Phylogenetic Analyses of Three Genes of *Pedinomonas noctilucae*, the Green
Endosymbiont of the Marine Dinoflagellate *Noctiluca scintillans*, Reveal its
Affiliation to the Order Marsupiomonadales (Chlorophyta, Pedinophyceae) under the
Reinstated Name *Protoeuglena noctilucae*

Lu Wang^a, Xin Lin^a, Joaquim I. Goes^b, and Senjie Lin^{a,c,1}

^aMarine Biodiversity and Global Change Research Center and State Key Laboratory
of Marine Environmental Science, Xiamen University, Xiamen 361005, China

^bLamont Doherty Earth Observatory at Columbia University, Palisades, New York
10964, USA

^cDepartment of Marine Sciences, University of Connecticut, Groton, CT 06340, USA

Submitted July 8, 2015; Accepted February 25, 2016

Monitoring Editor: Michael Melkonian

Running title: Phylogeny of Green Endosymbiont in *Noctiluca scintillans*

¹Corresponding author; e-mail senjie.lin@uconn.edu (S. Lin).

In the last decade, field studies in the northern Arabian Sea showed a drastic shift from diatom-dominated phytoplankton blooms to thick and widespread blooms of the green dinoflagellate, *Noctiluca scintillans*. Unlike the exclusively heterotrophic red form, which occurs widely in tropical to temperate coastal waters, the green *Noctiluca* contains a large number of endosymbiotic algal cells that can perform photosynthesis. These symbiotic microalgae were first described under the genus *Protoeuglena* Subrahmanyam and further transferred to *Pedinomonas* as *P. noctilucae* Sweeney. In this study, we used the 18S rDNA, *rbcL* and chloroplast 16S rDNA as gene markers, in combination with the previously reported morphological features, to re-examine the phylogenetic position of this endosymbiotic algal species. Phylogenetic trees inferred from these genes consistently indicated that *P. noctilucae* is distantly related to the type species of *Pedinomonas*. The sequences formed a monophyletic clade sister to the clade of *Marsupiomonas*, necessitating the placement of the algal symbionts as an independent genus within the family Marsupiomonadaceae. Based on the phylogenetic affiliation and ecological characteristics of this alga as well as the priority rule of nomenclature, we reinstate the genus *Protoeuglena* and reclassify the endosymbiont as *Protoeuglena noctilucae*.

Keywords: *Protoeuglena*; *Pedinomonas noctilucae*; flagellate; *Noctiluca*; symbiosis; molecular phylogeny; Arabian Sea.

49 **Introduction**

50 The heterotrophic dinoflagellate *Noctiluca scintillans* (Macartney) Kofoid is the first
51 described and the largest known dinoflagellate that exhibits bioluminescence when
52 disturbed. It is distributed worldwide and forms distinct red tides of the so-called red
53 *Noctiluca* in temperate coastal waters (Elbrächter and Qi 1998). It feeds by ingesting
54 small phytoplankton with its tentacle and usually appears transparent or pale pink in
55 color (Sweeney 1976). In the tropical western Pacific and Southeast Asia, however,
56 this species contains large numbers of green algal endosymbionts (Furuya et al. 2006;
57 Gomes et al. 2008). The symbiotic relationship appears to be species-specific and the
58 green color from the flagellate makes the host known as green *Noctiluca* (Furuya et al.
59 2006). The small green symbiotic algae may provide organic matter produced in its
60 photosynthesis but can also be ingested to ensure the survival of *N. scintillans* under
61 food limitation (Furuya et al. 2006; Saito et al. 2006; Sweeney 1976).

62 In recent years, massive blooms of the green *N. scintillans* have become the
63 norm occurring with predictable regularity during winter in the Arabian Sea. Outbreaks
64 of *N. scintillans* blooms coincide with the upward shoaling of subsurface hypoxic
65 waters, which led Gomes et al (2014) to posit that the survival of *N. scintillans* may be
66 tied to the extraordinary ability of its endosymbiont *P. noctilucae* to fix carbon under
67 hypoxic conditions. The emergence of the green bloom is often accompanied by a
68 decline in the species diversity of phytoplankton. Given the ecological importance of
69 *P. noctilucae* and its apparent ability to fix carbon dioxide under hypoxic conditions, it
70 became important to accurately determine its taxonomic identity and assess the
71 relationship of these symbionts to established taxa.

