J Mol Evol (1997) 44:258–271

JOURNAL OF MOLECULAR EVOLUTION

The Internal Transcribed Spacer 2 Exhibits a Common Secondary Structure in Green Algae and Flowering Plants

Jeffrey C. Mai, Annette W. Coleman

Department of Molecular and Cellular Biology and Biochemistry, Brown University, Providence, RI 02912, USA

Received: 29 July 1996 / Accepted: 27 August 1996

Abstract. Sequences of the Internal Transcribed Spacer 2 (ITS-2) regions of the nuclear rDNA repeats from 111 organisms of the family Volvocaceae (Chlorophyta) and unicellular organisms of the Volvocales, including *Chlamydomonas reinhardtii*, were determined. The use of thermodynamic energy optimization to generate secondary structures and phylogenetic comparative analysis of the spacer regions revealed a common secondary structure that is conserved despite wide intra- and interfamilial primary sequence divergence. The existence of this conserved higher-order structure is supported by the presence of numerous compensating basepair changes as well as by an evolutionary history of insertions and deletions that nevertheless maintains major aspects of the overall structure. Furthermore, this general structure is preserved across broad phylogenetic lines, as it is observed in the ITS-2s of other chlorophytes, including flowering plants; previous reports of common ITS-2 secondary structures in other eukaryotes were restricted to the order level. The reported ITS-2 structure possesses important conserved structural motifs which may help to mediate cleavages in the ITS-2 that occur during rRNA transcript processing. Their recognition can guide further studies of eukaryotic rRNA processing, and their application to sequence alignments may contribute significantly to the value of ITS-2 sequences in phylogenetic analyses at several taxonomic levels, but particularly in characterizing populations and species.

Key words: Angiosperms — *Chlamydomonas* — Ribosomal RNA — RNA secondary structure — Transcribed spacer

Correspondence to: A.W. Coleman

Introduction

Eukaryotic ribosomal RNA (rRNA) genes occur as tandem repeats and their primary transcript includes small-subunit (SSU), 5.8S, and large-subunit (LSU) rRNA separated by internal transcribed spacer (ITS) regions. During ribosome biogenesis in eukaryotic nuclei, a specific cleavage occurs within the Internal Transcribed Spacer 2 (ITS-2) of the primary transcript which leads to the eventual generation of the mature 5.8S 3′ and LSU 5′ termini (Veldman et al. 1981; Hadjiolova et al. 1984a,b; Nashimoto et al. 1988; Hadjiolova et al. 1994). Insertion/deletion mutagenesis studies of ribosomal genes have demonstrated the necessity of certain subsequences in the ITS-2 that are necessary for proper ribosomal RNA transcript processing. Work by Hadjiolova et al. (1994) points to the terminal 326 nt of the ITS-2 and the first 502 nt of the 28S as important *cis* signals in the human transcript. More recently, studies on yeast have elucidated specific regions of the secondary structure of the ITS-2 that, when altered, directly affect its processing (van der Sande et al. 1992; van Nues et al. 1995). Also, ribosomal DNA transfected from mouse into rat cells and human rDNA genes transfected into mouse cells seem to undergo normal processing despite marked primary structural divergence (Vance et al. 1985; Hadjiolova et al. 1994). These data suggest that the spacers possess information in their higher-order structures that participates in directing endonucleolytic enzymes to the proper cut sites.

Localization of the initial ITS-2 cut site appears to be mediated by the formation of thermodynamically stable hairpin loops within the ITS-2 that are recognized by the associated processing molecules. These higher-order structures have previously been visualized in electron microscopy studies positioned on the ITS-2 of human transcripts (Gonzalez et al. 1990). Comparative sequence analysis, as well as chemical and enzymatic probing and energy optimization, have been used to determine putative secondary structures of the ITS-2s for several other eukaryotic organisms (Michot et al. 1983, 1993; Kupriyanova and Timofeeva 1987; Nazar et al. 1987; Yeh and Lee 1990; Venkateswarlu and Nazar 1991; Kwon and Ishikawa 1992; Wesson et al. 1992; Schlötterer et al. 1994; Bakker et al. 1995; Baldwin et al. 1995). However, to date, among all eukaryotes only the ITS-2 secondary structure for *Saccharomyces cerevisiae* remains unambiguously resolved.

In yeast, structural data coupled with mutational and insertion/deletion studies have provided strong evidence for structural elements within the ITS-2 that are important for efficient ITS-2 processing (Yeh and Lee 1990; van der Sande et al. 1992; van Nues et al. 1995). For other eukaryotes, the bulk of structural ITS-2 determination has been inconclusive, with little supporting evidence of compensating basepair changes (CBCs) and other structural clues. Among animals, only in insects and trematodes have compensating basepair changes been detected to support proposed secondary structures in the ITS-2, with a number of CBCs detected in five trematode sequences, nine compensatory changes detected in seven mosquito sequences, and one CBC found in eight *Drosophila* sequences (Wesson et al. 1992; Michot et al. 1993; Schlötterer et al. 1994). Moreover, evidence of secondary structure in plants is circumstantial, at best.

The difficulty in arriving at a common ITS-2 secondary structure for a given group of eukaryotes is usually complicated by either a high rate of sequence divergence among distantly related organisms or a level of conservation that is too great to be structurally revealing within a group of close relatives. However, if a set of sequences can be obtained that spans the appropriate levels of conservation, then phylogenetic comparison can be invoked effectively to reveal secondary and even tertiary structural elements.

The phylogenetic comparative method of analysis is considered to be one of the most powerful tools for uncovering the higher-order structure of RNA, with information that can sometimes reveal subtle structural motifs that cannot be detected through single-strand-specific chemical modification of nucleotides or nuclease susceptibility experiments (Noller and Woese 1981; Michot et al. 1984; Gutell et al. 1985, 1992; Pace et al. 1989; Gutell and Woese 1990). Furthermore, such analysis yields biologically relevant foldings that may not be detectable in vitro.

Phylogenetic comparison appears to be a particularly attractive approach to studying the ITS-2 in eukaryotes, as the work of van Nues et al. (1995) has shown that the portions of the ITS-2 that are most important in transcript processing in yeast are those that are evolutionarily conserved. Thus, we have chosen this method of analysis in order to determine ITS-2 structures for the chlorophytes.

In this paper, we report the results of the analysis of over 100 ITS-2 regions from the green flagellate order Volvocales (Table 1). In addition, previously published ITS-2s from the macroalga *Acrosiphonia arcta* (van Oppen et al. 1993) and a full range of representative angiosperms including the dicots *Daucus* (Apiaceae), *Calycadenia* (Asteraceae), *Alnus* (Betulaceae), *Arabidopsis* and *Sinapis* (Brassicaceae), *Canella* (Canellaceae), *Dubautia* (Compositae), *Cucumis* (Cucurbitaceae), *Vicia* (Fabaceae), *Gossypium* (Malvaceae), *Epilobium* (Onagraceae), *Ranunculus* (Ranunculaceae), *Photinia* (Rosaceae), *Populus* (Salicaceae), *Lycopersicon* (Solanaceae), and *Arceuthobium* (Viscaceae) and the monocots *Yucca* (Liliaceae), *Vanilla* (Orchidaceae), and *Oryza* (Poaceae) have been included in this study. Through the use of thermodynamic energy optimization in conjunction with the phylogenetic comparative method, a secondary structure possessing several highly conserved structural motifs has been defined that can be used to guide analysis of the ITS-2 sequences of Chlorophyta.

Materials and Methods

Strains and Cultures and Growth Conditions. The 111 Volvocacean algae used here, obtained from the sources listed in Table 1, were grown in either Volvox medium or soil–water medium (Starr and Zeikus 1993). To avoid excessive redundancy, from any subgroup no more than two organisms known to be capable of interbreeding are listed.

Enzymes, Transformational Procedures, Plasmids, and Sequencing. For each organism, the nucleotide sequence of the nuclear rDNA ITS-2 was obtained as described previously (Coleman et al. 1994). Briefly, DNA extracts prepared using InstaGene Matrix solution (Bio-Rad, Hercules, CA) were used as template for polymerase chain reactions (PCRs) utilizing primers for sites near the 3′ end of the SSU and the 5′ end of the LSU, yielding the entire region between the SSU and LSU RNA. The product was ligated into pT7 Blue vector (Novagen, Madison WI) and used to transform *E. coli* cells. Minipreps (Sambrook et al. 1989) of plasmid DNA from transformants provided the material for manual sequencing of both DNA strands by the Sanger dideoxy method (Sanger et al. 1977), using the protocols and Sequenase 2.0 kit of US Biochemicals Co. (Cleveland, OH). External (vector) primers for sequencing are available commercially, and additional primers within conserved regions of the 5.8S gene were those used by Goff and Moon (1993).

MacVector and AssemblyLIGN software (Kodak, International Biotechnologies, Inc., New Haven, CT) were used to organize the sequences.

