Journal of Oceanology and Limnology
Vol. 39 No. 2, P. 508-524, 2021
**https://doi.org/10.1007/s00343-020-9304-5**

# Development of high-resolution chloroplast markers for intraspecific phylogeographic studies of *Phaeocystis globosa**

Qingchun ZHANG[1, 2, 3], Zhuang NIU[1, 2, 3, 4], Jinxiu WANG[1, 2, 3, 4], Chao LIU[1, 2, 3, 4], Fanzhou KONG[1, 2, 3], Xiaokun HU[1, 2, 3, 4], Jiayu ZHAO[1, 2, 3, 4], Rencheng YU[1, 2, 3, 4, **]

[1] *CAS Key Laboratory of Marine Ecology and Environmental Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China*

[2] *Laboratory for Marine Ecology and Environmental Sciences, Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao 266237, China*

[3] *Center of Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China*

[4] *University of Chinese Academy of Sciences, Beijing 100049, China*

Received Nov. 18, 2019; accepted in principle Jan. 23, 2020; accepted for publication Mar. 2, 2020


**Abstract** *Phaeocystis globosa* is an important harmful algal bloom causative species distributing widely in temperate and tropical coastal waters in the world. The morphological, physiological, and biochemical characteristics are different among geographic strains, which can not be distinguished with nuclear ribosomal DNA markers at present. Therefore, the genetic distance and phylogeographic relationships of nuclear 28S rDNA D1–D2 and ITS regions, and three chloroplast intergenic spacers (*pet*N-*trn*S1, *trn*M1-*psb*A, and *rbc*S-*rpl*27) were analyzed and compared among 13 strains of *P. globosa* isolated from the Pacific Ocean and Atlantic Ocean in this study. In addition, the nucleotide polymorphisms of 28S rDNA D1–D2, ITS, and *rbc*S-*rpl*27 regions were evaluated in two *P. globosa* strains. The various levels of nucleotide polymorphism were in the nuclear 28S rDNA D1–D2 region and ITS region, but no polymorphism was in the chloroplast *rbc*S-*rpl*27 intergenic spacer. A reasonable intraspecific phylogeographic relationship was presented by *rbc*S-*rpl*27 intergenic spacer, which had the strongest distinction to geographic strains compared to those of 28S rDNA D1–D2 and ITS regions. In the phylogenetic tree of *rbc*S-*rpl*27 intergenic spacer, the two strains from the North Sea of the Atlantic Ocean were divided firstly from the species of *P. globosa*, and then formed an independent clade, while the other Atlantic strains and all of Pacific strains joined up to build the other clade. It was implied that at least two genetically distant populations of *P. globosa* existed in the Atlantic coastal regions. This study provided a high-resolution chloroplast marker to analyze intraspecific phylogeographic populations of *P. globosa*, and preliminarily clarified the genetic relationships of the Pacific and Atlantic strains of *P. globosa*.

**Keyword**: *Phaeocystis globosa*; chloroplast; DNA marker; phylogeny

## 1 INTRODUCTION

Genus *Phaeocystis*, a major contributor of dimethylsulfide propionate (DMSP) enzymatically converted into dimethylsulfide (DMS) and acrylate (Stefels and Van Boekel, 1993; Stefels et al., 1995), is not only an important primary producer distributing from the Antarctic waters to the Arctic Ocean, but also an ecologically important member of phytoplankton with a significant role in biogeochemical cycles, including the global sulphur and carbon cycles (Charlson et al., 1987; Liss et al., 1994; Wassmann et al., 2005; Whipple et al., 2005a).

* Supported by the National Key Research and Development Plan Project, the Sino-Australian Centre for Healthy Coasts (No. 2016YFE0101500), the Science & Technology Basic Resources Investigation Program (No. 2018FY100206), and the National Natural Science Foundation of China (Nos. 41576121, 41776127)
** Corresponding author: rcyu@qdio.ac.cn

*Phaeocystis* contains some causative species of harmful algal blooms (HABs), during which ichthyotoxic and haemolytic substances can be often released with detrimental impacts on growth and reproduction of shellfish and macrozooplankton (Davidson and Marchant, 1992; Shen et al., 2004; Liu et al., 2007). To date, there are 10 species of the genus *Phaeocystis* recorded in the AlgaeBase (http://algaebase.org/), and three of them, *P. antarctica*, *P. globosa*, and *P. pouchetii*, are reported as the bloom-forming species. *P. antarctica* and *P. pouchetii* are cold-water species prevailing in the Antarctic and Arctic waters, respectively. While *P. globosa* is a warm-water species blooming in the nutrient-enrich (mostly nitrates) coastal areas of the north Atlantic, the North Sea, the Norwegian Sea, the Dutch coastal waters, the Arabian Sea, and the Southeast Asian including China, Vietnam, Thailand, etc. (Lancelot et al., 1987; Cadée, 1996; Qi et al., 2004; Schoemann et al., 2005; Hai et al., 2010; Rousseau et al., 2013; Smith et al., 2014b). Extensive research interests are focused on the species of *P. globosa* not only for its wide distribution, but also for its inclination to form gelatinous colonies encased by polysaccharide envelopes under suitable environmental conditions (Zingone et al., 1999; Jacobsen, 2002). Thousands of single cells are embedded in a colony in size from several millimeters (Rousseau et al., 1994; Peperzak et al., 2000), even to 3 centimeters in the blooms in China (Chen et al., 2002; Shen et al., 2004; Smith et al., 2014b). Therefore, their negative effects are indirectly resulted from the high abundance of colony cells and oxygen depletion upon its degradation in marine ecosystems (Smith et al., 2014b).

*Phaeocystis globosa* is supposed to be diverged from the cold-water species at approximately 30 million years ago (MYA) by molecular clock calculations (Medlin and Zingone, 2007), and now has spread in the temperate and tropical waters of the global ocean (Schoemann et al., 2005). Accordingly, the substantial diversities have been accumulated among the different geographic populations of *P. globosa* during the long time of evolution, such as the morphological characteristics, cell and colony sizes, pigment features, and even toxic impacts on marine organisms (Vaulot et al., 1994; Qi et al., 2004; Smith et al., 2014b; Wang et al., 2019). It had once been hypothesized that *P. globosa* might have sufficient time to develop detectable mutations at the nuclear ribosomal DNA (rDNA) locus to distinguish distinctly geographic populations and genetic divergence (Lange et al., 2002; Chen et al., 2003; Decelle et al., 2012). However, the fact is that although the phylogenetic relationships are close among *P. globosa* strains from same geographic region (Lange et al., 2002; Chen et al., 2003), there are no perspicuous phylogenetic relationships in different geographic populations of *P. globosa* based on nuclear 18S rRNA gene and ITS region, for a part reason that a high level of rDNA genetic overlaps and polymorphisms are present in different geographic strains of *P. globosa* (Lange et al., 2002; Chen et al., 2003; Qu et al., 2008; Decelle et al., 2012; Hu et al., 2019a). Moreover, the polymorphisms of rDNA sequences also exist in a colony, or in different clones from a single cell, which is a kind of phenomenon rarely occurred in the cold-water species *P. antarctica* (Lange et al., 2002; Medlin and Zingone, 2007). As a result, the existing of genetic overlaps and polymorphisms in rDNA sequences have greatly hampered the understanding of the phylogenetic relationships of *P. globosa* populations.

