

# A new measure to study phylogenetic relations in the brown algal order Ectocarpales: The “codon impact parameter”

SMARAJIT DAS<sup>a</sup>, JAYPROKAS CHAKRABARTI<sup>a,b,\*</sup>, ZHUMUR GHOSH<sup>a</sup>,  
SATYABRATA SAHOO<sup>a,c</sup> and BIBEKANAND MALLICK<sup>a</sup>

<sup>a</sup>Computational Biology Group, Theory Department, Indian Association for the cultivation of Science,  
Kolkata 700 032, India

<sup>b</sup>Biogyan, BF 286, Salt Lake, Kolkata 700 064, India

<sup>c</sup>Raidighi College, South 24, Paraganas, India

\*Corresponding author (Fax, 91-33-2473 2805; Email, tpjc@iacs.res.in)

We analyse forty-seven chloroplast genes of the large subunit of RuBisCO, from the algal order Ectocarpales, sourced from GenBank. Codon-usage weighted by the nucleotide base-bias defines our score called the codon-impact-parameter. This score is used to obtain phylogenetic relations amongst the 47 Ectocarpales. We compare our classification with the ones done earlier.

[Das S, Chakrabarti J, Ghosh Z, Sahoo S and Mallick B 2005 A new measure to study phylogenetic relations in the brown algal order Ectocarpales: The “codon impact parameter”; *J. Biosci.* **30** 699–709]

## 1. Introduction

Algae have grown in importance all over the world. Today they provide nutrition to millions. In this communication we study the family relations amongst an order of brown algae known as Ectocarpales. Brown algae are one of the major seaweeds. We analyse the chloroplast gene, the large subunit of D-ribulose 1,5-bisphosphate carboxylase oxygenase (RuBisCO), to get the codon usage pattern. The score codon-impact-parameter (CIP) is proposed. In the space of these scores the distance metric is defined and then used to estimate relations amongst Ectocarpales.

RuBisCO catalyses the carboxylation of ribulose 1,5-bisphosphate (RuBP) in the Calvin cycle. This bifunctional enzyme at low carbondioxide concentration and high oxygen level catalyses the oxygenation and cleavage of RuBP to form phosphoglycolate and 3-phosphoglycerate (Kellogg and Juliano 1997). The large subunit of RuBisCO, denoted *rbcL* is co-transcribed with small subunit denoted *rbcS*. A spacer separates the two. Because of

the nature of its specific function, *rbcL* is a fairly conserved gene. The relatively low sequence-divergence in *rbcL* and the importance of its function (Morton 1993, 1996) has led to the assumption that change in codon usage is infrequent and unlikely for this gene.

Among the eukaryotic algae, three major groups are recognized on the basis of photosynthetic pigmentation: viz. Chlorophyta, Rhodophyta and Chromophyta. The Chromophyta is further subdivided into many small lineages. The brown algae or Phaeophyceae belong to this group. Ectocarpales are the most examined order within Phaeophyceae. *rbcL* has been studied widely for algae (Hommersand *et al* 1994; McCourt *et al* 1995; Bailey and Freshwater 1997; Daugbjerg and Anderson 1997; Nozaki *et al* 1997; Bailey *et al* 1998; Siemer *et al* 1998; Kogame *et al* 1999; Hanyuda *et al* 2000; Daugbjerg and Guillou 2001). Ectocarpales have been chosen because of their simple structure (Vanden Hoek *et al* 1995). The cells of this family often have discoid chloroplast; the chlorophyll (a, c<sub>1</sub>, c<sub>2</sub>) is masked by other fucoxanthin pigment.

**Keywords.** Ectocarpales; Phaeophyceae; phylogeny; *rbcL* gene; weighted codon usage.

Abbreviations used: CIP, Codon-impact-parameter; IC, impact codon; PR, phylogenetic relations; RuBisCO, D-ribulose 1,5-bisphosphate carboxylase oxygenase; RuBP, ribulose 1,5-bisphosphate.

The earlier classification of the Phaeophyceae was based on thallus construction, mode of growth and type of life history. In the early studies (Kylin 1933; Vanden Hoek and Flinterman 1968; Henry 1984) the following orders were recognized in Phaeophyceae: (i) Ascoseirales (1 family), (ii) Cutleriales (1 family), (iii) Desmarestiales (2 families), (iv) Dictyotales (3 families) (De Clerck *et al* 2001; Lee and Bae 2002), (v) Ectocarpales (23 families), (vi) Fucales (about 8 families) (Rousseau and De Reviers 1999), (vii) Laminariales (8 families), (viii) Ralfsiales (2 families), (ix) Scytothamnales (2 families), (x) Sphaecelariales (about 4 families), (xi) Sporochnales (1 family), (xii) Syringodermatales (1 family) and (xiii) Tilopteridales (1 family).

A clustering approach to brown algal classification using morphology, reproduction and ecological characters (Russell and Fletcher 1975) was done earlier with 132 species. The significant outcome of this study was the suggestion to merge Chordariales, Dictyosiphonales, Tilopteridales and Scytosiphonales into Ectocarpales *Sensu lato*. Cytological characters, such as plastid structure, also play a major role for better understanding of the systematics of brown algae (Peters and Clayton 1998; Muller *et al* 1998). Molecular phylogenies of the morphologically simple brown algae based on nuclear r-DNA sequences indicate that the taxa with predunculated pyrenoids form the monophyletic group the 'Ectocarpales' (Rousseau and De Reviers 1999). Phylogenetic relations (PR) based on molecular analysis in Chromophyta (Bhattacharya *et al* 1992; Leipe *et al* 1994; Bhattacharya and Medlin 1995; Medlin *et al* 1997; Ben Ali *et al* 2001; Draisma *et al* 2001) have involved RuBisCO as well as 18s and 26s r-DNA. In recent years the molecular evidences using ribosomal genes and RuBisCO have divided Ectocarpales into five families namely, Chordariaceae, Actinetosporaceae, Adenocystaceae, Ectocarpaceae and Sytosiphonaceae.

Codon-usage data of the large subunit of RuBisCO has been studied to define the CIP. From the z-score of test statistic in the space of the CIP, the most significant codons have been identified; namely, the impact codons (IC). These IC provide a sufficiently sensitive measure to explore phylogenetic relations, PR. The relatively homogeneous order has been classified by us; namely, Ectocarpales, belonging to brown algae (Phaeophyceae) phylum using the CIP scores of IC of rbcL. Using the standard pair group average analysis and the city block distance analysis we have obtained PR among 47 Ectocarpales.