72 Subrahmanyam (1954) described the symbiotic green alga *Protoeuglena*
 73 *noctilucae* because of its similarity to euglenoids in flagellar structure and absence of a
 74 cellulose wall. Sweeney (1976) distinguished it from euglenoids and
 75 Prasinophyceae (*Micromonas* and *Mantoniella*) based on its starch
 76 storage, dissimilarities in color, flagellum orientation and subcellular structures, and
 77 reclassified it as a new species, *Pedinomonas noctilucae*, which was separated from *P.*
 78 *minor* and *P. tuberculata* by its deep flagellar groove and smooth cell surface. When
 79 Sweeney described *P. noctilucae*, the genus *Pedinomonas* was placed in the class of
 80 Prasinophyceae, one of the two classes (the other Loxophyceae) of “primitive green
 81 algae” established by Christensen (1962, 1966). The classification of the “primitive
 82 green algae” has since been a matter of debate. In 1991, based on the difference of
 83 behavior and the flagellar apparatus between *Pedinomonas mikron* Thronsen, 1969
 84 and the well characterized type species *P. minor* Korshikov, 1923, Moestrup created a
 85 new genus, *Resultor*, and renamed *P. mikron* as *Resultor mikron* Moestrup, 1991.
 86 Furthermore, Moestrup established a new class of Pedinophyceae to contain
 87 *Pedinomonas* and the new genus *Resultor* to represent an ancient group of non-scaly
 88 green algae. Because *R. mikron* was erected based on the basionym *P. mikron* and
 89 without designating a type specimen, Marin (2012) redescribed and renamed it as
 90 *Resultomonas moestrupii*. Meanwhile, Jones et al. (1994) created the genus of
 91 *Marsupiomonas*, which was characterized by a deep flagellum pit and the possession of
 92 a theca, and transferred *Pedinomonas tenuis* Masyuk into it. Although wide
 93 morphological and structural variations existed between these different genera, all
 94 three genera were still kept in the family of Pedinomonadaceae Korshikov, in the only
 95 order Pedinomonadales of the class Pedinophyceae. Recently, nuclear and
 96 plastid-encoded rRNA phylogenetic analyses indicated that the genera *Resultomonas*

and *Marsupiomonas* formed a newclade (Marsupiomonadales) sister to the order
Pedinomonadales, consistent with their habitat preference of marine
(Marsupiomonadales) versus freshwater (Pedinomonadales) (Marin 2012). It has been
contended that the assignment of the *Noctiluca* endosymbiont to *P. noctilucae* was
inconclusive and its marine habitat and flagellar pit suggested that the species might
be closely related to the genus *Marsupiomonas* (Jones et al. 1994; Moestrup 1991). The
accurate phylogenetic affiliation of *P. noctilucae* has thus remained unresolved due to
the lack of well-defined discriminative morphological characters and molecular data.

In this study, the 18S ribosomal RNA gene (rDNA), the chloroplast 16S rDNA
and the chloroplast RUBISCO large subunit gene (*rbcL*) were obtained from the green
symbionts in the *Noctiluca* population during a bloom in the Arabian
Sea. Phylogenetic trees were reconstructed based on these genes to determine the
affiliation of these endosymbiotic green algae with the existing taxonomic groups in
the class of Pedinophyceae. Our phylogenies consistently placed the symbiotic
flagellates within the order of Marsupiomonadales, as a taxon very close to
Marsupiomonas sp. Based on the phylogenies, previous morphological
characteristics and habitat preference, this microalga is closely related to, but clearly
independent from *Marsupiomonas*, indicating the need to transfer this endosymbiont
to the family of Marsupiomonadaceae in the order of Marsupiomonadales. Based on
the priority principle of nomenclatural codes, we further propose to reinstate the genus
name *Protoeuglena* to accommodate this species as *Protoeuglenanoctilucae*.

Results

Microscopic Observation and Bloom Characteristics

Seawater samples were collected from the surface of the Central Arabian Sea (Lat. 20.9°N and 66.9°E) in February 2009, where an extensive bloom of green *N. scintillans* was encountered (Fig. 1a). Seawater temperature at this location was 25.2°C and salinity was 36.5 psu. Under the light microscope, a large number of small green flagellates and few ingested diatoms could be seen inside the vacuole of individual *N. scintillans* cells (Fig. 1b-d). Within this bloom, the maximum *N. scintillans* concentration was up to 9,600 cells l⁻¹, the chlorophylla concentrations in the sea were around 25 mg m⁻³, and carbon fixation rates were 2 g C m⁻³ day⁻¹ at the surface and 12 g C m⁻³ day⁻¹ in the 33 m euphotic zone (Gomes et al. 2014). Living flagellate cells were fixed in 2% glutaraldehyde and examined microscopically. They are pear-shaped or oval-shaped, about 8-15 µm long and 5-10 µm broad, and somewhat laterally compressed (Fig. 1e-f). The outer edge of the flagellar pocket thickened to form a flagellar lip. There is a deep groove between the flagellar lip and the cell body, into which the flagellum is inserted. No red eyespot was observed in the fixed material.

Direct- and Clone-sequencing Results

The 18S rDNA PCR products we obtained were about 1800 bp long, as predicted for most eukaryotic algae. After excluding most of the gene clones showing *N. scintillans* restriction pattern, 31 clones were sequenced, yielding 29 good-quality sequences. BLASTn results showed that 16 of these clones belonged to Chlorophyta, 2 belonged to copepods, and 11 to *N. scintillans*. This result indicated that *N. scintillans* had ingested copepods (likely eggs or larvae) as food. For the chloroplast

genes encoding 16S rDNA and *rbcL*, direct-sequencing results corresponded to clone-based sequencing results, revealing the green alga species dominant in *N. scintillans* cells. Their sequence lengths were 421bp and 1039bp, respectively. BLASTn results showed that the 16S rDNA was most closely related to *Marsupiomonas* while the *rbcL* sequence was closely related to *Pedinomonas*, *Marsupiomonas* and genera in the classes Trebouxiophyceae and Chlorophyceae.