Compared to the nuclear ribosomal genes including 18S rDNA, 28S rDNA, and ITS region commonly used for phylogenetic analysis (Lange et al., 2002; Chen et al., 2003; Qu et al., 2008; Decelle et al., 2012; Hu et al., 2019a), haploidic organelle genes are often uni-parental inheritance, which exclude the impacts of genetic fusion during sexual reproduction on the geographic population analysis (Hu et al., 2019a). Previously, the mutation rates of *P. globosa* chloroplast genome had been believed to be lower than that in mitochondria and nucleus (Smith et al., 2014a), and the highly conservative chloroplast genes *psa*A, *psb*A, and RUBISCO intergenic spacer were often used as interspecific DNA markers of the genus *Phaeocystis* (Lange et al., 2002; Yang et al., 2004; Yang et al., 2015). Nowadays, more and more chloroplast genomes have been assembled and compared in fast-developing next-generation sequencing technology, and it is found that non-coding and coding regions of chloroplast genomes differ greatly in the evolution rate of many species of plants and algae (Yan et al., 2015). More hypervariable and richer variations such as substitutions, translocations, inversions, insertions, and deletions are found in non-coding region (Katayama et al., 2012), and the chloroplast intergenic spacers, for example, represent ideal segments for interspecific phylogenetic studies at low taxonomic levels and reconstruction of genetic relationships (Bakker et al., 1999; Kimura et al., 2003; Hu et al., 2011; Carbonell-

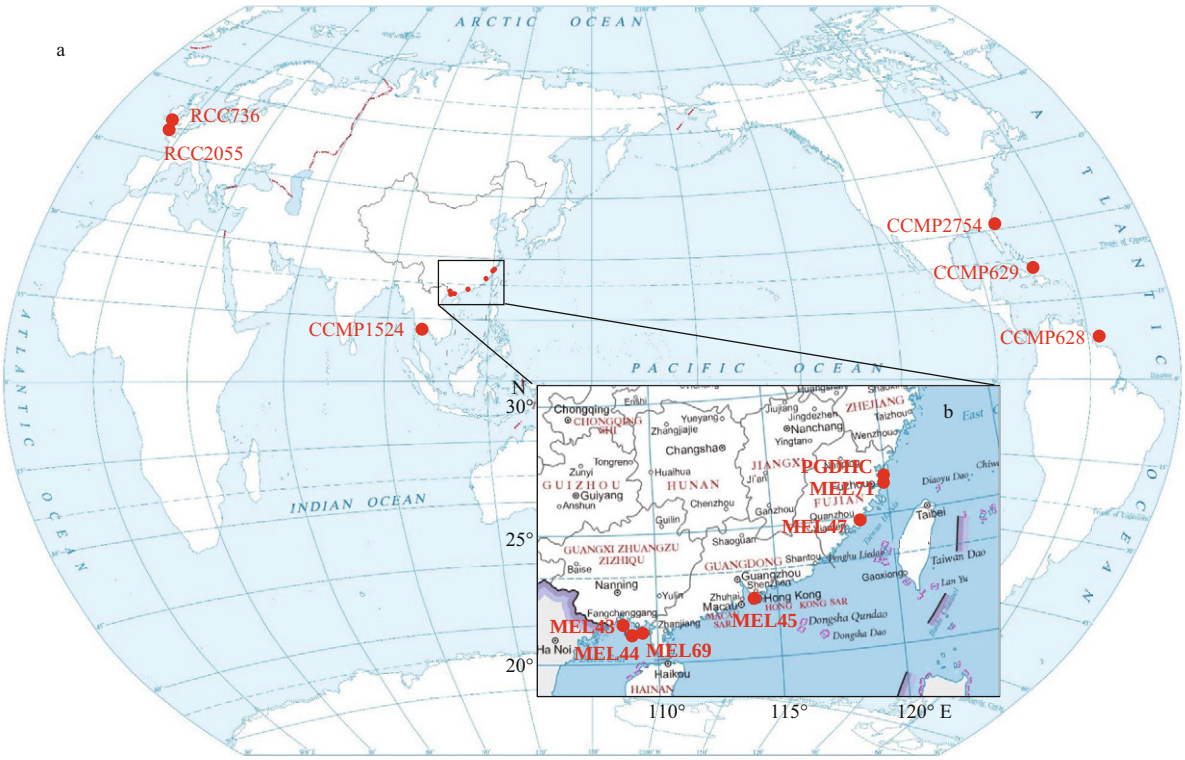



**Fig.1 Collection sites of thirteen strains of *Phaeocystis globosa***

a. world map (map drawing No. GS(2016)2962); b. collection sites of seven strains of *P. globosa* in China Sea (map drawing No. GS(2019)1708). ECS: the East China Sea; SCS: the South China Sea.

Caballero et al., 2015; Kholina et al., 2016; Raman et al., 2016; Zhang et al., 2016; Li et al., 2017; Vitales et al., 2019; Zhou et al., 2019). There is however no report on the evaluation and application of chloroplast intergenic markers for intraspecific phylogeographic studies on *P. globosa* at present.

In this study, the sequence polymorphisms, genetic distances and phylogeographic relationships of nuclear 28S rDNA D1–D2 and ITS regions, and three chloroplast intergenic spacers in different geographic strains of *P. globosa* from the Pacific and Atlantic coastal waters were compared and evaluated. The results presented a high-resolution chloroplast DNA marker to differentiate the geographic strains of *P. globosa*; therefore, much useful information was provided about the phylogenetic relationships of the Pacific and Atlantic populations.

## 2 MATERIAL AND METHOD

### 2.1 Algal culture

Seventeen strains of *Phaeocystis* genus were used in this study, including 13 strains of *P. globosa* isolated from the Pacific and Atlantic coastal waters (Fig.1), one strain of *P. antarctica*, one strain of *P. cordata*, one strain of *P. rex*, and one strain of *P. jahnii* (Table 1). They were all cultured in the laboratory at 18±1 °C with the light intensity of 100 μE/(m$^2$·s) and a light cycle of 14 h:10 h (light:dark), with the except of *P. antarctica* at 5±1 °C. Approximately 20 mL of algal cultures were filtered on a 0.4-μm HTTP membrane (Millipore, USA) under the vacuum level of 40 kPa, and the membranes were then kept at -80 °C until DNA extraction.