The algal cultures are clonal. Three to five PCR reactions from each

Table 1. Identities of the Volvocales for which the ITS-2 was sequenced[a]

	#	Source	Origin
"Unicells"			
Chlamydomonas reinhardtii	CC620	OGC	USA
C. incerta	7.73	SAG	Cuba
C. callosa	624	UTEX	Netherlands
C. zebra	229	UTEX	USA
C. cullieus	1057	UTEX	USA
C. asymmetrica	70.72	SAG	Czechoslovakia
C. lyengarii	221	UTEX	India
C. mexicana	729	UTEX	USA
C. mexicana	729	UTEX	USA
C. geitleri-noctigama	2289	UTEX	Czechoslovakia
C. allensworthii	LCA74	Starr	USA
C. sp.	266	Starr	USA
C. sp.	Flamin	Starr	Germany
C. sp.	21A	Starr	USA
C. komma	579	UTEX	Japan
C. gelatinosa	69.72	SAG	Czechoslovakia
Haematococcus droebakensis	55	UTEX	Sweden
Stephanosphaera sp.	2409	UTEX	China
Stephanosphaera pluvialis	771	UTEX	Finland
Dunaiella sp.		FHL	
Volvocaceae			
Pandorina morums			
Pandorina morum	876	UTEX	USA
P. morum	EP-2	Coleman	USA
P. morum	1726	UTEX	South Africa
P. morum	2331	UTEX	Korea
P. morum	2332	UTEX	USA
P. morum	SA-88	Coleman	Australia
P. morum	1732	UTEX	Nepal
P. morum	868	UTEX	USA
P. morum	FP-3	Coleman	USA
P. morum	862	UTEX	USA
P. morum	BI-1	Coleman	USA
P. morum	1734	UTEX	USA
P. morum	884	UTEX	Thailand
P. morum	865	UTEX	USA
P. morum	I-b	Coleman	Chile
P. morum	1724	UTEX	South Africa
Other *Pandorina* spp.			
Pandorina charkowiensis	Butler	Coleman	USA
P. charkowiensis	841	UTEX	USA
P. unicocca	2127	UTEX	USA
P. unicocca	2428	UTEX	Japan
P. charkowiensis	165	UTEX	Finland
Volvulinas			
Volvulina steinii	A	Coleman	Chile
V. steinii	1531	UTEX	USA
V. steinii	1525	UTEX	USA
V. steinii	Kris	Coleman	Argentina
V. pringsheimii	1020	UTEX	USA
V. compacta	7528	Starr	Japan
V. boldii	1761	UTEX	Australia
Pleodorinas			
Pleodorina californica	521	Starr	USA
P. californica	809	UTEX	USA
P. californica	198	UTEX	USA
P. japonica	2523	UTEX	Japan
P. indica	1991	UTEX	Mexico
Platydorina caudata	1658	UTEX	USA
Eudorinas			
Eudorina cylindrica	1197	UTEX	USA
E. sp.	Wuhan	Coleman	China
E. sp.	Cox	Starr	USA
E. unicocca	1221	UTEX	USA
E. sp.	Swed	Coleman	Sweden
E. sp.	Barb	Coleman	Barbuda, W.I.
E. elegans	1210	UTEX	USA
E. elegans	1205	UTEX	USA
E. elegans	1201	UTEX	USA
Volvoxes			
Volvox spermatosphaera	2273	UTEX	South Africa
V. spermatosphaera	62	Starr	Mexico
V. spermatosphaera	Stk	Coleman	USA
V. tertius	132	UTEX	England
V. africanus	1892	UTEX	South Africa
V. africanus	1889	UTEX	USA
V. africanus	1891	UTEX	Australia
V. africanus	1890	UTEX	Australia
V. carteri f. weismannia	1876	UTEX	Australia
V. carteri f. weismannia	1874	UTEX	USA
V. carteri f. kawasakiensis	Kawa	Starr	Japan
V. carteri f. nagariensis	Poona	Starr	India
V. carteri f. nagariensis	72-52	Starr	Japan
V. obversus	1865	UTEX	Australia
V. dissipatrix	1871	UTEX	Germany
V. rousselettii	1862	UTEX	South Africa
V. barberi	804	UTEX	USA
V. globator	955	UTEX	USA
V. capensis	zyg-6	Starr	South Africa
V. powersii	1864	UTEX	USA
V. gigas	1895	UTEX	South Africa
V. aureus	1899	UTEX	USA
Goniums			
Gonium pectorale	2570	UTEX	Canada
G. pectorale	16	Coleman	USA
G. quadratum	2	Starr	USA
G. quadratum	90-423	Starr	Nepal
G. sp.	956	UTEX	Italy
G. sociale	KG-4	Nozaki	Antarctic
G. sociale	21028	Nozaki	Japan
G. sociale	14	UTEX	England
G. sociale f. sociale	197	UTEX	Germany
G. sociale f. sacculiferum	936	UTEX	USA
G. sacculiferum	935	UTEX	USA
G. sacculiferum	822	UTEX	USA
G. multicoccum	2580	UTEX	Nepal
G. multicoccum	783	UTEX	USA
G. octonarium	842	UTEX	USA
G. viridastellatum	2520	UTEX	Japan
Astrephomenes			
Astrephomene gubernaculifera	2477	UTEX	USA
A. gubernaculifera	Zimbab	UTEX	Zimbabwe
A. gubernaculifera	2482	UTEX	USA
A. gubernaculifera	Wuhan	Coleman	China
A. gubernaculifera	SA88-10	Starr	Australia
A. gubernaculifera	Yoshi	Starr	Japan
A. gubernaculifera	J	Starr	Chile
A. gubernaculifera	2479	UTEX	USA
A. gubernaculifera	1392	UTEX	USA
A. perforata	2475	UTEX	Japan

Table 1. Continued

[a] "Unicells" (*Stephanosphaera* forms a type of colony) encompasses three families: Chlamydomonadaceae, Sphaerellaceae, and Polyblepharidaceae. Remaining groups are all members of the colonial family Volvocaceae. Sources include CGC = Chlamydomonas Genetics Center, SAG = Sammlung von Algen Kulturen, UTEX = University of Texas Algal Collection, FHL = Friday Harbor Marine Laboratory; Star = R. C. Starr, University of Texas at Austin, Coleman = A. W. Coleman, Brown University, Nozaki = H. Nozaki, University of Tokyo

culture were combined for ligation into vector; in most cases, at least two transformant clones were sequenced, and as discussed in Coleman (1994), such clones were usually identical in sequence or differed by one or two base pairs, usually in ITS-1 (All sequences listed in Table 1 have been submitted to the Genbank database.)

Other ITS-2 sequences utilized were *Acrosiphonia arcta* sequences from van Oppen et al. (1993), *Vanilla planifolia* from Dr. Mark Chase, and GenBank accessions X68138 (*Alnus incana*), X52320 (*Arabidopsis thaliana*), L25788 (*Arceuthobium americanum*), U04256 (*Calycadenia oppositifolia*), L03844 (*Canella winterana*), M36377 (*Cucumis melo*), X17534 (*Daucus carota*), N93790 (*Dubautia arborea*), L28012 (*Epilobium canum*), X74114 (*Funaria hygrometrica*), U12719 (*Gossypium hirsutum*), X52265 (*Lycopersicon esculentum*), M16845 (*Oryza sativa*), U16199 (*Photinia pyrifolia*), X64764 (*Populus deltoides*), X83848 (*Ranunculus enysii*), U16203 (*Rhaphiolepis indica*), X15915 (*Sinapis alba*), X17535 (*Vicia faba*), U24054 (*Yucca whipplei*).

Alignments and Thermodynamic Energy Optimization. The ITS-2 sequences were first aligned by eye in AssemblyLIGN by sequence conservation. In the Volvocales, the 5′ ends of the sequences were defined according to Thompson and Herrin (1994) and the 3′ ends were determined according the *Cladophora* LSU (Bakker et al. 1992). For the other sequences, the 5′ and 3′ ends of the ITS-2s were determined in accordance with the published or Genbank designations.

Initial screening for the secondary structure of the sequences was then conducted by folding the sequences either on a four-processor Sun SPARC station using the mFOLD program for sequences over 300 nt or on a Macintosh computer using the Mulfold program for temperature-dependent folding (Jaeger et al. 1989, 1990; Zuker 1989). Foldings were conducted at 20°C using a search within 20% of thermodynamic optimality set to yield 12–15 suboptimal folds. Since both mFOLD and Mulfold rely on the same RNA folding algorithm, when identical parameters were used, results were the same for either program. Output was visualized and edited using the LoopDLoop RNA editing program (Gilbert 1992) and the resulting secondary structures were screened for common motifs.

General sequence alignments were then refined in accordance to common secondary structures. In order to facilitate the alignments of all 111 sequences, the alignments were divided into smaller phylogenetically related groups. ITS-2 secondary structures from the RNA folding programs were then altered manually in order to achieve the most phylogenetically supported structures. The approach used is an iterative process in which possible structures are identified and repeatedly refined until the nucleotide sequence and covariation data and structures are in agreement.

The Search for Compensating Base-Pair Changes. The aligned sequences were searched for covariations. CBCs were defined to include Watson-Crick pairings and G·U pairs. Covariations were recorded to be those positions within *unambiguously alignable* subsequences that changed so as to maintain base pairing. Also, up to two consecutive, fully base-paired positions immediately adjacent to the end of a helical region of completely alignable positions could be included as CBCs. These criteria were established in order to reduce the number of nonhomologous, therefore nonrelevant, pairings flagged as covariations.

In addition, the search for positional covariances took place at two different phylogenetic levels; across the Volvocales and within subgroups of more closely related organisms, which generally comprised the same genus but represented a large family in the case of the Rosaceae. Compensating base-pair searches within groups of more closely related organisms permitted the detection of covariations that would have been lost at higher levels of genetic distance. Furthermore, such covariations are considered to be more significant than those detected among more distantly related organisms (Gutell et al. 1994).

