Phycological Research 1999; **47**: 33–38

Research note

Characteristics and utility of nuclear-encoded large-subunit ribosomal gene sequences in phylogenetic studies of red algae

D. Wilson Freshwater,[1]* Suzanne Fredericq[2] and J. Craig Bailey[3]

[1]Center for Marine Science Research, 7205 Wrightsville Avenue, Wilmington, NC, 28403, USA, [2]Department of Biology, University of Southwestern Louisiana, Lafayette, LA, 70504–2451, USA and [3]Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, ME, 04575–0475, USA

SUMMARY

Primer sequences are described for amplifying and sequencing a large fragment (approximately 2500 b.p.) of the nuclear-encoded large-subunit ribosomal RNA gene (LSU) from red algae. In comparison to RuBisCo large-subunit gene (*rbc*L) and nuclear-encoded small-subunit ribosomal RNA gene (SSU) sequence data, LSU sequence data was intermediate in the number of phylogenetically informative positions and sequence divergence. Parsimony analysis of LSU sequences for 16 Gelidiales species resolved some nodes unresolved in *rbc*L and SSU parsimony trees. An analysis of LSU sequences from 13 species of red algae classified in 11 orders suggests that this gene may be useful in studies of higher-level relationships of red algae.

Key words: Gelidiales, LSU, molecular evolution, phylogenetics, *rbc*L, Rhodophyta, SSU.

INTRODUCTION

Red algal phylogenetic studies involving DNA sequence analyses have predominantly used three sequence regions: the nuclear-encoded small-subunit ribosomal RNA gene (SSU), the internal transcribed spacer regions (ITS) of nuclear ribosomal DNA, and the chloroplast-encoded RuBisCO large-subunit gene (*rbc*L). Studies of species and higher level relationships have generally used analyses of either *rbc*L or SSU sequences to generate phylogenetic hypotheses (e.g. Freshwater and Rueness 1994; Hommersand *et al*. 1994; Ragan *et al*. 1994; Saunders and Kraft 1996), and recent studies have included both separate and combined (Bailey and Freshwater 1997; Vis *et al*. 1998) analyses. Not only have both these sequences proved useful for addressing phylogenetic questions, but the large number of publicly available *rbc*L and SSU sequences can greatly reduce the amount of sequence data which needs to be generated by individual researchers for a particular study.

For these reasons, *rbc*L and SSU sequences are obvious candidates for most phylogenetic studies in red algae. There can, however, be potential problems with the use of both genes. In comparison to SSU data, *rbc*L data sets generally exhibit a higher rate of mutational change (Bailey and Freshwater 1997). Although this rate of change provides a useful tool for resolving interspecific and intergeneric relationships (e.g. Freshwater *et al*. 1995; Fredericq and Ramírez 1996), it may lead to branch attraction problems when addressing phylogenetic questions at higher taxonomic levels (e.g. among orders; Felsenstein 1978; Graybeal 1998). For example, long branch attraction was cited by Freshwater *et al*. (1994, p. 7284) as the probable cause for the spurious position of the Rhodogorgonales within the Florieophycidae based upon *rbc*L analysis. Care must be taken when using *rbc*L sequence analyses for exploring higher level relationships to ensure sufficient sampling of taxa in order to prevent this problem. Furthermore, because of extinction and other stochastic causes, the problem of long branch attraction cannot always be ameliorated by including additional taxa.

In contrast, SSU sequence data sometimes does not exhibit enough variation to provide the necessary phylogenetic signal for addressing some lower level relationships. Whereas species level relationships for a set of Gelidiales taxa were well resolved by *rbc*L data, SSU data for those same taxa did not provide sufficient phylogenetic signal to resolve relationships at that level (Bailey and Freshwater 1997). Bailey and Freshwater (1997) did find that some higher level relationships which could not be determined with the *rbc*L data were resolved with the SSU data, and similar results were reported by Vis *et al*. (1998) in a study of the Batrachospermales.