## 2. Methods

Let  $PQR$  denote the codon.  $P$  is the nucleotide at the first codon position,  $Q$  the nucleotide in the second and  $R$  in

the third.  $P$ ,  $Q$  and  $R$  could be any of the four nucleotides A, C, G, T. The frequency,  $F_{PQR}$ , is given by

$$F_{PQR} = \frac{n_{PQR}}{\sum_{a=1}^{20} \sum_{s=1}^{K_a} X_{a,s}} \quad (1)$$

$n_{PQR}$  is the number of occurrences of the codon  $PQR$ . The index  $\alpha$  is the amino acid index going from 1 to 20, and the symbol  $s$  refers to synonymous codons. The sum over  $s$  goes from 1 to  $K_a$ , where  $K_a$  is the number of synonymous codons of the  $a$ th amino acid. Thus  $X_{a,s}$  is the number of occurrence of the  $s$ th synonymous codon of the  $a$ th amino acid.

The distribution of the nucleotides over the three-codon positions is different. The CIP score takes that into account. We define the CIP score of the codon  $PQR$  as:

$$(CIP)_{PQR} = \frac{F_{PQR}}{f_1(P)f_2(Q)f_3(R)}, \quad (2)$$

where  $f_1(P)$  is the frequency of the nucleotide  $P$  at the first codon position of the codon,  $f_2(Q)$  is the frequency of the nucleotide  $Q$  at the second codon position and  $f_3(R)$  refers to the frequency of the nucleotide  $R$  at the third codon position.

If  $\bar{X}$  and  $m$  denote the sample mean (mean of the CIP scores for a particular codon) and population mean (mean of sample mean) respectively; and  $s$  the population standard deviation, then  $z$  score of a test statistics is given by,

$$z = \frac{\bar{X} - m}{s / \sqrt{N}}, \quad (3)$$

where  $N$  is the total number of codons, i.e. 64.

The PR in our case is derived as follows: If  $d(C_l, C_m)$  is the distance between the Ectocarpales  $C_l$  and  $C_m$  ( $l, m$  could take any value from 1 to 47, see table 1) then:

$$d(c_l, c_m) = \sum |CIP(c_l) - CIP(c_m)| \quad (4)$$

Where the sum runs over all the ICs. The tree is generated by UPGMA method using arithmetic averages and city-block (Manhattan) distance analysis.

## 3. Results and discussions

The objective of the present study is to evaluate the codon usage of rbcL gene of Ectocarpales and reinvestigate PR within them to reassess the recently conceived family relations. For Ectocarpales, about 47 rbcL sequences are available. These are chosen from GenBank and are denoted as  $C_1, C_2, \dots, C_{47}$  (table 1).

For calculating the codon usage, it is observed that some codons appear more often than expected from the

underlying nucleotide composition of the gene. There are others that appear as expected and then there are others with lower than expected frequency (Ikemura 1981, 1985; Gour and Gautier 1982; Bennetzen and Hall 1982). The “codon preference bias” proposed by McLachlan *et al* (1984) quantify the degree of bias in codon usage and assess the relative merits of different codon from the viewpoint of translational efficiency. The “codon bias index” of Bennetzen and Hall (1982) calculate the frequency of optimal codons in a gene. Gribskov *et al* (1984) proposed another index the “codon preference statistics” based on the ratio of the likelihood of finding a particular codon in a highly expressed gene to the likelihood of finding the codon in a random sequence with the same base composition.

Every model has its own applicative advantage over others. We feel that CIP score of a codon of a gene, as defined above, contain somewhat more quantitative information since it considers codon usage as well as the base compositional bias. We quantify the degree of codon bias in such a way that comparisons can be made both within and between species. Our approach to this problem is to define a measure for assessing the degree of relative abundance of a set of codons taking account of DNA modification and mutational rates on the compositional patterns of genomic sequences. Our statistic has the advantage that it measures the frequency of optimal

codons from their underlying nucleotide composition of the gene. Moreover this codon bias statistic is defined in such a way that it has twin advantage of being simple to calculate yet making greater quantitative use of available information. We identify our impact codons, IC, chosen based on the level of significance from the  $z$  score of test statistic.

The ICs are selected based on the  $z$  score used in test statistic to establish and test hypothesis or the region of significance. If the null hypothesis that there is no significant difference is not rejected, there is no significant bias for the codon, and no over-representation of codons. One tailed test to examine the null hypothesis ( $H_0$ ), that CIP value of every codon is close to unity (as in random sequences) is performed. The CIP scores of the codons of RbcL differentiate the codons into different levels of significance. Three sets of codons identified at three different significance levels (0.01, 0.05, 0.1) and the alternative hypothesis, that a set of codons has a higher abundance with respect to their nucleotide composition, is accepted. These set of codons are IC. ICs are the ones that are significant above 90% level.

Classification of organisms based on molecular genomic methods depend crucially on the underlying arrangement/distribution of nucleotides or codons. If the biological group being studied is homogeneous, or the gene being studied for PR is well conserved (Kimura

**Table 1.** The 47 Phaeophyceae with their GenBank accession numbers.

Species	GenBank accession number	Species	GenBank accession number
<i>Delamarea attenuata</i> (C <sub>1</sub> )	AF055396	<i>Scytosiphon tenellus</i> (C <sub>25</sub> )	AB022241
<i>Dictyosiphon foeniculaceus</i> (C <sub>2</sub> )	AF055397	<i>Scytosiphon gracilis</i> (C <sub>26</sub> )	AB022240
<i>Giraudia sphacelarioides</i> (C <sub>3</sub> )	AF055399	<i>Scytosiphon canaliculatus</i> (C <sub>27</sub> )	AB022239
<i>Hummia onusta</i> (C <sub>4</sub> )	AF055402	<i>Scytosiphon lomentaria</i> (C <sub>28</sub> )	AB022238
<i>Myriotrichia clavaeformis</i> (C <sub>5</sub> )	AF055408	<i>Petalonia binghamiae</i> (C <sub>29</sub> )	AB022244
<i>Punctaria plantaginea</i> (C <sub>6</sub> )	AF055410	<i>Petalonia fascia</i> (C <sub>30</sub> )	AB022243
<i>Stictyosiphon soriferus</i> (C <sub>7</sub> )	AF055413	<i>Petalonia zosterifolia</i> (C <sub>31</sub> )	AB022242
<i>Striaria attenuata</i> (C <sub>8</sub> )	AF055415	<i>Rosenvingea intricate</i> (C <sub>32</sub> )	AB022232
<i>Coelocladia arctica</i> (C <sub>9</sub> )	AF055395	<i>Chnoospora implexa</i> (C <sub>33</sub> )	AB022231
<i>Isthmoplea sphaerophora</i> (C <sub>10</sub> )	AF055403	<i>Hydroclathrus clathratus</i> (C <sub>34</sub> )	AB022233
<i>Litosiphon pusillus</i> (C <sub>11</sub> )	AF055406	<i>Asperococcus fistulosus</i> (C <sub>35</sub> )	AY095321
<i>Pogotrichum filiforme</i> (C <sub>12</sub> )	AF055409	<i>Chordaria flagelliformis</i> (C <sub>36</sub> )	AY095324
<i>Ectocarpus siliculosus</i> (C <sub>13</sub> )	X52503	<i>Hecatonema</i> sp. 86 (C <sub>37</sub> )	AF207802
<i>Elachista fucicola</i> (C <sub>14</sub> )	AF055398	<i>Microspongium globosum</i> (C <sub>38</sub> )	AF207805
<i>Hecatonema</i> sp. (C <sub>15</sub> )	AF055401	<i>Myelophycus simplex</i> (C <sub>39</sub> )	AY095320
<i>Streblonema tenuissimu</i> (C <sub>16</sub> )	AF055414	<i>Myelophycus cavum</i> (C <sub>40</sub> )	AY095319
<i>Laminariocolax tomentosoides</i> (C <sub>17</sub> )	AF055404	<i>Polytretus reinboldii</i> (C <sub>41</sub> )	AF207809
<i>Myrionema strangulans</i> (C <sub>18</sub> )	AF055407	<i>Punctaria latifolia</i> (C <sub>42</sub> )	AY095322
<i>Sorocarpus micromorus</i> (C <sub>19</sub> )	AF055411	<i>Protectocarpus speciosus</i> (C <sub>43</sub> )	AF207810
<i>Pilayella littoralis</i> (C <sub>20</sub> )	X55372	<i>Feldmannia irregularis</i> (C <sub>44</sub> )	AF207800
<i>Colpomenia phaeodactyla</i> (C <sub>21</sub> )	AB022237	<i>Halothrix lumbricalis</i> (C <sub>45</sub> )	AF207801
<i>Colpomenia bullosa</i> (C <sub>22</sub> )	AB022236	<i>Hincksia hincksiae</i> (C <sub>46</sub> )	AF207803
<i>Colpomenia peregrina</i> (C <sub>23</sub> )	AB022235	<i>Mikrosyphar porphyrae</i> (C <sub>47</sub> )	AF207806
<i>Colpomenia sinuosa</i> (C <sub>24</sub> )	AB022234	–	–

