J Mol Evol (1995) 40:428–442

JOURNAL OF MOLECULAR EVOLUTION

Phylogenetic Relationship of the Green Alga *Nanochlorum eukaryotum* Deduced from Its Chloroplast rRNA Sequences

M. Schreiner,[1] M. Geisert,[1] M. Oed,[1] J. Arendes,[1] U. Güngerich,[1] H.-J. Breter[1], K. Stüber,[2] D. Weinblum[1]

[1] Institut für Physiologische Chemie und Pathobiochemie, Johannes Gutenberg-Universität, Duesbergweg 6, 55099 Mainz, Germany
[2] Max-Planck-Institut für Züchtungsforschung, 50933 Köln, Germany

Received: 6 January 1994 / Accepted: 24 May 1994

Abstract. The marine green coccoidal alga *Nanochlorum eukaryotum* (*N.e.*) is of small size with an average diameter of 1.5 μm. It is characterized by primitive-appearing biochemical and morphological properties, which are considerably different from those of other green algae. Thus, it has been proposed that *N.e.* may be an early developed algal form. To prove this hypothesis, DNA of *N.e.* was isolated by a phenol extraction procedure, and the chloroplast DNA separated by preparative CsCl density-gradient centrifugation. The kinetic complexity of the nuclear and of the chloroplast DNA was evaluated by reassociation kinetics to 3×10^7 bp and 9×10^4 bp, respectively. Several chloroplast genes, including the rRNA genes, were cloned on distinct fragments. The order of the rRNA genes corresponds to the common prokaryotic pattern. The 16S rRNA gene comprises 1,548 bases and is separated from the 23S rRNA gene with its 2,920 bases by a short spacer of 460 bases, which also includes the $tRNA^{Ile}$ and $tRNA^{Ala}$ genes. The 5S rRNA gene has not been found; it must start further than 500 bases downstream from the 3′-end of the 23S rRNA gene. From the chloroplast rRNA sequences, we have deduced secondary structures of the 16S and 23S rRNAs, which are in agreement with standard models. The rRNA sequences were aligned with corresponding chloroplast sequences; phylogenetic relationships were calculated by several methods. From these calculations, we conclude that *N.e.* is most closely related to *Chlorella vulgaris*. Therefore, *N.e.* does not represent an early developed algal species; the primitive-appearing morphological and biochemical characteristics of *N.e.* must rather be explained by secondary losses.

Key words: Algal phylogeny — Chloroplast phylogeny — Large-subunit rRNA — *Nanochlorum eukaryotum* — rRNA secondary structure — Small-subunit rRNA

Introduction

Green coccoidal algae of extremely small size, comparable to cyanobacteria, have been described (Andreoli et al. 1978; Dempsey et al. 1980; Johnson and Sieburth 1982; Turner and Gowen 1984; Thinh and Griffiths 1985). From morphological criteria, some of them have been identified as *Chlorella* species (*Chlorella nana,* Andreoli et al. 1978; *Chlorella minutissima,* Dempsey et al. 1980). However, their phylogenetic relationship to other algae has not been inferred from macromolecular sequences. Since small species may represent primitive algal forms, investigations of green microalgae may reveal useful information about the evolution of algae and algal plastids.

The marine green alga *Nanochlorum eukaryotum* (*N.e.*) (Wilhelm et al. 1982) is also of extremely small size (1.5 μm in diameter) and furthermore shows some features unusual for eukaryotic organisms (Zahn 1984). *N.e.* contains a single chloroplast and mitochondrion; histones and nucleosomes have not been found as yet. Upon division, chromosomes and spindle apparatus have

Correspondence to: D. Weinblum

not been observed. During mitosis, the nuclear membrane remains unchanged and forms two separate nuclei by pinching. Comparable mitotic characteristics have rarely been observed in green algae (Heath 1980; Margulis 1981). In particular, the absence of histones and thus nucleosomes and the small size have led to the assumption that *N.e.* is a "marginal" eukaryote (Zahn 1984). To either support or disprove this assumption, an even approximate knowledge of the phylogenetic relationship of this alga would be sufficient.

The phylogenetic relationship of organisms can be deduced in principle from the comparison of their macromolecular sequences. In particular, the rRNA genes, which are found not only in all prokaryotic and eukaryotic cells, but also in organelles, are well suited for such investigations (Cedergren et al. 1988; Van de Peer et al. 1990).

Thus, we have cloned and sequenced the chloroplast rRNA genes and have inferred the phylogenetic relationship of *N.e.* from these sequences.

Materials and Methods

Growth Conditions. N.e. was grown in continuous cultures as described by Zahn (1984).

DNA Isolation. Five grams (wet weight) of fresh *N.e.* cells were incubated overnight at 37°C in 2 ml lysis buffer (50 mmol/l Tris/HCl, pH 8, 250 mmol/l EDTA, 1 mmol/l aurintricarboxylic acid, 1.5% SDS, 2 mg/ml proteinase K). The mixture was centrifuged at 5,000*g* and the supernatant was extracted with an equal volume of phenol/chloroform (1:1/v:v). Nucleic acids were precipitated from the aqueous phase with isopropanol, redissolved in 4 ml TE buffer (50 mmol/l Tris/HCl, pH 8, 50 mmol/l EDTA), digested with RNase A, reextracted with phenol/chloroform, and repeatedly precipitated with isopropanol. About 0.5 mg total DNA was thus obtained. Yeast (*Saccharomyces cerevisiae*) DNA was isolated according to Cryer et al. (1975).

Analytical Density Gradient Centrifugation. Total DNA from *N.e.* was dissolved in 15 mmol/l Tris/HCl, pH 8, containing CsCl at a density of 1.7010 g/ml. Centrifugation was performed at 44,000 rpm in a Beckman model E centrifuge with digital data output in double sector cells. Densities were determined according to Szybalsky and Szybalsky (1971) with DNA from *Clostridium perfringens* and *Micrococcus lysodeicticus* as internal standards.

Preparative Density Gradient Centrifugation. Total DNA of *N.e.* was dissolved in CsCl/Tris/HCl (see above) at a concentration of 100 µg/ml and centrifuged for 20 h at 41,000 rpm in a Beckman VTi 50 rotor. The contents of the tubes were fractionated while reading the optical densities at 254 nm. To isolate the AT-rich DNA, the fractions containing approx. 5% of the total DNA at the "light side" of the peak were collected and recentrifuged three times until DNA of homogeneous density was obtained. Approx. 100 µg of DNA with a GC content of 34% was obtained from 10 mg of total DNA.

Reassociation Kinetics. DNA was dissolved in phosphate buffer (0.12 mol/l, pH 7), sonicated, and dialyzed against the same buffer for 24 h. Samples containing approx. 50 µg/ml DNA were gassed with helium, filled into the thermocuvette of a Gilford 250 spectrophotometer, and heated until no further increase in absorbancy at 260 nm was observed (approx. 95°C). Then the cuvette was cooled to 65°C within 1.5 min and the absorbancies of the samples at 260 nm were recorded on digital tape for 96 h (approx. 500 data points). Data were fitted to the equation

$$S/c_o = f(1 + k_1 f c_o t)^{-0.445} + (1-f)(1 + k_2(1-f)c_o t)^{-0.445} \quad (1)$$

(Britten and Davidson 1976) by a curve fit program (Minuit, CERN library), where S is the concentration of single-stranded DNA, c_o is the total DNA concentration, f is the fraction of the first component, and k_1 and k_2 are the reassociation constants of the two components, respectively.

Restriction Endonuclease Digestion and Hybridization with ctDNA Probes. Restriction endonuclease digestion, Southern blotting, and hybridization were done according to Sambrook et al. (1989). ctDNA probes from spinach (cytochrome f; 32-kDa protein from photosystem II [herbicide binding protein], ribulose-1,5-bisphosphate carboxylase/oxygenase, large subunit [rubisco]; photosystem I P700 apoprotein; ATP synthase, subunit alpha) were prepared according to Bolivar et al. (1977). Radioactive labeling of the probes was carried out as described by Rigby et al. (1977). A cDNA transcript of *Escherichia coli* 16S rRNA was obtained with the cDNA synthesis system from Amersham. All enzymes were purchased from Boehringer Mannheim; ^{32}P-labeled deoxyribonucleoside triphosphates came from Du Pont.

Construction of ctDNA Libraries and Mapping of the Genome. ctDNA fragments obtained from partial digestion with *Hin*dIII and from total digestion with *Cla*I were ligated into pBR322-DNA using standard methods (Sambrook et al. 1989). Transformation of *Escherichia coli* DH1 cells was carried out as described by Dagert and Ehrlich (1979). For the detection of ctDNA-containing clones, 10^5 colonies of the ctDNA library on nitrocellulose filters were lysed according to Sambrook et al. (1989). Colony hybridization with ctDNA probes was carried out under stringent conditions (68°C, 6× SSC). A radiolabeled cDNA transcript of purified *Escherichia coli* 16S rRNA was used to detect fragments carrying ribosomal genes on Southern blots of the restricted ctDNA. With these fragments, several *Hin*dIII clones were detected, which contained ribosomal genes.