The higher divergence rates, greater difficulty in aligning divergent sequences, and lesser constraints acting on higher-order structure in the ITS-2, in comparison to those of the genic rRNA sequences, precluded the a priori identification of CBCs that has been used to study the structures of other RNAs (Noller and Woese 1981; Michot et al. 1984; Gutell et al. 1985, 1992; Pace et al. 1989; Gutell and Woese 1990). This approach was attempted through the use of the Covariation HyperCard stack (Brown 1991); however, excessive background noise in the data obscured the detection of potential sites for positional covariations in the Volvocalean ITS-2s.

Alignments, analyses of all CBCs, and foldings not presented here are available from the authors upon request.

Results and Discussion

Gross Sequence Compositions and Lengths

The ITS-2 sequences of the 111 Volvocales fall generally within the range of 220 to 250 nt in length with a mean length of 250 nt (SD ± 63 nt). *Pleodorina californica* 809 possessed the longest spacer, 501 nt in length, and the ITS-2 of *Chlamydomonas* sp. 21A was the shortest at 190 nt. *Pleodorina* showed the greatest length heterogeneity within its genus, with ITS-2 sequences ranging from 236 nt in *Pleodorina indica* to 501 nt in *Pleodorina californica* 809. The mean G + C content of the Volvocalean ITS-2 sequences was 48.1% (SD ± 4.8%), with the highest G + C content detected in *Chlamydomonas callosa* (58.6%) and the lowest G + C content found in *Pleodorina japonica* (35.7%).

For the 20 diverse angiosperm ITS-2s (including only *Photinia* to represent the 23 Rosaceae discussed below) examined in this study, the ITS-2 lengths were relatively homogeneous compared to the Volvocales with a mean ITS-2 length of 219 nt (SD ± 16 nt) and a range of 187 to 249 nt. However, the average G + C was higher and more variable than the Volvocales with an average of 58.9% (SD ± 10.7%) and a range of 30 to 77%. (For a more extensive sampling of lengths and G + C values see Baldwin et al. 1995.) *Acrosiphonia arcta* Roscoff, at 183 nt (G + C = 51.9%), turned out to have the shortest chlorophyte ITS-2 examined.

Although there may be some lower limit to the length of the Volvocalean ITS-2, reflecting the core conserved secondary structures discussed below, there is no apparent upper limit. If we use the mean ITS-2 length of 250 nt, approximately 52% of the average Volvocalean ITS-2 consists of these conserved subsequences. Thus, about 48% of the typical Volvocalean ITS-2 appears to be free to diverge, within the limits that such changes do not disrupt formation of the core structure. For *C.* sp. 21A, which possesses the shortest Volvocalean ITS-2, conserved sequences constitute 68.4% of the entire spacer length.

In individual cases, e.g., species of *Pleodorina* and *Volvox*, spacer lengths can vary greatly, with no obvious upper limit. For *Pleodorina californica* 809 where the ITS-2 is 501 nt, conserved subsequences comprise only 26.0% of the entire spacer. Interestingly, it is the non-

conserved portions of the *P. californica* ITS-2s that have high A + U content, leading to the lowest overall G + C levels in the Volvocales (36.6–36.8% G + C).

The situation in the other plants seems to mirror the Volvocalean requirement for a minimum ''core'' ITS-2; the homogeneity in the length of the higher plants' ITS-2s stems from the preservation of underlying secondary structure (see below).

The Chlorophyte ITS-2 Contains Few Long-Range Interactions and May Reflect a Dynamic Folding Process

Figure 1 illustrates the secondary structure of *Chlamydomonas reinhardtii* and *Eudorina unicocca*, ITS-2 foldings typical of most Volvocales, as well as the more unusual variants *Platydorina caudata* and *Volvox spermatosphaera*. Most of the ITS-2 appears to be a self-contained folding complex, usually consisting of four distinct hairpin loops. No evidence for long-range interactions could be detected, including tertiary interactions, save for the putative 3′-5.8S and 5′-LSU base pairing suggested in Fig. 1 that may also involve the first and last few nucleotides of the ITS-2. The concept of the ITS-2 as a self-contained domain supports experimental evidence which indicates that excision of the ITS-2 from the transcript is a process largely separate from that of ITS-1 removal. Such studies have demonstrated that the main cleavage within the ITS-2 can occur even if processing in the ITS-1 is abolished (Musters et al. 1990; van der Sande et al. 1992; Hadjiolova et al. 1994; van Nues et al. 1994, 1995). It is also worth mentioning that Hadjiolova et al. (1994) have noted the possibility that the 5.8S-28S association, which has been previously hypothesized to associate during rRNA primary transcript processing (Veldman et al. 1981; Hadjiolova et al. 1984b), *may not* be involved in ribosome biogenesis.

The ITS-2 secondary structural pattern observed in the chlorophytes appears to reflect a dynamic folding process as it is synthesized 5′→3′. The localized hairpin loops with limited long-range pairing interactions may illustrate a compaction of the ITS-2 portion of the primary rRNA transcript as has been proposed by Venkateswarlu and Nazar (1991). Previous research has suggested that kinetic processes play an important role in the formation of functional structures in RNA (Emerick and Woodson 1994; Zarrinkar and Williamson 1994; Gultyaev et al. 1995; Konings and Gutell 1995).

Evolutionary Conserved Sequence Tracts Are Related to Higher-Order Structure in the Volvocales

The set of Volvocalean ITS-2 sequences spans an entire order of green algae and includes sufficient information on the population, species, genus, and family levels to reveal evidence of higher-order structure. When aligning ITS-2 sequences solely by primary structure, certain subsequences of the Volvocalean ITS-2 appear to be highly conserved whereas other regions are highly variable, to the point that they cannot be reliably aligned among members of even the same species designation. However, the highly conserved regions of the ITS-2 are readily alignable across *all* of the Volvocales. These 116 ''universally alignable'' positions are boldfaced and given subsequence **A–I** designations in Fig. 1. Within the order Volvocales, over half of these positions possess conservation levels (C) with $C \geq 1.30$ (Fig. 2), which indicates that these sites share the same nucleotide in greater than 83% of the sequences examined.

Furthermore, the positions of these subsequences in the phylogenetically supported secondary structure of the Volvocales makes it obvious why they are evolutionarily conserved. As shown in Table 2, 85.3% of the universally alignable positions fall into regions that base pair. Moreover, certain highly conserved subsequences orient themselves such that they pair on opposing sides of helix. Pairings of these regions are strongly supported by evidence of covariations within the Volvocalean ITS-2 (Fig. 1). The selective pressure to preserve certain pairings within the ITS-2 implies that the formation of these helices plays an important biological role during some stage of ITS-2 processing.

As shown in Fig. 2, the consensus Volvocalean ITS-2 secondary-structure model is composed of four major helices, each of which is discussed in detail in the legend. All numbers will refer to the *C. reinhardtii* ITS-2 nucleotide positions unless otherwise noted.

When primary structural conservation data are superimposed on a map of the ITS-2 secondary structure, an interesting pattern begins to emerge which characterizes the central core of the ITS-2 (Fig. 2). Near the bases of the four domains, there is an unusually high level of sequence conservation present within the putative single-stranded regions and at the proximal regions of the helices. There is a degree of uncertainty about the absolute Volvocalean ITS-2 structural conformation near the bases of helices III and IV and between nucleotides (1–3)·(239–241) since supporting covariation evidence is lacking, so more of the ITS-2 than illustrated may be single-stranded. In any case, it is clear that this area of the ITS is highly conserved in primary structure and at least some of the positions are single-stranded.

Furthermore, the ''palm'' of the ''four-fingered hand'' is purine-rich in content, especially in adenine, as eight out of the 12 universally unpaired adenines fall within the vicinity. As Table 2 illustrates, 47.8% of universally alignable adenines fall into unpaired regions, compared to no unpaired guanines in the same areas. This situation is reminiscent of the findings in the SSU rRNAs in which universally conserved positions, which were usually found in single-stranded regions, contained adenine twice as frequently as G, C, or U (Gutell et al.

