
Phaeophyceae					
Bodanella lauterbornii SAG 123.79	This study	XXXX	———	———	———
Costaria costata	Kuhsel et al. 1990	M55288	———	———	———
Dictyota dichotoma	Besendahl et al. 2000	AF182369	AJ287852	———	———
Pelagophycus porra	Miller et al. 2000	AF250314	AJ287858	———	———
Pylaiella littoralis SAG 2000	This study	XXXX	X55372	X14873	———
Scytosiphon lomentaria	This study	XXXX	AB022238	———	———
Chrysophyceae					
Ochromonas danica CBS	This study	———	———	———	———
Cryptophyta					
Chroomonas sp. SAG B980-1	Besendahl et al. 2000	———	———	———	———
Guillardia theta	Douglas and Penny 1999	———	———	———	———
Pyrenomonas helgolandii SAG B 28.87	Besendahl et al. 2000	———	———	———	———
Chilomonas CCAP 977/2A	This study	———	———	———	———
Haptophyta					
Chrysochromulina acantha CCMP 408	Besendahl et al. 2000	———	———	———	———
Emiliania huxleyi CCMP 318	Besendahl et al. 2000	———	———	———	———
Isochrysis UTEX LB 1292	This study	———	———	———	———
Pavlova lutheri CCMP 375	Besendahl et al. 2000	———	———	———	———
Phaeocystis globosa CCMP 374	Besendahl et al. 2000	———	———	———	———
Euglenozoa					
Astasia longa	Gockel and Hachtel 2000				
Euglena gracilis	Hallick et al. 1993	———	———	———	———
Apicomplexa					
Plasmodium falciparum	Wilson et al. 1996	X95276	———	———	———
Toxoplasma gondii	Beckers et al. 1995	U87145	———	———	———

*see McCourt 1995

The pattern we observed suggests that the universal retention of the intron in land plant chloroplasts is coupled with the pervasive loss of the intron in all nearly all other plastid lineages.

Plastids with more than two bounding membranes such as those of euglenoids, dinoflagellates, heterokonts, haptopytes, apicomplexa, cryptomonads, and the chlorarachniophytes arose by secondary endosymbiosis, in which a photosynthetic eukaryotic alga was engulfed and retained by a phagotroph. The plastids of the cryptophyta, haptophyta, and heterokonta are derived from red algal secondary endosymbioses (Oliveira and Bhattacharya 2000). The plastids of the euglenozoa and the apicomplexans on the other hand are thought to be derived from algal secondary endosymbioses in which a non-photosythetic eukaryote engulfed a green alga (McFadden 1999). The evolutionary origins of the apicomplexa plastid genomes are in dispute and alternatively proposed to be of green algal (Kohler et al. 1997) or red algal origin (Williamson et al. 1994). The plastids of *Euglena gracilis* (Hallick et al. 1993) and the colourless heterotrophic euglenoid flagellate *Astasia longa* (Gockel and Hachtel 2000) both lack the intron. Conversely, the remnant plastid genome of the non-photosynthetic apicomplexan parasites, *Plasmodium falciparum* (Wilson et al. 1996), and *Toxoplasma gondii* (Beckers et al. 1995) both contain remnants of tRNA-Leu introns inserted in a homologous position to all other tRNA-Leu introns. The *Plasmodium falciparum* and *Toxoplasma gondii* tRNA-Leu introns are extremely A + T rich (>90%) and are reduced in length (134 and 188 nt respectively). The intron sequences cannot be aligned with other tRNA-Leu introns due to its biased nucleotide content and the lack of conservation amongst the intron core, and therefore its ancestry has not been examined with phylogenetic methods.

Together this systematic survey of the tRNA-Leu intron in plastids revealed the presence of 70 introns in 138 taxa (Table 3.5.1). The numbers of introns interrupting the Streptophyta far exceeds this with many thousands of tRNA-Leu introns available in the Genbank sequence database (www.ncbi.nlm.nih.gov/entrez). For this reason the list is comprehensive of the major lineages of primary and secondary plastids but representative of the Streptophyta lineage.

3.5.3.2 Phylogeny of the tRNA-Leu (UAA) introns

The 234 aligned sequence positions of tRNA-Leu (UAA) intron dataset were submitted to minimum evolution distance anaylses using the maximum likelihood transformation based on the TrN+G model of DNA substitution (Fig. 3.5.1). Within the Streptophyta and the Chlorophyta, the tRNA-Leu (UAA) intron phylogeny generally agrees with the phylogeny of the green algae and the land plants based on nuclear, mitochondrial and plastid-encoded molecular markers (Manhart 1994, Lewis et al. 1997, Duff and Nickrent 1999, Nickrent et al. 2000, Pryer et al. 2001). Remarkably the phylogeny of the tRNA-Leu intron is congruent with what is currently known of land plant phylogeny, particularly the

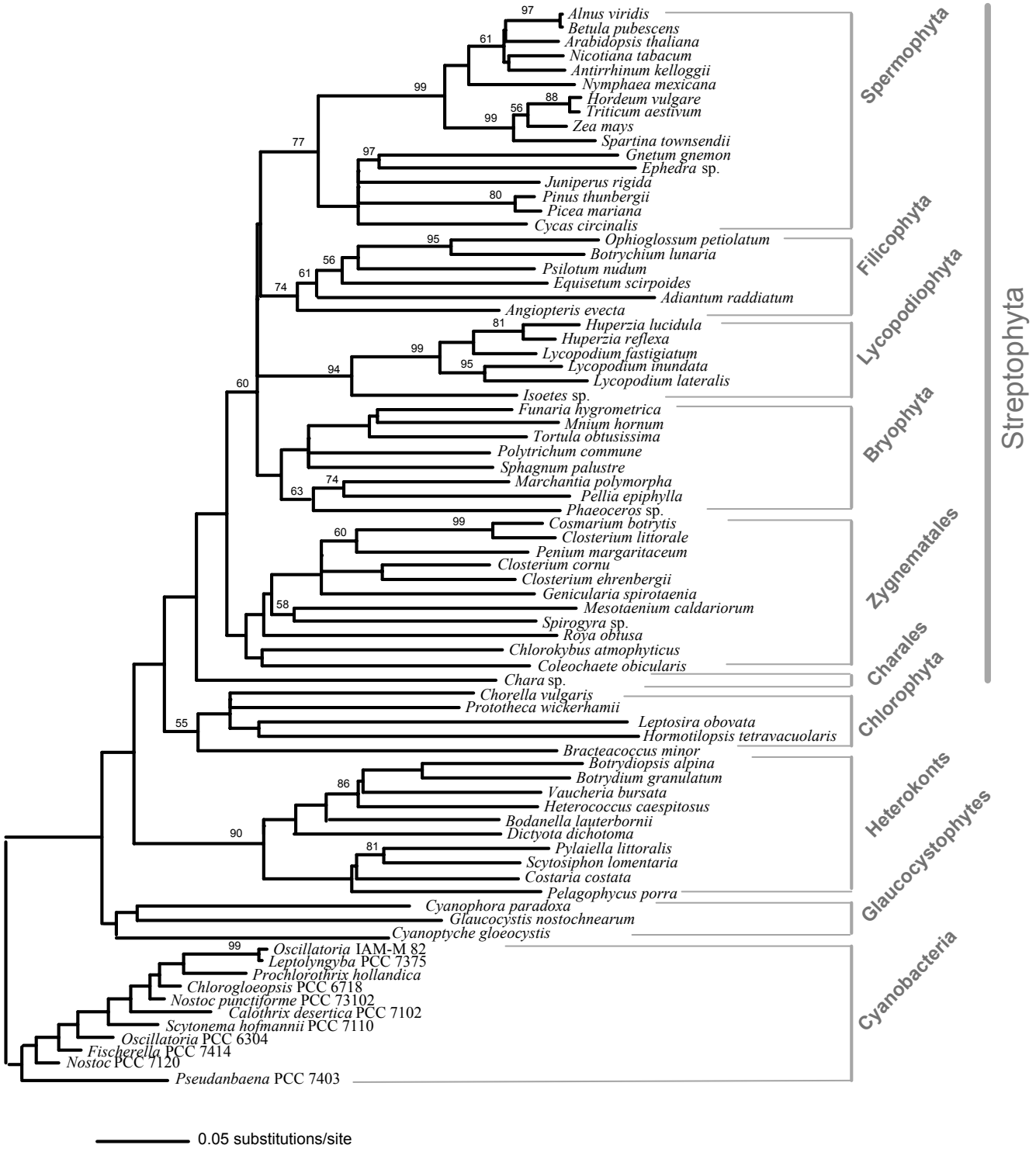

Figure 3.5.1. Phylogeny of the tRNA-Leu (UAA) intron from cyanobacteria, glaucocystophytes, heterokonts, green algae and the streptophyta. A minimum evolution tree with a maximum likelihood transformation based on the TrN+G model of DNA sustitution. Values at the nodes are bootstrap values from minimum evolution analyses using the maximum likeliihood tranformation and same model (1000 replications).

branching order of the earliest lineages based on the concurrence of multigene sequence analyses. The seed plants are monophyletic with moderate bootstrap support (77%). The introns from equisetophytes (horsetails), psilotophytes (whisk ferns) and all eusporangiate and leptosporangiate ferns examined in this study form a monophyltic group with moderate bootstrap support (74%). This monophyletic clades constituents are the closest relatives to seed plants, albeit with low bootstrap support (Fig. 3.5.1). This parallels recent findings which refuted the prevailing view that horsetails and ferns are transitional evolutionary grades between bryophytes and seed plants (Pryer et al. 2001). The introns from the lycophytes form a well supported monophyletic group (94%) which again parallels the findings of a number of recent studies (Nickrent et al. 2000, Pryer et al. 2001).

The introns from the Chlorophyta are monophyletic but with low bootstrap support (55%). Within the Chlorophyta there is no bootstrap support above 50% for any of the internodes (Fig. 3.5.1). Moderate bootstrap support for the monophyly of the Chlorophyta and Streptophyta lineages (74%) is retrieved. Likewise the analyses of the ten introns from the Xanthophyceae and Phaeophyceae revealed robust support for the monophyly of this group (90%). No support for the monophyly of the three glaucocystophyte introns was retrieved in these analyses but these sequences were basal to the remainder of the introns which is consistent with the idea that this is the earliest branching chloroplast lineage (Helmchen et al. 1995).

The tree shown in figure 3.51 is surprisingly consistent with the phylogeny of plastids inferred from plastid encoded small subunit rRNA or other plastid protein coding regions (Bhattacharya and Medlin 1995, Helmchen et al. 1995, Delwiche and Palmer 1997, Martin et al. 1998). The three primary plastid lineages identified by these studies are the green algal/land plant plastids (chloroplasts), the cyanelles of the glaucocystophytes, and the red algal plastids (rhodoplasts), all of which are likely to have arisen from a single common cyanobacterial endosymbiosis (Bhattacharya and Medlin 1995, Delwiche and Palmer 1997, Palmer and Delwiche 1996). Consistent with this idea, the tRNA-Leu (UAA) intron sequences, which represent all three lineages of primary plastids, form a monophyletic group. However, the support for the monophyly of the plastid introns is low in this study (below 50%). The monophyly of the introns from chloroplasts has not been previously questioned. Previous studies have found moderate support for the monophyly of plastid introns (Besendahl et al. 2000).

The distribution of the tRNA-Leu intron in the plastids of the Chlorophyta was reconstructed using the nuclear encoded small subunit rRNA (18S) from 35 taxa (Fig. 3.5.2). The distribution of the intron is surprisingly sporadic and contrasts to the almost universal retention of the intron in the Streptophyta (Table 3.5.1). The Prasinophyceae are polyphyletic in this analyses in keeping with the findings of the majority of phylogenetic analyses to date (e.g. Lemieux et al. 2000, Friedl 1997, Kohler et al. 1997).