**Table 2.** The CIP score of (12 + 1 + 1 = 14) impact codons (IC) of the ectocarpales are listed.

	Leu TTA	Arg CGT	Tyr TAC	Phe TTC	Met ATG	Asn AAC	Trp TGG	Gly GGT	Glu GAA	Ile ATC	Lys AAA	Gln CAA	Thr ACT	Pro CCA
C <sub>1</sub>	2.943	2.484	2.876	3.591	6.036	2.607	5.509	1.872	1.509	1.864	1.784	1.527	1.301	1.073
C <sub>2</sub>	2.942	2.336	3.121	3.478	6.085	2.311	5.563	1.944	1.484	1.416	1.843	1.544	1.328	1.266
C <sub>3</sub>	2.823	2.408	4.069	3.81	5.78	1.953	5.406	1.875	1.528	1.98	1.687	1.633	1.314	1.23
C <sub>4</sub>	2.67	2.419	3.647	3.026	5.318	3.126	5.265	1.93	1.493	1.321	1.909	1.47	1.33	1.208
C <sub>5</sub>	2.829	2.472	3.45	3.523	5.719	3.011	5.323	1.935	1.574	1.794	1.806	1.386	1.458	1.429
C <sub>6</sub>	2.925	2.505	3.195	2.937	5.919	2.889	5.359	1.831	1.5	1.879	1.787	1.517	1.322	1.066
C <sub>7</sub>	2.892	2.27	3.494	2.658	5.388	2.796	5.015	1.763	1.517	1.418	1.629	1.534	1.422	1.428
C <sub>8</sub>	2.898	2.471	3.763	3.524	6.036	3.01	5.618	2.049	1.501	2.05	1.784	1.507	1.312	1.235
C <sub>9</sub>	2.867	2.314	3.16	3.847	5.88	2.708	5.623	1.853	1.542	1.997	1.794	1.561	1.35	1.282
C <sub>10</sub>	2.94	2.171	2.309	3.243	5.567	2.748	5.132	1.82	1.531	1.872	1.861	1.588	1.296	1.281
C <sub>11</sub>	2.972	2.571	3.526	3.602	5.617	3.111	5.135	2.092	1.547	1.956	1.635	1.527	1.508	1.252
C <sub>12</sub>	2.596	2.185	2.56	2.963	5.669	2.469	5.374	2.034	1.421	2.286	1.558	1.451	1.251	1.359
C <sub>13</sub>	2.791	2.263	1.469	4.069	5.265	1.842	5.303	1.763	1.275	2.783	1.903	1.637	1.237	1.202
C <sub>14</sub>	2.905	2.482	3.916	3.362	5.571	2.679	5.081	2.011	1.594	1.729	1.653	1.536	1.405	1.072
C <sub>15</sub>	3.11	2.461	2.985	3.364	5.432	2.459	4.858	1.885	1.568	2.217	1.768	1.609	1.285	1.309
C <sub>16</sub>	2.795	2.514	3.499	3.574	6.036	3.032	5.619	2.085	1.555	2.144	1.784	1.375	1.483	1.235
C <sub>17</sub>	2.69	2.151	2.679	3.737	5.3	2.626	5.034	1.925	1.477	2.13	1.83	1.386	1.372	1.23
C <sub>18</sub>	3.072	2.427	3.371	3.703	5.437	2.543	5.011	1.954	1.458	1.172	1.718	1.332	1.476	1.334
C <sub>19</sub>	2.887	2.537	3.526	3.303	5.874	1.662	5.413	1.96	1.511	2.183	1.795	1.537	1.329	1.259
C <sub>20</sub>	2.873	2.284	2.619	3.567	5.616	2.393	5.031	1.941	1.256	1.711	1.675	1.491	1.161	1.233
C <sub>21</sub>	2.941	2.22	3.621	3.699	4.899	1.583	4.552	1.954	1.547	1.886	1.731	1.423	1.162	1.283
C <sub>22</sub>	2.941	2.248	3.726	3.49	4.899	1.493	4.552	1.979	1.547	2.034	1.731	1.423	1.177	1.283
C <sub>23</sub>	2.99	2.266	2.473	3.425	5.3	1.748	4.833	1.946	1.469	2.515	1.654	1.58	1.255	1.266
C <sub>24</sub>	2.795	2.21	2.107	3.638	5.3	1.445	4.83	1.904	1.468	2.668	1.562	1.528	1.132	1.235
C <sub>25</sub>	2.745	2.247	3.386	3.459	5.527	1.466	5.186	2.005	1.482	1.747	1.658	1.345	1.342	1.213
C <sub>26</sub>	2.87	2.225	3.416	3.489	5.389	1.493	4.954	2.016	1.509	1.78	1.689	1.527	1.113	1.083
C <sub>27</sub>	3.008	2.206	3.254	3.574	5.389	1.408	5.003	1.932	1.58	2.127	1.693	1.608	1.301	1.288
C <sub>28</sub>	2.876	2.257	2.966	3.61	5.013	1.398	4.751	1.945	1.54	1.891	1.815	1.405	1.187	1.247
C <sub>29</sub>	2.807	2.362	3.091	3.444	4.667	1.618	4.507	2.006	1.43	1.888	1.857	1.372	1.099	1.238
C <sub>30</sub>	2.756	2.218	3.731	3.494	5.719	1.724	5.377	1.995	1.413	1.761	1.741	1.319	1.277	1.19
C <sub>31</sub>	2.818	2.239	3.482	3.233	5.826	1.776	5.413	1.971	1.482	1.555	1.738	1.5	1.245	1.229
C <sub>32</sub>	2.899	2.178	1.783	3.541	5.351	1.684	4.855	1.966	1.426	2.628	1.654	1.517	0.999	1.043
C <sub>33</sub>	3.002	2.274	1.585	3.463	5.831	1.567	5.197	1.822	1.382	1.815	1.848	1.415	1.043	1.249
C <sub>34</sub>	3.008	2.114	1.53	2.774	5.094	2.266	4.516	1.925	1.397	1.775	1.719	1.252	1.076	1.06
C <sub>35</sub>	2.967	2.567	3.326	3.706	5.73	3.207	5.229	2.071	1.622	2.268	1.716	1.561	1.326	1.28
C <sub>36</sub>	2.858	2.243	1.799	3.675	5.583	2.18	5.145	2.177	1.503	1.979	1.696	1.5	1.349	1.23
C <sub>37</sub>	2.812	2.175	3.364	3.436	5.443	3.197	5.118	1.898	1.495	2.722	1.615	1.48	1.547	1.214
C <sub>38</sub>	2.8	2.52	3.506	3.581	6.048	3.039	5.63	2.089	1.558	2.149	1.787	1.372	1.486	1.237
C <sub>39</sub>	3.1	2.299	2.508	2.827	4.536	1.364	4.324	1.89	1.528	1.384	1.806	1.363	1.063	1.368
C <sub>40</sub>	3.147	2.289	2.257	2.943	4.444	1.578	3.971	1.881	1.554	1.067	1.856	1.188	1.134	1.342
C <sub>41</sub>	2.8	2.361	3.733	3.813	6.404	2.527	5.961	2.08	1.558	1.877	1.787	1.372	1.406	1.237
C <sub>42</sub>	3.136	2.456	3.053	3.441	5.311	2.516	4.749	1.936	1.581	2.268	1.783	1.474	1.282	1.32
C <sub>43</sub>	2.846	2.156	3.09	3.857	5.311	2.991	4.993	1.939	1.46	1.944	1.635	1.498	1.534	1.228
C <sub>44</sub>	2.856	2.377	3.633	2.855	4.808	3.112	5.145	2.041	1.406	1.816	1.534	1.467	1.311	1.203
C <sub>45</sub>	2.8	2.475	4.461	3.645	6.048	2.855	5.63	2.033	1.512	1.215	1.787	1.489	1.406	1.221
C <sub>46</sub>	2.96	2.175	1.666	3.403	5.838	2.697	5.37	1.946	1.504	2.204	1.486	1.549	1.242	1.269
C <sub>47</sub>	2.875	2.487	3.298	3.368	6.221	2.907	5.733	2.016	1.512	1.732	1.787	1.509	1.394	1.237