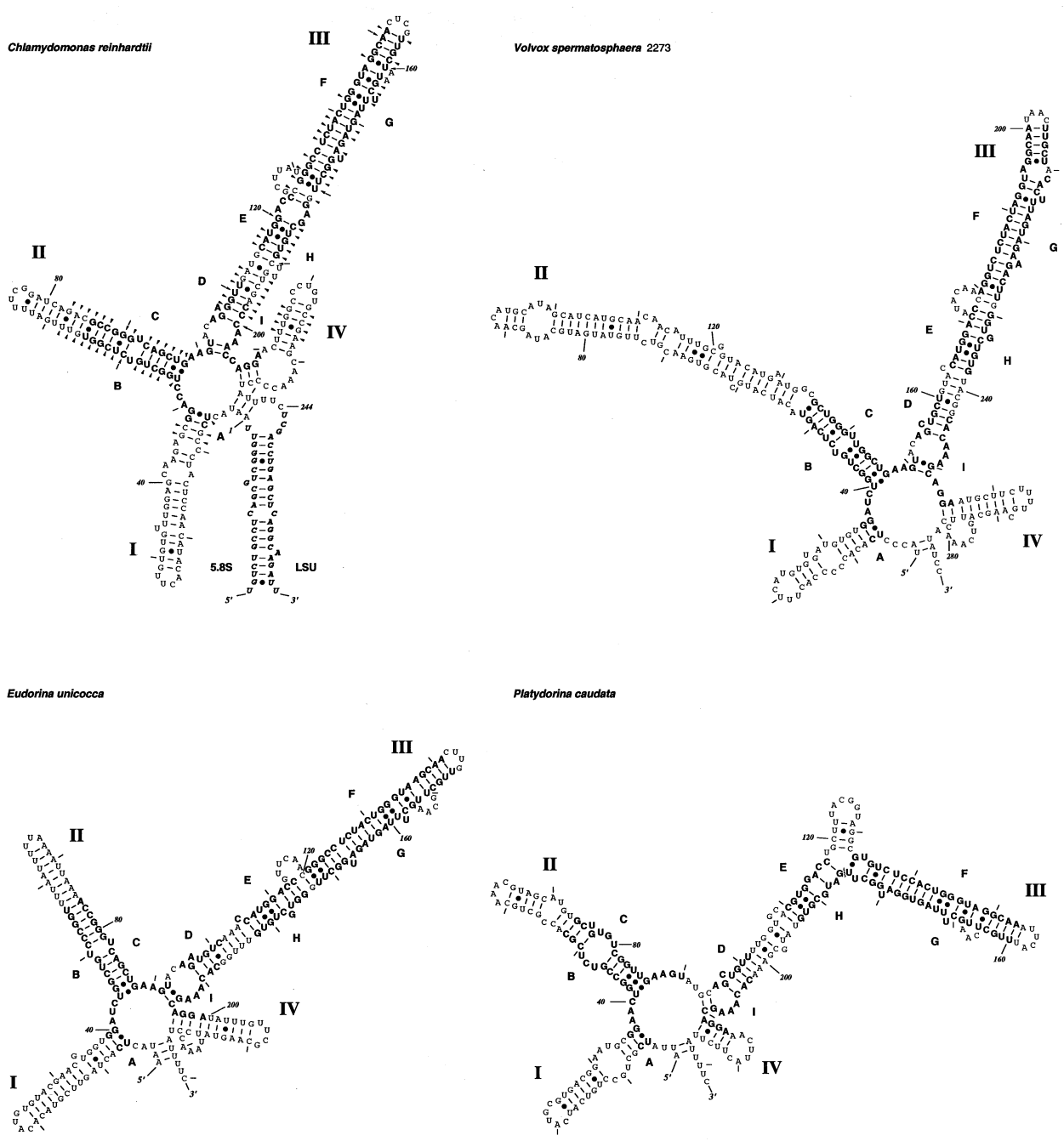

Fig. 1. Secondary structures of the ITS-2s for the organism designated as the primary model, *Chlamydomonas reinhardtii,* and representative Volvocaceae, *Volvox spermatosphaera* 2273, *Eudorina unicocca,* and *Platydorina caudata.* Structures are numbered every 40 nucleotides with *tic marks* at ten-nucleotide intervals. In all four diagrams, the 116 universally alignable nucleotide positions within the Volvocales are in *bold print.* The four major helices are labeled I–IV and the universally alignable subsequences are given corresponding letter designations. Placed on the *C. reinhardtii* diagram are *arrowheads* adjacent to nucleotide positions for which compensating base-pair changes found among various volvocales could be unambiguously mapped back to the ITS-2 of *C. reinhardtii.* Nucleotide positions with *arrowheads on opposing sides* of the helix all demonstrate positional covariations in at least one case (e.g., A:U→G:C) except 53:98, for which one-sided compensating changes (involving U·G pairs; U·G→C:G and U·G→U:A) occur on either side of the helix in separate lineages. Only in *C. reinhardtii* is the putative universal 3′-5.8S/5′-LSU interaction presented, *italicized and boldfaced.*

1985). The role that the adenine within single-stranded regions might play in ITS-2 processing is presently unclear.

The four universally alignable uracils that are ‘‘unpaired’’ in Table 2 *all* participate as conserved structural motifs, described earlier (noncanonical pair 59:92 and positions 166 and 175). It should be noted that no conserved A·G pairs have been detected in the Volvocalean

ITS-2, despite their presence in the large and small ribosomal RNA subunits (Gutell and Woese 1990; Gautheret et al. 1994; Gutell et al. 1994).

Preservation of Secondary Structural Motifs Is High Even in the Presence of Numerous Indels in the Volvocales

The ITS-2 of the Volvocales possesses a considerable level of length heterogeneity. However, the insertions and deletions that account for the large preponderance of variability in the ITS-2 do not seem to impede the formation of conserved structural elements.

As discussed above, nearly all of the major indels occur between the universally alignable subsequences and they pair in ways that still maintain the core structure. In fact, within the ITS-2 of a *C. reinhardtii* variant interfertile with the one included in this manuscript, the *only* difference between the two ITS-2 sequences is a ''clean'' and symmetrical 10-nt deletion at the tip of helix I that preserves the overall structure. Nearly all of the large insertions (>10 nt) are extensions of base-paired regions, similar to the observations made in the divergent domains of the LSU rRNAs (Michot et al. 1990).

The high toleration for indels within the Volvocalean ITS-2, particularly at ends of conserved helices, might help to explain not only how the ITS-2s of closely related organisms have managed to diverge rapidly but also how higher eukaryotic organisms (e.g., mammals) have evolved to have significantly longer ITS-2s than the ones observed here.

Types of Changes in Volvocalean ITS-2 Secondary Structure Reflect Differing Taxonomic Levels

Changes in Volvocalean ITS-2 primary and secondary structure vary at different levels in the classification hierarchy (Coleman and Mai, manuscript in preparation). Among organisms capable of intercrossing, the 116 conserved base positions in the ITS-2 are identical in the majority of cases; exceptions differ by at most three nucleotides, all of which are substitutions which occur in the single-stranded regions at the base of the helices or in the bulges of the hairpin stems. In rare cases, a single base substitution can be found in the last conserved pair (65:86) of helix II which still preserves the pairing (e.g., U·G→C:G). These observations have led to the recognition and subsequent confirmation of *Chlamydomonas incerta* as the nearest available relative of *C. reinhardtii* (Goodenough et al. 1996).

Above the biological species level, some species considered to be taxonomic sisters conserve the universally alignable subsequences but exhibit sequence variation in the stem-loops, particularly in helices I and IV. Other putative sister species display significant changes even in the 116 conserved positions, but these changes still preserve the basic secondary structure. Major differences in the variable regions, especially base-paired extensions of the conserved helices, coupled with compensating base-pair changes in the universally alignable subsequences, serve to define the genus level. Members of the same genus can usually be identified by virtue of shared secondary structural elements not part of the core structure. Finally, increased variability in the universally alignable subsequences typifies the family level and above.

ITS-2 Secondary Structure in Other Members of the Green Line of Evolution

The high level of conservation of specific primary and secondary structural elements within the Volvocales led us to extend our ITS-2 structural analysis to other taxa of the green line of evolution.

Macroalgal and Moss Spacers

First, the sequences of ten geographic isolates of the marine macroalga, *Acrosiphonia arcta,* were examined (van Oppen et al. 1993). *Acrosiphonia* was the most distant green alga from *Chlamydomonas* for which a set of complete ITS-2s illustrating positions of sequence variation was found.

These sequences were highly uniform, with nucleotide changes at only 11 positions, five of them leading to compensating base-pair changes. All the other differences from the sequence shown fell in terminal loops or conserved bulges. Consequently, the structure presented in Fig. 3 of *Acrosiphonia arcta* Roscoff is largely based on the phylogenetically derived Volvocalean model. The structure preserves the familiar four-helix domain, as well as the universally conserved pyrimidine mismatch in helix II and the GGU triplet 5′ to the apex of helix III. Furthermore, it is possible to assign homology of subsequences in *Acrosiphonia,* labeled **A, B, C, D,** and **I,** to those of the Volvocales with some degree of certainty. However, the subsequences marked **E, F, G,** and **H** have diverged to the point that they are no longer unambiguously alignable with their Volvocalean counterparts. It is notable that at this level of ITS-2 divergence, the highly conserved nucleotide bulge at position 175 in *C. reinhardtii* is no longer present, despite its conservation across the Volvocales. *Acrosiphonia* approaches the limits of where sequence homology to the Volvocalean ITS-2 can be directly assigned by comparison of primary structure. The ITS-2 of *Funaria hygrometrica,* a moss, was also examined, which led to the detection of a putative homologous region to the Volvocalean helix II (Fig. 4). Nevertheless, uncertainty about the 3′ end of the moss ITS-2 prevented identification of structure beyond this second stem-loop.