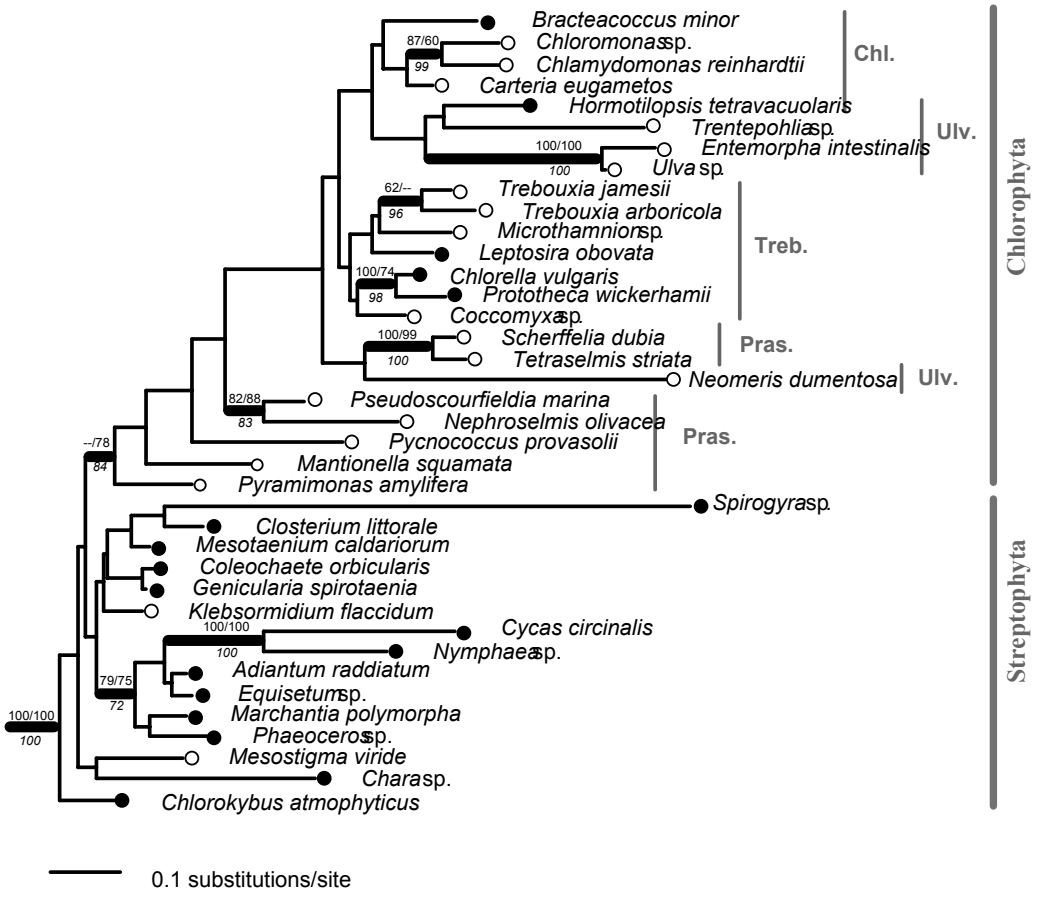

Figure 3.5.2. Distribution of the tRNA-Leu (UAA) intron in the choroplasts of green algae and land plants based on the nuclear small subunit rRNA (18S) gene. Maximum likelihood tree based on the TrN+I+G model of DNA sustitution (Ln=-15397.04). The taxa *Spirogyra* sp. and *Neomeris dumentosa* are on long branches and destablize the maximum likelihood tree. Thick lines mark internal nodes that were resolved in all distance, maximum parsimony and maximum likelihood trees (see text). Values above internal nodes are bootstrap values from 1000 replications as calculated from minimum evolution using the LogDet transformation (left) or the TrN+I+G model (right), values below branches were calculated from weighted parsimony. Only bootstrap values above 60% are shown.

Members of the Ulvophyceae, Chlorophyceae and Trebouxiophyceae all contain members which harbour tRNA-Leu introns in their chloroplasts (Fig. 2.5.2). The two members of the ulvophytes known to contain introns, *Derbesia marina* and *Bryopsis plumosa*, were not included here as no small subunit rRNA (18S) genes from these taxa were available. The order Ulvophyceae is non-monophyleic in this analysis and probably reflects an artefactual placement due to to long branch attraction. Apart from *Prototheca wickerhamii* and *Chlorella vulgaris* the relationships between the intron containing taxa is no well resolved. A split between Chlorophyceae and Trebouxiophyceae introns was to be expected but not found here (Fig. 3.5.1).

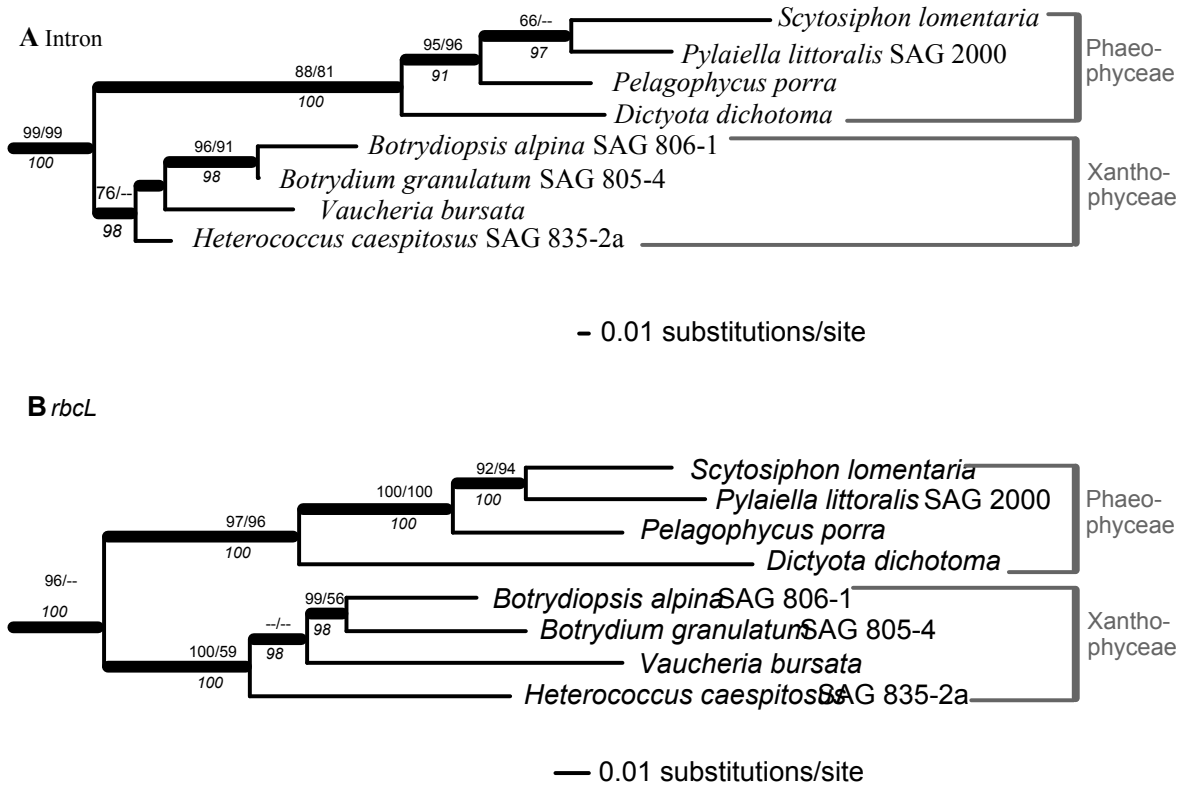

Figure 3.5.3. Congruence between the plastid encoded tRNA-Leu (UAA) intron and rbcL (ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit) gene from xanthophytes and Phaeophytes. (A) tRNA-Leu (UAA) intron maximum likelihood tree (LnL=-1353.55) based on the TrN+G model of DNA substitution. (B) Plastid encoded rbcL maximum likelihood tree (LnL=-4567.38) based on the GTR+G model of DNA substitution. Thick lines mark internal nodes that were resolved in all distance, maximum parsimony and maximum likelihood trees (see text). Values above internal nodes are bootstrap values from 1000 replications as calculated from minimum evolution using the LogDet transformation (left) or the TrN+G or GTR+G model respectively (right), values below branches were calculated from weighted parsimony. Only bootstrap values above 60% are shown.

None of the red algae examined to date based on the PCR analyses of 13 red algae from the classes Bangiophyceae and Florideophyceae presented here and previous data from 13 others (Glocker et al. 2000, Besendahl et al. 2000, Ohta 1997, Reith and Munholland 1995) contain intron containing tRNA-Leu genes. The tRNA-Leu (UAA) intron was almost certainly present in the common ancestor of the red algae. A horizontal transfer could be invoked as a more parsimonious solution to explain the sporadic distribution of the intron observed in red algae and its secondary derivatives. To address these

potential concerns we compared the phylogeny of the tRNA-Leu intron to the phylogeny of the plastid encoded *rbc*L gene of taxa which contain the intron from 4 members of the Xanthophyceae and 4 members of the Phaeophyceae (Fig. 3.5.3). The congruence observed between the plastid encoded *rbc*L marker and the intron is consistent with a history of vertical inheritance (Fig. 3.5.3).

Strict tests of coevolution between 16S rRNA and intron datasets (32 taxa) show that there is a large degree of congruence between the datasets right across the spectrum of plastids (P=0.125). Likewise the partition homogeneity test was used to estimate the degree of congruence between the intron and rbcL dataset at P=1.000. These results are in keeping with numerous micro-evolutionary studies in land plants which found no evidence to prevent the combination of tRNA-Leu intron sequence with other plastid encoded or nuclear encoded genes (e.g. Soliva et al. 2001, Asmussen and Chase 2001, Wallander and Albert 2000, Shaw and Allen 2000, Buck et al. 2000, Wikström and Kenrick 2000). This finding is highly consistent with the proposal that the intron has been stably inherited in land plants for over 470 million years (Kenrick and Crane 1997) and that the concerns over the reverse transcription of the intron into homologous positions in unrelated organisms (Daros and Flores 1996) seem to be largely unfounded. In summary, the phylogenetic and distribution data support the hypothesis that the tRNA-Leu (UAA) intron has been vertically inherited since the engulfment of the cyanobacteria which gave rise to plastids.

3.5.3.3 Sequence and secondary structure conservation of the tRNA-Leu (UAA) intron

All tRNA-Leu intron sequences determined in this study were included in an alignment with characterised from other from cyanobacteria and other plastids in previous studies. The P3, P4, P7, J8/7 regions that form the catalytic core of group I introns were used to guide this alignment (Cech 1988, Michel and Westhof 1990). These regions are conserved in all tRNA-Leu (UAA) group I introns as well as the more peripheral regions such as P2 and P5. Among the Streptophyta introns in our study, the alignable regions varied from 65% and 100% sequence identity. The only significant differences among these homologous introns was in the sequence composition and length of the peripheral loops and helices such as the P6 and P8 regions (Fig. 3.5.4). All tRNA-Leu (UAA) group I introns are the phylogenetically conserved core secondary structure required for intron removal and precise ligation of the exons (Cech 1990, Zuag et al. 1993).