Phylogenetic Analyses

Phylogenetic analyses were performed to determine more accurately which lineage the symbionts are affiliated with. In the 18S rDNA tree (Fig. 2), our 16 new sequences formed a monophyletic cluster positioned beside the sequences of *Marsupiomonas* spp. although four of them formed relatively long branches compared to other highly similar sequences. The branching pattern pertaining to the symbionts was the same in phylogenetic trees inferred using ML and Bayesian methods, although these methods differed in the branching order of different classes (Trebouxiophyceae, Ulvophyceae, Chlorophyceae, Chlorodendrophyceae and Pedinophyceae) of Chlorophyta (Fig. 2). All the sequences within the class Pedinophyceae formed two major clades with strong bootstrap support, one comprising of the 16 clone sequences we obtained from the *N. scintillans* samples in this study and four previously documented sequences of *Marsupiomonas* spp., the other consisting of nine previously reported sequences of *Pedinomonas* spp.

Analyses of the 16S rDNA and *rbcL* nucleotide data both strongly supported the embedment of the symbionts within the class Pedinophyceae (Figs 3 and 4). In both trees, all direct-sequenced and clone-based sequences formed a monophyletic group, branching with the sequences of *Marsupiomonas* spp. within the order of Marsupiomonadales, separated from Pedinomonadales. These gene trees exhibited similar tree topologies to each other and to the above 18S rDNA tree, with the only difference existing in the relationship among the different classes (Trebouxiophyceae, Ulvophyceae, Chlorophyceae, Chlorodendrophyceae and Pedinophyceae) of Chlorophyta. The class level branching order received relatively weak or no bootstrap support in some of the methods for phylogenetic tree reconstruction.

To assess whether the Arabian Sea population of *N. scintillans* is the same species as previously analyzed populations, we calculated genetic distances among the populations for which data were available. Internal transcribed spacer of ribosomal RNA gene (ITS) sequences were obtained from red *N. scintillans* from the USA (strain NS3) and Hong Kong as well as the green population from the Arabian Sea. P-distances between any two of these ITS clone sequences were 0.000-0.054, which is similar to the intra-species genetic distance (0.000-0.066) and much less than the inter-species genetic distance (0.134-0.216) in dinoflagellates previously studied (Litaker et al. 2007; Wang et al. 2014). Their highly similar sequences suggest that all these “red” and “green” populations of *N. scintillans* are the same species. The greatest divergence was found between the Arabian and American populations, which can be attributed to geographic isolation or ecotypic differences (symbiotic and non-symbiotic).

190 **Phylogenetic Position of the Symbionts within the Class Pedinophyceae**

191 Based on the 16S rDNA sequences (Fig. 5), the Pedinomonadales taxa displayed high
192 genetic diversity, but it was difficult to resolve families within this order because the
193 sequence data were limited. Within this order, the type species of *Pedinomonas* (*P.*
194 *minor*) branched with three strains of *Chlorochytridium*, one of which was reinstated
195 by Marin (2012). For the Marsupiomonadales order, the 16S rDNA phylogenies
196 comprised two major lineages with strong support, represented by the families
197 Resultomonadaceae and Marsupiomonadaceae (Fig. 5). The Resultomonadaceae family
198 contains two clearly resolved genera, *Resultomonas* and an unknown genus. The
199 subclades MARS A-C in the Marsupiomonadaceae family have been proposed to
200 represent three genera (Marin 2012); MARS A representing the *Marsupiomonas* genus
201 that currently consists of three species (*M. pelliculata*, strain NIES-1410 and
202 uncultured clones sequences from New Orleans), and MARS B and C likely
203 representing two unknown genera. Our new sequences formed an independent lineage
204 MARS D sister to MARS A and, likely a new genus that is most closely related to the
205 genus *Marsupiomonas*.