In this study, the characteristics and utility of sequence data for the nuclear-encoded, large-subunit,

*Email: freshwaterw@uncwil.edu

Communicating editor: S. M. Boo.

Received 14 July 1998; accepted 26 October 1998.

ribosomal RNA gene (LSU) in phylogenetic studies are outlined, and primer sequences for amplifying and sequencing portions of this gene are presented.

MATERIALS AND METHODS

Species and GenBank accession numbers of sequences used in this study are given in Table 1. The generation of *rbc*L and SSU sequence data was as described in Freshwater and Rueness (1994) and Bailey and Freshwater (1997). The LSU sequence data were generated using both manual and automated sequencing methods. Manual sequencing was performed as described in Freshwater and Bailey (1998).

For automated sequencing, amplification products were cleaned of excess primer, enzyme and dNTPs using either Wizard PCR preps (Promega, Madison, WI, USA) or by PEG precipitation (Hillis *et al*. 1996). Sequencing reactions were performed using the Big Dye sequencing kit and protocol (Perkin-Elmer, Foster City, CA, USA), and analyzed on an ABI Prism 310 Genetic Analyzer (PE Applied Biosystems, Foster City, CA, USA). The sequence data were compiled and edited with Sequencher (Gene Codes Corp., Ann Arbor, MI, USA), and aligned using SeqApp (Gilbert 1992). Sequence data sets were analyzed with PAUP (v. 3.1.1, Swofford 1993) and MacClade (v. 3.0, Maddison and Maddison 1992).

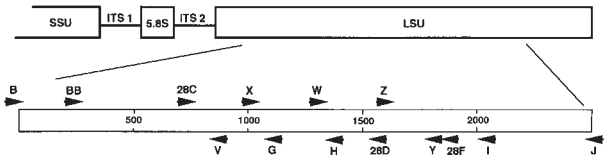
Table 1. Species and GenBank accession numbers of analyzed sequences

Species	Accessions LSU	*rbc*L	SSU
Gelidiales			
Capreolia implexa	AF039545	L22456	U60344
Gelidiella acerosa	AF039551	L22457	U60342
Gelidium americanum	AF039536	L22459	U60347
Gelidium caulacantheum	AF039544	U00103	U60343
Gelidium floridanum	AF039537	U00107	U60351
Gelidium latifolium	AF039540	U00112	U60350
Gelidium pusillum	AF039542	U00999	U60352
Gelidium 'pusillum'	AF039543	U00981	U60355
Gelidium serrulatum	AF039538	U01042	U60340
Gelidium sesquipedale	AF039539	L22071	U60354
Onikusa pristoides	AF039541	U01044	U60353
Pterocladia lucida	AF039550	U01048	U60349
Pterocladiella capillacea	AF039549	U01896	U60346
Pterocladiella melanoidea	AF039548	U01046	U60341
Ptilophora pinnatifida	AF039547	U16834	U60345
Ptilophora subcostata	AF039546	U16835	U60348
Hildenbrandiales			
Hildenbrandia sp.	AFO97880	na	na
Corallinales			
Amphiroa dilatata	AFO97876	na	na
Batrachospermales			
Paralemanea annulata	AFO97877	na	na
Nemaliales			
Cumagloia andersonii	AFO97878	na	na
Palmariales			
Palmaria palmata	AFO97879	na	na
Ceramiales			
Spermothamnion repens	AFO97881	na	na
Rhodymeniales			
Rhodymenia pseudopalmata	AFO97886	na	na
Plocamiales			
Plocamium cartilagineum	AFO97885	na	na
Bonnemaisoniales			
Bonnemaisonia asparagoides	AFO97882	na	na
Gigartinales			
Chondrus crispus	AFO97883	na	na
Weeksia sp.	AFO97884	na	na

na, sequence not analyzed in this study.