These cursory secondary structure comparisons revealed that the tRNA-Leu (UAA) intron from land plants have extensions in the structural elements which separate the 3' splice site and the catalytic core of the intron. This is consistent with these introns inability to form either the first or second step in

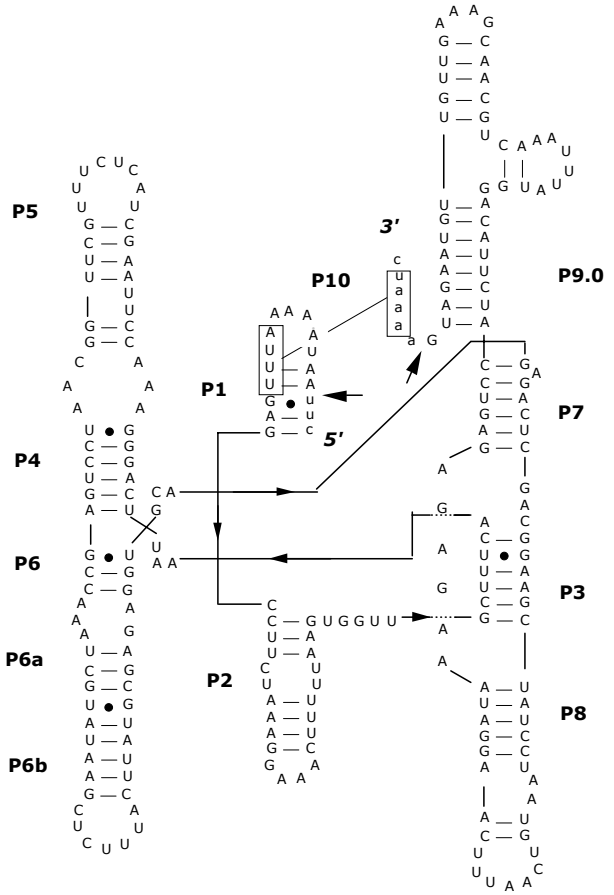

Figure 3.5.4. Secondary structure of the tRNA-Leu intron from the chloroplast of *Chlorokybus atmophyticus* showing the stem elements P1-P9 and a putative 4 basepair P10. Exon seqeucne from teh tRNA-Leu gene is show in lowercase.

vitro (Table 3.5.2). All of the tRNA-Leu (UAA) introns examined retain the conserved core structural elements required for intron removal and precise ligation of the exons and provides the catalytic center for the excision reaction *in vivo* and that all tRNA-Leu (UAA) introns are excised by the same ωG dependent pathway.

3.5.3.4 Self-Splicing

It is a long-standing hypothesis that the chloroplast introns are incapable of auto-excision and are dependent upon the host for removal (Xu et al. 1990). The splicing competence of the tRNA-Leu (UAA) intron from cyanobacteria, basal chloroplast lineages and higher plants was examined here. Introns from representatives of the three photosynthetic lineages and cyanobacteria (Table 3.5.2) were transcribed into RNA and incubated in a splicing buffer containing guanosine. As expected all cyanobacterial tRNA-Leu (UAA) introns were capable of self-splicing *in vitro* in the absence of proteins (Fig. 3.5.5). However, more surprisingly the introns from the glaucocystophytes, heterokonts and

Table 3.5.2. Strains of cyanobacteria and plants in which the tRNA-Leu (UAA) intron can complete either the first, second or neither steps of the two-step transesterification reaction that leads to excision of the intron in the absence of protein. Scoring of the northern blots was done on the basis of the expected size of splicing products. Conditions employed did not allow the sizing of ligated exons.

Name	pre-tRNA	First step	free intron
Cyanobacteria			
Anabaena PCC 7122	+	+	+
Arthrospira PCC 8005	+	+	+
Chlorogloeopsis fritschii PCC 6718	+	+	+
Chroococcidiopsis BB 79.2	+	+	+
Chroococcidiopsis BB 96.1	+	+	+
Fischerella muscicola PCC 7414	+	+	+
Leptolyngbya PCC 7375	+	+	+
Nodularia sphaerocarpa PCC 7804	+	+	+
Nostoc PCC 7120	+	+	+
Oscillatoria IAM-M82	+	+	+
Planktothrix SAG 3.92	+	+	+
Pseudanabaena PCC 7367	+	+	+
Synechocystis PCC 6301	+	+	+
Glaucocystophytes			
Cyanophora paradoxa UTEX 555	+	+	-
Cyanoptyche gloeocystis SAG 4.99	+	+	-
Glaucocystis nostochinearum SAG 45.88	+	+	-
Heterkontophyta			
Bodanella lauterbornii SAG 123.79	+	+	-
Pyliaella littoralis SAG 2000	+	+	-
Chlorophyta			
Chlorella saccharophila SAG 211-1a	+	+	-
Prototheca wickerhamii SAG 263-11	+	+	-
Streptophyta			
Angiopteris evecta	+	-	-
Chara sp.	+	+	-
Chlorokybus atmophyticus SAG 48.80	+	+	-
Coleochaete scutata SAG 110.80	+	+	-
Cycas circinalis	+	-	-
Gnetum gnemon	+	-	-
Isoetes sp.	+	-	-
Mesotaenium caldariorum SAG 230-1	+	-	-
Nicotiana tabacum	+	-	-
Phaeoceros sp.	+	-	-
Pinus strobus	+	-	-
Polytrichium formosum	+	-	-
Spirogyra sp.SAG 170.80	+	+	-

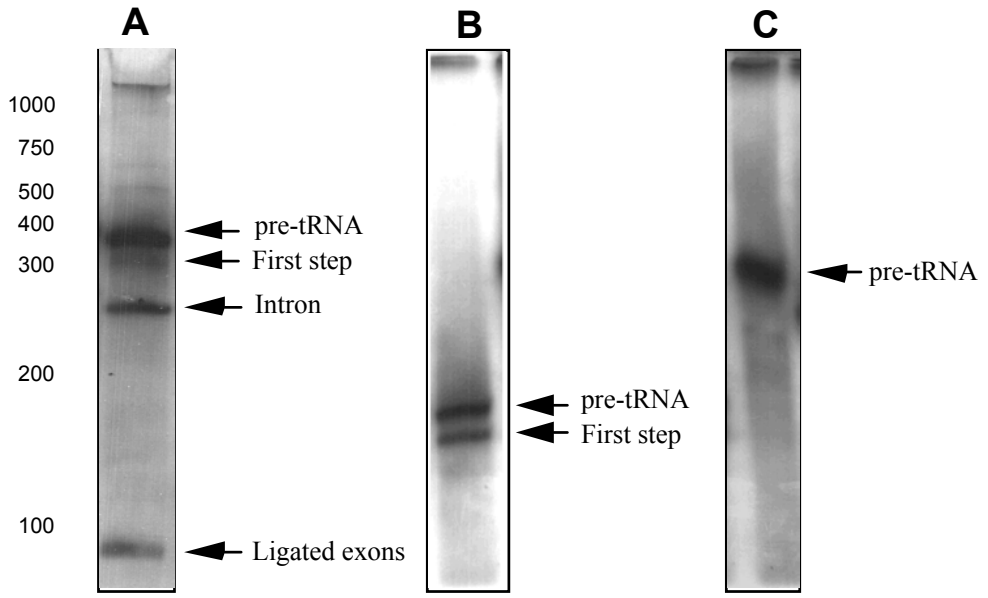

Figure 3.5.5 Examples of the *in vitro* splicing assays with the tRNA-Leu (UAA) intron from representatives of the Streptophyta, Chlorophyta, Heterokontophyta and cyanobacteria. The three lanes A, B and C show the products of splicing and patterns observed during the splicing assays listed in Table 3.5.2 (A) completion of the second of the two-step reaction leading to auto-excision (B) completion of the first step leading to auto-excision (C) no splicing activity. The RNA ladder is from BRL.

some green algae were able to complete the first of the two-step transesterification reaction which leads to auto-excision in group I introns (Cech 1990). None of the introns from land plants tested for splicing competence could complete either the first or second step of the two-step transesterification reaction. It is also possible that the *in vitro* conditions did not allow for proper folding of the intron RNA, thereby hindering self-splicing (Uhlenbeck 1995; Pan et al. 1997). In support of the results presented here, however, previous analyses of plastid tRNA-Leu introns showed that the intron of *Marchantia polymorpha* also cannot self-splice (Xu et al. 1990).

This pattern is consistent with the idea that splicing was lost after the endosymbiotic event which gave rise to plastids. However, given the ability of some of the earlier branching plastid introns to complete the first step we hypothesise that the tRNA-Leu (UAA) intron was capable of splicing at the beginning of endosymbiosis and gradually became more and more dependent on a protein mediated splicing pathway to the exclusion of the auto-catalytic pathway. In any case the cursory secondary structure comparison revealed that the terminal guanosine of the intron (omega G) is universally conserved in all tRNA-Leu (UAA) introns in higher plants suggesting that while the reaction catalytic center is provided by the intron in all plastids.

The self-splicing retention by extant cyanobacteria and the loss of splicing is consistent with the idea that splicing originated as a RNA-mediated reaction. The RNA structure creates the active site but is unable to function efficiently in vivo only when assisted by a protein complex. Consistent with this interpretation chloronucleoproteins have been demonstrated to associate with unspliced tRNA-Leu (UAA) transcripts in *Nicotiana tabacum* (Nakamura et al. 1999).

These cpRNPs associate with nascent pre-RNAs immediately after transcription in the nucleus and form RNA-protein complexes in the stroma. These cpRNP-RNA complexes confer stability and ribonuclease resistance to the RNAs. The complexes also act as a scaffold for the specific catalytic machinery involved in RNA splicing of the intron containing tRNAs (Nakamura et al. 2001). The biased distribution of the intron in chlorophytes and streptophytes suggests that the maturation process of tRNA-Leu (UAA) genes differs slightly in the two. If the maturation of the tRNA gene were to depend upon the intron being present in the tRNA-Leu (UAA) gene only in streptophytes this would explain the universal retention in this lineage which contrasts to the almost pervasive loss in all other chloroplasts lineages. This is highly speculative and experiments on the cpRNPs role in the splicing of the tRNA-Leu (UAA) intron currently underway may shed light upon this issue (M. Sugita pers. commun.).

These experiments have important general insights into the long-term evolution of catalytic RNAs in eukaryotic genomes and the loss of natural splicing abilities by group I introns. This work is important because it suggests that the ancient tRNA-Leu (UAA) intron was self-splicing at the time of endosymbiosis and gradually became dependent on the host for excision. We tentatively conclude that, as hypothesised, all chloroplast tRNA-Leu (UAA) introns receive enzymatic assistance *in vivo* during the excision process. Concomitant loss of splicing competence with endosymbiosis seems unlikely. Likewise this dependency is highly unlikely to predate the endosymbiotic event given the fact that all cyanobacteria examined are auto-catalytic. Drawing precise phylogenetic boundaries between self-splicing and host-mediated excision brings us one step nearer to identifying the splicing factor. Interestingly, this evolutionary pattern is observed multiple independent times when catalytic RNAs become stable genetic elements in eukaryotic cells. The ribosomes, for example, most certainly evolved from a RNA-based machinery (Cech 2000) that has over time recruited proteins to increase the efficiency of translation. Systematic analysis of the splicing evolution of the relatively simple tRNA-Leu (UAA) provides important general insights into the long-term evolution of catalytic RNAs in eukaryotic genomes.

4. Acknowledgements

I would like to thank my supervisor Prof. Dr. Thomas Friedl for his support, guidance and enthusiasm throughout work leading to this thesis. I am also grateful to Gert Helms, Dr. Maike Lorenz, Marlis Heinemann, Ilse Kunkle, Elke Zufall-Roth, Michaela Kitschke, Anke Behnke and Julia Müller for creating a warm and friendly working environment.

The vast bulk of the work presented in this thesis was carried out at the Albrecht-von-Haller Institute for Plant Science, Department of Botany, University of Göttingen during the years 1999-2001. Financial support for this project was provided by a DFG grant (905/7-1,2) to Prof. Dr. Thomas Friedl. A travel grant from the SmithKline and Beecham Foundation (205/2000 RKF) and a research grant from the Sigma Xi Grants-in-aid-of-Research program to myself permitted a research visit to the U.S. I would also like to thank the European Science Foundation, the EU Hochschulbüro and the Systematics Association for financial support to defray the costs of attending conferences and workshops.

I would like to thank Prof. Dr. Burkhard Büdel, Dr. Rosmarie Rippka and Dr. Michael Schwerdfeger who were kind enough to provide me with samples of cyanobacteria and land plants. I would also like to thank Marlis Heinemann, Ilse Kunkle and Elke Zufall-Roth for the endless supply of cyanobacteria and assistance with culturing. I am grateful to Prof. Dr. Debashish Bhattacharya and Dr. Rippka for invaluable discussion on aspects of the work presented here. I am grateful to Prof. Dr. Debashish Bhattacharya and Dawn Simon for assistance with the self-splicing assays and the systematic survey of chloroplast containing eukaryotes. I am also particularly indebted to Prof. Dr. Burkhard Büdel for providing me with strains of the genus *Chroococcidiopsis* from his personal collection and making available his help and expertise on aspects of my thesis work throughout my stay here in Germany.