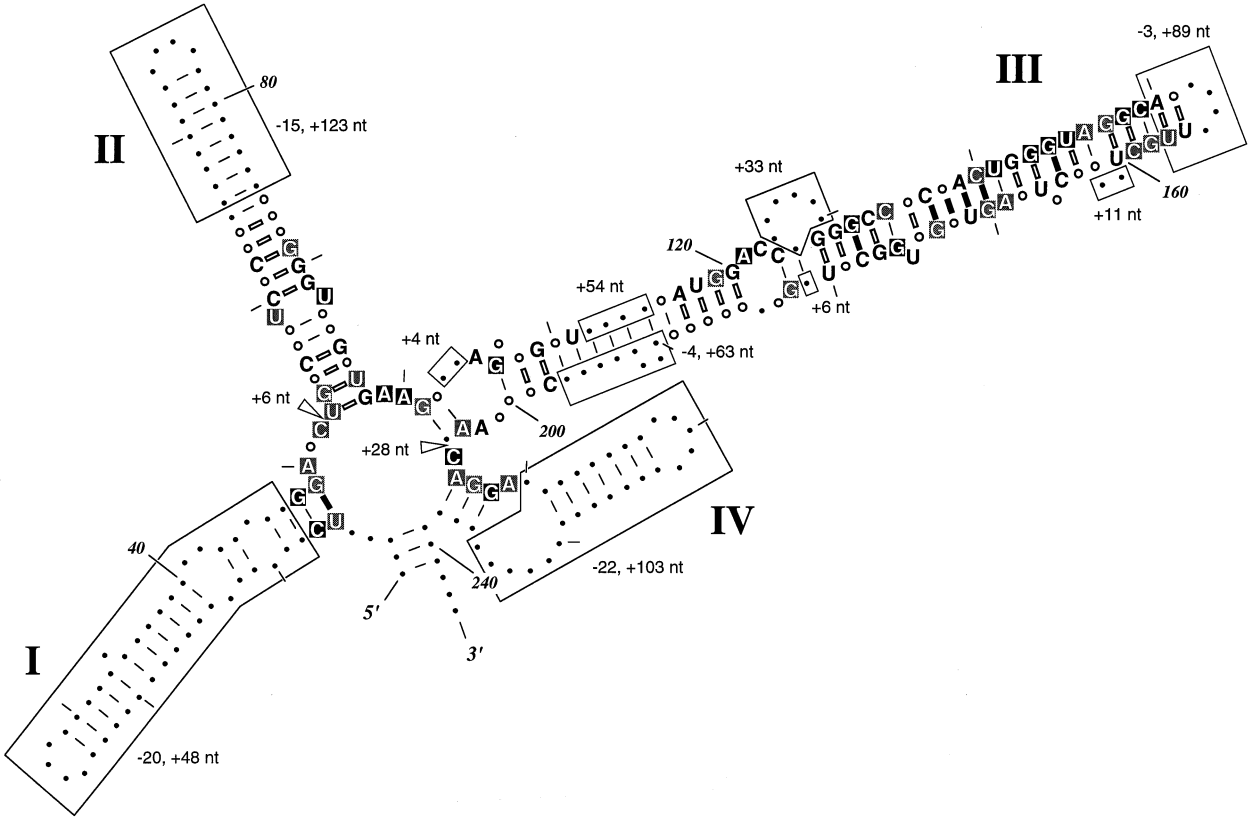

Fig. 2. Conserved volvocalean core ITS-2 superimposed on the secondary structure of *C. reinhardtii.* The structure is numbered every 40 nucleotides with *tic marks* every ten nucleotides. Nucleotides in *white lettering on a black background,* e.g., position 100, are absolutely conserved in all volvocales examined here ($C = 2.00$); nucleotides in *white lettering on a gray background* indicate ''very high'' conservation (two exceptions, at most, out of all 111 organisms; $2.00 > C \geq 1.83$), e.g., at position 203 where $C = 1.87$, A—98%, C—2%; *unboxed letters* indicate positions of ''high'' conservation ($1.83 > C \geq 1.30$), e.g., at position 118 where $C = 1.32$, U—86%, A—10%, *C*—4%; *open circles* indicate ''moderate'' conservation ($1.30 > C \geq 0.30$), e.g., at position 86 where $C = 0.31$, G—59%, A—15%, *C*—15%, U—10%, deletion—1%; *solid circles* indicate ''variable'' positions ($C < 0.30$). Nucleotides shown are chosen by majority rule at each position considered. The Conservation Value (*C*) at a given position for all 111 ITS-2 sequences was calculated using the following expression from Gutell et al. (1985):

$$C = \Sigma P_i \log_2 (4P_i) + P_\Delta \log_2 (P_\Delta)$$

where P_i = frequency of occurrence of base *i* at a given position and P_Δ = frequency of deletions occurring at that position. Within the universally alignable regions of the ITS-2, phylogenetically supported pairings that always have the potential to pair are denoted with *solid black bars.* Phylogenetically supported pairings with only a few exceptions in the volvocales (≤9 out of 111) are denoted with *open bars.* Approximate locations of the volvocalean indels of greater than two nucleotides are *boxed* or denoted by *hollow arrowheads* and the ranges of length variation are indicated with respect to the *C. reinhardtii* ITS-2. *Helix I.* Helix I runs from (6–24)·(29–49) and can be highly variable in both length and sequence from one organism to another. In *C. reinhardtii,* the domain has a $\delta G = -21.1$ kcal/mol for its 44 nt. Helix I variants still maintain fairly stable hairpin loops which preserve association of the basal 6:49 and 7:48 pairings, conserved throughout the volvocales. Although some of the covariations within helix I can be unambiguously mapped back to *C. reinhardtii* from other volvocales, the majority of CBCs detected are shared only within the same genus or species. *Helix II.* The paired nucleotides (53–73)·(78–98) form the region of the ITS-2 designated as helix II. Extensive covariation characterizes this part of the ITS-2, with very few insertions/deletions (indels) found within the base-pairing areas of subsequences **B** and **C** that lie within the domain. The two **G·U** pairs at the base of the helix are highly conserved throughout the volvocales. Most striking is the universal **U·U** juxtaposition (a **C·U** pairing in just two of the 111 volvocaleans) at position 59:92. The preservation of this motif is remarkable in the light of the fairly high level of nucleotide changes that occur in the surrounding sequences, suggesting an important function for this structure. Although the volvocalean core ITS-2 structure shows that helix II possesses few residues with $C \geq 1.83$, conservation of *pairing* is high since most of the nucleotide changes still result in pairings. Beyond the limits of the conserved subsequences **B** and **C,** toward the apex of the helix, substantial variability is present, but in a manner that preserves the base pairing of the conserved sequences. *Helix III.* Helix III, (101–151)·(156–204), is generally the longest helix in the volvocalean ITS-2 and, interestingly, exhibits in region **F** near the tip of its loop the highest degree of primary and secondary structural conservation within the ITS-2. The δG for the 104 residues that make up this domain in *C. reinhardtii* is approximately −43 kcal/mol. The two basal pairings (101–102)·(203–204) are not universally paired but have the potential to pair in most of the volvocales. Much of helix III, especially near the end of the loop, maintains base pairing across the volvocales such that the general secondary structure is preserved. The region of the helix from (137–140)·(170–173) is conserved to the extent that pairing potential is maintained for all 111 volvocales. Also, the longest absolutely conserved nucleotide sequence within the volvocales is observed near the apex of helix III, a GGU triplet from positions 143–145. Curiously, there is a higher level of conservation of primary structure on the 5′ side of the arm than on the 3′ side. This imbalance is mainly accounted for by a greater number of base changes on the 3′ side which are still compatible with base pairing (one-sided CBCs involving **G·U** pairings); however, the reasons underlying this bias are unknown. The helix contains several conserved motifs interrupting the 3′ side of the

Table 2. Nucleotide composition and distribution in the entire *C. reinhardtii* ITS-2 sequence (upper portion) and in the 116 universally alignable positions across Volvocales (lower portion)[a]

Base	Number	Percentage	Base paired	Unpaired	Percentage unpaired
			All *C. reinhardtii* positions		
A	51	20.9	28	23	45.1
C	61	25.0	42	19	31.1
G	66	27.0	61	5	7.6
U	66	27.0	47	19	28.8
Total	244		178	66	
	52% G+C		73.0%	27.0%	
			Universally alignable positions[b]		
A	23	19.8	12	11	47.8
C	26	22.4	24	2	7.7
G	38	32.8	38	0	0.0
U	29	25.0	25	4	13.8
Total	116		99	17	
	55.2% G+C		85.3%	14.7%	

[a] Universally alignable positions are the highly conserved subsequences marked in boldface in Fig. 1
[b] Nucleotide used in calculation at given a position chosen by majority rule

The Angiosperm ITS-2

In only a few cases were sufficient complete sequences available, reflecting the appropriate level of relatedness, to permit use of comparative modeling in other chlorophyte ITS-2s. These instances were confined to the flowering plants, where the majority of the plant ITS sequencing has been done to assess phylogenetic relationships. Sequences representing both monocots and dicots were obtained in order to survey overall features of the angiosperm ITS-2.

A set of closely related angiosperm ITS-2 was required first in order to obtain a more highly resolved secondary structure which could then be compared to other flowering plants. The sequences of 23 Rosaceae (Campbell et al. 1995), all but four belonging to the subfamily Maloideae, were chosen because of their especially high level of ITS-2 pairwise sequence divergence in comparison to other sequenced plant families (Baldwin et al. 1995).

As shown in Fig. 3, the representative folding for the Rosaceae, *Photinia pyrifolia*, possesses the conserved four-helix domain common to the green algae. Helix II contains the conserved pyrimidine mismatch pairing (two sequential pairs in the Rosaceae) and helix III contains the universally conserved GGU (positions 127–130 in *Photinia*) near its apex on the 5′ side. Helix III, which is generally the longest helix in the Volvocales, is also the longest helix within the rose family. Nucleotide conservation was rather high across the family in comparison the Volvocaceae; however, sufficient numbers of CBCs could be detected to substantiate portions of the representative folding (Fig. 3).