Sequence and Secondary Structure Determination of the rRNA Genes. For sequence determination, the *Hin*dIII, *Eco*RI, and *Eco*RI/*Hin*dIII subfragments of the rRNA operon were inserted into pBR322-DNA following standard methods (Sambrook et al. 1989). Plasmid sequencing was done by the chain termination method (Sanger et al. 1977) using the T7-sequencing kit from Pharmacia-LKB. Standard primers for pBR322-derived *Hin*dIII and *Eco*RI sequences from Pharmacia-LKB, or specific 16-bp oligonucleotide primers produced on a DNA synthesizer (Beckman type 200) were used. Computer sequence analyses were performed with the Microgenie software package (Beckman). The sequence is available from EMBL nucleotide sequence database under access number X76084 CHNERRNA.

To obtain secondary structures of the SS and LS rRNA, the sequences were arranged in analogy to the standard models of Gutell et al. (1985) and Gutell and Fox (1988). Secondary structures of several variable domains were calculated according to Zuker and Stiegler (1981), using energy values from Freier et al. (1986).

Sequence Alignment and Tree Construction. Sequence alignments were performed with the "Clustal Software Package" (Higgins and Sharp 1988). Phylogenetic trees were calculated with different programs from the "Phylip Software Package," release 3.2 (Felsenstein 1989).

Results

DNA Isolation

The cell wall of *N.e.* contains a layer of sporopollenin (Geisert et al. 1987). Since no enzyme for the digestion of sporopollenin is known we were unable to produce protoplasts and subsequently failed to isolate circular ctDNA. The same experience has been reported for *Chlorella* (Yamada and Sakaguchi 1981), in which protoplasts have been obtained only from sporopollenin-free *Chlorella* strains. Treatment of *N.e.* cells with proteinase K and SDS, followed by phenol extraction, yielded low amounts of DNA. The average size of this DNA was 25 kb as judged from gel electrophoresis and density gradient-centrifugation profiles. Mechanical opening of the cells by grinding with alumina or passage through a French press prior to enzymatic digestion yielded DNA of even lower molecular weight, which was unsuitable for density gradient-separation procedures.

Density Gradient Centrifugation of DNA

The analytical CsCl gradient-centrifugation profile of the total DNA of *N.e.* reveals one main band and one satellite band. The GC content of the main band DNA was 44%; the GC content of the satellite band DNA was 34%, as determined from internal standards. No DNA subcomponents with a defined, but different GC content were hidden under the main band DNA. These were not expected, since in lower eukaryotes such subclasses have rarely been identified (Macaya et al. 1976). The satellite band DNA comprises 3% of the total DNA. It was separated from the main band DNA by repeated cycles of preparative CsCl gradient centrifugation without the addition of any density difference-enhancing substances.

Reassociation Kinetics

Reassociation kinetics were determined by measuring the decreasing hyperchromicity at 260 nm as a function of time. Besides the single-copy DNA, eukaryotic DNA normally contains fast-reassociating repetitive DNA and fold-back sequences. Therefore, a two-component curve (eq. 1) was selected to describe the experimental data. The reassociation constants and the portion of the single-copy DNA were determined by a curve fit procedure. The reassociation constants, which are inversely proportional to the genetic complexity, are listed in Table 1; 95% of the main band DNA of *N.e.* reassociates with a constant of 0.10. Single-copy DNA from yeast, which was used as a reference, reassociates with a constant of 0.23. Using the complexity of the yeast genome of 1.4×10^7 bp (Mortimer et al. 1992), we calculated the genetic complexity of the *N.e.* genome to be approx. 3.2×10^7 bp. This value is slightly smaller than values determined for several *Chlorella* species (Dörr and Huss 1990). We conclude from the kinetic complexity of the main band DNA and from the total amount of DNA per cell of 6×10^{-14} g (Zahn 1984) that the genome of *N.e.* is haploid. The reassociation constant of 35.3 for the satellite band DNA corresponds to a genetic complexity of 9×10^4 bp.

Table 1. Reassociation constants of DNA from *N.e.* and yeast

	Reassociation constant ($mol^{-1}s^{-1}$)	Proportion of DNA which reassociates with this constant
N.e. main band DNA	0.10	95.3%
N.e. satellite band DNA	35.3	90.5%
Yeast total DNA	0.23	87.9%

No corrections for the difference in base ratios between reference DNA and sample DNA were made. The influence of the GC content upon reassociation velocity has been differently assessed: no change (Britten et al. 1974), a decrease (Gillis et al. 1970), and an increase (Wetmur and Davidson 1968) have been reported for increasing GC content. Since yeast (42% GC) and *N.e.* (44% GC) have almost identical base ratios a correction in any case would be negligible. Since the base ratio of the satellite band DNA (34% GC) differs considerably from the base ratio of the yeast reference DNA, the value of the kinetic complexity for the satellite band DNA (9×10^4 bp) may well be greater or smaller by a margin of up to 20%. The value of 9×10^4 bp is consistent with the data obtained from restriction endonuclease analysis. (See below.)

Identification of Chloroplast DNA in the Satellite Band DNA

The satellite band DNA, isolated by preparative density gradient centrifugation, was digested with several restriction enzymes, and the fragments were separated by gel electrophoresis (shown for *Pvu*II, *Sal*I, and *Xba*I in Fig. 1). Distinct from the background, a pattern of well-separated bands was obtained for each enzyme, indicating that this fraction preferentially contains DNA of low genetic complexity. The lengths of the fragments add up to 72,000 bp for the *Pvu*II digest.

The DNA fragments were blotted onto a nitrocellulose sheet and probed with several chloroplast genes from spinach. The DNA probes hybridized exclusively with the digests of the satellite band DNA, but not with the main band DNA. The genetic complexity of the satellite band DNA of approx. 9×10^4 bp, the base ratio of 34% GC, and the hybridization characteristics prove that the satellite band DNA predominantly consists of ctDNA. Since the satellite band DNA was isolated by a method selective for the base ratio, it probably contains

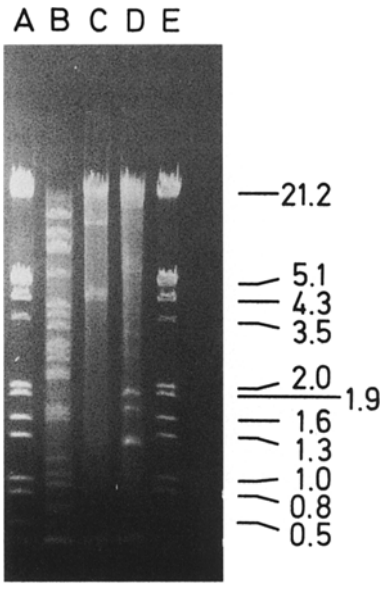

Fig. 1. Electrophoretic separation of restriction endonuclease digests of the satellite DNA band of *N.e.* Lanes **A** and **E:** marker DNA (lambda DNA, restricted with *Eco*RI + *Hin*dIII, fragment lengths in kb). Lane **B:** *Pvu*II digest. Lane **C:** *Sal*I digest. Lane **D:** *Xba*I digest.

some additional AT-rich nuclear DNA sequences of higher genetic complexity. This fact explains the background to be seen in Fig. 1.

The genetic complexity of the ctDNA compares to approx. 0.3% of the haploid nuclear DNA of *N.e.* (9×10^4 bp vs 3.2×10^7 bp). Since 3% of the total DNA is ctDNA, the chloroplast of *N.e.* contains approx. ten copies of the ct genome.

Partial Mapping of the Chloroplast Genome

Satellite-band DNA libraries were established in the vector pBR322 using fragments from partial (*Hin*dIII) and complete (*Cla*I) restrictions. Starting with the gene for the herbicide binding protein (psb A) from spinach, overlapping clones were detected by colony hybridization, which comprise a DNA segment of 10.9 kb. This segment contained the gene for the large subunit of rubisco as well as the gene for the ATP synthase subunit alpha (Fig. 2A). The same order has been found in *Codium fragile* (Manhart et al. 1989). However, in algal ctDNA the order of genes varies considerably, even in closely related species (Palmer 1985).

A 15.4-kb segment of the chloroplast genome was mapped by hybridization analysis using overlapping *Hin*dIII fragments from the satellite band DNA. On this segment a ribosomal operon was identified by hybridization analysis with *Escherichia coli* 16S cDNA. As shown in Fig. 2B, this rRNA operon is entirely located on a 6-kb fragment from the *Cla*I library.

Primary Structure of the rRNA Genes

The 6-kb ctDNA fragment from the *Cla*I library contains most of the chloroplast rRNA operon and was almost completely sequenced. The order of the rRNA and tRNA genes follows the common chloroplast pattern (Palmer 1985). No introns or internally transcribed spacer (ITS) sequences were found inside the genes. The distance between the 16S rRNA and the 23S rRNA genes comprises merely 459 bases, including the usual $tRNA^{Ile}$ and $tRNA^{Ala}$ genes and three spacers. This distance is comparably short, a shorter distance has only been observed in *Euglena gracilis* thus far (Palmer 1985). In contrast, the spacer sequence between the 23S rRNA and the 5S rRNA genes must be unusually long, since we sequenced 500 bases downstream from the 3′-end of the 23S rRNA gene without reaching the start of the 5S rRNA gene.