Parts of this thesis rely heavily upon the use of unpublished genome sequence data of cyanobacteria and outgroup taxa for which I acknowledge and thank the Department of Energy Joint Genome Institute (www.doe.jgi.gov), The Institute for Genomic Research (www.tigr.org) and The Kazusa DNA Research Institute (www.kazusa.or.jp).

I also wish to thank my family and friends for their patience, support and inspiration. Most of all I want to thank my wife Jaanika and my daughters Lea and Orla for their understanding and support provided throughout my time here in Göttingen.

4.1 Erklärung

Hiemit erkläre ich, daß ich die Arbeit selbständig verfaßt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe. Ferner erkläre ich, daß ich nicht anderweitig mit oder ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

Göttingen, den 09.11.01

_____

4.2 Lebenslauf

Name:	David Peter Fewer
Geburtsdatum:	27. Febuar 1975
Geburtsort:	Waterford
Staatsangehörigkeit:	Irisch
Familienstand:	verheiratet

1979-1987	Primary school, Waterford, Irland
1987-1992	Secondary School, Waterford, Irland
1992	Leaving Certificate exam
1992-1996	Studium Biologie, Galway, Irland
1996	Bachelor of Science, Universität Galway, Irland.
1996-1998	Arbeiten für Masters Thesis
1998	Masters of Science, Universität Galway, Irland.
1999-2001	Arbeiten für Promotion

5. References

Adams K. L., Clements M. J. and Vaughn J. C. 1998. The *Peperomia* mitochondrial cox I group I intron: timing of horizontal transfer and subsequent evolution of the intron. *J. Mol. Evol.* 46: 689-696.

Agnostidis K. and Komárek J. 1985. Modern approaches to the classification system of cyanophytes. (Suppl. 71). *Algol. Stud.* 38: 291-302.

Andersen R. A. , Morton S. L. and Sexton J. P. 1997. CCMP - Provasoli-Guillard National Center for culture of marine phytoplankton. *J. Phycol.* 33: 1-75.

Asmussen C. B. and Chase M. W. 2001. Coding and noncoding plastid DNA in palm systematics. *Am. J. Bot*. 88: 1103-1117.

Barinaga M. 1990. Introns in new places – What does it all mean? *Science* 250: 1512.

Bayer R. J., Puttock C. F. and Kelchner S. A. 2000. Phylogeny of South African Gnaphalieae (Asteraceae) based on two noncoding chloroplast sequences. *Am J Bot.* 87: 259-272.

Bechhofer D. H., Hue K. K. and Shub D. A. 1994. An intron in the thymidylate synthase gene of *Bacillus* bacteriophage beta 22: evidence for independent evolution of a gene, its group I intron and the intron open reading frame. *Proc. Natl. Acad. Sci. USA* 91: 11669-11673.

Beckers C. J., Roos D. S., Donald R. G., Luft B. J., Schwab J. C., Cao Y. and Joiner K. A. 1995. Inhibition of cytoplasmic and organellar protein synthesis in *Toxoplasma gondii*. Implications for the target of macrolide antibiotics. *J. Clin. Invest.* 95:367-76.

Belfort M. 1991. Self-splicing introns in prokaryotes: migrant fossils? *Cell* 64: 9-11.

Belfort M. and Perlman P. S. 1995. Mechanisms of intron mobility. *J. Biol. Chem.* 270: 30237-30240 .

Belfort M. and Weiner A. 1997. Another bridge between Kingdoms: tRNA splicing in Archaea and Eukaryotes. *Cell* 89: 1003-1006.

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

Bergsland K. J. and Haselkorn R. 1991. Evolutionary relationships among eubacteria, cyanobacteria and chloroplasts: evidence from the *rpoC1* gene of *Anabaena* sp. strain PCC 7120. *J Bacteriol.* 173: 3446-3455.

Besendahl A., Qiu Y-L., Lee J., Palmer J. D. and Bhattacharya D. 2000. The cyanobacterial origin and vertical transmission of the plastid $tRNA^{Leu}$ group-I intron. *Curr. Genet.* 37: 12-23.

Bhattacharya D. and Medlin, L. 1995. The phylogeny of plastids: a review based on comparisons of small subunit ribosomal RNA coding regions. *J. Phycol*. 31: 489-498.

Bhattacharya D., Friedl T. and Damberger, S. 1996. Nuclear-encoded rDNA group I introns: origin and phylogenetic relationships of insertion site lineages in the green algae. *Mol. Biol. Evol.* 13: 978-989.

Bhattacharya D., Friedl T. and Schmidt, H. 1999. The phylogeny of thermophiles and hyperthermophiles and the three domains of life. In *Enigmatic microorganisms and life in extreme environments*. Editited by J. Seckbach. Kluwer Academic Publishers. Pages. 291-304.

Bhattacharya D., Surek B., Rusing M., Damberger S. and Melkonian M. 1994. Group I introns are inherited through common ancestry in the nuclear-encoded rRNA of Zygnematales (Charophyceae). *Proc. Natl. Acad. Sci. USA* 91: 9916-9992.

Billi D. and Grilli-Caiola M., 1996. Effect of nitrogen limitation and starvation on *Chroococcidiopsis* sp. (Chroococcales). *New Phytol.* 133: 563-571.

Billi D., Friedmann E. I., Helm R.F. and Potts M. 2001. Gene transfer to the desiccation-tolerant cyanobacterium *Chroococcidiopsis*. *J. Bacteriol*. 183: 2298-305.

Billi D., Friedmann E. I., Hofer K. G., Grilli-Caiola M. and Ocampo-Friedmann R. 2000. Ionizing-Radiation resistance in the desiccation-tolerant cyanobacterium *Chroococcidiopsis*. *Appl. Environ. Microbiol*. 66: 1489-1492.

Biniszkiewicz D., Cesnaviciene E. and D. A. Shub. 1994. Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. *EMBO J.* 13: 4629-4635.

Bonen L. and Doolittle W. F. 1976. Partial sequences of 16S rRNA and the phylogeny of blue-green algae and chloroplasts. *Nature* 261: 669-673.

Bonocora R. P. and Shub D. A. 2001. A novel group I intron-encoded endonuclease specific for the anticodon region of $tRNA^{fMet}$ genes. *Mol. Microbiol.* 39: 1299-1306.

Boone D. R. and Castenholz R. W. 2001. Bergey's Manual of Systematic Bacteriology, second edition, vol. 1 (The Archaea and the deeply branching and phototrophic Bacteria), Springer-Verlag, New York, pp. 473-599.

Brocks J. J., Logan G. A., Buick R. and Summons R. 1999. Archaen molecular fossils and the early rise of the eukaryotes. *Science* 285: 1033-1036.

Brown J. R., Douady C. J., Italia M. J., Marshall W. E. and Stanhope M. J. 2001. Universal trees based on large combined protein sequence data sets. *Nat. Genet.* 28: 281-285.

Buck W. R., Goffinet B. and Shaw A. J. 2000. Testing morphological concepts of orders of pleurocarpous mosses (Bryophyta) using phylogenetic reconstructions based on *trnL-trnF* and *rps4* sequences. *Mol. Phylogenet. Evol.* 16: 180-198.

Büdel B. 1999. Ecology and diversity of rock inhabiting cyanobacteria in tropical regions. *Eur. J. Phycol*. 34: 361-370.

Büdel B. and Henssen A. 1983. *Chroococcidiopsis* (Cyanophyceae), a phycobiont in the lichen family Lichinaceae. *Phycologia* 22: 367-375.

Büdel B. and Wessels D. C. J. 1991. Rock inhabiting blue-green algae/cyanobacteria from hot arid regions. *Algological Studies* 64: 385-398.

Büdel B., Becker U., Follmann G. and Sterflinger K. 2000. Algae, fungi and lichens on Inselbergs. In *Inselbergs.* Edited by S. Porembski and W. Barthlott. Springer-Verlag, Heidelberg. Pages 69-90.

Bult C. J., White O., Olsen G. J., Zhou L., Fleischmann R. D., Sutton G. G., Blake J. A., FitzGerald L. M., Clayton R. A., Gocayne J. D. et al. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273: 1058-1073.

Burger-Wiersma T., Stal L. J. and Mur L. R. 1989. *Prochlorothrix hollandica* gen. nov. sp. nov. a filamentous oxygenic photoautotrophic prokaryote containing chlorophylls a and b: assignment to *Protrichaceae* fam. nov. and order Prochorales Florenzano, Balloni and Materassi 1986, with emendation of the ordinal description. *Int. J. Syst. Bacteriol*. 39: 250-257.

Burke, J. M. 1988. Molecular genetics of group I introns: RNA structure and protein factors required for splicing. *Gene* 73: 273-294.

Castenholz R. W. 1992. Species usage, concept and evolutionin the cyanobacteria (blue-green algae). *J. Phycol.* 28: 737-745.

Castenholz R. W. and Waterbury J. B. 1989. Group I cyanobacteria. In *Bergey's Manual of Systematic Bacteriology* Edited by J. T. Stanley, M. P. Byrant, N. Pfennig and J. G. Holt. Volume 3. Williams and Wilkins Co. Baltimore. Pages 1710-1728.

Caudales R., Wells J. M. and Butterfield J. E. 2000. Cellular fatty acid composition of cyanobacteria assigned to subsection II order Pleurocapsales. *Int. J. Syst. Bacteriol*. 50: 1029-1034.

Cavalier-Smith T. 1991. Intron phylogeny: a new hypothesis. *Trends Genet.* 7: 145-148.

Cech T. R. 1990. Self-splicing of group I introns. *Annu. Rev. Biochem.*59: 543-568.

Cech T. R. 2000. The ribosome is a ribozyme. *Science* 289: 878-879.

Cech T. R., Damberger S. H. and Gutell R. R. 1994. Representation of the secondary structure of group I introns. *Nat. Struct. Biol.* 1: 273-280.

Cho Y. and Palmer J. D. 1999. Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial cox1 gene during evolution of the Araceae family. *Mol. Biol. Evol*. 16: 1155-1165.

Cho Y., Qiu Y-L., Kuhlman P. and Palmer J. D. 1998. Explosive invasion of plant mitochondria by a group I intron. *Proc. Natl. Acad. Sci. USA* 95: 14244-14249.

Clouser C. R., Germano J. and Klein A.S. 1999. Sequence variation of three noncoding chloroplast regions between *Picea* species. Unpublished. Genbank: AF156801.

Cunningham C. W. 1997. Can three incongruence tests predict when data should be combined? *Mol. Biol. Evol*. 14: 733-40.

Dalgaard J. Z., Garrett R. A. and Belfort M. 1993. A site-specific endonuclease encoded by a typical archaeal intron. *Proc. Natl. Acad. Sci. USA* 12: 5414-5417.

Darnell J. E. and Doolittle W. F. 1986. Speculations on the early course of evolution. *Proc. Natl. Acad. Sci. USA* 83: 1271-1275.

Daros J. A. and Flores R. 1996. A group I plant intron accumulates as circular RNA forms with extensive 5' deletions in vivo. *RNA*. 2: 928-936.

Daugbjerg N. and Andersen R. A. 1997. Phylogenetic analyses of the *rbcL* sequences from haptophytes and heterokont algae suggest their chloroplasts are unrelated. *Mol. Biol. Evol.* 14:1242-1251.

Delahodde A., Goguel V., Becam A. M., Creusot F., Perea J., Banroques J. and Jacq C. 1989. Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. *Cell* 56: 431-441.

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

Delwiche, C. F. and Palmer J. D. 1997. The origin of plastids and their spread via secondary symbiosis. *Plant Syst. Evol.* (Suppl). 11: 53-86.

Diener TO. 1989. Circular RNAs: relics of precellular evolution? *Proc. Natl. Acad. Sci. USA* 86: 9370-9374.

Doolittle R. F., Feng D. .F, Tsang S., Cho G. and Little E. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* 271: 470-477.

Doolittle W. F. 1997. Fun with genealogy. *Proc. Natl. Acad. Sci. USA* 94: 12751-12753.