206 Discussion

207 Morphological Evidence to Reclassify *P. noctilucae*

208 The green symbiotic flagellate had been classified into the genus *Pedinomonas* based
 209 on the bright green colour and ultrastructural characteristics such as back-directed
 210 single flagellum with rhizoplast, the single chloroplast with prominent pyrenoid and
 211 starch sheath, the starch grains and multiple small mitochondria (Sweeney 1976).
 212 However, the long flagellum of *P. noctilucae* is inserted posteriorly in a deep
 213 asymmetric groove which is tunnel-shaped and much deeper than in the species of *P.*
 214 *minor* and *P. tuberculata*, the presence or absence of an eyespot was uncertain and the
 215 second basal body was not observed distinctly in living material (Subrahmanyam 1954;
 216 Sweeney 1976). Remarkably, the deep depression or groove in *P. noctilucae* was very
 217 similar to the deep pit described in the genus *Marsupiomonas*, which is characterized
 218 by its salient features of posterior theca and flagellar pit (Jones et al. 1994). In
 219 addition, species in the order of Pedinomonadales usually grow in freshwater or soil,
 220 whereas species in the order Marsupiomonadales, such as *M. pelliculata* and *R.*
 221 *moestrupii*, live in brackish water or seawater, which is in common with the marine
 222 habitat of *P. noctilucae* (Marin 2012). Previous studies suggested that *P. noctilucae* was
 223 classified incorrectly and its marine habitat and flagellum pit structure would place it
 224 in the genus *Marsupiomonas* (Jones et al. 1994; Moestrup 1991). However, there are
 225 still differences in cell shape, cell size, and living habitat between *P. noctilucae* and
 226 *M. pelliculata*; for instance, a theca is typically present in *Marsupiomonas* but was not
 227 found in *P. noctilucae*. These characteristics indicate that *P. noctilucae* should belong to
 228 the family of Marsupiomonadaceae in the order of Marsupiomonadales as a
 229 genus independent of *Marsupiomonas*. However, some distinguishing morphological

characteristics still remain to be described for *P. noctilucae*. For example, the eyespot opposite the flagellum insertion reported as a character of Pedinophyceae was not shown in transmission electron microscopy (SEM) images of *P. noctilucae* (Sweeney 1976). In addition, the short immature no.2 basal body without flagellum is also typically characteristic of Pedinophyceae, but has not been observed in *P. noctilucae*. Thus, more ultrastructural information using modern technology about the cell surface, flagellar transitional region and roots is needed. Unfortunately, no cell samples were available for such analysis in this study because of our inability to culture the green *Noctiluca*.

Molecular Evidence that *P. noctilucae* Belongs to a Genus Sister to *Marsupiomonas*

Previous studies have proposed that phagotrophy was facultative in green *Noctiluca* and is a requirement for its faster growth, probably contributing to bloom formation (Furuya et al. 2006; Saito et al. 2006). To ensure no algae outside of *N. scintillans* cells would contribute to the gene sequences obtained in this study, we thoroughly washed the surface of each *N. scintillans* cell in our sample collection for DNA extraction. We directly sequenced PCR products and checked the sequence chromatographs for possible existence of minor constituents of sequences. Furthermore, we also tested the universality of our common primers (18S, chloroplast 16S and *rbcL*) by using all downloaded sequences in this study making sure our primers were able to amplify all species included in the phylogenetic analysis. The clean and single chromatographs of the 16S and *rbcL* direct sequencing results from

the PCR products and the high consistency of 18S rDNA clone sequences indicated that the symbionts in *N. scintillans* belonged to one taxon.

Our phylogenetic trees consistently showed that the symbiont sequences were located in the same position beside the sequences of *Marsupiomonas* spp. This affiliation remained unchanged when different methods and models were used for phylogenetic tree reconstruction, even though these different methods and models changed the relationships among the major classes of Chlorophyta. The unstable branching pattern among classes is likely due to the usage of small single genes instead of multigene approaches using high numbers of characters (Marin 2012). Variable relationships among the classes or orders in Chlorophyta have been found depending on gene, taxon sampling and phylogenetic methods used (Fučíková et al. 2014; Leliaert et al. 2011, 2012; Lemieux et al. 2014). Thus far, attempts to reconstruct the deep phylogenetic relationships of Chlorophyta have seldom yielded a well-supported backbone of the phylogeny (Fučíková et al. 2014). For the purpose of the present study, however, we only need to focus on the taxonomic status of the symbiotic green algae in this study and its resolution at the family level. In all the phylogenetic trees that we obtained, the symbionts in *Noctiluca* were placed within the family of Marsupiomonadaceae and in the order of Marsupiomonadales rather than the previously assigned order Pedinomonadales. Based on the phylogenetic analyses we propose that this monophyletic group be recognized as an independent genus sister to *Marsupiomonas*. Including the other genus *Resultomonas*, the order of Marsupiomonadales now includes three genera.

A recent study including 16S rDNA from uncultured strains indicated the existence of three subclades (MARS A-C) in the Marsupiomonadaceae family, which potentially represented three genera (Marin 2012). Following the same analysis

including the reported data, our new sequences formed an independent lineage (MARS D) sister to one of those subclades (MARS A). Because some of the previously reported uncultured 16S rDNA sequences were short giving limited aligned positions, the new subclade formed by our new sequences needs to be validated by further analysis using more complete sequences or multiple genes in the future.