Secondary Structure of Chloroplast rRNA

Arranging rRNA sequences to secondary structure models is a prerequisite for alignment procedures. A secondary structure can be obtained from the SS rRNA of *N.e.* in analogy to the structure model of Gutell et al. (1985), which corresponds very well to other chloroplast SS rRNA structures (Fig. 3). Numbering of stems and variable regions follows the proposal of Dams et al. (1988). Noticeable differences among chloroplast sequences of *N.e.* and those from other species occur in two variable regions: (1) In SS rRNA of *N.e.*, the stem and loop structure 6 in V1 contains 40 bases while, for example, the length of this structure in the chloroplasts of *Zea mays* and *Chlamydomonas reinhardtii* comprises only 21 and 25 bases, respectively; (2) in SS rRNA of *N.e.*, the V7 region is shorter than in other chloroplasts. Recalculating the base pairing in the V7 region according to Zuker and Stiegler (1981) led to a secondary structure not showing stem 42; however, this region can be also arranged according to the standard model. Additional minor differences in the lengths of stems or loops, respectively, were exclusively found in variable regions. (See aligned sequences in Fig. 5.)

The chloroplast large ribosomal subunit (LS) rRNA from *N.e.* can be arranged with minor differences according to the LS rRNA secondary structure model of Gutell and Fox (1988) (Fig. 4). Stems, loops, and variable regions are designated according to Engberg et al. (1990). Two areas of high sequence variability are known in chloroplast LS rRNA. The first is D1, which in *N.e.* comprises the bases 257–443. Our calculations gave a secondary structure for this region with two additional stem-loop structures (D1a and D1d) as compared to the model of Gutell and Fox (1988). Helix D1a has also been found by Engberg et al. (1990) in *Tetrahymena pyriformis* LS rRNA. The second variable region is D7B, which in *N.e.* comprises the bases 1437–1639. For this domain in *N.e.*, our calculations led to a different secondary structure than proposed by Gutell and Fox (1988). Both regions show no homology to the corresponding sequences of other chloroplasts.

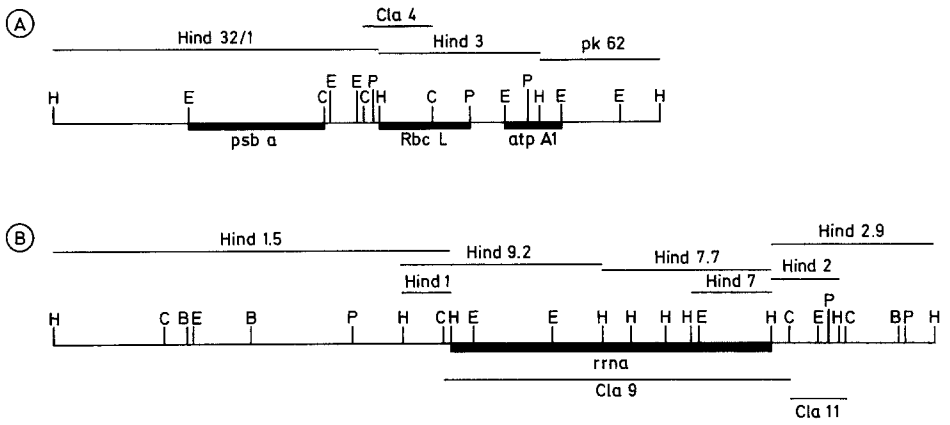

Fig. 2. Restriction map of the two consecutive segments of the ctDNA. **A** 10.9-kb segment. **B** 15.7-kb segment. *Bars* indicate adjacent or overlapping clones, respectively. Restriction sites: *B, Bam*HI; *C, Cla*I; *E, Eco*RI; *H, Hin*dIII. The approximate locations of psbA, RbcL, atp a1, and rRNA are indicated.

Michot et al. (1990), Bachellerie and Michot (1989), and Michot and Bachellerie (1987) have performed detailed comparisons of the variable domains D2, D3, and D8 and of the 3′-terminus of LS rRNA from many organisms. They observed specific conserved base motifs and secondary structures in these domains, which are characteristic for major phylogenetic groups or organelles, respectively. The LS rRNA domains of *N.e.* correspond well to the pattern characteristic for chloroplasts; for example, the variable domain 3 contains the additional short stem-loop structure (D3c), which is specific for chloroplast LS rRNA (Michot et al. 1990). This stem-loop structure is not contained in the general model of Gutell and Fox (1988).

Sequence Alignment

The chloroplast rRNA sequences from *N.e.* were aligned with the corresponding sequences from chloroplasts of algae and of some plants. The rRNA sequences from *Escherichia coli* and *Anacystis nidulans* were included as outgroups.

A satisfactory procedure for the alignment of only two DNA sequences is presently not known. (For a recent discussion of the problem see Thorne et al. 1991.) The alignment of rRNA sequences becomes particularly complicated due to the fact that rRNA genes consist of regions showing different degrees of conservation. For distantly related species, only the highly conserved regions can be aligned (Cedergren et al. 1988), while for closely related species it is even possible to align the variable regions (Lenaers et al. 1991).

We have used the "Clustal Program" (Higgins and Sharp 1988) for the alignment. With the aid of the secondary structure models, homologous, highly conserved regions were identified for all species, which could be aligned unambiguously by hand. Between such fixed regions sub-sequences several hundred bases in length were aligned with the Clustal Program. Under these conditions the program worked satisfactorily. The alignment obtained with the program was not further corrected by hand. Results are shown in Figs. 5 and 6.

Almost the complete SS rRNA sequences could be aligned, except that 21 bases from the *Escherichia coli* sequence in stem-loop structure 18 were deleted (Fig. 5). Poor alignment is observed, however, in the variable regions V1 and V7, mainly due to the greater length of the *Chlorella ellipsoidea* sequence.

The alignment of the LS rRNA was less satisfactory. Two short segments, comprising 14 and 25 nucleotides, respectively, had to be deleted from the *Escherichia coli* sequence (Fig. 6). Furthermore, for all species the variable domains D1 and D7B as well as the 3′-end with the 4.5S rRNA had to be excluded from the alignment procedure.

Phylogenetic Trees

We have calculated approx. 100 trees with the "Phylip Software Package," using parsimony, compatibility, distance matrix, and likelihood programs. Alternative tree rearrangements were employed by using the "global" option and the "Penny" algorithm. Predominantly, subsets of the aligned sequences have been used, including or excluding variable regions of uncertain alignment. The distance matrix trees calculated from the complete LS and SS rRNAs are shown as examples in Fig. 7.

From the different trees the following conclusions were drawn: On all trees the plant chloroplasts emerge on a single, defined branch in the known order; this branch always contains the chloroplasts of the *Chlorella* species, too. We could, however, not unambiguously define the position of the *Chlamydomonas reinhardtii* chloroplast. In the majority of calculations this species was located on the "plant branch." The branching point could not be determined precisely and varied with the method of calculation and the set of data used. On all SS rRNA trees we always found a close relationship be-