### 2.2 Design and selection of primers

The chloroplast genomes of *P. antarctica* (JN117275) and *P. globosa* Pg-G(A) (KC900889) were downloaded from the GenBank, and aligned by Clustal W. Three chloroplast intergenic spacers (*pet*N-*trn*S1, *trn*M1-*psb*A and *rbc*S-*rpl*27) with high variability were selected as the target sequences, and three primer pairs were designed accordantly by Primer 5.05: *pet*N-*trn*S1-F (5′-GTCCCCAAACT-ACTAATGCT-3′), *pet*N-*trn*S1-R (5′-GGTGGCTG-AGTGGTTGAAAG-3′); *trn*M1-*psb*A-F (5′-AGCA-GTTTGGTAGCTCGTCG-3′), *trn*M1-*psb*A-R (5′-TGTTGATTCTCAAGGTCGTGTA-3′); *rbc*S-*rpl*27-F (5′-GAAGCATACGCACCAAGC-3′), *rbc*S-*rpl*27-R (5′-CAGCCAGTAAAGGCAGGT-3′). The primer pair for 28S rDNA D1–D2 (forward primer LSU-F2: 5′-ASAGYCGCCTCCTGAATTGTAGTC-3′,

**Table 1 Information of microalgae in this study**

| Species | Strain | Source for culture | Sampling area | Sampling date |
|---|---|---|---|---|
| *P. globosa* | MEL43 | Single colony | Beibu Gulf, South China Sea, Pacific Ocean | 2018 |
| *P. globosa* | MEL44 | Single cell | Beibu Gulf, South China Sea, Pacific Ocean | 2017 |
| *P. globosa* | MEL45 | – | Zhujiang (Pearl) River estuary, South China Sea, Pacific Ocean | – |
| *P. globosa* | MEL47 | Single cell | Quanzhou coast, East China Sea, Pacific Ocean | 2017 |
| *P. globosa* | MEL69 | Single colony | Beibu Gulf, South China Sea, Pacific Ocean | 2017 |
| *P. globosa* | MEL71 | Single cell | Sansha Bay, East China Sea, Pacific Ocean | 2019 |
| *P. globosa* | PGDHC | Single colony | Sansha Bay, East China Sea, Pacific Ocean | 2019 |
| *P. globosa* | RCC736 | – | North Sea, Atlantic Ocean | 1991 |
| *P. globosa* | RCC2055 | – | English Channel, Atlantic Ocean | 2003 |
| *P. globosa* | CCMP628 | – | Caribbean Sea, Atlantic Ocean | 1965 |
| *P. globosa* | CCMP629 | – | Gulf of Mexico, Atlantic Ocean | 1982 |
| *P. globosa* | CCMP1524 | – | Gulf of Thailand, Pacific Ocean | 1992 |
| *P. globosa* | CCMP2754 | – | North America, Atlantic Ocean | 2003 |
| *P. antarctica* | CCMP1374 | – | McMurdo Sound, Antarctica | 1991 |
| *P. cordata* | CCMP3104 | – | Mediterranean Sea, Atlantic Ocean | 1991 |
| *P. rex* | CCMP2000 | – | Arabian Sea, Indian Ocean | 1995 |
| *P. jahnii* | CCMP2496 | – | Mediterranean Sea, Atlantic Ocean | 1996 |

"–": information unknown.

reverse primer LSU-R3: 5′-TCGAGCTTGCCACTCTAGTACTC-3′) (Hu et al., 2019a), and the primer pair for ITS region (forward primer Euk P18S (1705-)-F: 5′-GCCGGACGCGACGCTCC-3′, and reverse primer ITS-2R: 5′-GCTTATTGATATGCTTAAGTTCAGCGGGT-3′) (Blomster et al., 1998; Lange et al., 2002) were applied as references.

### 2.3 Genomic DNA extraction, PCR amplification, and sequencing

Total genomic DNA of microalgal samples was extracted with the modified method described by Winnepenninckx et al. (1993), and stored at -80 °C until analysis. PCRs were executed with a final volume of 20 μL containing 8.2-μL $ddH_2O$, 4 μL 5×rTaq Buffer ($Mg^{2+}$), 4 μL 5×PCR solution buffer, 1.6-μL dNTP Mix (2.5 mmol/L), 1 μL of DNA template (10 ng DNA template in each reaction), 0.4 μL of each PCR primer (10 pmol/mL), and 0.4-μL super-fidelity Taq DNA polymerase (5 U/mL, TransGen Biotech Co., Ltd., China). The amplifications were performed with an initial denaturation temperature at 98 °C for 2 min, 32 cycles of 98 °C for 20 s, 58 °C for 20 s, and 72 °C for 30 s, and a final elongation at 72 °C for 5 min. The PCR products were confirmed by electrophoresis in 1% agar gel, and the targeted DNA bands were purified and then sequenced from both ends by Sangon Biotech (Shanghai). The obtained forward and reversed fragments were assembled by Invitrogen 11.0 ContigExpress.

### 2.4 Clone library construction

Because the PCR productions of ITS regions from 10 strains of *P. globosa* could not be assembled due to the nucleotide differences, the clone libraries were constructed respectively by using the purified PCR fragments to obtain their full sequences, and to show the polymorphisms in the target genes. The fragments were cloned into blunt simple cloning vectors (TransGen Biotech Co., Ltd., China), and the vectors were transformed into *Trans*1-T1 competent *Escherichia coli* cells (TransGen Biotech Co., Ltd., China), which were placed on LB ampicillin plates containing final concentration of 100 μg/mL ampicillin. Three colonies were random selected from each plate and sequenced from both ends.

In addition, the clone libraries of 28S rDNA D1–D2 region, ITS region and *rbc*S-*rpl*27 intergenic spacer of *P. globosa* MEL43 and MEL47 were constructed as the above-mentioned, except that 30–50 colonies from each library were selected and sequenced from both ends.