Doolittle W. F. 2000. The nature of the universal ancestor and the evolution of the proteome. *Curr. Opin. Struct. Biol.* 10: 355-358.

Dor I., Carl N. and Baldinger I. 1991. Polymorphism and salinity tolerance as a criterion for differentiation of three new species of *Chroococcidiopsis* (Chroococcales*). Algol. Studies* 64: 411-421.

Douglas S. E. 1998. Plastid evolution: origins, diversity, trends. *Curr. Opin. Genet. Dev*. 8: 655-661.

Douglas S. E. and Penny S. L. 1999. The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. *J. Mol. Evol.* 48: 236-244.

Douglas S. E., Murphy C. A., Spencer D. F. and Gray M. W. 1991. Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* 350: 148-151.

Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L. T., Wu X., Reith M., Cavalier-Smith T. and Maier U. G. 2001. The highly reduced genome of an enslaved algal nucleus. *Nature* 410: 1091-1096.

Duff R. J. and Nickrent D. L. 1999. Characterization of mitochondrial small-subunit ribosomal RNAs from holoparasitic plants. J. Mol. Evol. 45: 631-639.

Dujon B. 1980. Nucleotide Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the omega and rib-1 loci. *Cell* 20: 185-97.

Dujon B. 1989. Group I introns as mobile genetic elements: facts and mechanistic speculations *Gene*. 82 91-114.

Ecarot-Charrier B. and Cedergren R. J. 1976. The preliminary sequence of Met-tRNA-f from Anacystis nidulans compared with other initiator tRNAs. *FEBS Lett*. 63: 287-290.

Eisen J. A. 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res*. 8: 163-167.

Elenkin A. A. 1949. Sinezelenye vodorosli SSSR (Monographia algarum cyanophycearum aquidulcium et terrestrium in finibus URSS inventarum). Sytematic part 1. Moscow, Lennigrad. Akad. Nauk. SSSR.

Evrard J.L., Kuntz M., Straus N. A. and Weil,J. H. 1988. A class-I intron in a cyanelle tRNA gene from *Cyanophora paradoxa*: phylogenetic relationship between cyanelles and plant chloroplasts *Gene* 71: 115-122.

Fangan B. M., Stedje B., Stabbetorp O. E., Jensen E. S. and Jakobsen K. S. 1994. A general approach for PCR-amplification and sequencing of chloroplast DNA from crude vascular plant and algal tissue *BioTechniques* 16: 484-494.

Farris J. S., Källersjö M., Kluge A. G. and Bult C. 1994. Testing significance of incongruence. *Cladistics* 10: 315-319.

Feng D. F., Cho G. and Doolittle R. F. 1997. Determining divergence times with a protein clock: update and reevaluation. *Proc. Natl. Acad. Sci. USA* 94: 13028-13033.

Ferris C., King R. A. and Gray A. J. Molecular evidence for the hybrid origin of *Spartina anglica* C.E. Hubbard based on DNA sequencing of chloroplast DNA. *Mol. Ecol*. 6: 185-187.

Ferris M. and Palenik B. 1998. Niche adaptation in ocean cyanobacteria. *Nature* 396: 226-228.

Fewer D. P., Friedl T. and Büdel B. 2002. *Chroococcidiopsis* and hetercyst forming cyanobacteria are each others closest living relatives. *Mol. Phylogenet. Evol.* (in press).

Foster C. B., Reed J. D. and Wicander R. 1989. *Gloeocapsamorpha prisca* Zalessky, 1917: a new study part I: taxonomy, geochemistry and paleoecology. *Geobios* 22: 735-759.

Fox G. E., Pechman D. R. and Woese C. R. 1977. Comparative catologing of the 16S ribosomal ribonucleic acid: a molecular approach to procaryotic systematics. *Int. J. Syst. bacteriol*. 27: 44-57.

Freeman E. 1997. Intron of chloroplast trnL (UAA) gene. Unpublished. Genbank: AF034886

Freshwater D. W., Fredericq S., Butler B. S., Hommersand M. H. and Chase M. W. 1994. A gene phylogeny of the red algae (Rhodophyta) based on plastid *rbcL*. *Proc. Natl. Acad. Sci. USA* 91: 7281-7285.

Friedl T. 1997. Evolution of green algae. In Origins of algae and their plastids (Bhattacharya D. Ed.) Spriner-Verlag, Wein. Pages 87-102.

Friedl T. and Schlösser U. G. 2000. Additions to the culture collection of algae at Göttingen since 1997. *Nova Hedwigia*. 71: 243-262.

Friedmann E. I. 1962. The ecology of the atomphytic nitrate alga *Chroococcidiopsis kashaii* Friedmann. Studies on cave algae from Israel. IV. *Archiv für Mikrobiol*. 42: 42-45.

Friedmann E. I. and Ocampo-Friedmann R. 1976. Endolithic blue-green algae in the dry valleys: primary producers in the Antarctic desert ecosystem. *Science* 193: 1247-1249.

Friedmann E. I. and Ocampo-Friedmann R. 1984. Endolithic microorganisms in extreme dry environments: Analysis of a lithobiontic microbial habitat. *In Current perspectives in Microbial Ecology*. Edited by M. J. Klug and C. A. Reddey. American Society for Microbiology, Washington, D.C. Pages. 177-185.

Friedmann E. I. and Ocampo-Friedmann R. 1995. A Primitive Cyanobacterium as Pioneer Microorganism for Terraforming Mars. *Adv. Space Res*. 15: 243-246.

Friedmann E. I., Ocampo-Friedmann R. and Hua M. 1994. *Chroococcidiopsis*, the most primitive living cyanobacterium? *Origins Life Evol. Biosphere*. 24: 269-270.

Fritsch F. E. 1959. The structure and reproduction of algae. 1. Cambridge: Cambridge University Press.

Fukuda T., Yokoyama J. and Ohashi H. 2001. Phylogeny and biogeography of the genus Lycium (Solanaceae): inferences from chloroplast DNA sequences. *Mol. Phylogenet. Evol.* 19: 46-28.

Galtier N., Tourasse N. and Gouy M. 1999. A nonhyperthermophilic common ancestor to extant lifeforms. *Science* 283: 220-221.

Garcia-Pichel F., Nubel U. and Muyzer G. 1998. The phylogeny of unicellular, extremely halotolerant cyanobacteria. *Arch. Microbiol*. 169: 469-482.

Gargas A., DePriest P. T. and Taylor J. W. 1995. Positions of multiple insertions in SSU rDNA of lichen-forming fungi. *Mol. Biol. Evol.* 12: 208-218.

Gatesy J., O'Grady P. and Baker R. H. 1999. Corroboration among data sets in simultaneous analysis: Hidden support for phylogenetic relationships among higher level Artiodactyl taxa. *Cladistics* 15: 271-313.

Geitler L. 1932. Cyanophyceae Rabenhorsts *Krytogammenflora von Deutschland, Österreich und der Schweiz.* 14. Leipzig: Akademische Verlagsgesellschaft.

Geitler L. 1979. Einige kritische Bemerkungen zu neuen zusammenfassenden Darstekkungen der Morphoogie und Systematik der Cyanophyceen. *Plant Syst. Evol.* 132: 153-160.

Geitler, L. 1933. Diagnosen neuer Blaualgen von den Sunda-Inseln.Arch. *Hydrobiol. Suppl*. 8: 622.

Gielly L. and Taberlet P. 1994. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. *Mol. Biol. Evol.* 11: 769-777.

Gielly L. and Taberlet P.1995. Chloroplast DNA polymorphism at the intrageneric level: incidences in resolving plant phylogenies. *C. R. Acad. Sci. III, Sci. Vie* 317: 685-692.

Gilbert W. 1986. The RNA world. *Nature* 319: 618.

Gimble F. S. 2000. Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiology Letters* 185: 99-107.

Giovannoni S. J., Turner S., Olsen G. J., Barns S., Lane D. J. and Pace, N. R. 1988. Evolutionary relationships among cyanobacteria and green chloroplasts. *J. Bacteriol.* 170: 3584-3592.

Glockner G., Rosenthal A. and Valentin K. 2000. The structure and gene repertoire of an ancient red algal plastid genome. *J. Mol. Evol.* 51: 382-390.

Gockel G. and Hachtel W. 2000. Complete gene map of the plastid genome of the nonphotosynthetic euglenoid flagellate *Astasia longa*. *Protist*. 151: 347-351.

Goddard M. R. and Burt A. 1999. Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci. USA* 96: 13880-13885.

Gogarten J. P., Olendzenski L., Hilario E., Simon C. and Holsinger K. E. 1996. Dating the cenancester of organisms. *Science* 274: 1750-1751.

Golubic S. 1976. Taxomomy of extant stromatolite building cyanophytes. In *Stromatolites.* Walter M. R. (Ed.). New York. Elsevier Developments in Sedimentology 20: 127-140.

Golubic S. 1979. Cyanobacteria (blue green algae) under the bacteriological code? An ecological objection. *Taxon* 28: 387-389.

Golubic S. and Hofmann K. 1976. Comparasion of Holocene and mid Precambrian Entophysalidaceae (Cyanophyta) in stromatolite mats: cell division and degradation. *J. Palentol*. 50: 1074-1082.

Golubic S., Sergeev V. N. and Knoll A. H. 1995. Mesoproterozoic *Archaeoellipsoides*: akinetes of heterocystous cyanobacteria. *Lethaia* 28: 285-198.

Gray M. W. 1988. Organelle origins and ribosomal RNA. *Biochem. Cell Biol.* 66: 325-348.

Gray M. W. and Doolittle W. F. 1982. Has the endosymbiont hypothesis been proven? Microbiol Rev. 46: 1-42.

Gray M. W. and Spencer D. F. 1996. In *Evolution of Microbial Life* (Roberts D. M., Sharp P., Alderson G. and Collins M., Eds.). Cambridge University Press, Cambridge, pp. 107-126.

Gray M. W., Burger G. and Lang F. 1999. Mitochondrial Evolution. *Science* 283:1476-81.

Grilli-Caiola M., Ocampo-Friedmann R. and Freidmann E. I. 1993. Cytology of long-term dessication in the desert cyanobacterium Chroococcidiopsis (Chroococcales). *Phycologia* 32: 315-322.

Gupta R. S., Bustard K., Falah M. and Singh D. 1997. Sequencing of heat shock protein 70 (DnaK) homologs from *Deinococcus proteolyticus* and *Thermomicrobium roseum* and their integration in a protein-based phylogeny of prokaryotes. J. Bacteriol. 179: 345-357.

Gutell R. R. 1993. Collection of small subunit ribosomal (16S and SSU-like) ribosomal RNA structures. *Nuc. Acids Res*. 21: 3051-3054.

Hall T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids. Symp. Ser*. 41: 95-98.

Hallick R. B., Hong L., Drager R. G., Favreau M. R., Monfort A., Orsat B., Spielmann A. and Stutz E. 1993. Complete sequence of Euglena gracilis chloroplast DNA. *Nucleic Acids Res.* 21: 3537-3544.

Hardy C. M. and Clark-Walker G. D. 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.

Hasegawa M. and Fitch W. M. 1996. Dating the cenancester of organisms. *Science* 274: 1750.

Healy F. G., Bukhalid R. A. and Loria R. 1999. Characterization of an insertion sequence element associated with genetically diverse plant pathogenic *Streptomyces* spp. *J. Bacteriol.* 181: 1562-1568.

Helmchen T. A., Bhattacharya D. and Melkonian M. 1995. Analyses of ribosomal RNA sequences from glaucocystophyte cyanelles provide new insights into the evolutionary relationships of plastids. *J. Mol. Evol.* 41: 203-210.

Hensgens L. A., Bonen L., de Haan M., van der Horst G., and Grivell L. A. 1983. Two intron sequences in yeast mitochondrial COX1 gene: homology among URF-containing introns and strain-dependent variation in flanking exons. *Cell* 32: 379-389.

Herdman M., Janvier M., Rippka R. and Stanier R. Y. 1979. Genome size of cyanobacteria. *J. Gen. Microbiol.* 111: 73-85.