First 12 IC with 99% level of confidence. Thirteenth IC (Thr ACT) appears when the level of significance relaxed to 0.05. Pro CCA appears as 14th IC when the significance level is 0.1.

1981), the resolution of the individual entities from its nearest neighbours becomes ambiguous. Usually several alternate methods have to be used to map out the evolutionary pathway. The most common approaches to the tree construction involve UPGMA (Sokal and Michener 1958), neighbour-joining (Saitou and Nei 1987), maximum parsimony (Farris 1970) and maximum likelihood (Cavalli-Sforza and Edward 1967).

We focus here on the distance method for tree building. The distance-based algorithms of tree building begin with a set of distances between each pair of sequences in a given dataset. There are many different ways of defining distance: Euclidean distance, City-Block distance, Chebychev distance, Power distance etc. The Chebychev distance measure may be appropriate in cases when one wants to define two objects as 'different', if they are different on any one of the dimensions. Sometimes one may want to increase or decrease the progressive weight that is placed on dimensions on which the respective objects are very different. This can be accomplished via the Power distance. On the other hand, the City-Block (Manhattan)

distance is the average difference across dimensions. In most cases this distance measure yields results similar to simple Euclidean distance. This City-Block distance is computed from raw data. The advantage is that the distance between any two species is not affected by the addition of new species into the analysis. It needs to be emphasized that the different codon bias scores or the alignment algorithms, which estimate the differences between heterogeneous genes, fail for almost homologous genes. Here *rbcL* is well conserved both at amino acid level and at codon level. For instance, some impact codons like Met ATG and Trp TGG are absolutely conserved within *rbcL* gene. But CIP values of these codons are different for different species due to unequal base composition. The CIP scores of IC of *rbcL* give us 47 matrices corresponding to 47 Ectocarpales (table 2). Using the un-weighted pair group average analysis and the City Block (Manhattan) distance analysis we obtain the PR amongst the 47 samples. We compare these with the presently conjectured relations (Draisma *et al* 2002). Interestingly, we find the matrices of CIP scores of just the

**Table 3.** Z-values for codons.

Codon	Corresponding		Codon	Corresponding	
	$\bar{X}$ values	Z values		$\bar{X}$ values	Z values
GCT	1.20268	1.21132	CGG	0	-7.36538
GCA	1.19424	1.15114	ATT	0.98822	-0.31805
GCC	0.87116	-1.15285	ATA	0.24232	-5.63732
GCG	0.83103	-1.43903	ATC	1.91191	<b>6.26908</b>
GGT	1.95517	<b>6.57758</b>	TAT	0.97882	-0.38509
GGA	0.55458	-3.41049	TAC	3.05117	<b>14.39353</b>
GGC	0.73613	-2.11579	TCT	0.83832	-1.38704
GGG	0.57971	-3.23128	AGT	0.41407	-4.41251
CTT	0.83107	-1.43874	TCA	0.68229	-2.49974
TTA	2.89406	<b>13.27312</b>	AGC	0.1678	-6.16874
CTA	0.74133	-2.07871	TCC	0.08189	-6.78140
TTG	0.00772	-7.31033	TCG	0.13728	-6.38639
CTC	0.00846	-7.30505	AAA	1.7352	<b>5.00890</b>
CTG	0.13397	-6.41000	AAG	0.62251	-2.92605
GTT	0.90499	-0.91159	TTT	0.51028	-3.72640
GTA	0.52789	-3.60082	TTC	3.44738	<b>17.21904</b>
GTC	0.03873	-7.08919	CCT	1.15579	0.87694
GTG	0.84426	-1.34468	CCA	1.24198	<b>1.49159</b>
ACT	1.29806	<b>1.89151</b>	CCC	0.84826	-1.31615
ACA	1.05971	0.19176	CCG	0.27211	-5.42487
ACC	0.20436	-5.90802	ATG	5.51917	<b>31.99367</b>
ACG	0.4669	-4.03576	AAT	0.3442	-4.91078
GAT	0.99533	-0.26735	AAC	2.32	<b>9.17931</b>
GAC	0.50881	-3.73689	CAA	1.46921	<b>3.11204</b>
GAA	1.49518	<b>3.29724</b>	CAG	0.73894	-2.09575
GAG	0.24278	-5.63404	CAT	0.78297	-1.78176
CGT	2.33334	<b>9.27444</b>	CAC	0.84628	-1.33027
AGA	0.92972	-0.73524	TGT	0.86632	-1.18736
CGA	0.76312	-1.92332	TGC	0.01852	-7.23331
AGG	0.32997	-5.01226	TGG	5.12951	<b>29.21487</b>
CGC	0.15507	-6.25953	-	-	-

IC are highlighted.