### 2.5 Genetic distances and phylogenetic analysis

The sequences of 28S rDNA region and ITS region of *P. globosa*, *P. antarctica*, *P. cordata* were downloaded from the GenBank (Suppl. Table S1). Together with the sequences of 28S rDNA D1–D2 region, ITS region and *pet*N-*trn*S1, *trn*M1-*psb*A, and *rbc*S-*rpl*27 intergenic spacers obtained in this study, the genetic distances of the different target sequences were estimated by Mega 10 pairwise distance method with bootstrap replications of 1 000 (Kumar et al., 2018). The maximum-likelihood trees of these five genes and intergenic spacers were built by Mega 10 variance estimation method with the bootstrap replications of 1 000 and the best-fitting nucleotide substitution models of T92, GTR+G, T92+G, T92 and K2+G were selected, respectively (Kumar et al., 2018). The Bayesian analysis of these five genes and intergenic spacers was constructed by MrBayes 3.2.6 (Ronquist et al., 2012) with generations of 5 000 000 and sampling frequencies of 100, and the best-fitting nucleotide substitution models of HKY+F+I, GTR+F+G4, HKY+F+I, HKY+F+G4, and K2P+G4, were selected using Modelfinder (Kalyaanamoorthy et al., 2017). The sequences of *P. antarctica* and *P. cordata* were as the outside group. Additionally, the sequences of 28S rDNA D1–D2 region, ITS region and *rbc*S-*rpl*27 intergenic spacer resulted from the clone libraries of *P. globosa* MEL43 and MEL47 were aligned by Clustal W. The multiple aligned sequences were used to analyze the relative frequency of the nucleic acid at each position by the online software WebLogo 3 (http://weblogo.berkeley.edu/logo.cgi) (Crooks et al., 2004).

## 3 RESULT

### 3.1 Primers design and their performances

The chloroplast genomes of *P. globosa*, and *P. antarctica* were downloaded and aligned, and three chloroplast intergenic spacers with a relatively high nucleotide variation level were identified and acted as the aimed regions, which were *pet*N-*trn*S1, *trn*M1-*psb*A, and *rbc*S-*rpl*27 intergenic spacers. Three primer pairs were designed and used for the amplification of the targeted chloroplast intergenic spacers from all the 15 *Phaeocystis* strains in this study in a different sequence length, and the amplified sequences were submitted to GenBank (ID MN807922–MN807935, MN928971–MN928984, MN935489–MN935501, and MN956829, the detailed information is shown in Suppl. Table S2). In detail, the *pet*N-t*rn*S1 spacers of *P. antarctica* CCMP1374 and *P. cordata* CCMP3104 were 368 bp and 341 bp, respectively, while for the strains of *P. globosa*, the *pet*N-t*rn*S1 spacer of RCC736 was 347 bp and the other 12 strains were 528 bp. The *trn*M1-*psb*A spacers of *P. antarctica* CCMP1374 and *P. cordata* CCMP3104 were 549 bp and 418 bp, respectively, and that of *P. globosa* were from 648 bp to 682 bp. The *rbc*S-*rpl*27 spacers of *P. antarctica* CCMP1374 and *P. cordata* CCMP3104 were 1 115 bp and 709 bp, respectively. Within the species of *P. globosa*, the *rbc*S-*rpl*27 spacer of RCC736 was 599 bp and the other 12 strains were 609 bp.

### 3.2 Genetic distances and sequence polymorphism of nuclear 28S rDNA D1–D2 region of *P. globosa*

The sequences of 28S rDNA D1–D2 region with the length of 522 bp were amplified by the primer pair LSU-F2 and LSU-R3 from the 13 isolated strains of *P. globosa* and submitted to GenBank (ID MN602538–MN602550, the detailed information is shown in Suppl. Table S2). Combined with other 17 *P. globosa* 28S rDNA D1–D2 region sequences downloaded from the GenBank (Suppl. Table S1), the genetic distances of 28S rDNA D1–D2 region from different strains of *P. globosa* were analyzed. As shown in Table 2, the obtained sequences could be categorized into nine groups with one to eleven same members in each group, and the values of genetic distances between these groups ranged from 0.000 0 to 0.013 7, while the grouping of these sequences failed to be directly related to the geographic regions.

To examine the potential polymorphisms that might be neglected by the directly assemble of the forward and reverse PCR fragments, the representative strains of MEL43 and MEL47 were chosen for the clone library construction, and 34 sequences (GenBank ID MN602900–MN602933, the detailed information is shown in Suppl. Table S3) and 40 sequences (GenBank ID MN610012–MN610051, the detailed information is shown in Suppl. Table S3) of 28S rDNA D1–D2 region were obtained. The alignment result showed that there were seven and four SNP sites existed in MEL43 and MEL47 clones, respectively (Suppl. Figs.S1 & S2), and 7 SNP sites between the two strains (Fig.2), indicating a relatively low level of the gene variations in this region.

**Table 2 Genetic distances between 28S rDNA D1–D2 region sequences of *Phaeocystis globosa***

| Seq No.* | 1–11 | 12 | 13–15 | 16–24 | 25–26 | 27 | 28 | 29 |
|---|---|---|---|---|---|---|---|---|
| 1–11 | 0.000 0 | | | | | | | |
| 12 | 0.001 9 | | | | | | | |
| 13–15 | 0.003 8 | 0.005 8 | 0.000 0 | | | | | |
| 16–24 | 0.003 8 | 0.005 8 | 0.007 7 | 0.000 0 | | | | |
| 25–26 | 0.005 8 | 0.007 7 | 0.009 7 | 0.001 9 | 0.000 0 | | | |
| 27 | 0.005 8 | 0.007 7 | 0.009 7 | 0.001 9 | 0.003 8 | | | |
| 28 | 0.005 8 | 0.007 7 | 0.009 7 | 0.001 9 | 0.003 8 | 0.003 9 | | |
| 29 | 0.007 7 | 0.009 7 | 0.003 8 | 0.011 6 | 0.013 6 | 0.013 6 | 0.013 6 | |
| 30 | 0.009 7 | 0.011 7 | 0.005 8 | 0.011 7 | 0.013 6 | 0.013 7 | 0.013 7 | 0.005 8 |

*: 1: MEL43 (SCS Pacific); 2: MEL44 (SCS Pacific); 3: MEL45 (SCS Pacific); 4: MEL47 (ECS Pacific); 5: MEL69 (SCS Pacific); 6: CCMP628 (Caribbean Sea Atlantic); 7: CCMP629 (Gulf of Mexico Atlantic); 8: CCMP1524 (Gulf of Thailand Pacific); 9: RCC575 (JX660933) (Atlantic); 10: RCC576 (JX660932) (Atlantic); 11: RCC2055 (English Channel Atlantic); 12: ZX28-0.2-29 (SCS Pacific) KT390053; 13: RCC540 (North Atlantic) JX660930; 14: RCC1398 (China Sea Pacific) JX660936; 15: *Phaeocystis* sp. RCC1000 (Pacific) JX660940; 16: *Phaeocystis* sp. Ros1 symbiont JX660804; 17: *Phaeocystis* sp. JD-2012 Ros2 symbiont JX660805; 18: *Phaeocystis* sp. PCC679 JX660929; 19: RCC64 (English Channel Pacific) JX660934; 20: *Phaeocystis* sp. AC618 EU502882; 21: K1321 (Portugal Atlantic) JX660935; 22: NIES388 (Pacific) JX660937; 23: RCC678 (North Sea Atlantic) JX660939; 24: CCMP2754 (North America Atlantic); 25: PGDHC (SCS Pacific); 26: MEL71 (SCS Pacific); 27: RCC736 (North Sea-Atlantic); 28: *Phaeocystis* sp. Ind12 symbiont JX660785; 29: ZX14-0.2-12 (SCS Pacific) KT390022; 30: ZX14-0.2-15 (SCS Pacific) KT390023.