RESULTS AND DISCUSSION

Two different LSU data sets were produced. One included 16 Gelidiales species and the other 13 species representing 11 orders which will be referred to as the 'interordinal' data set. The LSU gene in red algae is approximately 3300 b.p. in length. Partial LSU sequences were determined in this study using primer sequences either described by Hamby *et al*. (1988), designed from an analysis of six aligned vascular plant sequences (*Arabidopsis thaliana* [X52320], *Sinapsis alba* [X57137], *Citrus limon* [X05910], *Fragaria ananassa* [X58118], *Lycopersicon esculentum* [X13557], *Oryza sativa* [M11585]), or constructed via primer walking (Fig. 1). All primers listed in Fig. 1 have worked with at least some red algal species but those described by Hamby *et al*. (1988) and those designed from vascular plant sequences have not been used successfully with all taxa included here. For example, because of mismatches in the primer sequence, sequencing reactions with the 28D primer only worked



Primers	Sequence	Gelidiales alignment position
B	CATTGTAGTTGAATATTCTTGG	—
BB	GAACCTGAAACCGTTACG	223
28C[a]	CCGAGTTTCCCTCAGGATAGC	713
X	GCAGGACGGTGGCCATGGAAGT	991
W	GAGTCGTTCCTAAGCCGAGTG	1285
Z	CGTACCGATCACCGCATCAGG	1575
V	CGTATCGCCAGTTCTGCTTACC	892
G	CACCACGTCCTCCTACTC	1129
H	CAACTCTGCCCAGCGCCATC	1399
28D[a]	CTTGGAGACCTGCTGCGG	1586
Y	CGGATTCCCTTTGCCCGTTGC	1828
28F[a]	CAGAGCACTGGGCAGAAATCAC	1902
I[b]	GTGGTTTCGCTGGATAGTAGAC	2058
J[b]	CGCTATGAACGCTTGGCTGC	—

Fig. 1. LSU primer sequences and position within the Gelidiales alignment. [a]Primer sequence from Hamby *et al*. (1988). [b]Primer sequence from alignment of six vascular plant LSU sequences.

if that primer was also used to amplify the template fragment. The automated sequencing scheme included amplification of three separate fragments using primer pairs B-G, 28C-28F and Z-J. Sequence contigs resulting from sequencing reactions using these same primers showed good overlap and were easily assembled to provide data for both strands except at the terminal ends

Gelidiales data set

From 2525 to 2537 b.p. of LSU sequence were determined for 16 Gelidiales species. An alignment of 2519 sites was generated for these sequences. Manual alignment of these sequences was facilitated by the small size of the insertion/deletion mutations (indels), and the conserved nature of sequences adjacent to these sites. Of the 18 indels in this alignment, only an apomorphic insertion found in *Ptilophora subcostata* involved more than three sites. The position of these indels within the alignment indicates that they occur most often in portions of the sequence experiencing the highest rate of point mutations (Fig. 2a).

An exact alignment of these sequences with a secondary structure model for the red algal LSU is presently not possible because none is available, but a comparison was made with the secondary structure model of the land plant, *Arabidopsis thaliana* (De Rijk *et al*. 1997). The basic LSU structure is a multibranched central loop. The structures branching from the central loop are composed of helical paired strands called stems, and unpaired loops. By convention, these major branch structures are labeled A through I consecutively around the central loop (De Rijk *et al*. 1997). The Gelidiales sequences contain a portion of major branch B, all of C, D, E, F, and a portion of major branch G (Fig. 2a).