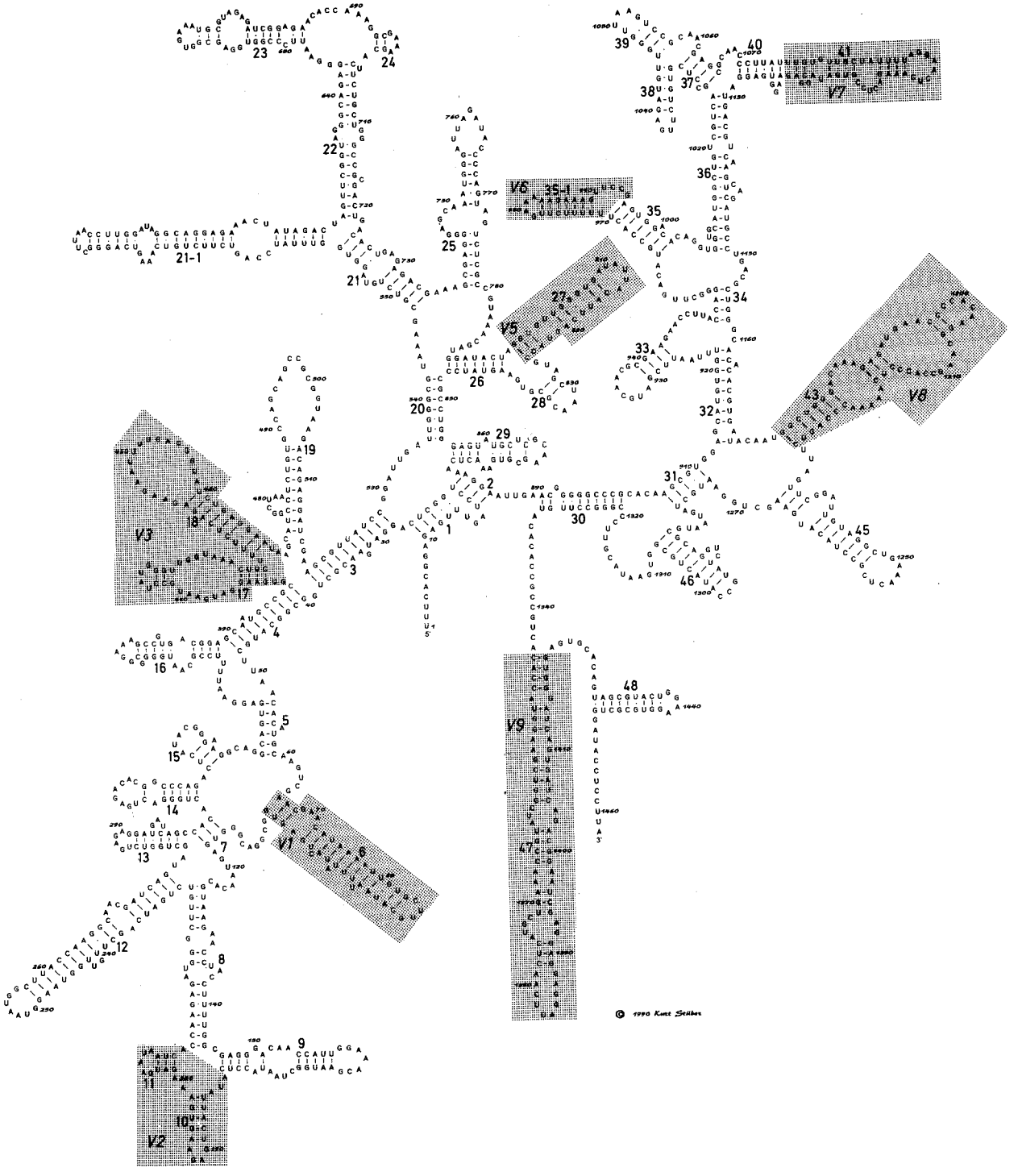

Fig. 3. Secondary structure of chloroplast SS rRNA of *N.e.* Stem-loop structures and variable domains are numbered according to Dams et al. (1988).

tween *Pylaiella littoralis* and *Euglena gracilis* chloroplasts. We could not determine, however, whether the branch containing *Pylaiella littoralis* and *Euglena gracilis* emerges separately from the cyanobacterial branch or whether it is contained in a monophyletic chloroplast tree. These results are almost identical to chloroplast phylogenetic trees recently constructed from SS rRNA (Turner et al. 1989; Markowicz and Loiseaux-de Goer 1991; Douglas 1992). Our LS rRNA tree corroborates these results; the branching order of the plant and the *Chlorella* chloroplasts is identical to the SS rRNA tree, but like in the SS rRNA tree the branching position of *Chlamydomonas reinhardtii* and of *Euglena gracilis* chloroplasts cannot exactly be determined.

Concerning the position of *N.e.*, we obtained identical results from all calculations: the chloroplast of *N.e.* was

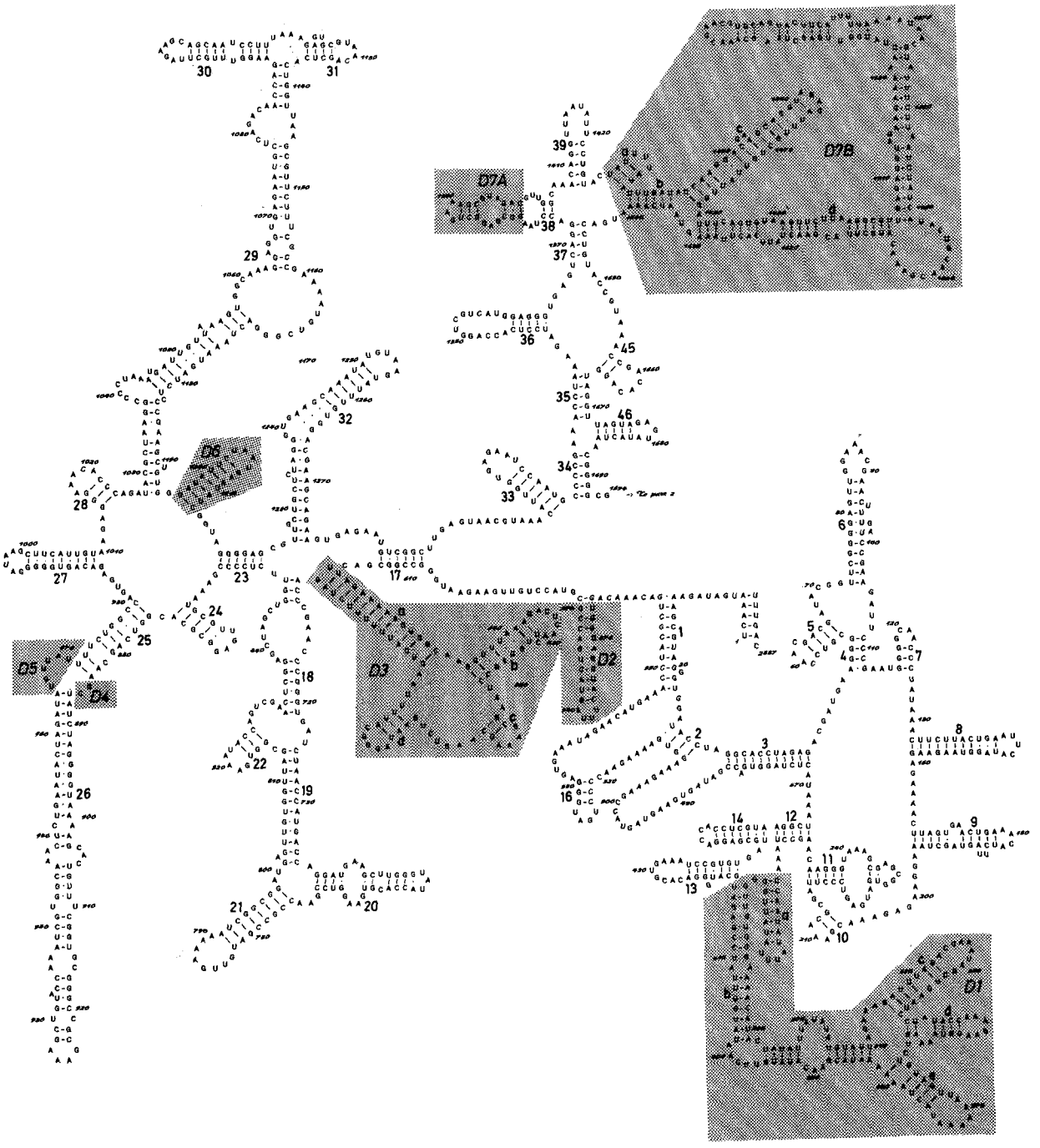

Fig. 4. Secondary structure of chloroplast LS rRNA of *N.e.* Stem-loop structures and variable domains are numbered according to Engberg et al. (1990).

always located on a subbranch, which also contained the *Chlorella* species. This was valid for both SS and LS rRNA and for all methods of phylogenetic calculations, which were employed. On the SS rRNA tree, which included two *Chlorella* species, we always obtained an even closer relationship of *Chlorella vulgaris* with *N.e.* than of *Chlorella vulgaris* with *Chlorella ellipsoidea*. Bootstrap probability for the close relation among *N.e.* and *Chlorella vulgaris* was 98%.

Discussion

Inference of phylogenetic relationships from sequence data is a notoriously difficult task (for a thorough discussion of the problem see Felsenstein 1988), and thus conflicting results are numerous in the recent literature. Ambiguous results in tree construction may be due to the fact that the particular data set used may not contain sufficient information in a statistical sense to support a