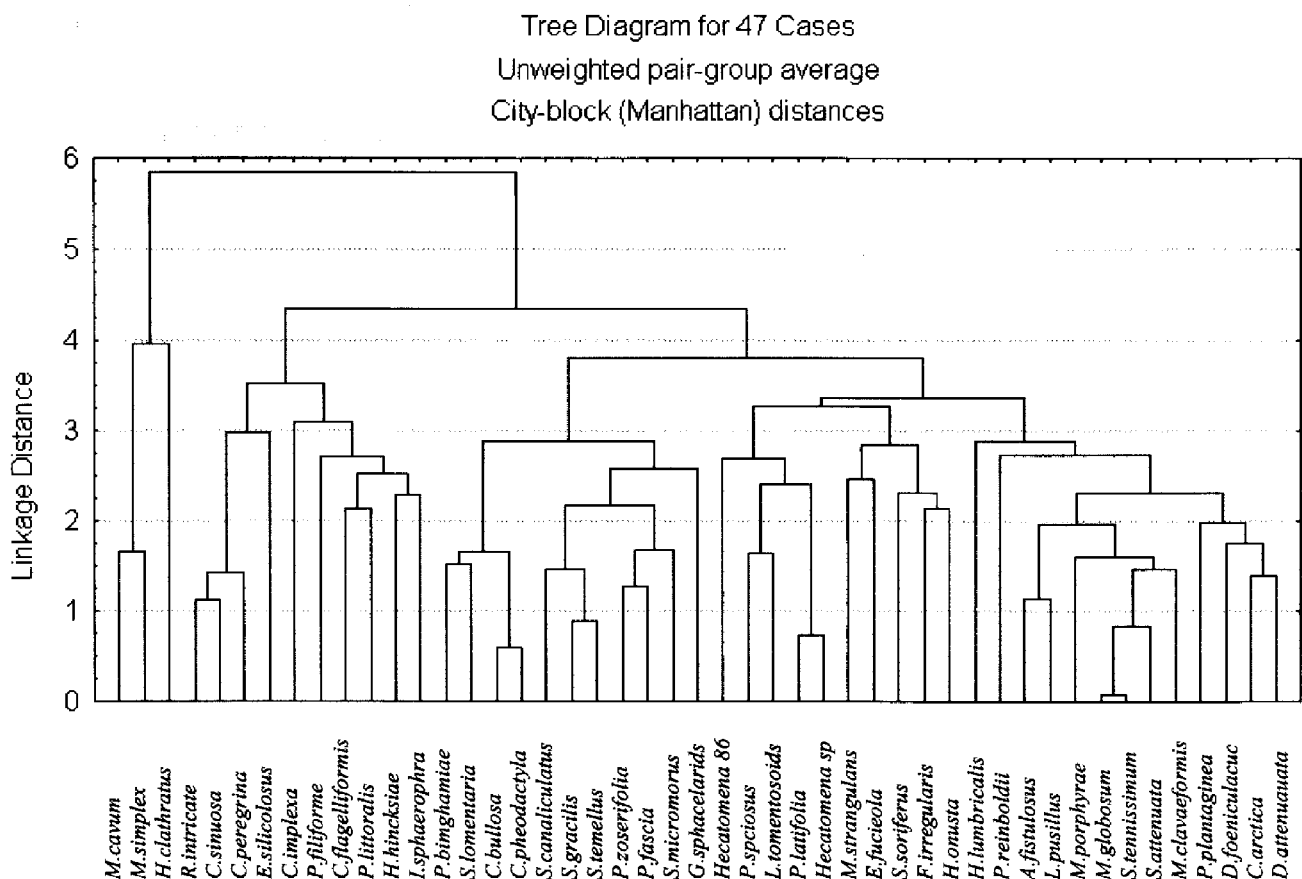
IC largely reproduce the classification for the Ectocarpales. The few differences that we have are discussed.

The *rbcL* gene in all 47 cases is of length 1467 bp. Table 2 gives the CIP scores (Eq. 2) for the impact codons of *rbcL*. The  $z$  score (Eq. 3) of the CIP-score-sample-statistic is significant; the null hypothesis rejected at 0.01 levels and IC are obtained. Thus at the 99% level of confidence (table 3) we have the following IC; Tyr TAC, Phe TTC, Met ATG, Trp TGG, Leu TTA, Arg CGT, Asn AAC, Gly GGT, Ile ATC, Lys AAA, Gln CAA, and Glu GAA. These are 12 in number.

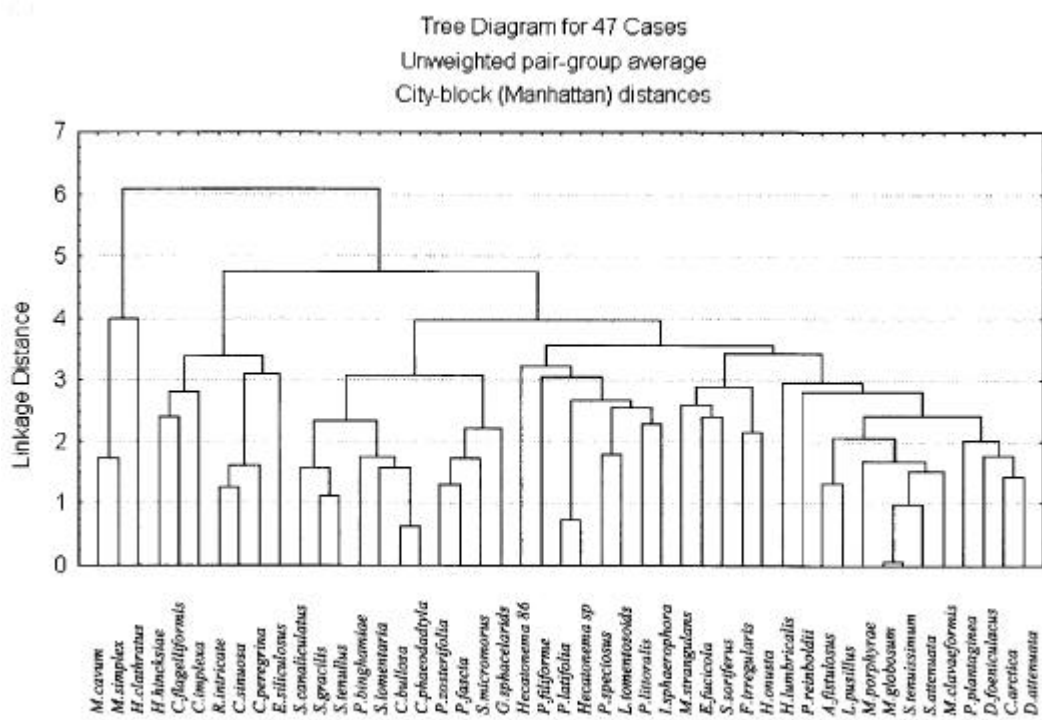
If we relax the level of significance and allow for  $z$  to be 0.05 (instead of 0.01), we have to add to the above list one extra codon, namely Thr ACT. That takes IC number to 13 at 95% level of confidence. And, if we relax our selection criteria a notch lower,  $z$  score (Eq. 3) set at 0.1, we get one more codon in our list, namely Pro CCA. We use Eq. 4 to measure the distance between the matrices and then apply UPGMA to generate figure 1. Similarly for  $z$  at 0.05, the 47 matrices all have 13 rows and one