Reinstatement of the *Protoeuglena* as the New Genus Name

Subrahmanyam (1954) first described the symbiotic flagellate in *Noctiluca* and observed a definite depression in the middle of the cell body, which is similar to the deep depression described in Sweeney (1976) and in accordance with the characteristic observed in this study (Fig. 1e-f). Given that the phylogenetic and morphological analysis, in concert with the ecological habitat, indicate that this symbiotic alga belongs to a sister genus of *Marsupiomonas* and is distantly related to Pedinomonadales, we suggest to abolish the current name *Pedinomonas noctilucae*. As the original designation of this alga as *Protoeuglena noctilucae* (Subrahmanyam 1954) has priority in nomenclature, we further reinstate *Protoeuglenanoctilucae*, as the name of the symbiotic flagellates in *Noctiluca*.

Protoeuglenanoctilucae Subrahmanyam 1954, Proc. Indian Acad. Sci. 39 (Sect. B): p.

123

Emended diagnosis: Marine habitat, usually growing in the dinoflagellate *Noctiluca scintillans*. Cells small, compressed laterally, measuring about 5-10 μm long, without theca, periplast smooth. Single flagellum very long and inserted in a deep groove, more than twice the cell length. Chloroplast bright green, single and lateral. Mobile actively and generally rapid backward movements.

-
- 303 Lectotype (designated here): fig. 8 in Subrahmanyam (1954)
- 304 Synonym: *Pedinomonas noctilucae*(Subrahmanyam) Sweeney 1976, J. Phycol. 12: p.
- 305 463

306 **Methods**

307 **Sample collection and initial processing:** Samples of *N. scintillans* were collected on
 308 February 17, 2009 at 20.9° N, 66.17° E on board the Fisheries Oceanographic
 309 Research Vessel Sagar Sampada, belonging to the Ministry of Earth Sciences, Govt.
 310 of India(station 6 of the cruise reported in Gomes et al. 2014). Cells were observed
 311 microscopically for the presence of endosymbiotic algae. About ten *Noctiluca* cells
 312 were washed 6times in 0.45µm-filteredseawater and pipetted onto a pre-combusted
 313 (450°C) GF/F membrane to eliminate contamination from the filter.The filters were
 314 then individually placed in a cryoprotective cassette andimmediately frozen in liquid
 315 nitrogen. Once transported to the laboratory, the filters were individually immersed in
 316 DNA lysis buffer (10mM Tris pH 8.0; 100 mM EDTA pH 8.0; 0.5% SDS)and stored
 317 at -20°Cuntil DNA extraction.

318 We attempted to isolate and grow *N.scintillans* in seawater obtained from
 319 locations where we encountered blooms of the organism. A known number of cells of
 320 *N.scintillans* were carefully picked out under a stereo-microscope with either a
 321 Pasteur pipette or a polycarbonate dropper, and then carefully transferred into 50 and
 322 100ml culture vessels containing either plain filtered(0.2mm pore size Nucleopore®
 323 polycarbonate filters) or in 0.2-mm-filtered seawater amended with f/20nutrients. The
 324 culture vessels containing isolated cells of *N. scintillans* were then placed either in a
 325 deck incubator with running seawater, or in a laboratory incubator maintained at 27°C.
 326 Invariably after showing signs of positive growth during the first week and a half, the
 327 isolates would not survive.

328 **DNA extraction:** DNA was extracted following a recently developed protocol
 329 (Yuan et al. 2015). Briefly, sampleswere amended with Proteinase K (100
 330 µg/ml)andincubated at 55°Cfor three days, then 4/5 of the supernatant was removed

into a new tube after centrifuging samples at ~13,000 x g for 10 min. Ceramic beads (0.5 mm diameter) were added to the remaining 1/5 lysate and cells remaining intact after the incubation were disrupted through a FastPrep-24 bead mill (MP Biomedicals, USA) at 6 m/s for 1 min. The above beads-beating process was repeated three times and samples were placed on ice during the intervals. At the end, each sample was mixed with its corresponding supernatant removed earlier and incubated at 55 °C for 30 min. Subsequent steps of DNA extraction followed a CTAB method coupled with DNA Clean-up & Concentration column (Zhang et al. 2005). DNA was eluted in 25 µL 10 mM Tris-HCl buffer (pH 8.0), quantified using NanoDrop D2000 (ThermoFisher Scientific, Inc., Wilmington, DE, USA), and stored at -20 °C.

PCR, cloning, restriction enzyme digestion and sequencing: To analyze the 18S rDNA of *Noctiluca* and other organisms that might be associated with it, nearly full-length 18S rDNA was amplified using eukaryotic universal primers of 18ScomF1 and 18ScomR1 (Table 1), PCR amplification was carried out with the aforementioned DNA samples (3-10 ng) as the template under the following conditions: 94 °C for 1 min, 35 cycles of 15 s at 95 °C, 30 s at 58 °C and 45 s at 72 °C, followed by an additional step of extension at 72 °C for 5 min. The resultant amplicons were purified using Clean & Concentrator column (Zymo Research, Orange, CA) and cloned into the T-vector. 96 clones were randomly picked from the 18S library and plasmid DNA was isolated. To maximize the chance to detect all organisms associated with the *N. scintillans* cells, clones with inserts as shown by PCR were digested with restriction enzymes XbaI and EcoRI (Fermentas) at 37 °C for 2 h to screen out clones that showed the restriction pattern of *Noctiluca* 18S rDNA insert. The *N. scintillans* 18S rDNA clones produced three fragments (about 1750 bp, 220 bp and 130 bp), while other insert-containing clones that showed fragment patterns different from this were considered as indication of other organisms. Representative clones of non-*Noctiluca* 18S rDNA were selected

for sequencing (Beijing Genome Institute, Shanghai, China). PCR products of ITS were also purified and cloned into the T-vector, ten positive clones were collected for sequencing.