A comparison of the LSU data set with *rbc*L and SSU data sets generated from the same species shows it to be intermediate in both information content and sequence divergence between taxa (Table 2). The percentage of parsimony informative sites for the *rbc*L, LSU and SSU data sets were 22, 5.8 and 2.2%, respectively. The range of uncorrected LSU sequence divergence values between species was 0.4–2.2%, and between the genera *Gelidium*, *Gelidiella*, *Pterocladia* and *Pterocladiella*, 3.5–5.4%. These values are less than those reported by Freshwater and Bailey (1998) for a LSU data set of only 1032 sites, but are still intermediate in comparison to the sequence divergence values for *rbc*L and SSU data sets (Table 2). The discrepancy with the Freshwater and Bailey (1998) values is the result of adding a more conserved portion of the LSU gene to the alignment. Of the approximate 1500 additional sites (*vis-à-vis* Freshwater and Bailey 1998), approximately 750 were added to the 5′ end and approximately 750 were added to the 3′ end of the original 1032 site alignment. Few fixed changes were found in the additional 3′ end sites for the included

Table 2. Total sites, number and percentage of parsimony informative sites, and percentage sequence divergence values at different taxonomic levels for *rbc*L, SSU, and LSU data sets of 16 Gelidiales species

	*rbc*L	SSU	LSU
Total sites	1467	1646	2519
Informative sites	324 (22%)	36 (2.2%)	146 (5.8%)
Sequence divergence (%)			
Between species	3.1–11.5	0.0–0.4	0.4–2.2
Between genera*	10.2–15.7	0.3–2.9	3.5–5.4

*Due to the uncertain taxonomic status of *Onikusa*, *Capreolia*, and *Ptilophora*, comparisons presented are based only on *Gelidiella*, *Gelidium*, *Pterocladia*, and *Pterocladiella*.

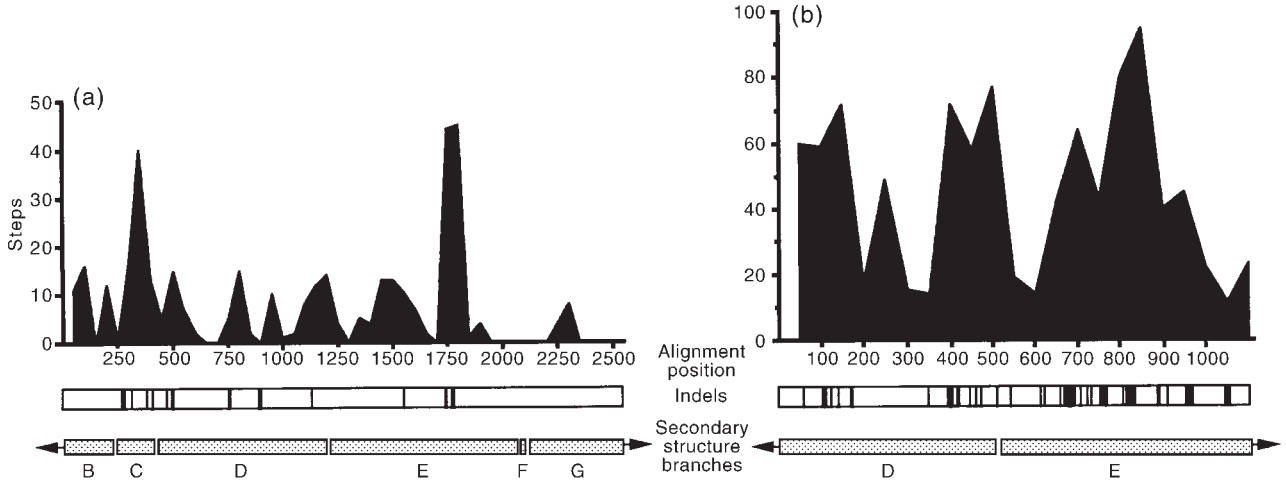


Fig. 2. Mutational changes (measured as steps on minimal parsimony trees) per non-overlapping 50 site segments, position of indels, and secondary structure major branches for LSU gene sequences in two different data sets. (a) 16 Gelidiales species. (b) 13 species representing 11 red algal orders.

species (Fig. 2). The differences in number of fixed mutational changes in different portions of the LSU sequence are also reflected in the number of parsimony informative sites. Whereas 7.7% of the first 1800 sites were informative, only 0.98% of the remaining sites were informative. This information could be important for phylogenetic studies when the question of adding additional sequence or additional taxa to answer a particular question is addressed.