column, we get figure 2. Figure 3 is for  $z$  at 0.1 level of significance. Figure 4 is the bootstrap consensus dendrogram. It has been drawn by UPGMA method using Mega3 software.

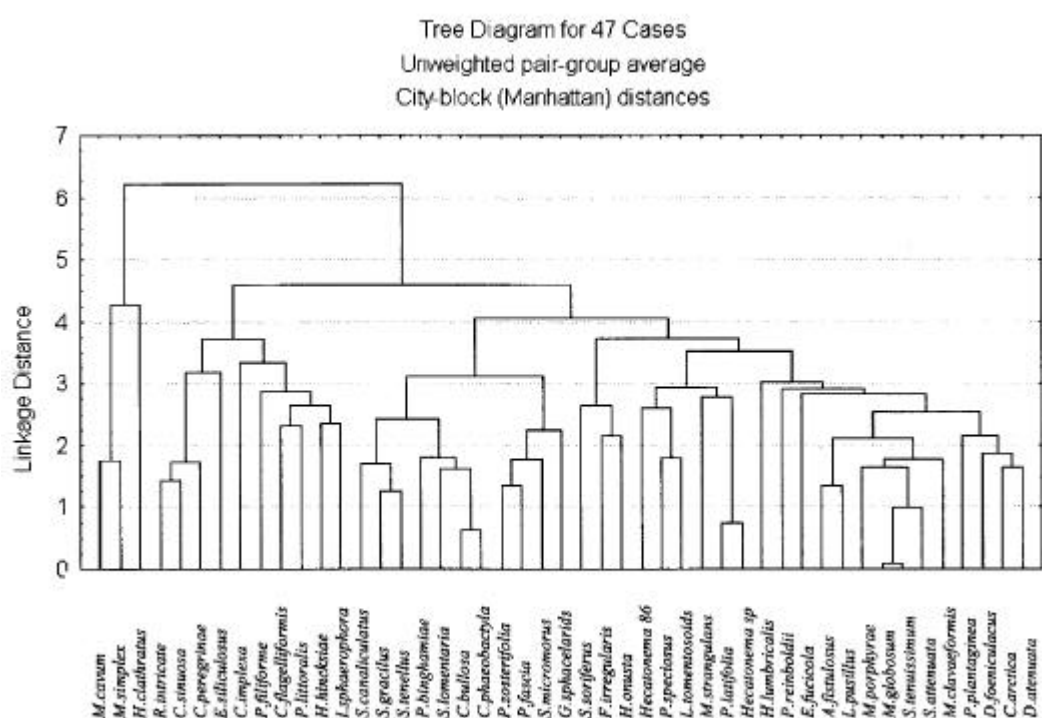
In the bootstrap consensus tree, figure 4, the position of the species largely agree with the present classification, figure 5, of Ectocarpales (Peters and Ramirez 2001; Draisma *et al* 2002). Some differences between the previous classification and our molecular tree are there for the genera *Colpomena*, *Scytosiphon* and *Petalonia*. Species of these genera, specially  $C_{23}$  and  $C_{24}$ , do not clade with  $C_{21}$  and  $C_{22}$  in our molecular phylogeny (figures 1–3) and are scattered in the overall consensus bootstrap tree. One reason could be the phylogenetic tree built with just codon usage in *rbcL* is not sufficient. Other genes have to be considered as well. With just *rbcL* the amino acid content and the IC configurations of  $C_{24}$  are identical with  $C_{32}$ . Hence these cluster together. A close relationship between *Scytosiphon*, *Petalonia* along with *Giraudia* and *Sorocarpus* is suggested by our data.



**Figure 1.** Phylogenetic analysis of 47 Ectocarpales species. The tree drawn according to unweighted pair group method using arithmetic mean (UPGMA) and city-block (Manhattan) distance analysis at 0.01 level of significance. 12 codons have been used.



**Figure 2.** Phylogenetic analysis of 47 Ectocarpales species. The tree drawn according to unweighted pair group method using arithmetic mean (UPGMA) and city-block (Manhattan) distance analysis at 0.05 level of significance. 13 codons have been used.



**Figure 3.** Phylogenetic analysis of 47 Ectocarpales species. The tree drawn according to unweighted pair group method using arithmetic mean (UPGMA) and city-block (Manhattan) distance analysis at 0.1 level of significance. 14 codons have been used.