To specifically determine the identity of the endosymbiotic algae, chloroplast 16S rDNA and *rbcL* nucleotide genes were amplified using primer sets of CYA361f /CYA785r and RubI-forward /RubI-reverse (Table 1), using the same PCR protocols as described above, except for the annealing temperatures shown in Table 1. The amplicons were purified by Clean & Concentrator column and directly sequenced with corresponding primers. To ensure the sequencing results at the end of fragments were accurate, the PCR products of 16S rDNA and *rbcL* were also separately cloned into T-vectors and 2-4 clones were sequenced for each gene.

Alignment and phylogenetic analysis: Additional sequences of Trebouxiophyceae, Ulvophyceae, Chlorophyceae, Chlorodendrophyceae and Pedinophyceae were obtained from GenBank by taxonomic searches and from blastn by using the sequences obtained in this study as queries. The selection of the downloaded sequences (18S rDNA, *rbcL* and 16S rDNA gene) was partially based on the phylogenetic trees reported in Marin (2012). Outgroup taxa were chosen from prasinophytes to root the phylogenetic trees. These sequences combined with our new data formed 18S rDNA, *rbcL* and 16S datasets separately. For each dataset, alignment was performed using CLUSTAL X (Thompson et al. 1994) with adjusted alignment parameters (Gap Opening: 10.00-15.00, Gap Extension: 6.66) and manually edited. In addition, ITS sequences of *N. scintillans* obtained from Arabian Sea (8 clone sequences, KU055475-KU055482), Hong Kong (GQ380592) and Port Aransas, Texas, USA (8 clone sequences, KR607075-KR607082) formed the *Noctiluca* dataset for estimating the divergence rates. Genetic distance was calculated using the Molecular Evolutionary Genetic Analysis (MEGA) program based on the p-distance model. Maximum likelihood (ML) phylogenetic analysis of 18S rDNA, 16S and *rbcL* was performed with a PhyML web server.

(<http://www.atgc-montpellier.fr/phyml/>) using the general time reverse (GTR) model with estimated values of the proportion of invariants (I) and the gamma shape parameter (G). The branch supports of the ML tree were inferred using Shimodaira-Hasegawa-like (SH) aLRT and were estimated by performing 100 non-parametric bootstrap replicates with the same parameters mentioned above. To compare our new sequences with all existing sequences, we added our 16S rDNA sequences to a published dataset (Marin 2012) including sequences of uncultured strains (Fig. 5), and constructed a phylogenetic tree with PhyML 3.0 as implemented in SeaView 4.3 (Gouy et al. 2010) using the GTR model as described above. SeaView 4.3 was used for this dataset because PhyML web server shown above failed to generate a robust tree, likely because of the short aligned regions caused by the short sequences available for some of the previously reported uncultured strains. Bayesian analyses were conducted in MrBayes 3.1.2 (Hueslenbecket al, 2001) and were run for 1,000,000 generations (16S and *rbcL*) and 2,000,000 (18S) generations respectively with trees sampled every 1000 cycles (three independent runs for different gene), and the 'burn-in' determined by the convergence criterion.

Acknowledgements

We are thankful to Dr. Shailesh Nayak, former Secretary, Ministry of Earth Sciences, Govt. of India and Dr. Sanjeevan, former Director of CMLRE for facilitating our research work in the Arabian Sea on board *FORV Sagar Sampada*, to Dr. Fernando Gomez from the Oceanographic Institution, University of Sao Paulo, for advice on nomenclature, and to Ms. Brittany Sprecher from Department of Marine Sciences, University of Connecticut, USA, for help on English grammar. We also thank the three anonymous reviewers for their constructive comments, especially the detailed advice on taxon description from one of them. The work was supported by the Natural Science Foundation of China grants 41176091, 41129001, and 41330959 (to S.L.), the

US National Science Foundation grants #EF-0629624 (to S.L.) and OCE-1121022 (to J.G.) and NASA grant NNX11AE22G (to J.G.). The sequences obtained in this study have been deposited in GenBank under the accession numbers KR82593-KR822616, KR607075- KR607082, KU055475-KU055482.