**Table 3 Genetic distances of ITS region sequences from the clones of ten strains of *P. globosa***

| Strain | MEL43 | MEL44 | MEL45 | MEL47 | MEL69 | RCC2055 | CCMP628 | CCMP629 | CCMP1524 | CCMP2754 |
|---|---|---|---|---|---|---|---|---|---|---|
| Evolutionary divergence | 0.000 0–0.008 4 | 0.001 4–0.005 6 | 0.011 2–0.015 4 | 0.000 0–0.005 6 | 0.001 4–0.009 8 | 0.011 2–0.019 7 | 0.005 6–0.005 6 | 0.001 4–0.011 2 | 0.000 0–0.030 9 | 0.000 0–0.040 7 |

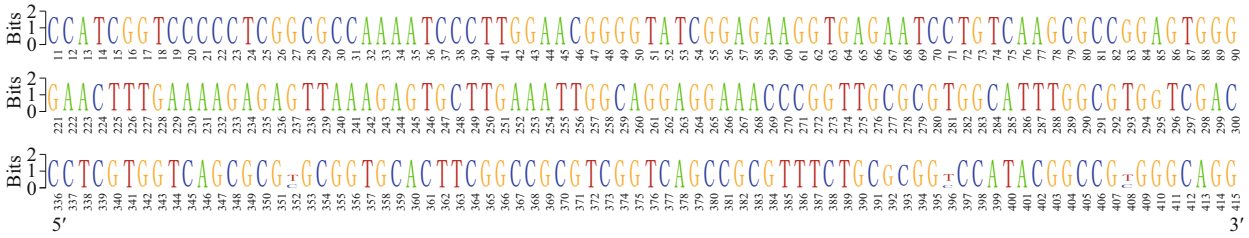



**Fig.2 WebLogo of 28S rDNA D1–D2 region between two strains of *Phaeocystis globosa* MEL43 and MEL47**

### 3.3 Genetic distances and sequence polymorphism of nuclear ribosomal ITS region of *P. globosa*

The sequences of nuclear ribosomal ITS region were successfully amplified by the primer pair Euk P18S and ITS-2R from all thirteen strains of *P. globosa*, but for only three strains of MEL71, PGDH2019C and RCC736, the forward and reverse PCR fragments could be directly assembled with the length of 871–872 bp. For the other ten strains, ITS clone libraries were constructed, and then three clones from each library were randomly sequenced from both ends to get the full gene length of 895–925 bp. Accordantly with the direct assemble result of PCR productions, ITS sequence variations from the clone libraries were observed in every strain of *P. globosa* with the divergent rate of 0.001 4–0.040 7 (Table 3), indicating a high polymorphism level in this region. The genetic distances were analyzed for the obtained ITS sequences (GenBank ID MN603007–MN603009, MN603057–MN603086, the detailed information in Suppl. Tables S2 & S4) and the other nineteen sequences of *P. globosa* downloaded from GenBank (Suppl. Table S1). Their values of genetic distances ranging 0.000 0–0.067 4 (Suppl. Table S5), were much higher than that of 28S rDNA D1–D2 region.

To further evaluate ITS polymorphism level in different strains of *P. globosa*, the clone libraries sampled from strain MEL43 and MEL47 were constructed, and 28 sequences (ID MN603081–MN603083, MN603166–MN603190, the detailed information is shown in Suppl. Table S4) and 35 sequences (ID MN603084–MN603086, MN603191–MN603222, the detailed information in Suppl.

**Table 4 Genetic distances between sequences of *pet*N-*trn*S1 intergenic spacer of fourteen strains of *Phaeocystis globosa***

| Strain | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 MEL43 (SCS Pacific) | | | | | | | | | | | | | |
| 2 MEL44 (SCS Pacific) | 0.000 0 | | | | | | | | | | | | |
| 3 MEL45 (SCS Pacific) | 0.000 0 | 0.000 0 | | | | | | | | | | | |
| 4 MEL69 (SCS Pacific) | 0.000 0 | 0.000 0 | 0.000 0 | | | | | | | | | | |
| 5 MEL47 (ECS Pacific) | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | | | | | | | | | |
| 6 CCMP1524 (Gulf of Thailand Pacific) | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | | | | | | | | |
| 7 CCMP628 (Caribbean Sea Atlantic) | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | | | | | | | |
| 8 CCMP629 (Gulf of Mexico Atlantic) | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | | | | | | |
| 9 RCC2055 (English Channel Atlantic) | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | | | | | |
| 10 CCMP2754 (North America Atlantic) | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | | | | |
| 11 MEL71 (ECS Pacific) | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.006 4 | | | |
| 12 PGDHC (ECS Pacific) | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.003 2 | 0.006 4 | 0.000 0 | | |
| 13 RCC736 (North Sea Atlantic) | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.118 6 | 0.118 6 | |
| 14 Pg-G(A) (North Sea Atlantic) | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.115 4 | 0.118 6 | 0.118 6 | 0.000 0 |