Figure 3 shows the results of separate parsimony analyses with the *rbc*L, SSU, and LSU data sets. Gaps were treated as missing in these analyses. The results were all similar with respect to the overall arrangement of species, but differ in the resolution of different portions of the trees. The *rbc*L tree is fully resolved but there is no bootstrap support for the arrangement of some basal lineages. Parsimony analysis of the SSU data results in 22 minimal length trees, the consensus of which is poorly resolved. The SSU tree does provide strong bootstrap support for positioning *Pterocladiella* among the early diverging Gelidialean lineages. The minimal length LSU tree does not resolve all the branches seen in the *rbc*L tree, but does resolve more than are found by analyses of the SSU data. The arrangement at the base of the LSU tree, which indicates that *Gelidiella* and *Gelidium*-like species are sister taxa, and that *Pterocladia* and *Pterocladiella* are sister taxa, is supported by high bootstrap values (95–100%). In analyses of all three data sets, the order of branching for the *Capreolia implexa*/*Gelidium caulacantheum* and *Ptilophora pinnatifida*/*P. subcostata* lineages is not resolved (Fig. 3). It is uncertain if this is a result of insufficient taxon sampling or phylogenetic history. For a discussion of this question and other taxonomic implications of these trees, see Bailey and Freshwater (1997) and Freshwater and Bailey (1998).

Interordinal data set

The interordinal data set was produced from sequences of the LSU gene fragment amplified using primer pair 28C–28F only. Successful amplifications and sequencing reactions for some taxa were also made with primer pair BB–G, but are not included here. From 1141 to 1240 b.p. of sequence were determined for each species. An alignment of 1073 sites was constructed which excluded the 3′ ends of the sequenced fragments because they could not be confidently aligned. This alignment contains portions of the D and E major branches of the secondary structure model (Fig. 2b). As in the Gelidiales data set, indels are predominantly found in regions of high mutational change (Fig. 2b). There are 60 indels in the alignment which vary from 1 to 36 sites in size. As expected for an alignment of a few phylogenetically diverse taxa, the assignment of gaps is difficult but should be made easier as more closely related species are added. Uncorrected interordinal sequence divergence values for these taxa range from 4.1 to 24.9% when sites where gaps occur are removed. Some of the interordinal divergence values such as the 4.1% between *Cumagloia andersonii* and *Palmaria palmata*, and 5.4% between *Chondrus crispus* and *Bonnemaisonia asparagoides* are low in comparison to the intergeneric differences seen in the Gelidiales (Table 2). A low interordinal sequence divergence between the Nemaliales and Palmariales relative to the intergeneric differences in the Gelidiales is also seen in the SSU distance tree of Saunders and Bailey (1997, p. 1441).