When the simple procedure of measuring the number of nucleotide changes at a given position across the 23 sequences is applied (a simple form of conservation analysis), a pattern of variable sequence conservation in the ITS-2 becomes apparent. The spatial positioning, lev-

←

arm. Between residues 160 and 163 lies an area that forms a small bulge (usually one-sided) present in most organisms. It can range in size from 1 nt in some unicells to 13 nt in *Pleodorina californica.* A few bases away, toward the base of the helix, there is usually an unpaired nucleotide on the 3′ side (position 166; usually uracil). However, the most prominent and highly conserved motif is an unpaired nucleotide that is almost always uracil, but never guanine at position 175 ($C = 1.44$). The presence of the bulged nucleotide across all of the volvocales suggests that it might represent an important signal needed for excision of the spacer. This helix also contains a number of variable regions that lie between its conserved subsequences. The nucleotides from 124–130 in *C. reinhardtii* form a one-sided bulge, which, when extended in *Volvox aureus* and *Platydorina caudata,* can base pair to form a ''sidearm'' that juts out between the conserved subsequences E and F (Fig. 1). Also, at the tip of helix III, insertions up to +89 nt compared to *C. reinhardtii* have been observed which seem to associate stably into base-paired extensions of the helix that do not interfere with the conserved pairings. *Helix IV.* Helix IV is a short region from (206–217)·(223–238) which has the highest variability of all of the ITS-2 helices. Its high divergence rate, even among sequences from organisms classified as the same species (e.g., the 16 *Pandorina morum* sequences), prevents most of the positional covariances from being mapped back to *C. reinhardtii.* Nevertheless, there is sufficient compensating base-change evidence available to be confident of its existence in most organisms. Despite the high divergence characterizing most of this ITS-2 substructure, residues 206–209 are very conserved ($C \geq 1.83$). Although they are shown as paired in Fig. 1 due to minimum free-energy calculations, nucleotides 206–208 may actually be single-stranded. (The evidence for association is largely ''neutral''; many potential pairing partners are present, but there are no certain CBCs to confirm pairing.) In *C. reinhardtii,* the helix has a theoretical minimum free energy of −10.9 kcal/mol for its 33 nt. In other taxa, helix IV can be considerably longer and can generate a more stable stem-loop. With respect to *C. reinhardtii, Volvox carteri f. weismannia* 1874 has an insertion of 103 nt, which associates into an extended helix; however, *Chlamydomonas culleus* has a deletion of 22 nt that nearly removes the entire substructure.

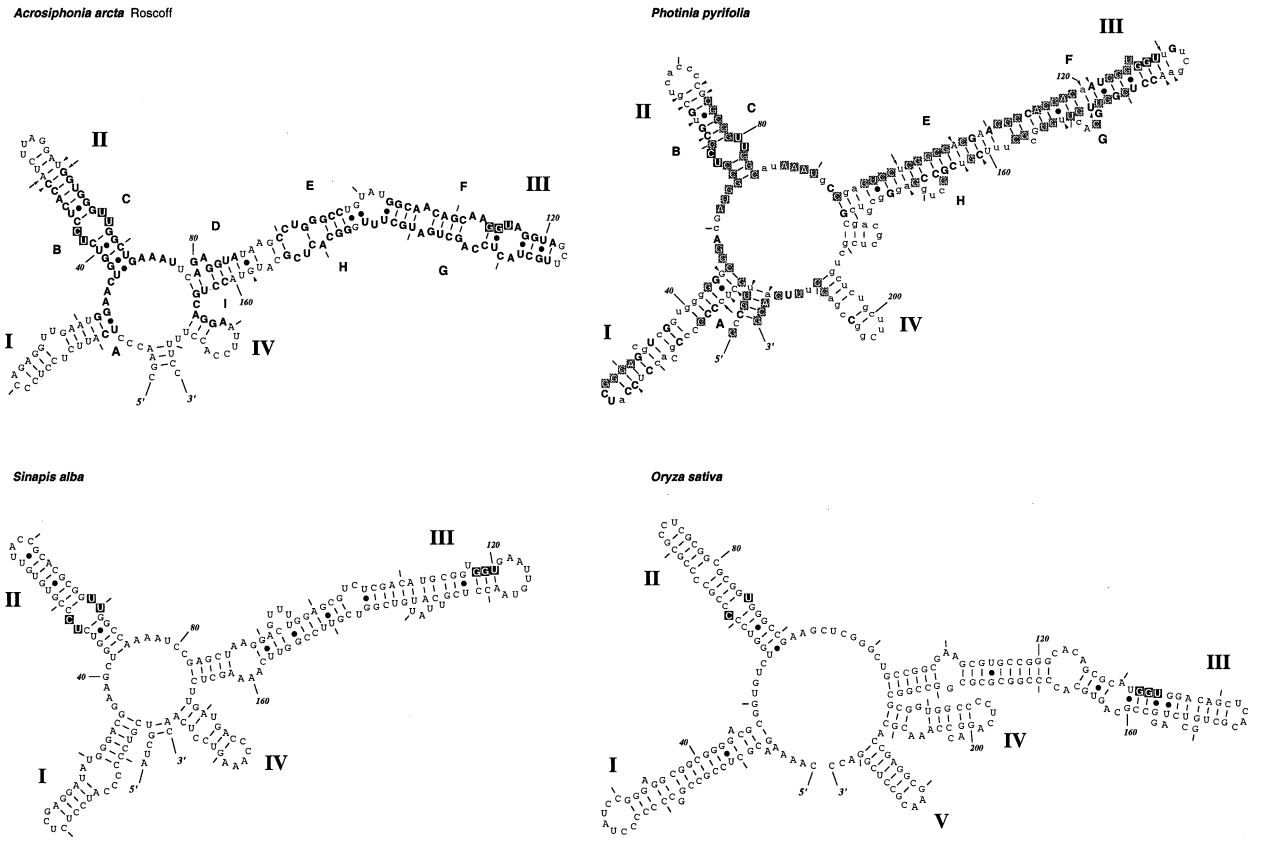

Fig. 3. Secondary structure of the ITS-2 region of selected nonvolvocalean green organisms; the marine macroalga, *Acrosiphonia arcta* Roscoff, and three flowering plants, *Photinia pyrifolia, Sinapis alba,* and *Oryza sativa.* Structures are numbered every 40 nucleotides with *tic marks* at ten-nucleotide intervals. Helices are labeled according to assumed homology to the volvocalean model. (Note that *Oryza* can have two helices in the position typically occupied by helix IV.) All four diagrams illustrate the universally conserved pyrimidine–pyrimidine pairing(s) in helix II and the conserved trinucleotide motif in helix III. (These conserved nucleotides are marked with *white lettering on a black background.*) The regions in *Acrosiphonia* of putative subsequence homology to the volvocales are indicated in *boldface and labeled.* Similarly, subsequence regions in *Photinia* possibly homologous to the conserved volvocalean subsequences are denoted with the corresponding *letter designations.* Additionally, in *Photinia,* family conservation of nucleotides is represented on its ITS-2 structure; nucleotides in *white lettering on a gray background* have no changes at that position among the 23 Rosaceae sequences examined; nucleotides in *bold, uppercase lettering* have one change at that position out of the 23 sequences; nucleotides in *just uppercase lettering* have two exceptions; and nucleotides in *lowercase* have three or more exceptions at that position. *Arrowheads* on *Acrosiphonia arcta* indicate the positions where CBCs were found among the ten geographic isolates, and *arrowheads* on *Photinia pyrifolia* indicate CBCs among the 23 Rosaceae sequences. The ITS-2s of *Sinapis* and *Oryza* are based upon minimum free-energy calculations. Of the two, only *Sinapis* has been modified to illustrate the potential pairing between its 5′ and 3′ ends.

els of conservation, and pairing of the conserved subsequences mirror the core Volvocalean ITS-2 closely enough to be assigned putative homologous subsequence designations in Fig. 3. Subsequences **D** and **I,** which lie at the base of helix III in the Volvocales, could not be assigned with confidence to the rose family ITS-2.

As in the Volvocales, helices II and III exhibit the least variation among the taxa. Helix I is more variable and helix IV possesses the highest level of variability, with an additional small helix appearing between helices III and IV in a few species. A further resemblance to the Volvocalean ITS-2 is observed, with the 5′ side of helix III of the Rosaceae exhibiting a higher degree of conservation than the 3′ side. The final parallelism between the two is that the sequences connecting the four helices have the potential to be single-stranded and are rich in conserved adenine residues.

The close correspondence in features of the Rosacean and Volvocalean ITS-2s does not arise from similarities in nucleotide sequences. Even with the aid of alignment by secondary structure, regions of sequence similarity are not obvious (except for the conserved GGU motif). Thus, it should be assumed that assignments of homology across such large evolutionary distances should be taken as provisional.

The conserved four-helix domain motif also consistently reemerged across a wide spectrum of other angiosperm ITS-2s from thermodynamic predictions. Secondary structures of *Sinapis* and *Oryza*, in addition to *Photinia*, are shown in Fig. 3 to illustrate the general structure found in the angiosperms. The ITS-2s of some