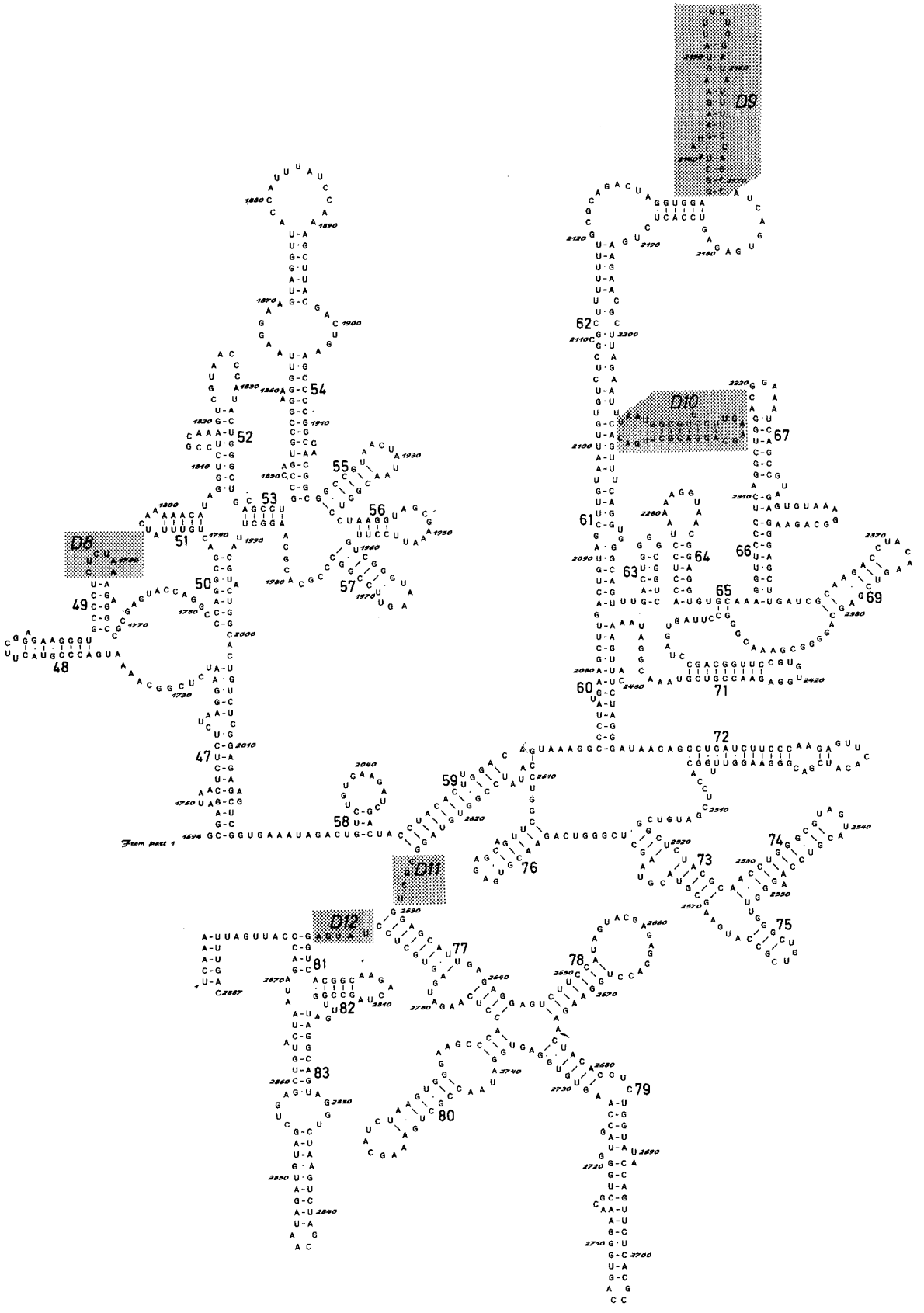

Fig. 4. Continued.

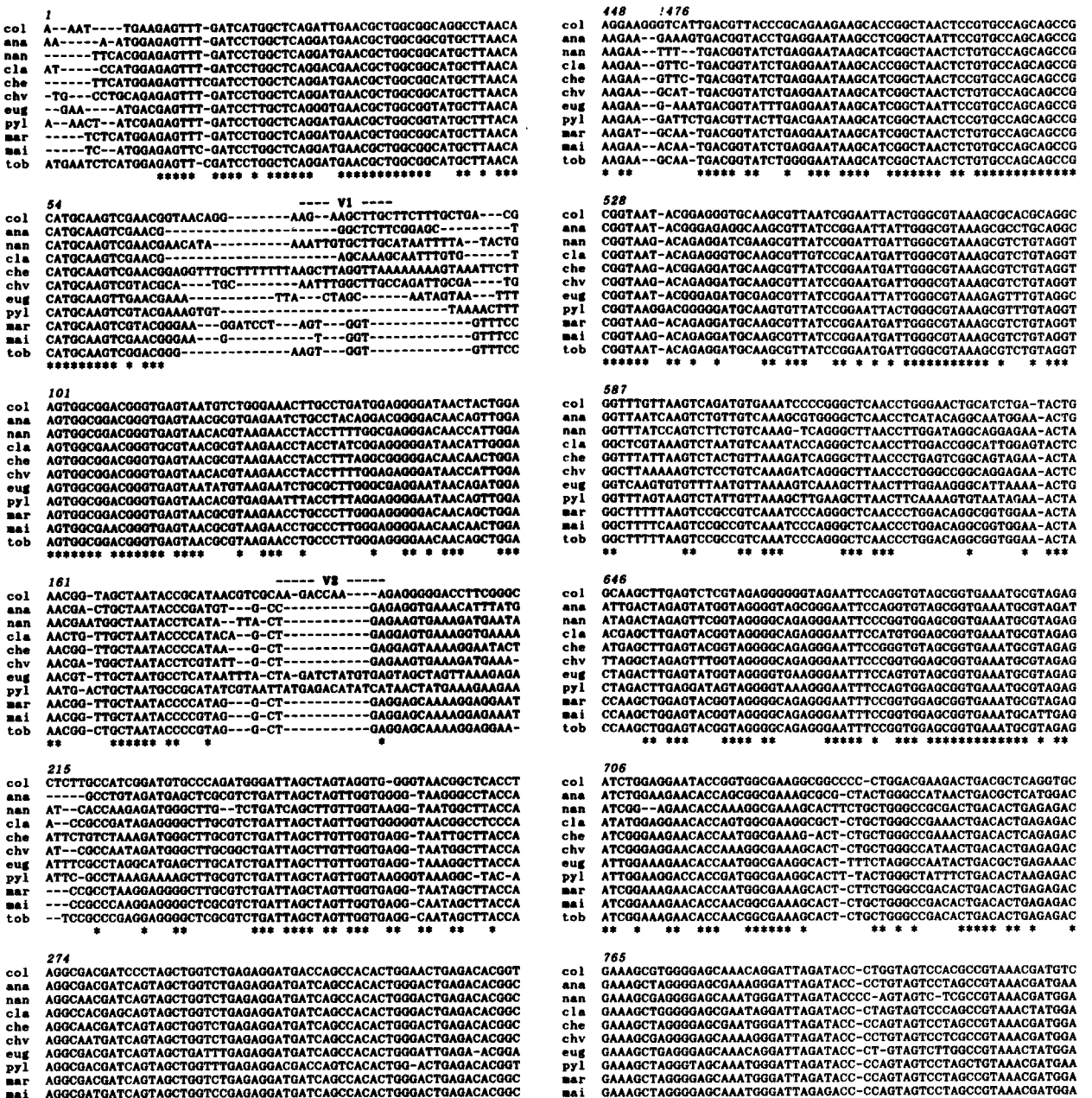

Fig. 5. Alignment of chloroplast SS rRNA sequences from 11 species. *Nucleotide numbers* refer to the *Escherichia coli* sequence. One deletion in this sequence is indicated by an *exclamation mark.* Variable regions are indicated and refer to Fig. 3. Abbreviations: *col, Escherichia coli; ana, Anacystis nidulans; nan, Nanochlorum eukaryotum; cla, Chlamydomonas reinhardtii; che, Chlorella ellipsoidea; chv, Chlorella vulgaris; eug, Euglena gracilis; pyl, Pylaiella littoralis; mar, Marchantia polymorpha; mai, Zea mays; tob, Nicotiana tabacum.*

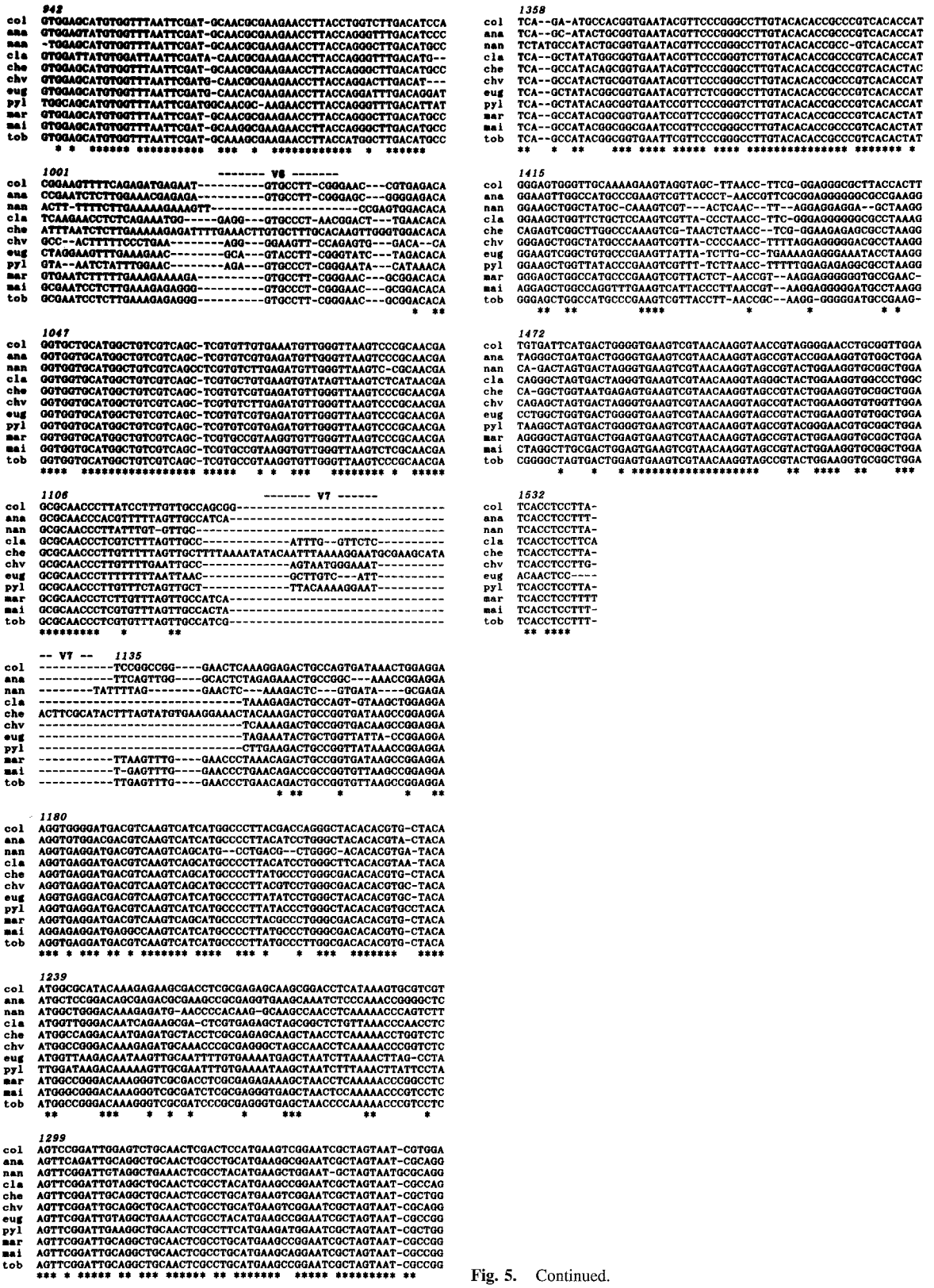

Fig. 5. Continued.