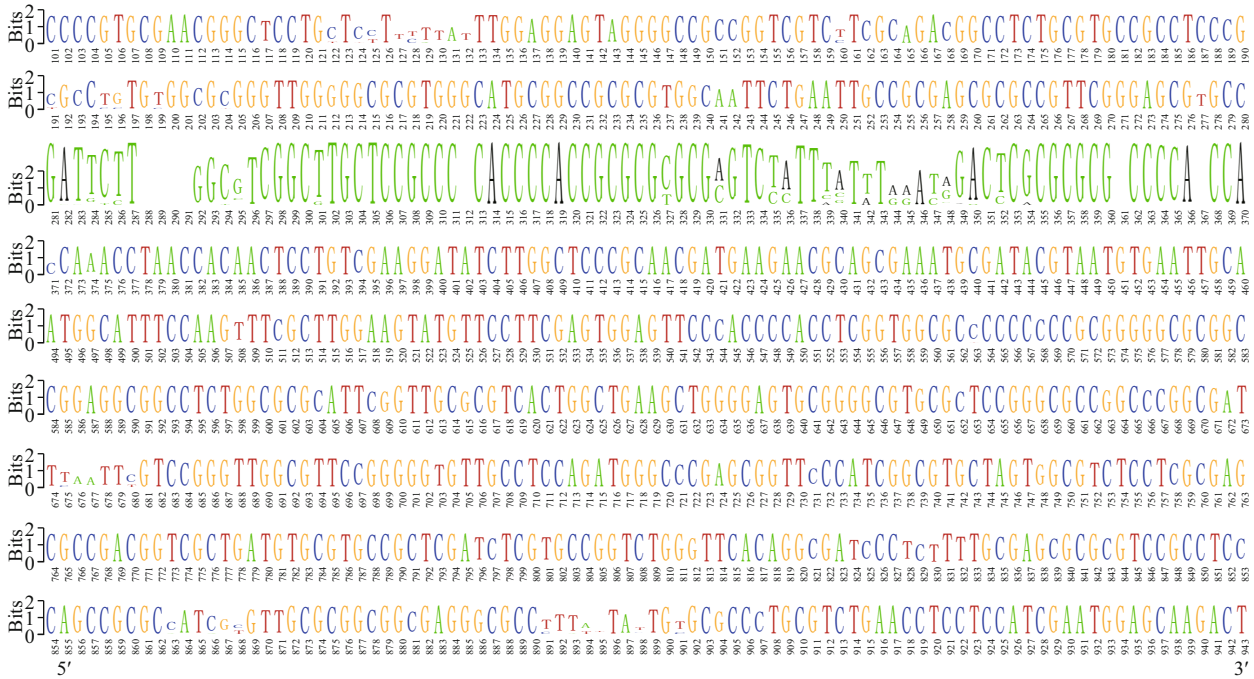



**Fig.3 WebLogo of ITS region between two strains of *P. globosa* MEL43 and MEL47**

Table S5) were sequenced and assembled respectively. The webLogo analysis revealed 160 and 119 sites nucleotide variation, insertion and deletion in the MEL43 and MEL47 clone libraries (Suppl. Figs.S3 & S4). In addition, there were 177 nucleotide sites varying between the two strains, which accounted for 19.13% sites of the full sequences (Fig.3). The results displayed a rather high level of mutation rate in the ITS region of *P. globosa*.

### 3.4 Genetic distances and sequence polymorphism of the chloroplast intergenic spacers of *P. globosa*

The genetic distances of three chloroplast intergenic spacers of *P. globosa* were analyzed and compared. As shown in Table 4, the *pet*N-*trn*S1 spacer sequences had no difference in the nine strains from SCS, ECS and Gulf of Thailand Pacific, and Caribbean Sea, Gulf of Mexico and English Channel Atlantic, while the

**Table 5 Genetic distances between sequences of *trn*M1-*psb*A intergenic spacer of fourteen strains of *Phaeocystis globosa***

| Strain | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 MEL43 (SCS Pacific) | | | | | | | | | | | | | |
| 2 MEL44 (SCS Pacific) | 0.000 0 | | | | | | | | | | | | |
| 3 MEL45 (SCS Pacific) | 0.000 0 | 0.000 0 | | | | | | | | | | | |
| 4 MEL69 (SCS Pacific) | 0.000 0 | 0.000 0 | 0.000 0 | | | | | | | | | | |
| 5 MEL47 (ECS Pacific) | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | | | | | | | | | |
| 6 CCMP1524 (Gulf of Thailand Pacific) | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | 0.000 0 | | | | | | | | |
| 7 CCMP628 (Caribbean Sea Atlantic) | 0.011 6 | 0.011 6 | 0.011 6 | 0.011 6 | 0.011 6 | 0.011 6 | | | | | | | |
| 8 CCMP629 (Gulf of Mexico Atlantic) | 0.011 6 | 0.011 6 | 0.011 6 | 0.011 6 | 0.011 6 | 0.011 6 | 0.000 0 | | | | | | |
| 9 RCC2055 (English Channel Atlantic) | 0.011 6 | 0.011 6 | 0.011 6 | 0.011 6 | 0.011 6 | 0.011 6 | 0.009 9 | 0.009 9 | | | | | |
| 10 CCMP2754 (North America Atlantic) | 0.013 2 | 0.013 2 | 0.013 2 | 0.013 2 | 0.013 2 | 0.013 2 | 0.011 6 | 0.011 6 | 0.014 9 | | | | |
| 11 MEL71 (ECS Pacific) | 0.021 5 | 0.021 5 | 0.021 5 | 0.021 5 | 0.021 5 | 0.021 5 | 0.026 4 | 0.026 4 | 0.023 1 | 0.021 5 | | | |
| 12 PGDHC (ECS Pacific) | 0.021 5 | 0.021 5 | 0.021 5 | 0.021 5 | 0.021 5 | 0.021 5 | 0.026 4 | 0.026 4 | 0.023 1 | 0.021 5 | 0.000 0 | | |
| 13 RCC736 (North Sea Atlantic) | 0.318 5 | 0.318 5 | 0.318 5 | 0.318 5 | 0.318 5 | 0.318 5 | 0.316 8 | 0.316 8 | 0.320 1 | 0.316 8 | 0.320 1 | 0.320 1 | |
| 14 Pg-G(A) (North Sea Atlantic) | 0.316 8 | 0.316 8 | 0.316 8 | 0.316 8 | 0.316 8 | 0.316 8 | 0.315 2 | 0.315 2 | 0.318 5 | 0.315 2 | 0.318 5 | 0.318 5 | 0.001 7 |