References

- Bai FY, Zhao JH, Takashima M, Jia JH, Boekhout T, Nakase T**(2002) Reclassification of the *Sporobolomyces roseus* and *Sporidiobolus pararoseus* complexes, with the description of *Sporobolomyces phaffii* sp. nov. Int J SystEvolMicrobiol **52**:2309-2314
- Christensen T** (1962) Alger. In Bocher TW, Lange M, Sorensen(eds) Botanik, Vol. 2, Systematisk Botanik. Munksgaard, Copenhagen, 178 p
- Christensen T** (1966) Alger. In Bocher TW, Lange M, Sorensen (eds) Botanik, Vol. 2, Systematisk Botanik, 2nd ed.Munksgaard, Copenhagen, 180 p
- Do Rosário Gomes H, Goes JI, Matondkar SP, Parab SG, Al-Azri AR, Thoppil PG**(2008) Blooms of *Noctiluca miliaris* in the Arabian Sea—An in situ and satellite study. Deep Sea Res Part I: Oceanogr Res Papers**55**:751-765
- Do Rosário Gomes H, Goes JI, Matondkar SGP, Buskey EJ, Basu S, Parab S, Thoppil P**(2014) Massive outbreaks of *Noctiluca scintillans* blooms in the Arabian Sea due to spread of hypoxia. Nature Commun **5**:4862
- Elbrächter M, Qi Z**(1998) Aspects of *Noctiluca* (Dinophyceae) Population Dynamics. In Anderson DM, Cembella D, Hallegraeff GM (eds) Physiological Ecology of Harmful Algal Blooms. Springer Verlag, New York, pp 315–336

- 435 **Furuya K, Saito H, Sriwoon R, Omura T, Furio EE, Borja VM, Lirdwitayaprasit**
 436 **T**(2006) Vegetative growth of *Noctiluca scintillans* containing the
 437 endosymbiont *Pedinomonas noctilucae*. *AfrJ Mar Sci* **28**:305-308
- 438 **Fučíková K, Leliaert F, Cooper ED, Škaloud P, D'hondt S, De Clerck O, Gurgel**
 439 **CFD, Lewis LA, Lewis PO, Lopez-Bautista JM, Delwiche CF and**
 440 **Verbruggen H**(2014) New phylogenetic hypotheses for the core Chlorophyta
 441 based on chloroplast sequence data. *Front Ecol Evol* **2**: 63
- 442 **Gouy M, Guindon S, Gascuel O**(2010) SeaView version 4: a multiplatform
 443 graphical user interface for sequence alignment and phylogenetic tree building.
 444 *Mol Biol Evol* **27**:221-224
- 445 **Harrison PJ, Furuya K, Glibert PM, Xu J, Liu HB, Yin K, Lee JHW, Anderson**
 446 **DM, Gowen R, Al-Azri AR, Ho AYT** (2011) Geographical distribution of
 447 red and green *Noctiluca scintillans*. *Chin J. Oceanol Limnol* **29**:807–831
- 448 **Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP** (2001) Bayesian inference of
 449 phylogeny and its impact on evolutionary biology. *Science* **294**:2310–2314
- 450 **Jones HLJ, Leadbeater BSC, Green JC**(1994) An ultrastructural study of
 451 *Marsupiomonas pelliculata* gen. et sp. nov., a new member of the
 452 Pedinophyceae, *Eur J Phycol* **29**:171-181
- 453 **Leliaert F, Verbruggen H, Zechman FW**(2011) Into the deep: new discoveries at the
 454 base of the green plant phylogeny. *Bioessays* **33**:683-692
- 455 **Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, De**
 456 **Clerck O**(2012) Phylogeny and molecular evolution of the green algae. *Crit*
 457 *Rev Plant Sci* **31**:1–46
- 458 **Lemieux C, Otis C, Turmel M** (2014) Six newly sequenced chloroplast genomes
 459 from prasinophyte green algae provide insights into the relationships among

- 460 prasinophyte lineages and the diversity of streamlined genome architecture in
461 picoplanktonic species. BMC Genomics **15**:857
- 462 **Marin B (2012)** Nested in the Chlorellales or independent class? Phylogeny and
463 classification of the Pedinophyceae (Viridiplantae) revealed by molecular
464 phylogenetic analyses of complete nuclear and plastid-encoded rRNA operons.
465 Protist **163**:778-805
- 466 **McManus GB, Zhang H, Lin S (2004)** Marine planktonic ciliates that prey on
467 macroalgae and enslave their chloroplasts. Limnol Oceanogr **49**:308-313
- 468 **Moestrup Ø (1991)** Further studies of presumed primitive green algae, including
469 the description of Pedinophyceae class. nov. and *Resultor* gen. nov. J Phycol
470 **27**:119-133
- 471 **Mühling M, Woolven-Allen J, Murrell JC, Joint I (2008)** Improved group-specific
472 PCR primers for denaturing gradient gel electrophoresis analysis of the genetic
473 diversity of complex microbial communities. ISME J **2**:379-392
- 474 **Saito H, Furuya K, Lirdwitayaprasit T (2006)** Photoautotrophic growth of
475 *Noctiluca scintillans* with the endosymbiont *Pedinomonas noctilucae*.
476 Plankt Benth Res **1**:97-101
- 477 **Subrahmanyam R (1954)** A new member of the Euglenineae, *Protoeuglena*
478 *noctilucae* gen. et sp. nov., occurring in *Noctiluca miliaris* Suriray, causing
479 green discoloration of the sea off Calicut. In Proc Ind Acad Sci, Sect B, Indian
480 Academy of Sciences **39**:118-127
- 481 **Sweeney BM (1976)** *Pedinomonas Noctilucae* (Prasinophyceae), the flagellate
482 symbiotic in Noctiluca (dinophyceae) in Southeast Asia. J Phycol **12**:460-464
- 483 **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011)** MEGA5:
484 molecular evolutionary genetics analysis using maximum likelihood,