single "best" tree. Statistical methods to estimate the phylogenetic information content of sequence data are currently being developed (Archie 1989; Hillis and Huelsenbeck 1992; Goldman 1993). Furthermore, the methods for tree construction which are presently employed may not be optimal and thus better methods might produce less-ambiguous results from the same set of data (Penny et al. 1992). On the other hand, "robust" trees with identical branching order and similar branch lengths are frequently obtained from different data sets

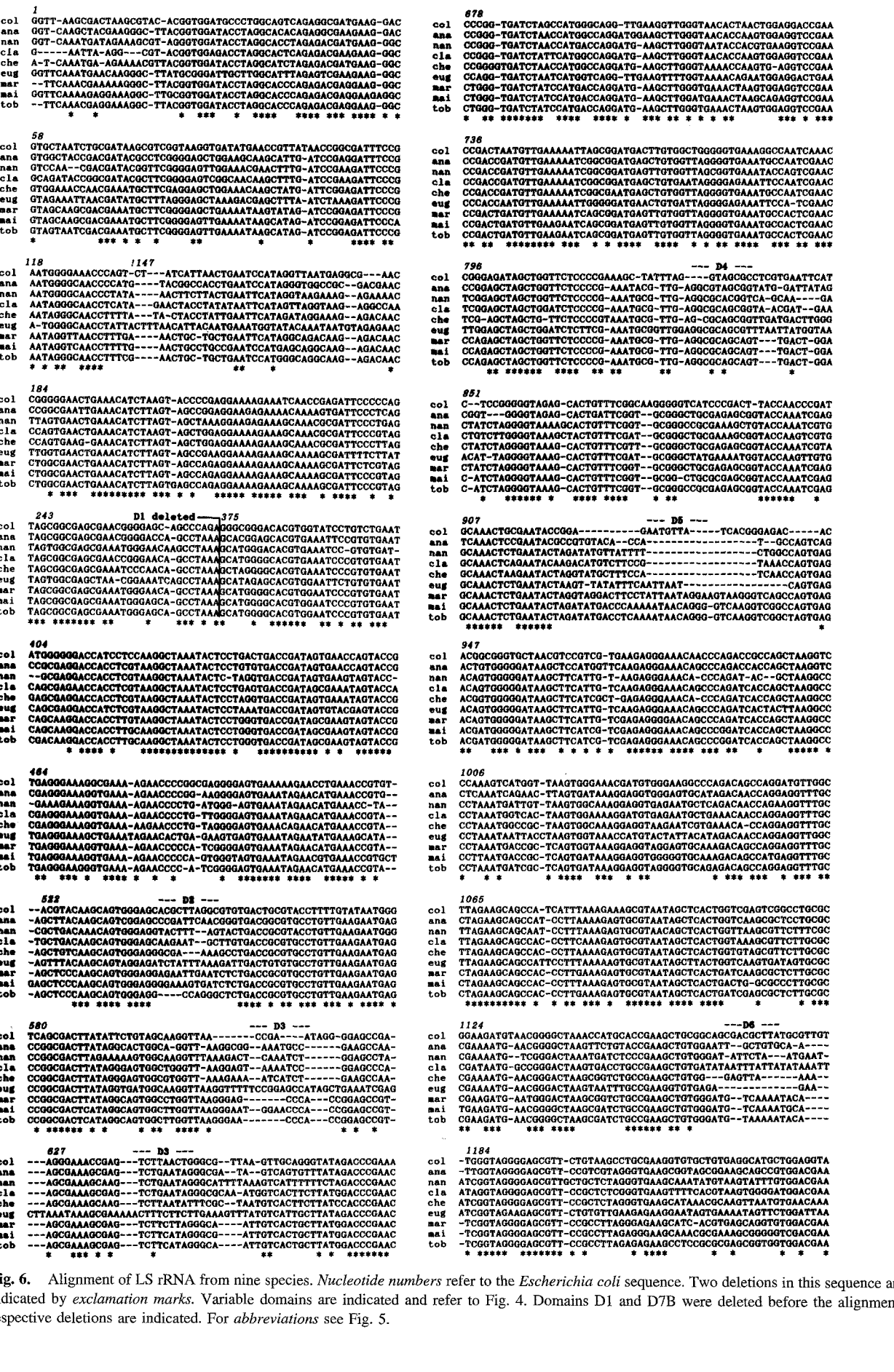

Fig. 6. Alignment of LS rRNA from nine species. *Nucleotide numbers* refer to the *Escherichia coli* sequence. Two deletions in this sequence are indicated by *exclamation marks.* Variable domains are indicated and refer to Fig. 4. Domains D1 and D7B were deleted before the alignment; respective deletions are indicated. For *abbreviations* see Fig. 5.

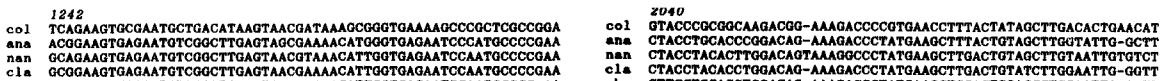

Fig. 6. Continued.

```
      2627
col   CTGGAGAACTGAGGGGGGCTGCTCCTAGTACGAGAGGACCGGAGTGGAC-GCATCACTGG
ana   TAAGAGTATTGAGAGGATTTCTCCCTAGTACGAGAGGACCGGGAGGAAC-GCACCGCTGG
nan   CTGGAGCATTGAGAGGAGTCTTCCATAGTACGAGAGGACCTGGAAGAACTACACCTCTGG
cla   TTAGAGCATTGAGAGTAGCCTTTCATAGTACGAGAGGACCTGAAAGGAC-ATGCCAATTG
che   TAAGAGCATTGAGAGGAGTCTTCCATTGTACGAGAGGACCTGGAAGAAC-GCACCACTAG
eug   TTAGAGTATTGAAAGGCGCTTTCCTTAGTACGAGAGGACCGGGAAGGAC-GCACCACTGA
mar   TTAGAGCATTGAGAGGACCTTTCCCTAGTACGAGAGGACCGGGAAGGAC-GCACCTCTGG
mai   TTAGAGCATTGAGAGGACCTTTCACTAGTACGAGAGGACCGGGAAGGAC-GCACCTCTGG
tob   TTAGAGCATTGAGAGGACCTTTCCCTAGTACGAGAGGACCGGGAAGGAC-GCACCTCTGG
         *** * *** *          * ********** *   * **     *  *

      2686
col   TGTTCGGGTTGTCATGCCAATGGCAC-TGCCCGGTAGCTAAATGCGGAAGAGATAAGTGC
ana   TGTACCAGTTATCGTGCCAACGGTAAACGCTGGGTAGCTACGTGTGGAGTGGATAACCGC
nan   TATACCAGTTCTCACGCCAGTGGGAAACGCTGGGTAGCCAAGTGTGGAGTGGATAACCGC
cla   TGTACCAGTTCTCATTCCAATGGGAAACGCTGGGTAGCTACGCATGGATAGAATA-CTGC
che   TATACCAGTTCTCCTGCCAAGGGGATACGCTGGGTAGCCAAGTGCGGAGCGGATAACTGC
eug   TGTACCAGTTTTTGTGCCAACAGAATACGCTGGGTAGTCAAGTGCGGAGTGGATAACTGC
mar   GTTACCAGTTATCGTGCCCACGGTAAACGCTGGGTAGCCAAGTGCGGA-CGGATAACTGC
mai   TGTACCAGTTATCGTGCCTACGGTAAACGCTGGGTAGCCAAGTGTGGAGAGGATAACTGC
tob   TTGACCAGTTATCGTGCCCACGGTAAACGCTGGGTAGCCAAGTGCGGAGCGGATAACTGC
         *  *** *    **    * *   **  *****  *     ***    ***   **

      2745
col   TGAAAGCATCTAAGCACGAAACTTGCCCCGAGATGAGTTCTCCC
ana   TGAAAGCATCTAAGTGGGAAGCCCACCTCAAGATGAGTACTCTC
nan   TGAAAGCATCTAAGTGGGAAGCCCACCTCAAGATGAGTGCTCCT
cla   TGAAAGCATCTAAGTAGGAAGCTAAACTCAAGATGAGTGCTCTC
che   TGAAAGCATCTAAGTAGGAAGCCCACCTCAAGATGAGTGCTCTC
eug   TGAAAGCATATAAGTAGGAAGCCCACCTTAAGATAAGTACTCTT
mar   TGAAAGCATCTAAGTAGGAAGCCCACCTCAAGATGAGTGCTCTC
mai   CGAAAGCATATAAGTAGTAAGCCCACCCCAAGATGAGTGCTCTC
tob   TGAAAGCATCTAAGTAGTAAGCCCACCCCAAGATGAGTGCTCTC
      ******** ****    ** *   *   **** *** ***
```

Fig. 6. Continued.

using various methods of calculation. Although methods to prove the correctness of such trees are still lacking (Hillis and Huelsenbeck 1992), these trees are considered to be correct.