**Table 6 Genetic distances between sequences of *rbc*S-*rpl*27 intergenic spacer of 14 strains of *Phaeocystis globosa***

| Strain | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 MEL43 (SCS Pacific) | | | | | | | | | | | | | |
| 2 MEL44 (SCS Pacific) | 0.000 0 | | | | | | | | | | | | |
| 3 MEL45 (SCS Pacific) | 0.000 0 | 0.000 0 | | | | | | | | | | | |
| 4 MEL69 (SCS Pacific) | 0.000 0 | 0.000 0 | 0.000 0 | | | | | | | | | | |
| 5 MEL47 (ECS Pacific) | 0.003 6 | 0.003 6 | 0.003 6 | 0.003 6 | | | | | | | | | |
| 6 CCMP1524 (Gulf of Thailand Pacific) | 0.003 6 | 0.003 6 | 0.003 6 | 0.003 6 | 0.000 0 | | | | | | | | |
| 7 CCMP628 (Caribbean Sea Atlantic) | 0.005 3 | 0.005 3 | 0.005 3 | 0.005 3 | 0.005 3 | 0.005 3 | | | | | | | |
| 8 CCMP629 (Gulf of Mexico Atlantic) | 0.005 3 | 0.005 3 | 0.005 3 | 0.005 3 | 0.005 3 | 0.005 3 | 0.000 0 | | | | | | |
| 9 RCC2055 (English Channel Atlantic) | 0.014 2 | 0.014 2 | 0.014 2 | 0.014 2 | 0.014 2 | 0.014 2 | 0.008 9 | 0.008 9 | | | | | |
| 10 CCMP2754 (North America Atlantic) | 0.024 9 | 0.024 9 | 0.024 9 | 0.024 9 | 0.024 9 | 0.024 9 | 0.019 5 | 0.019 5 | 0.023 1 | | | | |
| 11 MEL71 (ECS Pacific) | 0.030 2 | 0.030 2 | 0.030 2 | 0.030 2 | 0.030 2 | 0.030 2 | 0.024 9 | 0.024 9 | 0.028 4 | 0.019 5 | | | |
| 12 PGDHC (ECS Pacific) | 0.030 2 | 0.030 2 | 0.030 2 | 0.030 2 | 0.030 2 | 0.030 2 | 0.024 9 | 0.024 9 | 0.028 4 | 0.019 5 | 0.000 0 | | |
| 13 RCC736 (North Sea Atlantic) | 0.087 0 | 0.087 0 | 0.087 0 | 0.087 0 | 0.087 0 | 0.087 0 | 0.081 7 | 0.081 7 | 0.083 5 | 0.087 0 | 0.090 6 | 0.090 6 | |
| 14 Pg-G(A) (North Sea Atlantic) | 0.087 0 | 0.087 0 | 0.087 0 | 0.087 0 | 0.087 0 | 0.087 0 | 0.081 7 | 0.081 7 | 0.083 5 | 0.087 0 | 0.090 6 | 0.090 6 | 0.000 0 |

other two strains from ECS Pacific (MEL71 and PGDHC) were different and showed a moderate variation compared to other ECS Pacific strains, but same with the one from the North America Atlantic (CCMP2754). The highest divergence was observed in two strains from the North Sea Atlantic (RCC736 and Pg-G(A)), indicating their early separation from other strains.

As shown in Table 5, there was no nucleotide difference in the *trn*M1-*psb*A spacer sequences of six strains from SCS, ECS, and Gulf of Thailand in the Pacific Ocean, and these strains had genetic distance of 0.011 6 with three strains from Caribbean Sea, Gulf of Mexico, and English Channel Atlantic (CCMP628, CCMP629 and RCC2055), respectively. The CCMP2754 strain and the other two strains from ECS Pacific (MEL71 and PGDHC) had further genetic distance with the value of 0.011 6–0.021 5. And the highest divergence (0.315 2–0.320 1) was observed in two strains from the North Sea Atlantic (RCC736 and Pg-G(A)).

As shown in Table 6, the relationships of genetic

distances for *rbc*S-*rpl*27 intergenic spacer sequences were similar with that for *trn*M1-*psb*A. But it was worth noting that the four SCS strains generated genetic divergence in *rbc*S-*rpl*27 when compared to other strains isolated ECS and Gulf of Thailand, which exemplified that *rbc*S-*rpl*27 provided better performance to distinguish the different geographic strains of *P. globosa* than the other two markers of *pet*N-*trn*S1 and *trn*M1-*psb*A. The polymorphisms of this region were then further testified by clone sequencing sampled from MEL43 and MEL47, respectively. The results showed that there was no nucleotide variation of *rbc*S-*rpl*27 in 40 sequences from MEL43 (GenBank ID MN990301–MN990340, the detailed information in Suppl. Table S6) and 32 sequences from MEL47 (GenBank ID MN990269–MN990300, the detailed information in Suppl. Table S6).

### 3.5 Phylogenetic relationship of different geographic strains of *P. globosa*

Thirty sequences of 28S rDNA D1–D2 region of *P. globosa* strains from the Pacific Ocean and Atlantic Ocean, were used to build the phylogenetic tree, with the sequences from *P. antarctica* and *P. cordata* as the outgroup branch. These strains of *P. globosa* were mainly divided into two clades (Fig.4): the sequences of 11 strains of *P. globosa* isolated from the Pacific Ocean and the Atlantic Ocean formed one clade with the other six strains sampled from China Sea, the Pacific Ocean, and the North Atlantic Ocean. In the other clade, 9 strains (*Phaeocystis* sp. Ros1, *Phaeocystis* sp. JD-2012 Ros2, *Phaeocystis* sp. RCC679, *Phaeocystis* sp. AC618, *P. globosa* CCMP2754, *P. globosa* RCC64, *P. globosa* RCC678, *P. globosa* K1321, and *P. globosa* NIES388) had no nucleotide difference, and then gathered with PGDHC, MEL71, RCC736, AC618 and three endosymbiotic strains (*Phaeocystis* sp. Ind12, *Phaeocystis* sp. Ros1, and *Phaeocystis* sp. JD-2012 Ros2). It was obvious that the phylogenetic tree of 28S rDNA failed to differentiate the *P. globosa* strains from distinct regions.

Thirty-three ITS region sequences of *P. globosa* obtained in this study and the other nineteen *P. globosa* ITS region sequences downloaded from the GenBank were used to construct the phylogenetic tree. The phylogenetic relationship of different geographic *P. globosa* strains was more complex than that in the 28S rDNA tree. The MEL71 and PGDHC isolated from the Sansha Bay of ECS stayed together, and then gathered with two sequences isolated from a *P. globosa* bloom in the Bohai Sea in the year of 2005 (Qu et al., 2008), staying at the base position of *P. globosa* clade. They were close to the three sequences of CCMP2754 clones, and were followed by RCC736 and seven sequences of *P. globosa* isolated from the English Channel. The clone sequences of nine strains of *P. globosa* (MEL43, MEL44, MEL69, MEL45, MEL47, CCMP628, CCMP629, CCMP1524, and RCC2055) stayed together, and formed a big branch with the sequences of the different clones of CCMP627 isolated from the Gulf of Mexico in the Atlantic Ocean, as well as Santou 97 and JD-2012 collected from SCS (Fig.5). As a result, the phylogenetic tree of ITS region interlaced, and the phylogenetic relationships of *P. globosa* were inconsistent with their geographic distribution (Fig.5).