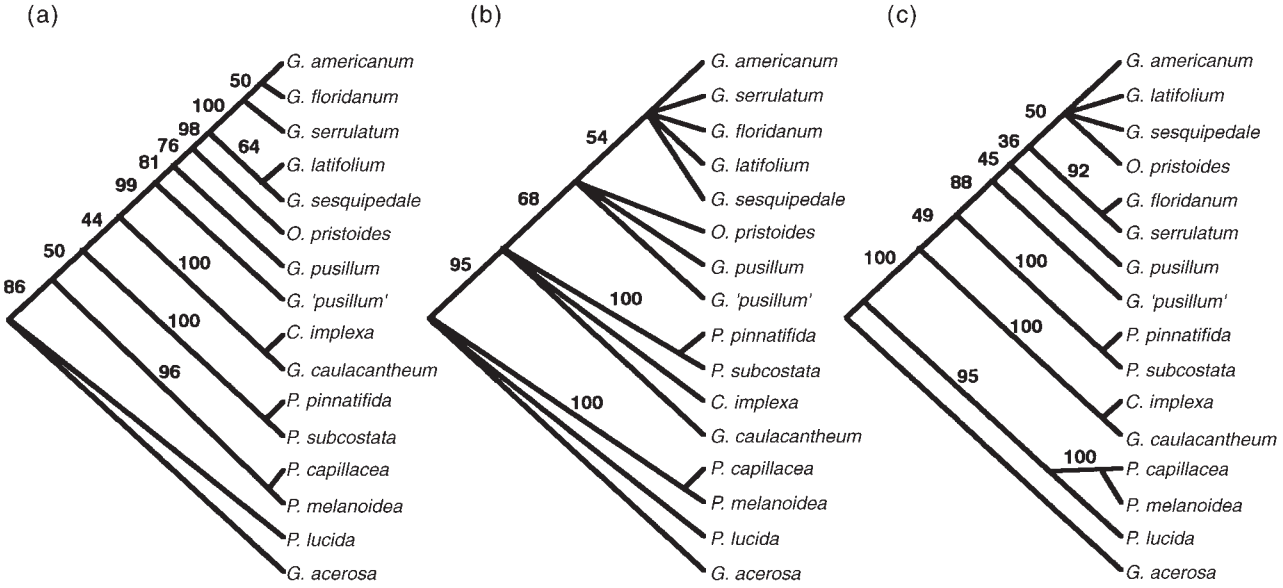


Fig. 3. Trees resulting from parsimony analyses of (a) *rbc*L, (b) SSU and (c) LSU data sets for 16 Gelidiales species: (a) one tree, L = 1086, CI = 0.50, RI = 0.52; (b) 22 trees, L = 93, CI = 0.73, RI = 0.81; (c) one tree, L = 258, CI = 0.65, RI = 0.75. SSU and *rbc*L trees are from Bailey and Freshwater (1997). Bootstrap proportion values are shown above resolved internodes.