Obviously, due to reasons cited above, the phylogeny of plastids has not been elucidated conclusively, although many investigations have been performed on this subject, using either rRNA sequences (Hori and Osawa 1987; Woese 1987; Cedergren et al. 1988; Giovannoni et al. 1988; Turner et al. 1989; Van de Peer et al. 1990; Markowicz and Loiseaux-de Goer 1991; Douglas 1992) or protein sequences (Schwartz and Dayhoff 1978; Janssen et al. 1989; Grace 1990; Kraus et al. 1990; Morden et al. 1992) or gross genome organization (Shivji et al. 1992).

Most of these authors agree that plastids originated from cyanobacteria and are of monophyletic origin. However, a "robust" tree comprising all plastids has not been found yet. The branching order of most algal chloroplasts could not be determined precisely. Only the tree containing exclusively the chloroplasts of plants was found to be "robust." This tree is also in agreement with "classical" taxonomy, based on morphological criteria.

Recently, the monophyletic origin of the plastids has, however, been questioned (Boczar et al. 1989; Assali et al. 1990). High homology has been found between the ribulosebisphosphate carboxylase gene of the chloroplast of *Pylaiella littoralis* and β-purple bacteria (Assali et al. 1991; Markowicz and Loiseaux-de Goer 1991) and between the same gene of the chloroplast of *Ectocarpus siliculosus* and hydrogen bacteria (Valentin and Zetsche 1990). However, the rRNA genes of these algae show the usual homology to cyanobacteria. Thus, it has been proposed that the plastid genomes from Chromophyta, Rhodophyta, and Cryptophyta are of a composite phylogenetic origin. Furthermore, plastids have also originated from secondary symbiotic events. It has been shown that several algae are intertaxonic chimaeras generated from different protozoa (Douglas et al. 1991; Eschbach et al. 1991; Douglas 1992).

Considering the preliminary status of sequence derived chloroplast phylogeny, we will cautiously assess the significance of our results. The position of *N.e.* on the tree with respect to the positions of the plant and *Chlorella* chloroplasts was "robust," which means it is identical under all conditions. Consequently, we conclude that *N.e.* is most closely related to *Chlorella vulgaris* and therefore belongs to the order Chlorococcales. This con-

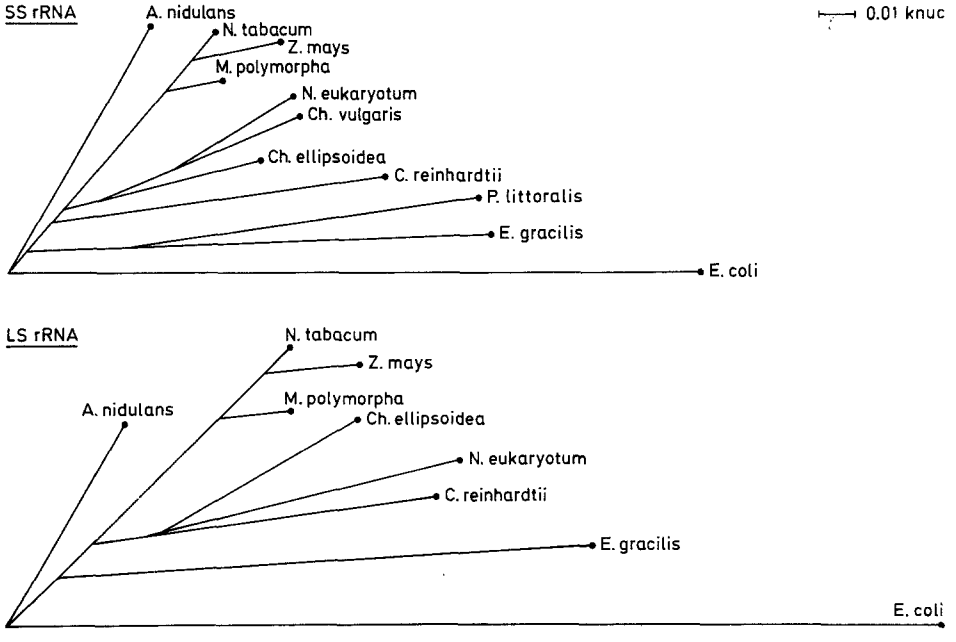

Fig. 7. Phylogenetic trees inferred from SS rRNA and LS rRNA sequences. The complete, aligned sequences from Figs. 5 and 6 were used. No distance corrections were used. Distance matrices were calculated according to Kimura (1980) with the program DNADIST, trees with the program FITCH, both from the Phylip software package. Lengths of vectors are proportional to evolutionary distances.

clusion is supported by investigations using the nuclear SS rRNA sequence of *N.e.* (Sargent et al. 1988), which has been included in calculations of the phylogeny of algae (Rausch et al. 1989; Eschbach et al. 1991; Hendriks et al. 1991; Douglas 1992). In all investigations, *N.e.* was found to be closely related to *Chlorella.*

Even if this taxonomic position is only approximately correct, we certainly can exclude that *N.e.* belongs to an ancestral algal lineage. Consequently, the primitive morphological and biochemical appearance of *N.e.* must be due to reduction. A comparable phylogenetic misinterpretation of unusual biochemical properties has occurred in dinoflagellates. In particular, the absence of histones has led to the assumption that dinoflagellates are an "ancestral" taxon (Herzog et al. 1984). However, phylogenetic analysis, based on LS rRNA sequences, places dinoflagellates close to ciliates and yeast in the middle of the kingdom of unicellular eukaryotes (Lenaers et al. 1989).

Thus, it is obviously unwarranted to draw phylogenetic conclusions from biochemical peculiarities, even if they differ significantly from established standards. The phylogenetic position of an organism can be inferred only from a thorough analysis using sequences of several suitable genes.

Acknowledgments. The authors are indebted to Dr. Zahn from our institute for samples of *N.e.*, to Dr. Herrmann, München, for providing the ctDNA probes from spinach, and to Dr. S. Suhai, Heidelberg, for help in computer sequence analysis.

References

Andreoli C, Rascio N, Casodoro G (1978) *Chlorella nana* sp. nov. (Chlorophyceae): a new marine *Chlorella.* Bot Mar 21:253–256

Archie JW (1989) Phylogenies of plant families: a demonstration of phylogenetic randomness in DNA sequence data derived from proteins. Evolution 43:1796–1800

Assali NE, Martin WF, Sommerville CC, Loiseaux-de Goer S (1991) Evolution of the rubisco operon from prokaryotes to algae: structure and analysis of the rbcS gene of the brown alga *Pylaiella littoralis.* Plant Mol Biol 17:853–863

Assali NE, Mache R, Loiseaux-de Goer S (1990) Evidence for a composite phylogenetic origin of the plastid genome of the brown alga *Pylaiella littoralis* (L.) Kjellm. Plant Mol Biol 15:307–315

Bachellerie JP, Michot B (1989) Evolution of large subunit rRNA structure: the 3′-terminal domain contains elements of secondary structure specific to major phylogenetic groups. Biochimie 71:701–709

Boczar BA, Delaney T, Cattolico RA (1989) The gene for the ribulose-1,5-bisphosphate carboxylase (Rubisco) small subunit protein of the marine chromophyte *Olisthodiscus luteus* is similar to that of a chemoautotrophic bacterium. Proc Natl Acad Sci USA 86:4996–4999

Bolivar F, Rodriguez RL, Betlach MC, Boyer HW (1977) Construction and characterization of new cloning vehicles I. Ampicillin-resistant-derivatives of the plasmid pMB9. Gene 2:75–93

Britten RJ, Davidson EH (1976) Studies on nucleic acid reassociation kinetics: empirical equations describing DNA reassociation. Proc Natl Acad Sci USA 73:415–419

Britten RJ, Graham DE, Neufeld BR (1974) Analysis of repeating DNA sequences by reassociation. In: Grossman L, Moldave K (eds) Methods in enzymology, vol XXIX. Academic Press, New York, pp 363–418

Cedergren R, Gray MW, Abel Y, Sankoff D (1988) The evolutionary relationship among known life forms. J Mol Evol 28:98–112

Cryer DF, Eccleshall R, Marmur J (1975) Isolation of yeast DNA. In: Prescott DM (ed) Methods in cell biology, vol XII. Academic Press, New York, pp 39–44