- 485 evolutionary distance, and maximum parsimony methods. Mol Biol Evol
486 **28**:2731-2739
- 487 **Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG**(1997) The
488 CLUSTAL-X windows interface: flexible strategies for multiple sequence
489 alignment aided by quality analysis tools. Nucleic Acids Res **25**:4876–4882
- 490 **Wang L, Zhuang Y, Zhang H, Lin S**(2014) DNA barcoding species in *Alexandrium*
491 *tamarense* complex using ITS and proposing designation of five species.
492 Harmful Algae **31**:100-113
- 493 **Yuan J, Li M, Lin S** (2015) An improved DNA extraction method for efficient and
494 quantitative recovery of phytoplankton diversity in natural assemblages PLoS
495 ONE**10**: e0133060
- 496 **Zhang H, Bhattacharya D, Lin S**(2005) Phylogeny of dinoflagellates based on
497 mitochondrial cytochrome B and nuclear small subunit rDNA sequence
498 comparisons. J Phycol **41**:411-420

499 **Figure Legends**

500 **Figure 1.** *N. scintillans* with endosymbiont and a green bloom formed in the Arabian
 501 Sea. (a) Thick surface bloom of *N. scintillans* in the Arabian Sea on 17 February 2009.
 502 (b) Micrograph of single cells of *N. scintillans* with the green algal symbionts inside.
 503 (c) Higher magnification of *N. scintillans* showing the large number of small green
 504 flagellates. (d) Close-up of *N. scintillans* containing ingested diatom and other
 505 phytoplankton. (e-f) Unicellular flagellate fixed in 2% glutaraldehyde, the arrows
 506 showed the deep groove where the flagellum inserted. Scale = 10 μ m.

507 **Figure 2.** Phylogenetic tree of 18S rDNA sequences obtained in this study
 508 (boldface) and downloaded sequences from 49 taxa. The Prasinophyte sequences were
 509 used as outgroup. Numbers near branches are SH-aLRT values ≥ 0.5 (left) and the
 510 standard nonparametric bootstrap values ≥ 50 (right) for ML analysis, and posterior
 511 probabilities ≥ 0.9 (middle) for Bayesian analysis. Each sequence name consists of the
 512 Genbank accession numbers, the species designation and strain identification. The
 513 gray triangle represents 12 highly similar clone sequences. The corresponding class
 514 names are shown to the right of the tree. Scale bar denotes 0.05 substitutions per site.

515 **Figure 3.** A chloroplast 16S rDNA phylogenetic tree. Clone and direct sequencing
 516 results obtained in this study are shown in boldface. Tree is rooted by four
 517 prasinophyte sequences. Support of branches is shown as SH-aLRT values ≥ 0.5
 518 (left) and the standard nonparametric bootstrap values ≥ 50 (right) from ML
 519 analysis, and the posterior probabilities ≥ 0.9 (middle) from Bayesian analysis. The
 520 corresponding class names are shown to the right of the tree. Scale bar denotes
 521 0.05 substitutions per site.

522 **Figure 4.** A phylogenetic tree based on *rbcL* nucleotide sequences. Clone and direct
 523 sequencing results obtained in this study are shown in boldface. Tree is rooted with
 524 three prasinophyte sequences. Support of branches is shown as SH-aLRT

values ≥ 0.5 (left) and the standard nonparametric bootstrap values ≥ 50 (right) from the ML analysis, and the posterior probabilities ≥ 0.9 (middle) from Bayesian analysis. The corresponding class names are shown to the right of the tree. Scale bar denotes 0.05 substitutions per site.

Figure 5. Chloroplast 16S rDNA phylogeny of Pedinophyceae. The sequence of *Scherffelia dubia* from the class Chlorodendrophyceae was defined as outgroup. Numbers near branches are SH-aLRT values by phyML analyses, only values ≥ 0.5 at critical nodes are shown. Clone sequences obtained in this study were shown in red letter, the seven cultivated type species in Pedinophyceae are in bold. In the Marsupiomonadales, environmental sequences reveal a well-supported branching into two families and further subclades, termed MARS A-C (Marin, 2012). The new subclade MARS D found in this study is also shown. Scale bar denotes 0.05 substitutions per site.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5