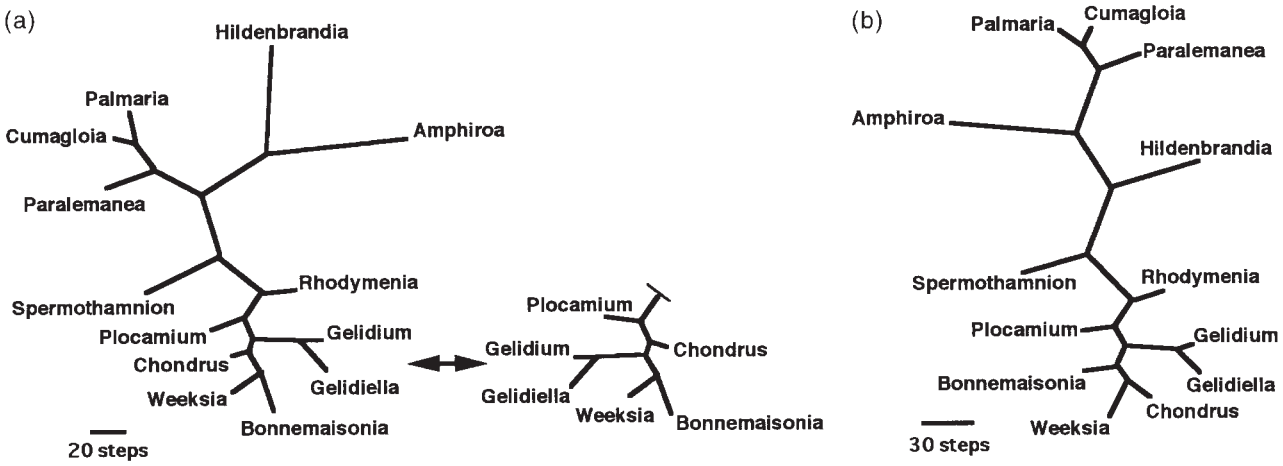


Fig. 4. Unrooted maximum parsimony trees resulting from analyses of LSU sequence data for 13 species representing 11 red algal orders. (a) One of two minimal trees and the placement of taxa in the alternative topology resulting from analysis of the data set when sites where gaps occur were removed (L = 595, CI = 0.58, RI = 0.53). (b) Single minimal tree resulting from analysis of the data set when gap data were coded separately and added to the sequence data (L = 772, CI = 0.59, RI = 0.55). Gelidium = *Gelidium sesquipedale* and Gelidiella = *Gelidiella acerosa*.

Due to the difficulty in exactly aligning these sequences, parsimony analyses were performed on: (i) a data set from which all sites where gaps occur were removed; and (ii) a data set for which gap data were coded separately and combined with the sequence data. Results from analyses of these data sets are shown in Fig. 4. These trees are topologically similar except for the position of *Amphiroa dilatata* and *Chondrus crispus*. Even with the small number of species included in these analyses, major groupings of taxa are consistent with those found in the large SSU trees of Saunders and Bailey (1997). This includes a clade containing the Batrachospermales, Nemaliales, Palmariales, and when gap data are included, the Corallinales. These orders all possess pit plugs with two cap layers (Pueschel 1990, 1994). These trees also resolve the Ceramiales as an early diverging lineage within the clade containing orders having pit plugs with a cap membrane and no outer cap (Fig. 4).

CONCLUSIONS

Primers described for the amplification and sequencing of a majority of the LSU gene have been shown to be widely applicable for a phylogenetically diverse set of red algal species. Sequences generated for this gene are intermediate in the number of informative sites for phylogenetic analyses, and levels of sequence divergence between taxa, as compared to *rbc*L and SSU sequences. LSU sequence data may provide resolution for phylogenetic problems where SSU sequences are uninformative and the sampling requirements for using *rbc*L sequences are prohibitive. Analyses of LSU sequence data in combination with *rbc*L and SSU data have also been effective in producing robust phylogenetic hypotheses (Freshwater and Bailey 1998). Known variation in the fixation of mutational changes across the length of the LSU gene allows researchers to restrict the generation of sequence data to those regions exhibiting the rate of change most appropriate for answering their specific phylogenetic question. The characteristics of this gene indicate that it will be a useful tool, alone and in combination with other genes, for molecular evolutionary studies of red algae.

ACKNOWLEDGEMENTS

The authors wish to thank Dr J. Merrit, director of the Center for Marine Science Research, and the Radiology Department of Columbia Cape Fear Memorial Hospital for facilities and technical support. This project was funded in part by NC SeaGrant project no. R/MER-28 and NSF grant DEB-9726170 to DWF, US Department of Energy grant DE-FG02–97ER12220 and Louisiana Board of Regents grants LEQSF (1997–99)-RD-A-30, and LESQSF (1997–98)-ENH-TR-86 to SF, and NSF grant DEB-9423636 to JCB. DWF also thanks the University of Southwestern Louisiana, Department of Biology for hosting a research sabbatical stay during 1998. This is CMSR contribution number 203.

REFERENCES

Bailey, J. C. and Freshwater, D. W. 1997. Molecular systematics of the Gelidiales: inferences from separate and combined analyses of plastid *rbc*L and nuclear SSU gene sequences. *Eur. J. Phycol.* **32**: 343–52.

De Rijk, P., Van de Peer, Y. and De Wachter, R. 1997. Database on the structure of large ribosomal subunit RNA. *Nucl. Acids Res.* **25**: 117–22.

Felsenstein, J. 1978. Cases in which parsimony or combatibility methods will be positively misleading. *Syst. Zool.* **27**: 401–10.

Fredericq, S. and Ramírez, M. E. 1996. Systematic studies of the Antarctic species of the Phyllophoraceae (Gigartinales, Rhodophyta) based on *rbc*L sequence analysis. *Hydrobiologia* **326/327**: 137–43.