Hibbett D. S. 1996. Phylogenetic evidence for horizontal transmission of group I introns in the nuclear ribosomal DNA of mushroom-forming fungi. *Mol. Biol. Evol.* 13: 903-917.

Holland H. D. and Beukes N. J. 1990. A paleoweathering profile from Griqualand West, South Africa: evidence for a dramatic rise in atmospheric oxygen between 2.2 and 1.9 bybp. *Am. J. Sci.* 290-A:1-34.

Holst-Jensen A., Vaage M., Schumacher T. and Johansen S. 1999. Structural characteristics and possible horizontal transfer of group I introns between closely related plant pathogenic fungi. *Mol. Biol. Evol.* 16: 114-126.

Honda D., Yokota A. and Sugiyama J. 1999. Detection of seven major evolutionary lineages in cyanobacteria based on 16S rRNA gene sequence analyses with new sequences of five marine *Synechococcus* strains. *J. Mol. Evol*. 48: 723-739.

Ishida T., Watanabe M. M., Sugiyama J. and Yokota A. 2001. Evidence for polyphyletic origin of the members of the orders of Oscillatoriales and Pleurocapsales as determined by 16S rDNA analysis. *FEMS Microbiol. Lett.* 201: 79-82.

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

Kajita T., Kamiya K., Nakamura K., Tachida H., Wickneswari R., Tsumura Y., Yoshimaru H., Yamazaki T. 1998. Molecular phylogeny of Dipetrocarpaceae in Southeast Asia based on nucleotide sequences of *matK*, *trnL* intron, and *trnL-trnF* intergenic spacer region in chloroplast DNA. *Mol. Phylogenet. Evol.* 10: 202-209.

Keeling P. J. and Palmer J. D. 2001. Lateral transfer at the gene and subgenic levels in the evolution of eukaryotic enolase. *Proc. Natl. Acad. Sci. USA* 98: 10745-10750.

Kenrick P. and Crane P. R. 1997. The origin and early evolution of plants on land. *Nature* 389: 33-39.

Kenyon C. N. 1972. Fatty acid composition of unicellular strains of blue green algae. *J. Bacteriol.* 109: 827-834.

King J. L. and Hanner R. 1998. Cryptic species in a "living fossil" lineage: taxonomic and phylogenetic relationships within the genus Lepidurus (Crustacea: Notostraca) in North America. *Mol. Phylogenet. Evol*. 10: 23-36.

Kishino H. and Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data and the branching order in *Hominoidea*. *J. Mol. Evol.* 29: 170-179.

Knauf U. 2001. A 22 kb fragment of the 53 kb plastid genome of the colourless alga *Prototheka wickerhamii* containing atp-, rpl-,rps-, rrn-, and trn-genes. Unpublished. Genbank AJ245645.

Knoll A. H. and Holland H. D. 1995. In *Effects of Past Global Change on Life*. (Ed. Stanley S. M.). National Academy Press. Washington DC. 21-33.

Knoll A. H., Golubic S., Green J. andSwett K. 1986. Organically preserved microbial endoliths from the late Proterozoic of East Greenland. *Nature* 321: 856-857

Kohler S., Delwiche C. F., Denny P. W., Tilney L G., Webster P., Wilson R. J., Palmer J. D. and Roos D. S. 1997. A plastid of probable green algal origin in Apicomplexan parasites. *Science* 275: 1485-1489.

Komárek J. and Anagnostidis K. 1999. Cyanoprokaryota, 1. Teil Chroococcales. In *Süßwasserflora von Mitteleuropa 19/1*. Edited by H. Ettl, G. Gärtner, H. Heynig, and D. Mollenhauer. Gustav Fischer Verlag, Jena Sturrgart Lübnek Ulm. Pages 1-548.

Kowallik K.V., Stoebe B., Schaffran I., Kroth-Pancic P. and Freier U. 1995. The Chloroplast Genome of a chlorophyll a+c- containing Alga, *Odontella sinensis*. *Plant Mol. Biol. Rep*. 13: 336-342.

Kruger K., Grabowski P. J. , Zaug A. J., Sands J., Gottschling D. E. and Cech T. R. 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* 31: 147-57.

Kuhsel M. G., Strickland R. and Palmer J. D. 1990. An ancient group-I intron shared by eubacteria and chloroplasts. *Science* 250: 1570-1573.

Kunkel D. D. 1984. Cell division in baeocyte producing cyanobacteria. *Protoplasma* 123: 104-115.

Kusumi J., Tsumura Y., Yoshimaru H. and Tachida H. 2000. Phylogenetic relationships in Taxodiaceae and Cupressaceae *sensu stricto* based on *matK* gene, *chlL* gene, *trnL-trnF* IGS region, and *trnL* intron sequences. *Am. J. Bot.* 87: 1480-1488.

Kyrpides N. C. and Olsen G. J. 1999. Archaeal and bacterial hyperthermophiles: horizontal gene exchange or common ancestry? *Trends Genet.*15: 298-299.

La Farge C., Mishler B. D., Wheeler J. A., Wall D. P., Johannes K., Schaffer S., and Shaw A. J. 2000. Phylogenetic relationships within the Haplolepideous mosses. *The Bryologist* 103: 257-276.

Lambert D. H., Bryant D. A., Stirewalt V. L., Dubbs J. M., Stevens S. E. Jr., and Porter R. D. 1985. Gene map for the *Cyanophora paradoxa* cyanelle genome. *J. Bacteriol.* 164: 659-664.

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

Lang B. F. 1984. The mitochondrial genome of the fission yeast *Schizosaccharomyces pombe*: highly homologous introns are inserted at the same position of the otherwise less conserved cox1 genes in *Schizosaccharomyces pombe* and *Aspergillus nidulans*. *EMBO J.* 3: 2129-2136.

Lemieux C, Otis C. and Turmel M. 2000. Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature* 403: 649-652.

Lewin B. 1997. Genes IV. Oxford University Press. Oxford, New York. Pages 1-1260.

Lewin R. A 1979. Formal taxonomic treatment of cyanophytes. *Int. J. Sytem. Bacteriol.* 29: 411-412.

Lewin R. A 1989. Group II order Prochlorales In *Bergeys Manual of Systematic Bacteriology.* Edited by Stanley J. T. Byrant M. P. Pfennig N. Holt J. G. Volume 3. Baltimore: Williams and Wilkins. Pages. 1805-1806.

Lewis L. A., Mishler B. D. and Vilgalys R. 1997. Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene *rbcL*. *Mol. Phylogenet. Evol*. 3: 377-393.

Li W-H. and Graur D. 1991. Fundamentals of Molecular Evolution. Sinauer, Sunderland.
Liu X. Q. 1991. An ancient intron in eubacteria: new light on intron origins. *Bioessays*. 13: 185-186.
Lockhart P. J, Steel M. A., Hendy M. D. and Penny D. 1994 Recovering Evolutionary Trees under a More Realistic Model of Sequence. *Mol. Biol. Evol.* 11: 605-612.
Lockhart P. J., Howe C., Bryant D. A., Beanland T. J. and Larkum A.W. D. 1992. Substitutional bias confounds inference of cyanelle origins from sequence data. *J. Mol. Evol*. 34: 153-162.
Logsdon J. M. and Faguy D. M. 1999. *Thermotoga* heats up lateral gene transfer. *Curr. Biol.* 9: 747-751.
Maier R. M., Neckermann K., Igloi G. L. and Kossel H. 1995. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol*. 251: 614-628.
Manhart J. R. 1994. Phylogenetic analysis of green plant *rbc*L sequences. *Mol. Phylogenet. Evol*. 114-127.
Margulis L. 1970. Origin of eukaryotic cells. Yale University press. New Haven. pages 1-394.
Martin W., Stoebe B., Goremykin V., Hapsmann S., Hasegawa M. and Kowallik K. V. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393: 162-165.
McCourt R. M. 1995. Green algal phylogeny. *Trends in Ecology and Evolution* 10: 159-163.
McFadden G. I. 1999. Plastids and protein targeting. *J. Eukaryot. Microbiol*. 46: 339-346.
Medlin L. K., Cooper A., Hill C., Wrieden S. and Wellbrock U. 1995. Phylogenetic position of the Chromista plastids based on small subunit rRNA coding regions. *Curr Genet.* 28: 560-565.
Meißner K., Frahm J.-P., Stech M. and Frey W. 1998. Molecular divergence patterns and infrageneric relationship of Monoclea (Monocleales, Hepaticae). Studies in austral temperate rain forest bryophytes 1. *Nova Hed*. 67: 289-302.
Mereschowsky C 1905. Über Natur undUrsprung der Chromatophoren in Pflanzenreide. *Biol.Centr.* 25 593-604.
Michel F. and Westhof E. 1990. Modelling the three-dimensional arcitecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol*. 216: 585-610.
Miller K. A., Olsen J. L. and Stam W. T. 2000. Genetic divergence correlates with morphological and ecological subdivision in the deep-water elk kelp, *Pelagophycus porra* (Phaeophyceae). *J. Phycol.* 36: 862-870.
Miyamoto M. M. and Fitch W. M. 1996. Constraints on protein evolution and the age of the eubacteria/eukaryote split. *Syst Biol.* 45: 568-575.
Mohr G. and Lambowitz A. M. 1991. Integration of a group I intron into a ribosomal RNA sequences promoted by a tyrosyl-tRNA synthetase. *Nature* 354: 164-167.
Moore M. J., Query C. C. and Sharp P. A. 1993. Splicing of precursors to mRNA by the spliceosome. In *The RNA world*. Edited by Gesteland R. F. and Atkins J. F. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. Pages 303-357.
Morden C. W., Delwiche C. F., Kuhsel M. and Palmer J. D. 1992. Gene phylogenies and the endosymbiotic origin of plastids. *Biosystems*. 28: 75-90.
Morell V. 1996. New mammals discovered by biology's new explorers. *Science* 273: 1491.
Mota E. M. and Collins R. A. 1988. Independent evolution of structural and coding regions in a *Neurospora* mitochondrial intron. *Nature* 332: 654-6.
Muscarella D. E. and Vogt V. M. 1989. A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*. *Cell* 56: 443-454.
Nakamura T, Ohta M, Sugiura M. and Sugita M. 1999. Chloroplast ribonucleoproteins are associated with both mRNAs and intron-containing precursor tRNAs. *FEBS Lett*. 460: 437-441.
Nakamura T., Ohta M., Masahiro S. and Sugita M. 2001. Chloroplast ribonucleoproteins function as a stabilizing factor of ribosome-free mRNAs in the stroma. *J. Biol. Chem*. 267: 147-152.

Nelissen B., de Peer Y. V., Wilmotte A. and de Wachter R. 1995. An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences. *Mol. Biol. Evol* 12: 1166-1173.

Nelson K. E., Clayton R. A., Gill S. R., Gwinn M. L., Dodson R. J., Haft D. H. et al. 1999. Evidence for lateralgene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399: 323-329.

Newton A. E., Cox C. J., Duckett J. G., Wheeler J. A., Goffinet B., Hedderson T. A. J., and Mishler B. D. 2000. Evolution of the major moss lineages: phylogenetic analyses based on multiple gene sequences and morphology. *The Bryologist* 103: 187-211.

Nickrent D. L., Parkinson C. L., Palmer J. D. and Duff R. J. 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Mol. Biol. Evol*. 17: 1885-9185.

Nikoh N. and Fukatsu T. 2001. Evolutionary dynamics of multiple group I introns in nuclear ribosomal RNA genes of endoparasitic fungi of the genus *Cordyceps*. *Mol. Biol. Evol*. 8: 1631-1642

Nishida H., Blanz P. A. and Sugiyama J. 1993. The higher fungus *Protomyces inouyei* has two group I introns in the 18S rRNA gene. *J. Mol. Evol.* 37: 25-28.

Nishida H., Tajiri Y. and Sugiyama J. 1998. Multiple origins of fungal group I introns located in the same position of nuclear SSU rRNA gene. *J. Mol. Evol.* 46: 442-448.

Nubel U., Bateson M. M., Madigan M. T., Kuhl M. and Ward D. M. 2001. Diversity and Distribution in Hypersaline Microbial Mats of Bacteria Related to *Chloroflexus* spp. *Appl. Environ. Microbiol*. 9: 4365-4371.