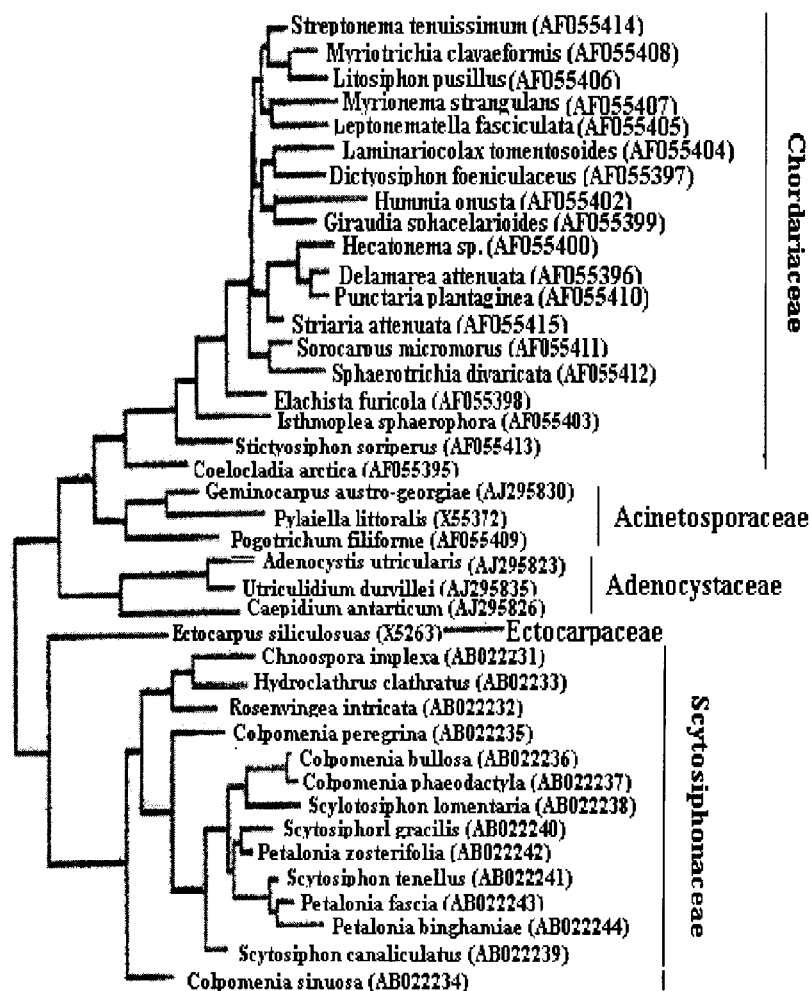


Figure 4. Consensus bootstrap phylogeny of 47 Ectocarpales.

The *Delamerea*, *Dictyosiphon*, *Punctaria* and *Coleocladia* (C<sub>1</sub>, C<sub>2</sub>, C<sub>6</sub> and C<sub>9</sub>) clade is highly supported in all analysis (figures 1–3). It is noted that both *Delamereae* and *Punctaria* have hecatonematoid microthalli and true Phaeophyceae hairs with basal sheaths. Our study suggests that *Delamereaceae* is closely related to *Punctariaceae* (Siemer *et al* 1998) along with *Dictyosiphonaceae* and *Coleocladaceae*. Consensus bootstrap tree supports our data that C<sub>1</sub>, C<sub>2</sub> and C<sub>6</sub> are closely clustered, but diverge for C<sub>9</sub>. This divergence is noticeable for *Dictyosiphonaceae* in figure 5. One reason is that the whole set of RbCL genes of Ectocarpales were not available earlier (Draisma *et al* 2002). But the divergence does not destroy the family structure of Chordariaceae. Chordariaceae is understood here as the largest and most diverse clade within Ectocarpales. Chordariaceae includes several species of so far accepted families such as *Hummia*, *Myrio-*

*trichia*, *Stictyosiphon*, *Striaria*, *Isthmoplea*, *Litosiphon*, *Elachista*, *Hecatonema*, *Stebulonema* etc. Our data however suggest that there is one exception. *Myelophycus* (C<sub>39</sub>, C<sub>40</sub>) clusters with (C<sub>34</sub>), which belong to *Scytosiphonaceae*.

Acinetosporaceae contains taxa with discoid plastids, a filamentous to parenchymatous thallus structure. In our analysis, at 99% level of confidence, we find this clade contains the genera *Pogotrichum* (C<sub>12</sub>), *Pilayella* (C<sub>20</sub>), *Feldmannia* (C<sub>44</sub>) and *Hincksia* (C<sub>46</sub>). Here in our analysis three of these (C<sub>12</sub>, C<sub>20</sub> and C<sub>46</sub>) belong to the clade along with *Isthmoplea* (C<sub>10</sub>) and *Chordaria* (C<sub>36</sub>). C<sub>44</sub> is in a separate cluster. Whereas at 95% level of confidence C<sub>12</sub>, C<sub>20</sub> and C<sub>44</sub> cluster but C<sub>46</sub> is in a separate group. Therefore, Acinetosporaceae overlaps with Chordariaceae and Scytosiphonaceae. Our phylogenetic analysis suggests that a further analysis, perhaps with other genes, is necessary to segregate these two families.



**Figure 5.** Present classification of Ectocarpales (Draisma *et al* 2002). See text for details.

*Ectocarpus silliculosus* (C<sub>13</sub>), the sole member of Ectocarpaceae (Draisma *et al* 2002) shows a clear divergence from Chordariaceae. In our analysis it is closely related to Acinetosporaceae and the genera *Colpomenia* (C<sub>23</sub>, C<sub>24</sub>), *Rosenvinegea* (C<sub>32</sub>) and *Chnoospora* (C<sub>33</sub>) of Scytosiphonaceae. This finds support in figure 4. In figure 5 we notice the distinction between these three families.

We conclude that IC from CIP scores have the potential to sketch the PR among closely related species. The potential of the phylogenetic pattern is reduced if we introduce non-impact codons [ $z_{\text{critical}} < 1.28$ ]. Although the CIP score are not independent, but a single mutation does have negligible impact on the average distances between the species in our analysis. We have checked the single nucleotide mutational effect at different positions of *rbcL* gene for every species and find infinitesimal deviation in average distance calculation. Also it is to be noted that if we plot the dendrogram using just the non-impact codons i.e. the ones not in the set above – the PR bear little resemblance to what is known about classification in the Ectocarpales.