Freshwater, D. W. and Bailey, J. C. 1998. A multigene phylogeny of the Gelidiales including nuclear large-subunit rRNA sequence data. *J. Appl. Phycol.* (in press).

Freshwater, D. W., Fredericq, S., Butler, B. S., Hommersand, M. H. and Chase, M. W. 1994. A gene phylogeny of the red algae (Rhodophyta) based on plastid *rbc*L. *Proc. Natl Acad. Sci. USA* **91**: 7281–5.

Freshwater, D. W., Fredericq, S. and Hommersand, M. H. 1995. A molecular phylogeny of the Gelidiales (Rhodophyta) based on analysis of plastid *rbc*L nucleotide sequences. *J. Phycol.* **31**: 616–32.

Freshwater, D. W. and Rueness, J. 1994. Phylogenetic relationships of some European *Gelidium* (Gelidiales, Rhodophyta) species based on *rbc*L nucleotide sequence analysis. *Phycologia* **33**: 187–94.

Gilbert, D. 1992. SeqApp. Computer program provided by the author. Indiana University, Bloomington, IN.

Graybeal, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* **47**: 9–17.

Hamby, R. K., Sims, L., Issel, L. and Zimmer, E. 1988. Direct ribosomal RNA sequencing: Optimization of extraction and sequencing methods for work with higher plants. *Pl. Mol. Biol. Rep.* **6**: 175–92.

Hillis, D. M., Mable, B. K., Larson, A., Davis, S. K. and Zimmer, E. A. 1996. Nucleic acids IV: Sequencing and cloning. *In* Hillis, D. M., Moritz, C. & Mable, B. K. (Eds). *Molecular Systematics, 2nd ed.* Sinauer Associates, Sunderland, MA, pp. 321–81.

Hommersand, M. H., Fredericq, S. and Freshwater, D. W. 1994. Phylogenetic systematics and biogeography of the Gigartinaceae (Gigartinales, Rhodophyta) based on sequence analysis of *rbc*L. *Bot. Mar.* **37**: 193–203.

Maddison, W. P. and Maddison, D. R. 1992. *MacClade: Analysis of Phylogeny and Character Evolution*. Sinauer Associates, Sunderland, MA, 398pp.

Pueschel, C. M. 1990. Cell structure. *In* Cole, K. M. & Sheath, R. G. (Eds) *Biology of the Red Algae*. Cambridge University Press, New York, pp. 7–41.

Pueschel, C. M. 1994. Systematic significance of the absence of pit-plug cap membranes in the Batrachospermales (Rhodophyta). *J. Phycol.* **30**: 310–15.

Ragan, M. A., Bird, C. J., Rice, E. L., Gutell, R. R., Murphy, C. A. and Singh, R. K. 1994. A molecular phylogeny of the marine red algae based on the nuclear small-subunit rRNA gene. *Proc. Natl Acad. Sci. USA* **91**: 7276–80.

Saunders, G. W. and Bailey, J. C. 1997. Phylogenesis of pit-plug-associated features in the Rhodophyta: inferences from molecular systematic data. *Can. J. Bot.* **75**: 1436–47.

Saunders, G. W. and Kraft, G. T. 1996. Small-subunit rRNA gene sequences from representatives of selected families of the Gigartinales and Rhodymeniales (Rhodophyta). 2. Recognition Halymeniales Ord. Nov. *Can. J. Bot.* **74**: 694–707.

Swofford, D. L. 1993. PAUP: Phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign, IL.

Vis, M. L., Saunders, G. W., Sheath, R. G., Dunse, K. and Entwisle, T. J. 1998. Phylogeny of the Batrachospermales (Rhodophyta) inferred from *rbc*L and 18S ribosomal DNA gene sequences. *J. Phycol.* **34**: 341–50.