Ohman-Heden M A., Ahgren-Stalhandske S., Hahne R. and Sjoberg B. M. 1993. Translation across the 5'splice site interfers with autocatalytic splicing. *Mol. Microbiol.* 7: 975-982.

Ohta N. 1997. Analysis of a plastid gene cluster reveals a close relationship between *Cyanidioschyzon* and *Cyanidium*. *J. Plant Res.* 110: 235-245.

Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S., Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H. and Ozeki H. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322: 572-574.

Oliveira M. C. and Bhattacharya D. 2000. Phylogeny of the Bangiophycidae (Rhodophyta) and the secondary endosymbiotic origin of algal plastids. *Am. J. Bot.* 87: 482-492.

Orgel L. E.1986. RNA catalysis and the origins of life. *J. Theor. Biol.* 123: 127-49.

Oyaizu H, Debrunner-Vossbrinck B, Mandelco L, Studier JA, Woese CR. 1987. green non-sulfur bacteria: a deep branching in the eubacterial line of descent. *Syst. Appl. Microbiol.* 9: 47-53.

Pace N. R. and Marsh T. K. 1985. RNA catalysis and the origin of life. *Origins of Life* 16: 97-116.

Paerl H. W. 1996. A comparison of cyanobacterial bloom dynamics in freshwater, estuarine and marine environments. *Phycologia* 35 (Suppl. 6): 25-35.

Page R. D. M. 1994. Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Syst. Biol*. 43: 58-77.

Page R. D. M. and Holmes E. C. 1998. Molecular Evolution: A Phylogenetic Approach. Blackwell Science, Oxford. Pages. 1-346.

Palenik B. and Haselkorn R. 1992. Multiple evolutionary origins of prochlorophytes, the chlorophyll b-containing prokaryotes. *Nature* 355: 265-267.

Palmer J. D. and Delwiche C. F. 1996. Second-hand chloroplasts and the case of the disappearing nucleus. *Proc. Natl. Acad. Sci. USA* 93: 7432-7435.

Palmer J. D. and Logsdon J. M. 1991. The recent origin of introns. *Curr. Opin. Genet. Dev*. 1: 470-477.

Pan J., Thirumalai D. and Woodson S. A. 1997. Folding of RNA involves parallel pathways. *J. Mol. Biol*. 273: 7-13.

Paquin B., Heinfling A. and Shub D. A. 1999. Sporadic distribution of tRNA(Arg)CCU introns among alpha-purple bacteria: evidence for horizontal transmission and transposition of a group I intron. *J. Bacteriol.*181: 1049-1053.

Paquin B., Kathe S. D., Nierzwicki-Bauer S. A. and Shub D. A. 1997. Origin and evolution of group I introns in cyanobacterial tRNA genes. *J. Bacteriol.*179: 6798-6806.

Paquin B., Laforest M. J. and Lang B. F. 1994. Interspecific transfer of mitochondrial genes in fungi and creation of a homologous hybrid gene. *Proc. Natl. Acad. Sci. USA* 25: 11807-11810.

Paracer S. and Ahmadjian V. 2000. Symbiosis: An introduction to Biological Associations. 2nd Edition. Oxford University Press. Pages 1-291.

Paulsrud P. and Lindblad P. 1998. Sequence variation of the tRNA(Leu) intron as a marker for genetic diversity and specificity of symbiotic cyanobacteria in some lichens. *Appl Environ Microbiol.* 64: 310-5.

Pel H. J. and Grivell L. A. 1993. The biology of yeast mitochondrial introns. *Mol Biol Rep*. 1: 1-13.

Perotto S., Nepote-Fus P., Saletta L., Bandi C. and Young J. P. 2000. A diverse population of introns in the nuclear ribosomal genes of ericoid mycorrhizal fungi includes elements with sequence similarity to endonuclease-coding genes. *Mol. Biol. Evol.* 17: 44-59.

Poe S. and Swofford D.L. 1999. Taxon sampling revisited. *Nature* 398: 299-300.

Posada D. and Crandall K. A. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

Potter D., Saunders G. W. and Andersen R. A. 1997. Phylogenetic relationships of the Raphidophyceae and Xanthophyceae as inferred from nucleotide sequences of the 18S ribosomal RNA gene. *Am. J. Bot*. 84: 966-972.

Prager E. M. and Wilson A. C.1988. Ancient origin of lactalbumin from lysozyme: analysis of DNA and amino acid sequences. *J. Mol. Evol.* 27: 326-35.

Priscu J. C., Fritsen C. H., Adams E. E., Giovannoni S. J., Paerl H. W., McKay C. P., Doran P. T., Gordon D. A., Lanoil B. D. and Pinckney J. L. 1998. Perennial Antarctic lake ice: an oasis for life in a polar desert. *Science* 280: 2095-2098.

Pronk L. M. and Sanderson K. E. 2001. Intervening sequences in rrl genes and fragmentation of 23S rRNA in genera of the family Enterobacteriaceae. *J. Bacteriol*. 183: 5782-5787

Pryer K. M., Schneider H., Smith A. R., Cranfill R., Wolf P. G., Hunt J. S. and Sipes S. D. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* 409: 618-622.

Quirk S. M., Bell-Pederson D. and Belfort M. 1989. Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames. *Cell* 56: 455-465.

RajBhandary T. L. 1994. Initiator transfer RNAs. *J. Bacteriol.*176: 547-552.

Raven P. H. 1970. A multiple origin for plastids and mitochondria. *Science* 169: 641-646.

Reeves R. H. 1996. 16S ribosomal RNA and the molecular phylogeny of the Cyanobacteria *Nova Hedwigia* 112: 55-67.

Reinhold-Hurek B. and Shub DA. 1992. Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* 357: 173-176.

Reith M.E. and Munholland J. 1995. Complete nucleotide sequence of the Porphyra purpurea chloroplast genome. *Plant Mol. Biol. Rep.* 13: 333-335.

related plant pathogenic fungi. *Mol. Biol. Evol.* 16: 114-126.

Rippka R. 1988. Recognition and identification of cyanobacteria. *Meth. Enzymol*. 167: 28-67.

Rippka R. and Herdman M. 1992. Pasteur Culture collection of Cyanobacteria Catologue and Taxonomic Handbook 1. Catologue of Strains. Institut Pasteur, Paris.

Rippka R., Deruelles J., Waterbury J. B., Herdman M. and Stanier R.Y. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiology*. 111: 1-61

Rippka R., Waterbury J. B., Herdman M. and Castenholz R. W. 2001. Subsection II. (Formerly Pleurocapsales Geitler 1925, emend. Waterbury and Stanier 1978) In *Bergeys Manual of Systematic* Edited by Boone, D. R.and Castenholz R. W. Volume 3. Springer-Verlag New York, Berlin, Heidelberg. Pages 1746-1770.

Rodríguez F. J., Oliver J. L., Marín A. and Medina J. R. 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* 142: 485-501.

Roman J. and Woodson S. A. 1998. Integration of the *Tetrahymena* group I intron into bacterial rRNA by reverse splicing in vivo. *Proc. Natl. Acad. Sci. USA* 95: 2134-2139.

Rudi K. and Jakobsen K. S. 1999. Complex evolutionary patterns of tRNA Leu (UAA) group I introns in the cyanobacterial radiation. *J. Bacteriol.*181: 3445-3451.

Rudi K. and Jakobsen K.S. 1997. Cyanobacterial tRNALeu (UAA) group I introns have a polyphyletic origin. *FEMS Microbiol. Lett*. 156: 293-298.

Rudi K., Skulberg O. M., Larsen F. and Jakobsen K. S. 1997. Strain characterization and classification of oxyphotobacteria in clone cultures on the basis of 16S rRNA sequences from the variable regions V6, V7 and V8. *Appl. Environ. Microbiol*. 63: 2593-2599.

Rzhetsky A. and Nei M. 1992. Statistical properties of the ordinary least-squares, generalized least-squares and minimum-evolution methods of phylogenetic inference. *J. Mol. Evol.* 35: 367-375.

Sagan L. 1967. On the origin of mitosing cells. *J. Theoret. Biol*. 14: 225-274.

Saitou N. and Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. 4: 406-425.

Salzberg S. L., White O., Peterson J. and Eisen J. A. 2001. Microbial Genes in the Human Genome: Lateral Transfer or Gene Loss? *Science* 292: 1903-1906

Sambrook J., Fritsch E. F. and Maniatis T. 1989. Molecular cloning: A laboratory manual, 2nd Edition. Cold Spring Harbour Laboratory Press, New York.

Schimper A. F. W. 1883. Über die Entwicklung der Chlorophyll Körner und Farb-Körner. *Bot. Zeit.* 41: 105-114.

Schlösser U. G. 1994. SAG - Sammlung von Algenkulturen at the University of Göttingen. Catalogue of strains. *Bot. Acta.* 107: 113-186.

Schlösser U.G. 1997. Additions to the culture collection of algae since 1994. *Bot. Acta.* 110: 424-429.

Schopf J. W. 1993. Microfossils of the Early Archean Apex chert: new evidence of the antiquity of life. *Science* 260: 640-646

Schopf J. W. 1994. Disparate rates, differing fates: tempo and mode of evolution changed from the Precambrian to the Phanerozoic. *Proc. Natl. Acad. Sci. USA* 91: 6735-42.

Schopf J. W. 2000. Solution to Darwin's dilemma: discovery of the missing Precambrian record of life. *Proc. Natl. Acad. Sci. USA* 97 6947-6953.

Schopf J. W. and Packer B. M. 1987. Early Archaen (3.3 billion to 2.5 billion-year-old) microfossils from Warrawoona Group, Australia. *Science* 237: 70-73.

Schopf J. W. and Walter M. R. 1982. Origin and early evolution of cyanobacteria: the geological evidence. In *Biology of Cyanobacteria*. Edited by Carr N.G. and Whitton B. A. Blackwell Science Oxford. Pages 543-564

Sellem C. H. and Belcour L. 1997. Intron open reading frames as mobile elements and evolution of a group I intron. *Mol. Biol. Evol.* 5: 518-26.

Sharma M., Ellis R. L. and Hinton D. M. 1992. Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. *Proc. Natl. Acad. Sci. USA* 14: 6658-6662.

Shaw A. J. and Allen B. 2000. Phylogenetic relationships, morphological incongruence, and geographic speciation in the fontinalaceae (Bryophyta). *Mol. Phylogenet. Evol.* 16: 225-37.

Shimada H. and Sugiura M. 1991. Fine structural features of the chloroplast genome: comparison of

Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N., Matsubayashi T., Zaita N., Chunwongse J., Obokata J., Yamaguchi-Shinozaki K., Ohto C., Torazawa K., Meng B.Y., Sugita M., Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A., Tohdoh N., Shimada H. and Sugiura M. 1986.The complete nucleotide sequence of tobacco chloroplast genome:its gene organization and expression. *EMBO J.* 5: 2043-2049.

Shub D. A. 1991. The antiquity of tRNA introns. *Curr. Opin. Genes Devel.* 1: 478-484.

Shub D. A., Gott J. M., Xu M. Q., Lang B. F., Michel F., Tomaschewski J., Pedersen-Lane J. and Belfort M. 1988. Structural conservation among three homologous introns of bacteriophage T4 and the group I introns of eukaryotes. *Proc. Natl. Acad. Sci. USA* 4: 1151-1155.

Skulberg O. M. 1990. Norwegian Institute for Water Research. Culture collection of algae. Catalogue of strains. Norwegian Institute for Water Research, Oslo, Norway. Pages 1-39.

Smith J. F. 2000. Phylogenetic signal common to three data sets: combing data which initially appear heterogeneous. *Plant Syst. Evol*. 221: 179-198.

Sogin M. L., Ingold A., Karlok M., Nielsen H. and Engberg J. 1986. Phylogenetic evidence for the acquisition of ribosomal RNA introns subsequent to the divergence of some of the major *Tetrahymena* groups. *EMBO J.* 5: 3625-3630.