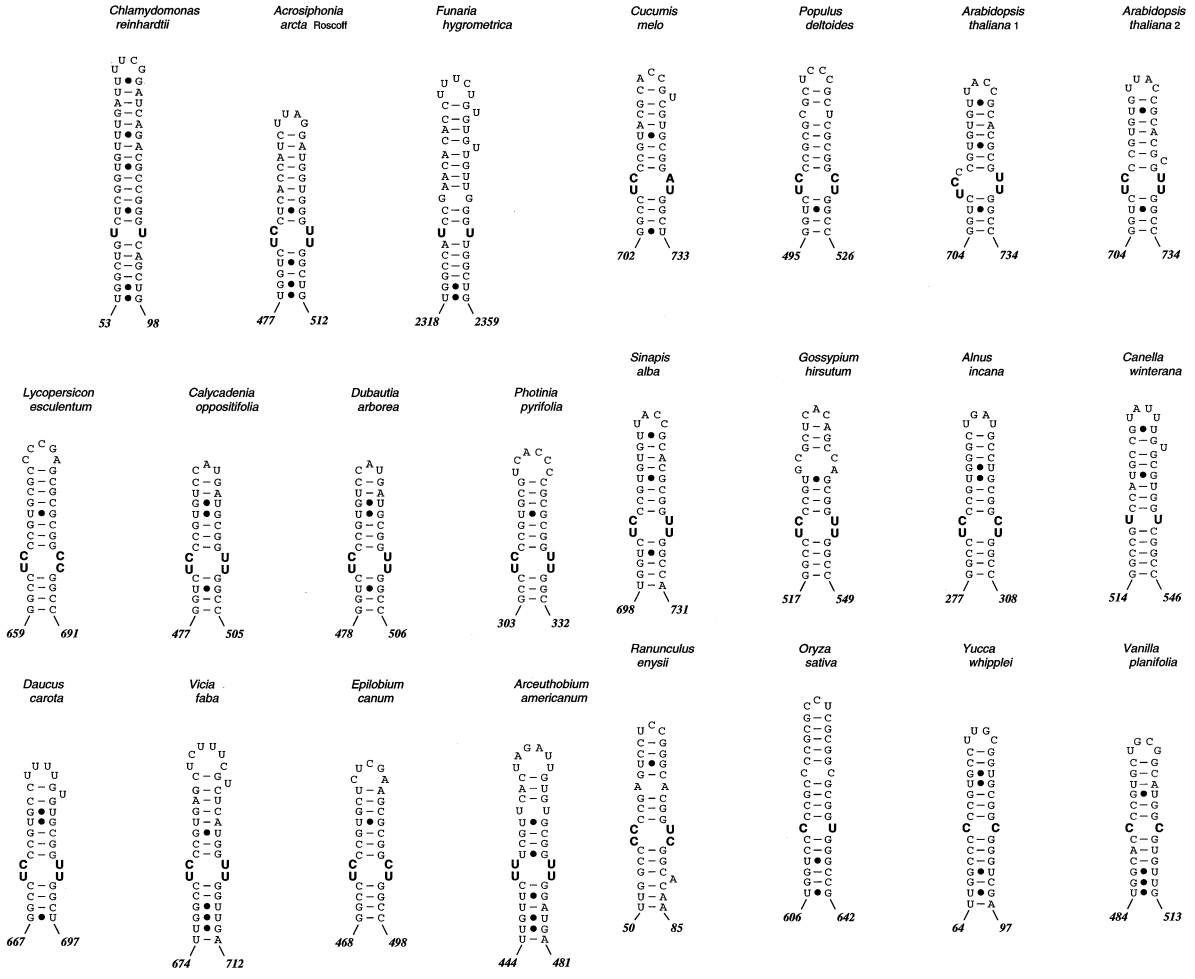

Fig. 4. Structure of helix II in representative green algae, a moss, and flowering plants. The universally conserved pyrimidine–pyrimidine pairing(s) is marked in *boldface. Numbers* refer to the position of the 5′ and 3′ termini of the helix according to Fig. 1 for *C. reinhardtii* or according to the sequence deposited in GenBank. Of the angiosperm sequences examined to date, only in *Cucumis* is a purine, adenine, present in the conserved internal bulge. (*Cucumis melo* has one adenine in the bulge and its relative, *Cucumis sativus,* has two adenines in the bulge.) In addition, two alternative foldings for *Arabidopsis thaliana* are shown. There is a deletion in a normally conserved region of the helix in *Arabidopsis* which causes the bulge to become asymmetric, as shown in helix 1 (leftmost). However, a potential alternative pairing is possible, helix 2 (rightmost), in which a cytosine (727) is unpaired on the 3′ side of the stem-loop, leading to the formation of the symmetrical pyrimidine–pyrimidine twin pairings observed in most dicots.

angiosperms had the potential to form three to six helices, although the equivalent of Volvocalean helices I, II, and III were always present. In spite of high variability at the interfamilial level, the two most conserved features of the Volvocalean ITS-2, the universally conserved pyrimidine mismatch (double pyrimidine–pyrimidine pairings in many cases) in helix II (Fig. 4) and the GGU triplet on the 5’ apex of helix III (Table 3) were found in all of the angiosperms studied. The conserved UGGU motif was previously noted in the ITS-2 of angiosperms (Liu and Schardl 1994), but was not linked to secondary structure. Once again, helices II and III are the most highly conserved helices across the angiosperms, as in the Volvocales, with helix III always the longest stem-loop in the species studied.

Previous searches for secondary structure in the flowering plants have anticipated some of the observations presented here. The putative secondary structure of *Calycadenia oppositifolia* shown in Baldwin et al. (1995) predicts helices I, II, and III that are nearly identical to our own estimates. This seems to be due to the use of the program FOLD, which uses the same algorithm that is employed in Mulfold and mFOLD (Jaeger et al. 1989, 1990; Zuker 1989). The Zuker algorithm is able to predict the foldings of the ITS-2s rather well. Also, by use of maximum base-pairing methods, Venkateswarlu and Nazar were able to suggest a folding for the *Nicotiana rustica* ITS-2 which recognized stem-loops comparable to the termini of our helices II, III, and IV (1991). However, in neither case was evidence uncovered for covariations to support the structural estimates of the angiosperm ITS-2s.

Table 3. Nucleotides common to the green algae flowering plants examined[a]

Chlamydomonas reinhardtii	CUACUG**GGU**AG
Acrosiphonia arcta Roscoff	CAGCAA**GGU**AG
Lycopersicon esculentum	AGUGGU**GGU**UG
Calycadenia oppositifolia	AGUGGU**GGU**UG
Dubautia arborea	UGUGGU**GGU**UG
Photinia pyrifolia	AUCGGU**GGU**UG
Daucus carota	AUCGGU**GGU**UG
Vicia faba	GAUGGU**GGU**UG
Epilobium canum	CGCGGU**GGU**UG
Arceuthobium americanum	UGUUUU**GGU**UA
Cucumis melo	UACGGU**GGU**UG
Populus deltoides	AGCGGU**GGU**UG
Arabidopsis thaliana	UGCGGU**GGU**GA
Sinapis alba	UGCGGU**GGU**GA
Gossypium hirsutum	AUCGGU**GGU**AA
Alnus incana	AUCGGU**GGU**UG
Canella winterana	AGUGGU**GGU**GG
Ranunculus enysii	AGCGGU**GGU**UG
Oryza sativa	GCGCAU**GGU**GG
Yucca whipplei	GCGAGU**GGU**GG
Vanilla planifolia	UGUUGU**GGU**GU

[a] This 11-nucleotide sequence is found on the 5′ side of helix III. The **GGU** triplet is universally conserved; note increased level of divergence in flanking sequences. The **UGGU** motif is preserved across all flowering plants.

Conclusions and Prospects

The conservation of primary sequence elements and of basic secondary structure points toward a significant functional role for the chlorophyte ITS-2 during rRNA primary transcript processing. The region on the 5′ side near the apex of helix III, with its sequence and position conserved across the green line of evolution, would seem to be a possible candidate for recognition by some element of the nucleolar processing machinery. Additionally, the high evolutionary conservation of pyrimidine–pyrimidine pairings in helix II, and perhaps the conserved unpaired single nucleotides within the Volvocalean ITS-2s, indicate that subtle secondary structural motifs also participate in the ITS-2 excision process, spatially positioned by the conserved framework of helices.

The close correlations in ITS-2 conservation patterns and structure between the algae and the flowering plants may reflect the similarity of processing requirements and mechanisms retained since the divergence of the angiosperms from their aquatic relatives. Given the surprising conservation of overall features of the ITS-2 despite the large evolutionary distance separating the angiosperms and green algae, it is possible that most of the green line of evolution, the Chlorobionta (Kranz et al. 1995), share similar ITS-2 conformations and processing mechanisms for the removal of the spacer from the primary transcript. Further study of ITS-2s from other green algae, bryophytes, and tracheophytes will help to provide evidence to support or refute this notion.

The common Volvocalean ITS-2 core structure suggests areas of the ITS-2 to focus upon for future studies of rRNA transcript processing in eukaryotes. Since the structure that we have proposed for the ITS-2 does not closely resemble that of yeast, further investigations should help to determine which, if any, of the various domains found in yeast and the Volvocales/flowering plants are indeed homologous and whether they are processed by the same mechanisms. ITS-2s of the various major eukaryotic lines of evolution may have evolved independently, to the point where processing motifs and mechanisms differ. Even if this is the case, the unexpected uniformity of ITS-2 structure in widely separated organisms such as green unicells and flowering plants raises the hope that other series of sequences can be used to resolve ITS-2 secondary structure both within and across major phylogenetic lines.

The higher-order structure of the ITS-2 also has important implications for phylogenetic analysis. The ITS-2 is commonly used for phylogenetic inference, so knowledge of the structural constraints acting upon this spacer is crucial because of its direct effect on both the alignments and the weightings used in phylogenetic analyses (Hillis and Dixon 1991; Dixon and Hillis 1993). Variation in distal positions of the hairpins can be expected even among interbreeding organisms. Further investigations into this matter should help to define clearly the secondary structural constraints shaping the sequence of this spacer region. Even without considerations of weighting, the use of higher-order structures should at least serve to improve the accuracy of alignments of sequences that include the ITS-2, and consequently, their phylogenetic utility. Undoubtedly, more remains to be learned concerning our understanding of the details of an RNA structure, which appears to be subject to remarkably fine selective pressure in this nongenic region.