### References

- Bailey J C and Freshwater D W 1997 Molecular systematics of the Gelidiales: inferences from separate and combined analyses of plastid *rbcL* and nuclear SSU gene sequences; *Eur. J. Phycol.* **32** 343–352
- Bailey J C, Bidigare R R, Christensen S J and Andersen R A 1998 Phaeothamniophyceae classis nova: A new lineage of Chromophytes based upon photosynthetic pigments, *rbcL* sequence and ultra structure; *Protist* **149** 245–263
- Ben Ali A, Baere R de, Auwera G V, Wachter R de and Pee Y van de 2001 Phylogenetic relationships among algae based on complete large-subunit rRNA sequences; *Int. J. Syst. Evol. Microbiol.* **51** 737–749
- Bennetzen J L and Hall B D 1982 Codon selection in yeast; *J. Biol. Chem.* **257** 3026–3031
- Bhattacharya D and Medlin L 1995 The phylogeny of plastids: A review based on comparisons of small-subunit ribosomal RNA coding regions; *J. Phycol.* **31** 489–498
- Bhattacharya D, Medlin L, Wainright P O, Ariztia E V, Bibeau C, Stickle S K and Sogin M L 1992 Algae containing chlorophylls a + c are paraphyletic: molecular evolutionary analysis of the Chromophyta; *Evolution* **46** 1801–1817
- Cavalli-Sforza L L and Edward A W F 1967 Phylogenetic analysis: Models and estimation procedures; *An. J. Hum. Genet.* **19** 233–257
- Daugbjerg N and Guillou L 2001 Phylogenetic analysis of Bolidophyceae (Heterokontophyta) using *rbcL* gene sequences support their sister group relationship to diatoms; *Phycologia* **40** 153–161
- Daugbjerg N and Andersen R A 1997 A molecular phylogeny of the heterokont algae based on analysis of chloroplast-encoded *rbcL* sequence data; *J. Phycol.* **33** 574–585
- De Clerck O, Devos P, Gillis M and Coppejans E 2001 Molecular systematics in the genus Dictyota (Dictyotales, Phaeophyta): a first attempt based on restriction patterns of the Internal Transcribed Spacer 1 of the rDNA (ARDRA-ITS1); *Syst. Geogr.* **71** 25–35
- Draisma S G A, Prud'homme van Reine W F, Stam W T and Olsen J L 2001 A reassessment of phylogenetic relationships within the Phaeophyceae based on RUBISCO large subunit and ribosomal sequences; *J. Phycol.* **37** 586–603
- Draisma S G A, Olsen J L, Stam W T and Prudhomme van Reine W F 2002 Phylogenetic relationships within the Sphaecelariales (Phaeophyceae): *rbcL*, RuBisCO spacer and morphology; *Eur. J. Phycol.* **37** 385–401
- Farris J S 1970 Methods for computing Wagner trees; *Syst. Zool.* **34** 21–34
- Gour M and Gautier C 1982 Codon usage in bacteria: correlation with gene expressivity; *Nucleic Acids Res.* **10** 7055–7073
- Gribskov M, Devereux J and Burgess R R 1984 The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expressions; *Nucleic Acids Res.* **12** 539–549
- Hanyuda T, Arai S and Ueda K 2000 Variability in the *rbcL* introns of caulerpalean algae (Chlorophyta, Ulvophyceae); *J. Plant. Res.* **113** 403–413
- Henry E C 1984 Syringidermatales ord. nov and *Syringoderma floridana* sp. nov. (Phaeophyceae); *Phycologia* **23** 419–426
- Hommersand M H, Frederique S and Freshwater D W 1994 Phylogenetic systematics and biogeography of the Gigartinales (Gigartinales, Rhodophyta) based on sequence analysis of *rbcL*; *Bot. Mar.* **37** 193–203
- Ikemura T 1981 Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system; *J. Mol. Biol.* **151** 389–409
- Ikemura T 1985 Codon usage and t-RNA content in unicellular and multicellular organisms; *Mol. Biol. Evol.* **2** 13–34
- Kellogg E A and Juliano N D 1997 The structure and function of RuBisCO and their implication for systematic studies; *Am. J. Bot.* **84** 413–428
- Kimura M 1981 Estimation of evolutionary distances between homologous nucleotide sequences; *Proc. Natl. Acad. Sci. USA* **78** 454–458
- Kogame K, Horiguchi T and Masuda M 1999 Phylogeny of the order Scytosiphonales (Phaeophyceae) based on DNA sequences of *rbcL*, partial *rbcS* and partial LSU nrDNA; *Phycologia* **38** 496–502
- Kylin H 1933 Uber die Entwicklungsgeschichte der Phaeophyceen; *Acid univ. Land* **29** 1–102
- Lee W J and Bae K S 2002 Phylogenetic relationships among several genera of Dictyotaceae (Dictyotales, Phaeophyceae) based on 18S rRNA and partial *rbcL* gene sequences; *Mar. Biol.* **140** 1107–1115
- Leipe D D, Wainright P O, Gunderson J H, Porter D, Patterson D J, Valois F, Himmerich S and Sogin M L 1994 The stramenopiles from a molecular perspective: 16S-like rRNA sequences from *Labyrinthuloides minuta* and *Cafeteria roenbergensis*; *Phycologia* **33** 369–377
- McCourt R M, Karol K G, Kaplan S and Hoshaw R W 1995 Using *rbcL* sequences to test hypothesis of chloroplast and thallus evolution in conjugating green algae (Zygnematales, Charophyceae); *J. Phycol.* **31** 989–995
- McLachlan A D, Staden R and Boswell D R 1984 A method for measuring the non-random bias of a codon usage table; *Nucleic Acid Res.* **12** 9567–9575
- Medlin L K, Kooistra W H C F, Potter D, Saunders G W and Andersen R A 1997 Phylogenetic relationships of the 'golden algae' (haptophytes, heterokont chromophytes) and their plastids; *Plant. Syst. Evol. (Suppl.)* **11** 187–219

- Morton B R 1993 Chloroplast DNA codon use: Evidence for selection at the psbA locus based on tRNA availability; *J. Mol. Evol.* **37** 273–280
- Morton B R 1996 Selection on the codon bias of *Chlamydomonas reinhardtii* chloroplast genes and the plant psb A gene; *J. Mol. Evol.* **42** 28–31
- Muller D G, Parodi E R and Peters A F 1998 *Asterocladon lobatum* gen. et sp. nov., a new brown algae with stellate chloroplast arrangement, and its systematic position judged from nuclear rDNA sequences; *Phycologia* **37** 425–432
- Nozaki H, Ito M, Sano R, Uchiba H, Watanabe M M, Takahashi H and Kurowai T 1997 Phylogenetic analysis of *Yamagishiella* and *Platydorina* (Volvocaceae, Chlorophyta) based on rbcL gene sequences; *J. Phycol.* **33** 272–278
- Peters A F and Clayton M N 1998 Molecular and morphological investigations of three brown algal genera with stellate plastids: evidence for Scytothamiales ord nov (Phaeophyceae); *Phycologia* **37** 106–113
- Peters A F and Ramirez M E 2001 Molecular Phylogeny of small brown algae, with special reference to the systematic position of *Caepidium antarcticum* (Adenocystaceae, Ectocarpales); *Cryptogamie Algol.* **22** 187–200
- Rousseau F and De Reviere B 1999 Circumscription of the order Ectocarpales (Phaeophyceae): bibliographical synthesis and molecular evidence; *Cryptogamie Algol.* **20** 5–18
- Russell G and Fletcher R L 1975 A numerical taxonomic study of the British Phaeophyta; *J. Mar. Biol. Assoc. UK* **55** 763–783
- Saitou N and Nei M 1987 The neighbour joining method, a new method for reconstructing phylogenetic trees; *Mol. Biol. Evol.* **4** 406–425
- Siemer B L, Stam W T, Olsen J L and Pedersen P M 1998 Phylogenetic relationships of the brown algal orders Ectocarpales, Chordariales, Scytosiphonales and Tilopteridales (Phaeophyceae) based on RuBisCO large subunit and spacer sequences; *J. Phycol.* **34** 1038–1048
- Sokal R R and Michener C D 1958 A statistical method for evaluating systematic relationships; *Univ. Kans. Sci. Bull.* **28** 1409–1438
- Vanden Hoek C and Flinterman A 1968 The life history of *Sphaecelaria furcigera* Kutz (Phaeophyceae); *Blumea* **16** 193–242
- Vanden Hoek C, Mann D G and Jahns H M 1995 *Algae: An introduction to phycology* (Cambridge, UK: Cambridge University Press) Vol. 14, pp 623

MS received 2 September 2005; accepted 16 August 2005

ePublication: 10 October 2005

Corresponding editor: R GEETA