Dagert M, Ehrlich SD (1979) Prolonged incubation in calcium chloride improves the competence of E. coli cells. Gene 6:23–28

Dams E, Hendriks L, van de Peer Y, Neefs JF, Smits G, Vandenbempt I, De Wachter R (1988) Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res 16:r87–r173

Dempsey GP, Lawrence D, Cassie V (1980) The ultrastructure of *Chlorella minutissima* Fott et Novakova (Chlorophyceae, Chlorococcales). Phycologia 19:13–19

Dörr R, Huss VAR (1990) Characterization of nuclear DNA in 12 species of Chlorella (Chlorococcales, Chlorophyta) by DNA reassociation. Biosystems 24:145–155

Douglas SE (1992) Eukaryote-eukaryote endosymbioses: insights from studies of a cryptomonad alga. Biosystems 28:57–68

Douglas SE, Murphy CA, Spencer DF, Gray MW (1991) Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. Nature 350:148–151

Engberg J, Nielsen H, Lenaers G, Fujitani H, Higashinakagawa T (1990) Comparison of primary and secondary 26S rRNA structures in two Tetrahymena species: evidence for a strong evolutionary and structural constraint in expansion segments. J Mol Evol 39:514–521

Eschbach S, Wolters J, Sitte P (1991) Primary and secondary structure of the nuclear small subunit ribosomal RNA of the Cryptomonad *Pyrenomonas salina* as inferred from the gene sequence: evolutionary implications. J Mol Evol 32:247–252

Felsenstein J (1989) Phylip manual 3.2. University of California, Berkeley

Felsenstein J (1988) Phylogeny from molecular sequences: inference and reliability. Annu Rev Genet 22:521–565

Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH (1986) Improved free-energy parameters for predictions of RNA duplex stability. Proc Natl Acad Sci USA 83:9373–9377

Geisert M, Rose T, Bauer W, Zahn RK (1987) Occurrence of carotenoids and sporopollenin in *Nanochlorum eukaryotum.* Biosystems 20:133–142

Gillis M, DeLey J, De Cleene M (1970) The determination of molecular weight of bacterial genome DNA from renaturation rates. Eur J Biochem 12:143–153

Giovannoni SJ, Turner S, Olsen GJ, Barns S, Lane DJ, Pace NR (1988) Evolutionary relationship among cyanobacteria and green chloroplasts. J Bacteriol 170:3584–3592

Goldman N (1993) Statistical tests of models of DNA substitution. J Mol Evol 36:182–198

Grace SC (1990) Phylogenetic distribution of superoxid dismutase supports an endosymbiotic origin for chloroplasts and mitochondria. Life Sci 47:1875–1886

Gutell RB, Fox GE (1988) A compilation of large subunit RNA sequences presented in a structural format. Nucleic Acids Res 16:r175–r269

Gutell RB, Weiser B, Woese CR, Noller HF (1985) Comparative anatomy of 16S-like ribosomal RNA. Prog Nucleic Acid Res Mol Biol 32:155–216

Heath I (1980) Variant mitosis in lower eukaryotes, indication of the evolution of mitosis. Int Rev Cytol 64:1–80

Hendriks L, De Baere R, Van de Peer Y, Neefs J, Goris A, De Wachter R (1991) The evolutionary position of the rhodophyte *Porphyra umbilicalis* and the basidiomycete *Leucosporidium scottii* among other eukaryotes as deduced from complete sequences of small ribosomal subunit RNA. J Mol Evol 32:167–177

Herzog M, von Boletsky S, Soyer MO (1984) Ultrastructural and bio-

chemical nuclear aspects of eukaryote classification: independent evolution of the dinoflagellates as a sister group of the actual eukaryotes? Orig Life 13:205–312

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene 73:237–244

Hillis DM, Huelsenbeck JP (1992) Signal, noise and reliability in molecular phylogenetic analysis. J Hered 83:189–195

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

Janssen I, Jakowitsch J, Michalowski CB, Bohnert HJ, Löffelhardt W (1989) Evolutionary relationship of *psbA* genes from cyanobacteria, cyanelles and plastids. Curr Genet 15:335–340

Johnson PW, Sieburth J (1982) In situ morphology and occurrence of eukaryotic phototrophs of bacterial size in the picoplankton of estuarian and oceanic waters. J Phycol 18:318–327

Kimura M (1980) A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kraus M, Götz M, Löffelhardt W (1990) The cyanelle *str* operon from *Cyanophora paradoxa:* sequence analysis and phylogenetic implications. Plant Mol Biol 15:561–573

Lenaers G, Scholin Ch, Bhaud Y, Saint-Hilaire D, Herzog M (1991) A molecular model of Dinoflagellate protist (Pyrrhophyta) inferred from the sequence of 24S rRNA divergent domains D1 and D8. J Mol Evol 32:53–63

Lenaers G, Maroteaux L, Michot B, Herzog M (1989) Dinoflagellates in evolution. A molecular phylogenetic analysis of large subunit ribosomal RNA. J Mol Evol 29:40–51

Macaya G, Thiery JP, Bernardi G (1976) An approach to the organization of eukaryotic genomes at a macromolecular level. J Mol Biol 108:237–254

Manhart JR, Kelly K, Dudock BS, Palmer JD (1989) Unusual characteristics of *Codium fragile* chloroplast DNA revealed by physical and gene mapping. Mol Gen Genet 216:417–421

Margulis L (1981) Symbiosis in cell evolution. WH Freeman and Co., San Francisco, CA, pp 233–283

Markowicz Y, Loiseaux-de Goer S (1991) Plastid genomes of the Rhodophyta and Chromophyta constitute a distinct lineage which differs from that of the Chlorophyta and have a composite phylogenetic origin, perhaps like that of the Euglenophyta. Curr Genet 20:427–430

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, Bachellerie JP (1987) Comparisons of large subunit rRNA reveal some eukaryote-specific elements of secondary structure. Biochimie 69:11–23

Morden W, Delwichen CF, Kuhsel M, Palmer JD (1992) Gene phylogenies and the endosymbiotic origin of plastids. Biosystems 28: 75–90

Mortimer RK, Contopoulou CR, King JS (1992) Genetic and physical maps of *S. cerevisiae,* edition 11. Yeast 8:817–902

Palmer JD (1985) Comparative organization of chloroplast genomes. Annu Rev Genet 19:325–354

Penny D, Hendy MD, Steel MA (1992) Progress with methods for constructing evolutionary trees. Trends Ecol Evol 7:73–79

Rausch H, Larsen N, Schmitt R (1989) Phylogenetic relationship of the green alga *Volvox carteri* deduced from small-subunit ribosomal RNA comparisons. J Mol Evol 29:255–265

Rigby PW, Dieckmann M, Rhodes C, Berg P (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA-Polymerase I. J Mol Biol 113:237–251

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

Sargent M, Zahn RK, Walters B, Gupta R, Kain B (1988) Nucleotide sequence of the 18S rRNA from the microalga *Nanochlorum eukaryotum.* Nucleic Acids Res 16:4156

Schwartz RM, Dayhoff MO (1978) Origin of prokaryotes, eukaryotes, mitochondria, and chloroplasts. Science 199:395–403

Shivji MS, Li N, Cattolico RA (1992) Structure and organization of rhodophyte and chromophyte plastid genomes: implications for the ancestry of plastids. Mol Gen Genet 232:65–73

Szybalski W, Szybalski EH (1971) Equilibrium density gradient centrifugation. In: Cantoni GL, Davis DR (eds) Procedures in nucleic acid research, vol II. Harper and Row, New York, pp 333–343

Thinh LV, Griffiths DJ (1985) A small marine Chlorella from the waters of a coral reef. Bot Mar 28:41–46

Thorne JL, Kishino H, Felsenstein J (1991) An evolutionary model for maximum likelihood alignment of DNA sequences. J Mol Evol 33:114–124

Turner MF, Gowen RJ (1984) Some aspects of the nutrition and taxonomy of fourteen small green and yellow-green algae. Bot Mar 27:249

Turner S, Burger-Wiersma T, Giovannoni SJ, Mur LR, Pace NR (1989) The relationship of a prochlorophyte *Prochlorothrix hollandica* to green chloroplasts. Nature 337:380–385

Valentin K, Zetsche K (1990) Rubisco genes indicate a close phylogenetic relation between the plastids of Chromophyta and Rhodophyta. Plant Mol Biol 15:575–584

Van de Peer Y, Neefs JM, De Wachter R (1990) Small ribosomal subunit RNA sequences, evolutionary relationship among different life forms, and mitochondrial origins. J Mol Evol 30:463–476

Wetmur JG, Davidson N (1968) Kinetics of renaturation of DNA. J Mol Biol 31:349–370

Wilhelm C, Eisenbeis G, Wild A, Zahn RK (1982) *Nanochlorum eukaryotum:* a very reduced coccoid species of marine chlorophyceae. Z Naturforsch 37C:107–114

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271

Yamada T, Sakaguchi K (1981) Protoplast induction in Chlorella species. Agric Biol Chem 45:1905–1909

Zahn RK (1984) A green alga with minimal eukaryotic features: *Nanochlorum eukaryotum.* Orig Life 13:289–303

Zuker M, Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary informations. Nucleic Acids Res 9:133–148