References

Bakker FT, Olsen JL, Stam WT, van den Hoek C (1992) Nuclear ribosomal DNA internal transcribed spacer regions (ITS 1 and ITS 2) define discrete biogeographic groups in *Cladophora albida* (Chlorophyta). J Phycol 28:839–845

Bakker FT, Olsen JL, Stam WT (1995) Evolution of nuclear rDNA ITS sequences in the *Cladophora albida/sericea* clade (Chlorophyta). J Mol Evol 40:640–651

Baldwin B, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann MO Bot Gard 82:247–277

Brown JW (1991) Phylogenetic comparative analysis of RNA structure on Macintosh computers. Comput Appl Biosci 7:391–393

Campbell CS, Donoghue MJ, Baldwin BG, Wojciechowski MF (1995) Phylogenetic relationships in Maloideae (Rosaceae): evidence from sequences of the internal transcribed spacers of nuclear ribosomal DNA and its congruence with morphology. Am J Bot 82:903–918

Coleman AW, Suarez A, Goff LJ (1994) Molecular delineation of species and syngens in volvocacean green algae (Chlorophyta). J Phycol 30:80–90

Dixon MT, Hillis DM (1993) Ribosomal RNA secondary structure:

compensating mutations and implications for phylogenetic analysis. Mol Biol Evol 10:256–267

Emerick VL, Woodson SA (1994) Fingerprinting the folding of a group I precursor RNA. Proc Nat Acad Sci USA 91:9675–9679

Gautheret D, Konings D, Gutell RR (1994) A major family of motifs involving G·A mismatches in ribosomal RNA. J Mol Biol 242:1–8

Gilbert DG (1992) *loopDloop*, a Macintosh program for visualizing RNA secondary structure. Published electronically on the Internet, available via anonymous ftp to ftp.bio.indiana.edu

Goff LJ, Moon DA (1993) PCR amplification of nuclear and plastid genes from algal herbarium specimens and algal spores. J Phycol 29:381–384

Gonzalez IL, Chambers C, Gorski JL, Stambolian D, Schmickel RD, Sylvester JE (1990) Sequence and structure correlations of human ribosomal transcribed spacers. J Mol Biol 212:27–35

Goodenough U, Ferris P, Woessner J, Armbrust EV (1996) Genes of the mating type locus in *Chlamydomonas*. J Phycol 32:s19

Gultyaev AP, van Batenburg FHD, Pleij CWA (1995) The computer simulation of RNA folding pathways using a genetic algorithm. J Mol Biol 250:37–51

Gutell RR, Woese CR (1990) Higher order structural elements in ribosomal RNAs: pseudo-knots and the use of noncanonical pairs. Proc Natl Acad Sci USA 87:663–667

Gutell RR, Weiser B, Woese CR, Noller HF (1985) Comparative anatomy of 16-S-like ribosomal RNA. Prog Nucleic Acid Res Mol Biol 32:155–215

Gutell RR, Power A, Hertz GZ, Putz EJ, Stormo GD (1992) Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. Nucleic Acids Res 20:5785–5795

Gutell RR, Larsen N, Woese CR (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiol Rev 58:10–26

Hadjiolova KV, Georgiev OI, Hadjiolov AA (1984a) Excess 5′-terminal sequences in the rat nucleolar 28S ribosomal RNA. Exp Cell Res 153:266–269

Hadjiolova KV, Georgiev OI, Nosikov VV, Hadjiolov AA (1984b) Localization and structure of endonuclease cleavage sites involved in the processing of the rat 32S precursor to ribosomal RNA. Biochem J 220:105–116

Hadjiolova KV, Normann A, Cavaillé J, Soupéne E, Mazan S, Hadjiolov AA, Bachellerie J-P (1994) Processing of truncated mouse or human rRNA transcribed from ribosomal minigenes transfected into mouse cells. Mol Cell Biol 14:4044–4056

Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. Q Rev Biol 66:411–453

Jaeger JA, Turner DH, Zuker M (1989) Improved predictions of secondary structures for RNA. Proc Natl Acad Sci USA 86:7706–7710

Jaeger JA, Turner DH, Zuker M (1990) Predicting optimal and suboptimal secondary structure for RNA. Methods Enzymol 183:281–306

Konings DAM, Gutell RR (1995) A comparison of thermodynamic foldings with comparatively derived structures of 16S and 16S-like rRNAs. RNA 1:559–574

Kranz HD, Miks D, Siegler M-L, Capesius I, Sensen CW, Huss VAR (1995) The origin of land plants: phylogenetic relationships among Charophytes, Bryophytes, and vascular plants inferred from complete small-subunit ribosomal RNA gene sequences. J Mol Evol 41:74–84

Kupriyanova NS, Timofeeva MY (1987) Primary structure of the internal transcribable spacer of ribosomal genes of the loach *Misgurnus fossilis* L. Similarity in the organization of the major processing site in eukaryotes. Mol Biol 21:830–838

Kwon O-Y, Ishikawa H (1992) Nucleotide sequence and presumed secondary structure of the internal transcribed spacers of rDNA of the pea aphid, *Acyrthosiphon pisum*. Comp Biochem Physiol 103: 651–655

Liu J-S, Schardl CL (1994) A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. Plant Mol Biol 26:775–778

Michot B, Bachellerie J-P, Raynal F (1983) Structure of the mouse rRNA precursors. Complete sequence and potential folding of the spacer regions between 18S and 28S rRNA. Nucleic Acids Res 11:3375–3391

Michot B, Hassouna N, Bachellerie J-P (1984) Secondary structure of mouse 28S rRNA and general model for the folding of the large rRNA in eukaryotes. Nucleic Acids Res 12:4259–4279

Michot B, Qu LH, Bachellerie JP (1990) Evolution of large-subunit rRNA structure: the diversification of divergent D3 domain among major phylogenetic groups. Eur J Biochem 188:219–229

Michot B, Despres L, Bonhomme F, Bachellerie J-P (1993) Conserved secondary structures in the ITS2 of trematode pre-rRNA. FEBS Lett 316:247–252

Musters W, Boon K, van der Sande CAFM, van Heerikhuizen H, Planta RJ (1990) Functional analysis of transcribed spacers of yeast ribosomal DNA. EMBO J 9:3989–3996

Nashimoto M, Ogata K, Mishima Y (1988) *In vitro* sequence-specific cleavage in transcribed spacer of mouse precursor ribosomal RNA. J Biochem 103:992–997

Nazar RN, Wong WM, Abrahamson JLA (1987) Nucleotide sequence of the 18–25 S ribosomal RNA intergenic region from a thermophile, *Thermomyces lanuginosus*. J Biol Chem 262:7523–7527

Noller HF, Woese CR (1981) Secondary structure of 16S ribosomal RNA. Science 212:403–411

Pace NR, Smith DK, Olsen GJ, James BD (1989) Phylogenetic comparative analysis and the secondary structure of ribonuclease P RNA—a review. Gene 82:65–75

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schlötterer C, Hauser M-T, von Haeseler A, Tautz D (1994) Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. Mol Biol Evol 11:513–522

Starr RC, Zeikus JA (1993) UTEX—the culture collection of algae at the University of Texas at Austin. J Phycol 29 (Suppl):1–106

Thompson AJ, Herrin DL (1994) A chloroplast group I intron undergoes the first step of reverse splicing into host cytoplasmic 5.8 S rRNA. J Mol Biol 236:455–468

van der Sande CAFM, Kwa M, van Nues RW, van Heerikhuizen H, Raué HA, Planta RJ (1992) Functional analysis of Internal Transcribed Spacer 2 of *Saccharomyces cerevisiae* ribosomal DNA. J Mol Biol 223:899–910

van Nues RW, Rientjes JMJ, van der Sande CAFM, Zerp SF, Sluiter C, Venema J, Planta RJ, Raué HA (1994) Separate structural elements within Internal Transcribed Spacer 1 of *Saccharomyces cerevisiae* precursor ribosomal RNA direct the formation of 17S and 26S rRNA. Nucleic Acids Res 22:912–919

van Nues RW, Rientjes JMJ, Morré SA, Mollee E, Planta RJ, Venema J, Raué HA (1995) Evolutionarily conserved structural elements are critical for processing of Internal Transcribed Spacer 2 from *Saccharomyces cerevisiae* precursor ribosomal RNA. J Mol Biol 250: 24–36

van Oppen MJH, Olsen JL, Stam WT, van den Hoek C, Wiencke C (1993) Arctic-antarctic disjunctions in the benthic seaweeds *Acrosiphonia arcta* (Chlorophyta) and *Desmarestia viridis/willii* (Phaeophyta) are of recent origin. Marine Biol 115:381–386

Vance VB, Thompson EA, Bownman LH (1985) Transfection of mouse ribosomal DNA into rat cells: faithful transcription and processing. Nucleic Acids Res 13:7499–7513

Veldman GM, Klootwijk J, van Heerikhuizen H, Planta RJ (1981) The nucleotide sequence of the intergenic region between the 5.8S and 26S rRNA genes of the yeast ribosomal RNA operon. Possible implications for the interaction between 5.8S and 26S rRNA and the processing of the primary transcript. Nucleic Acids Res 9:4847–4862

Venkateswarlu K, Nazar R (1991) A conserved core structure in the 18–25S rRNA intergenic region from tobacco, *Nicotiana rustica*. Plant Mol Biol 17:189–194

Wesson DM, Porter CH, Collins FH (1992) Sequence and secondary structure comparisons of ITS rDNA in mosquitoes (Diptera: Culicidae). Mol Phyl Evol 1:253–269

Yeh L-CC, Lee JC (1990) Structural analysis of the internal transcribed spacer 2 of the precursor ribosomal RNA from *Saccharomyces cerevisiae*. J Mol Biol 211:699–712

Zarrinkar PP, Williamson JR (1994) Kinetic intermediates in RNA folding. Science 265:918–924

Zuker M (1989) On finding all suboptimal foldings of an RNA molecule. Science 244:48–52