Soliva M., Kocyan A. and Widmer A. 2001. Molecular phylogenetics of the sexually deceptive orchid genus Ophrys (Orchidaceae) based on nuclear and chloroplast DNA sequences. *Mol. Phylogenet. Evol*. 20: 78-88.

Springer M. S., Cleven G. C., Madsen O., de Jong W. W., Waddell V. G., Amrine H. M. and Stanhope M. J. 1997. Endemic African mammals shake the phylogenetic tree. *Nature* 388:61-64.

Stanhope M. J., Lupas A., Italia M. J., Koretke K. K., Volker C. and Brown J. R. 2001. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* 411: 940-944.

Stanier R. Y. and van Niel C. B. 1962. The concept of a bacterium. *Arch. Mikrobiol*. 42: 17-45.

Stanier R. Y., Kunisawa R., Mandel M. and Cohen-Bazire, G. 1971. Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriological Reviews.* 35: 171-205.

Starr R. C. and Zeikus J. A. 1993. The Culture Collection of Algae at the University of Texas at Austin. *J. Phycol*. 29 (suppl.): 1 – 106.

Steel M. A., Lockhart P. J. and Penny D. 1993. Confidence in evolutionary trees from biological sequence data. *Nature* 364: 440-442.

Steinmetz A., Gubbins E. J. and Bogorad L. 1982. The anticodon of the maize chloroplast gene for tRNA Leu UAA is split by a large intron. *Nucleic Acids Res.* 10: 3027-3037.

Stiller J. W. and Hall B. D. 1997. The origin of red algae: implications for plastid evolution. *Proc. Natl. Acad. Sci. USA* 94: 4520-4225.

Stirewalt V. L., Michalowski C. B., Luffelhardt W., Bohnert H. J. and Bryant D.A. 2000. Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. Unpublished. Genbank NC_001675.

Strehl B., Holtzendorff J., Partensky F. and Hess W. R. 1999. A small and compact genome in the marine cyanobacterium Prochlorococcus marinus CCMP 1375: lack of an intron in the gene for tRNA(Leu)(UAA) and a single copy of the rRNA operon. *FEMS Microbiol Lett*. 181: 261-266.

Sugita M., Luo L., Ohta M., Itadani H., Matsubayashi T. and Suguira M. 1995. Genes encoding the group I intron-containing tRNA(Leu) and subunit L of NADH dehydrogenase from the cyanobacterium *Synechococcus* PCC 6301. *DNA Res*. 2: 71-76.

Summons R. E., Jahnke L. L., Hope J. M. and Logan G. A. 1999. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* 400 :554-557.

Swofford D. L. and Olsen G. L. 1990. Phylogeny reconstruction. *In* Hillis DM and Moritz C (Eds.) *Molecular Systematics*. Sinauer, Sunderland, Mass. Pages 411-501.

Swofford D. L., Olsen G. J., Wadell P. J. and Hillis D. M. 1996. Phlogenetic Inference. *In Molecular Systematics*. Edited by D. M. Hillis, C. Moritz and B. K. Mable. Sinauer, Sunderland M. A. Pages 407-514.

Swofford, D. L. 2001. PAUP*. Phylogenetic analysis using Parsimony (*and other methods). Version 4.0b8. Sinauer, Sunderland MA.

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

Takeuchi M., Hamana K. and Hiraishi A. 2001. Proposal of the genus *Sphingomonas* sensu stricto and three new genera, *Sphingobium*, *Novosphingobium* and *Sphingopyxis*, on the basis of phylogenetic and chemotaxonomic analyses. *Int. J. Syst. Evol. Microbiol*. 51: 1405-1417.

Tamura K. and Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol*. 10: 512-526.

Taylor T. N., Hass H., Remy W. and Kerp H. 1995. The oldest fossil lichen. *Nature* 378: 244.

Teeling E. C., Scally M., Kao D. J., Romagnoli M. L., Springer M. S. and Stanhope M. J. 2000. Molecular evidence regarding the origin of echolocation and flight in bats. *Nature* 403: 88-92.

Templeton A. R. 1983. Phylogenic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* 37: 221-244.

Thuret G. 1875. Essai de classification des Nostochinees. *Annals Sci. Nat. Bot*. 6: 372-382.

Toledo G., Palenik B. and Brahamsha B.1999. Swimming marine Synechococcus strains with widely different photosynthetic pigment ratios form a monophyletic group. *Appl. Environ. Microbiol.* 65: 5247-5251.

Tsudzuki J., Ito S., Tsudzuki T., Wakasugi T. and Sugiura M. 1994. A new gene encoding tRNA(Pro) (GGG) is present in the chloroplast genome of black pine: a compilation of 32 tRNA genes from black pine chloroplasts *Curr. Genet.* 26: 153-158.

Turmel M., Cote V., Otis C., Mercier J. P., Gray M. W., Lonergan K. M. and Lemieux C. 1995. Evolutionary transfer of ORF-containing group I introns between different subcellular compartments (chloroplast and mitochondrion) *Mol. Biol. Evol.* 12: 533-545.

Turmel M., Otis C. and Lemieux C. 1999. The complete chloroplast DNA sequence of the green alga Nephroselmis olivacea: insights into the architecture of ancestral chloroplast genomes. *Proc. Natl. Acad. Sci. USA* 96: 10248-10253.

Turner S. 1997. Molecular systematics of oxygenic photosynthetic bacteria. *Plant Syst. Evol*. [Suppl.] 11: 13-52

Turner S., Pryer K. M., Miao V. P. and Palmer J. D. 1999. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J. Euk. Microb.* 46: 327-338.

Uhlenbeck O. C. 1995. Keeping RNA happy. *RNA* 1: 4-6.

Vaughn J. C., Mason M. T., Sper-Whitis G. L., Kuhlman P. and Palmer J. D. 1995. Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric CoxI gene of *Peperomia*. *J. Mol. Evol.* 41: 563-572.

Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S., Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T., Inamura A., Yoshinaga K. and Sugiura M. 1997. Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: the existence of genes possibly involved in chloroplast division. *Proc. Natl. Acad. Sci. USA* 11: 5967-5972

Wallander E. and Albert V. A. 2000. Phylogeny and classification of Oleaceae based on rps16 and *trnL*-F sequence data. Am. J. Bot. 87: 1827-1841.

Walter M. R. 1983. In *Earth's Earliest Biosphere*. Edited by Schopf J. W. Princeton University Press. Pages 187-213.

Watanabe K. I., Ehara M., Inagaki Y., Ohama T. 1998. Distinctive origins of group I introns found in the COXI genes of three green algae. *Gene* 213:1-7.

Waterbury J. B. 1989. Subsection II Order Pleurocapsales. In *Bergeys Manual of Systematic Bacteriology*. Edited by R. E. Buchanan and N. E. Gibbons. Volume 3:. Williams and Wilkins, Baltimore. Pages 1746-1770.

Waterbury J. B. and Rippka R. 1989 Subsection I order Chroococcales Wettstein 1924 emend, Waterbury and Stanier 1978. In *Bergeys Manual of systematic bacteriology.* Edited by Stanley J. T., Byrant M. P., Pfennig N. and Holt J. G. Volume 3. Baltimore: Williams and Wilkins. Pages 1746-1770.

Waterbury J. B. and Stanier R.Y. 1978. Patterns of growth and development in Pleurocapsalean cyanobacteria. *Microbiol. Rev*. 42: 2-44.

Weber U., Beier H. and Gross H. J. 1996. Another heritage from the RNA world: self-excision of intron sequence from nuclear pre-tRNAs. *Nucleic Acids Res.* 24: 2212-2219.

Wenzlau J. M., Saldanha R. J., Butow R. A. and Perlman P. S. 1989. A latent intron-encoded maturase is also an endonuclease needed for intron mobility. *Cell* 56: 421-430.

Wikström N. and Kenrick P. 2001. Evolution of Lycopodiaceae (Lycopsida): estimating divergence times from *rbcL* gene sequences by use of nonparametric rate smoothing. *Mol. Phylogenet. Evol*. 19:177-86.

Wikström N., Kenrick P. and Chase M. 1999. Epiphytism and terrestrialization in tropical Huperzia (Lycopodiaceae). *Pl. Syst. Evol.* 218: 221–243.

Williamson D. H., Gardner M. J., Preiser P., Moore D. J., Rangachari K., Wilson R. J. 1994. The evolutionary origin of the 35 kb circular DNA of *Plasmodium falciparum*: new evidence supports a possible rhodophyte ancestry. *Mol. Gen. Genet.* 243: 249-252.

Wilmotte A., Neefs J. M. and De Wachter R. 1994. Evolutionary affiliation of the marine nitrogen-fixing cyanobacterium *Trichodesmium* sp. strain NIBB 1067, derived by 16S rRNA ribosomal RNA sequence analysis. *Microbiology*. 140: 2159-2164.

Wilmotte A., Van der Auwera G. and De Wachter R. 1993. Structure of the 16S ribosomal RNA of the thermophilic cyanobacterium *Chlorogloeopsis* HTF (*'Mastigocladus laminosus* HTF') strain PCC 7518 and phylogenetic analysis. *FEBS Lett.* 317: 96-100.

Wilson R. J., Denny P. W., Preiser P. R., Rangachari K., Roberts K., Roy A., Whyte A., Strath M., Moore D. J., Moore P. W. and Williamson D.H. 1996. Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J. Mol .Biol.* 261: 155-172.

Wittzell H. 1999. Chloroplast DNA variation and reticulate evolution in sexual and apomictic sections of dandelions. *Mol Ecol.* 8: 2023-2035.

Woese C. R. Sogin M. L. Bonen L. and Stahl D. 1975. Sequence studies on 16S ribosomal RNA from blue green algae. *J. Mol. Evol*. 4: 307-315.

Woese C. R., Sogin M. L., Bonen L. and Stahl D. 1975. Sequence studies on 16S ribosomal RNA from a blue-green alga. *J. Mol Evol.* 4: 307-315.

Wolk C. P., Ernst A. and Ehlai J. 1994. Heterocyst metabolism and Development. *In The Molecular Biology of Cyanobacteria*. Edited by D. A. Byrant. Kluwer Academic Publishers, Holland. Pages. 769-823

Woodson S. A. 1998. Ironing out the kinks: splicing and translation in bacteria. *Genes Dev.* 12: 1243-1247.

Woodson S. A. and Cech T. R. 1989. Reverse self-splicing of the *Tetrahymena* group I intron: implication for the directionality of splicing and for intron transposition. *Cell* 57: 335-345.

Wright D., Prickett T., Helm R. F. and Potts M. 2001. Form species *Nostoc commune* (Cyanobacteria). *Int J Syst Evol Microbiol.* 51: 1839-1852.

Xu M. Q., Kathe S. D., Goodrich-Blair H., Nierzwicki-Bauer S. A. and Shub D. A. 1990. Bacterial origin of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria. *Science* 250: 1566-1570.

Yabuuchi E, Yano I, Oyaizu H, Hashimoto Y, Ezaki T, Yamamoto H. 1990. Proposals of *Sphingomonas paucimobilis* gen. nov. and comb. nov., *Sphingomonas parapaucimobilis* sp. nov., *Sphin-*

gomonas yanoikuyae sp. nov., *Sphingomonas adhaesiva* sp. nov., *Sphingomonas capsulata* comb. nov., and two genospecies of the genus *Sphingomonas*. *Microbiol. Immunol*. 34: 99-119.

Yang D., Oyaizu Y., Oyaizu H., Olsen G. J. and Woese C. R. 1985. Mitochondrial origins. *Proc. Natl. Acad. Sci. USA* 82: 4443-4447.

Yang J., Zimmerly S., Perlman P. S. and A. M. Lambowitz 1996. Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* 381: 332-335.

Zhu S., Sobolev A. Y. and Wek R. C. 1996. Histidyl-tRNA synthetase-related sequences in GCN2 protein kinase regulate in vitro phosphorylation of eIF-2. *J. Biol. Chem.* 271: 24989-24994.

Zuag A. J., McEnvoy M. M. and Cech T. R. 1993. Self-splicing of the Group I intron from Anabaena pre-tRNA: requirement for base-pairing of the exons in the anticodon stem. *Biochemistry*. 32: 7946-7953.