The phylogenetic trees of *pet*N-*trn*S1, *trn*M1-*psb*A, and *rbc*S-*rpl*27 intergenic spacers were similar (Figs.6–8), all of *P. globosa* strains stayed together, closing with *P. antarctica*, and far away from *P. cordata*. However, the branch of *P. rex* and *P. jahnii* had greater genetic distances with *P. globosa* than *P. antarctica* and *P. cordata* in the tree of *rbc*S-*rpl*27. In the three trees, all of *P. globosa* were divided into two independent clades, respectively. Collected from the Netherlands coastal waters of the Atlantic Ocean (Smith et al., 2014a), the two strains of *P. globosa* (RCC736 and Pg-G(A)) early separated from other strains and formed a clade staying at the base position. The *rbc*S-*rpl*27 tree had the highest resolution in the other clade, for that it could distinguish the four SCS strains (MEL43, MEL44, MEL69 and MEL45) from the two strains collected from the adjacent coastal waters of ECS and the Gulf of Thailand (MEL47 and CCMP1524), however all of them still stayed in a same branch in both *pet*N-*trn*S1 and *trn*M1-*psb*A trees. Moreover, a common feature of this clade was that the Pacific and Atlantic strains were generally clustered separately, except that MEL71 and PGDHC collected from ECS gathered with the Atlantic strain CCMP2754 with a bit difference.

## 4 DISCUSSION

Diverging from cold-water species at 30 MYA, *P. globosa* distributes in both temperate and tropical oceans worldwide (Schoemann et al., 2005; Medlin and Zingone, 2007). Diverse geographic distribution patterns had urged the existence of evolutionary

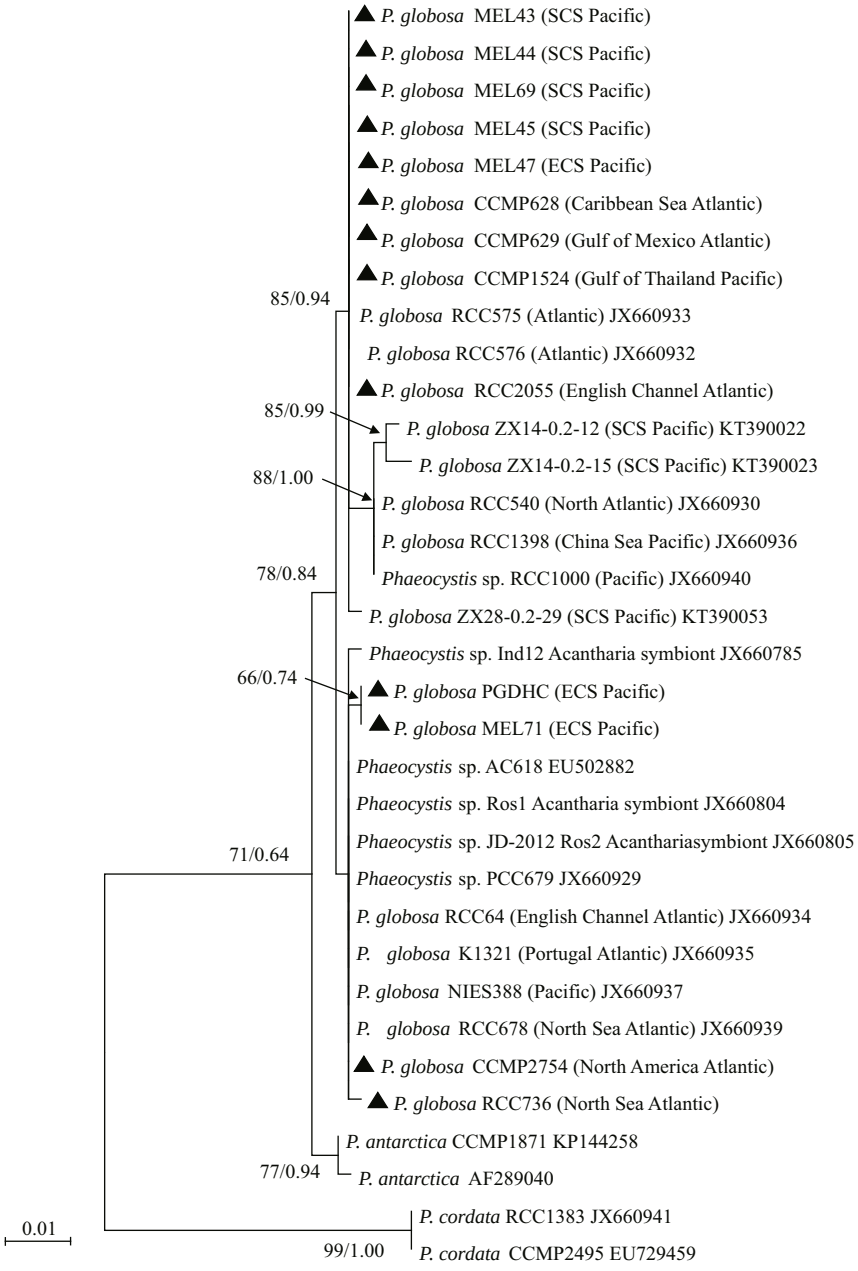



**Fig.4 Phylogenetic tree of 28S rDNA D1–D2 region of *Phaeocystis globosa***

Numbers indicated the bootstrap values from maximum-likelihood (left) and Bayesian (right) analyses. The sequences with a triangle were obtained in this study.

divergence, and genetic differentiation had taken place within the *P. globosa* taxon (Lange et al., 2002; Chen et al., 2003; Hu et al., 2019a). The convincing genetic divergence among different geographic strains is confirmed by a lot of studies on the basis of sequences of the nuclear 18S rDNA (Lange et al., 2002; Qu et al., 2008), 28S rDNA (Hu et al., 2019a), ITS (Lange et al., 2002; Chen et al., 2003; Qu et al., 2008; Liu et al., 2010; Hu et al., 2019a), and the concatenation of 18S rDNA, 28S rDNA, *psb*A and *rbc*L genes (Decelle et al., 2012). Morphological characteristics and pigment features also provide reliable evidences (Vaulot et al., 1994; Qi et al., 2004; Hu et al., 2019b; Wang et al., 2019), such as the giant colonies of Southeast Asian strains (Qi et al., 2004; Smith et al., 2014b), and the change of concentration of characteristic pigments 19'-hexanoyloxyfucoxanthin (Hex-fuco) and 19'-butanoyloxyfucoxanthin (But-fuco) of different geographic strains (Vaulot et al., 1994; Antajan et al., 2004; Zapata et al., 2004; Seoane et al., 2009; Wang et al., 2019). In brief, the differentiation of morphological features, biochemical characteristics, and molecular

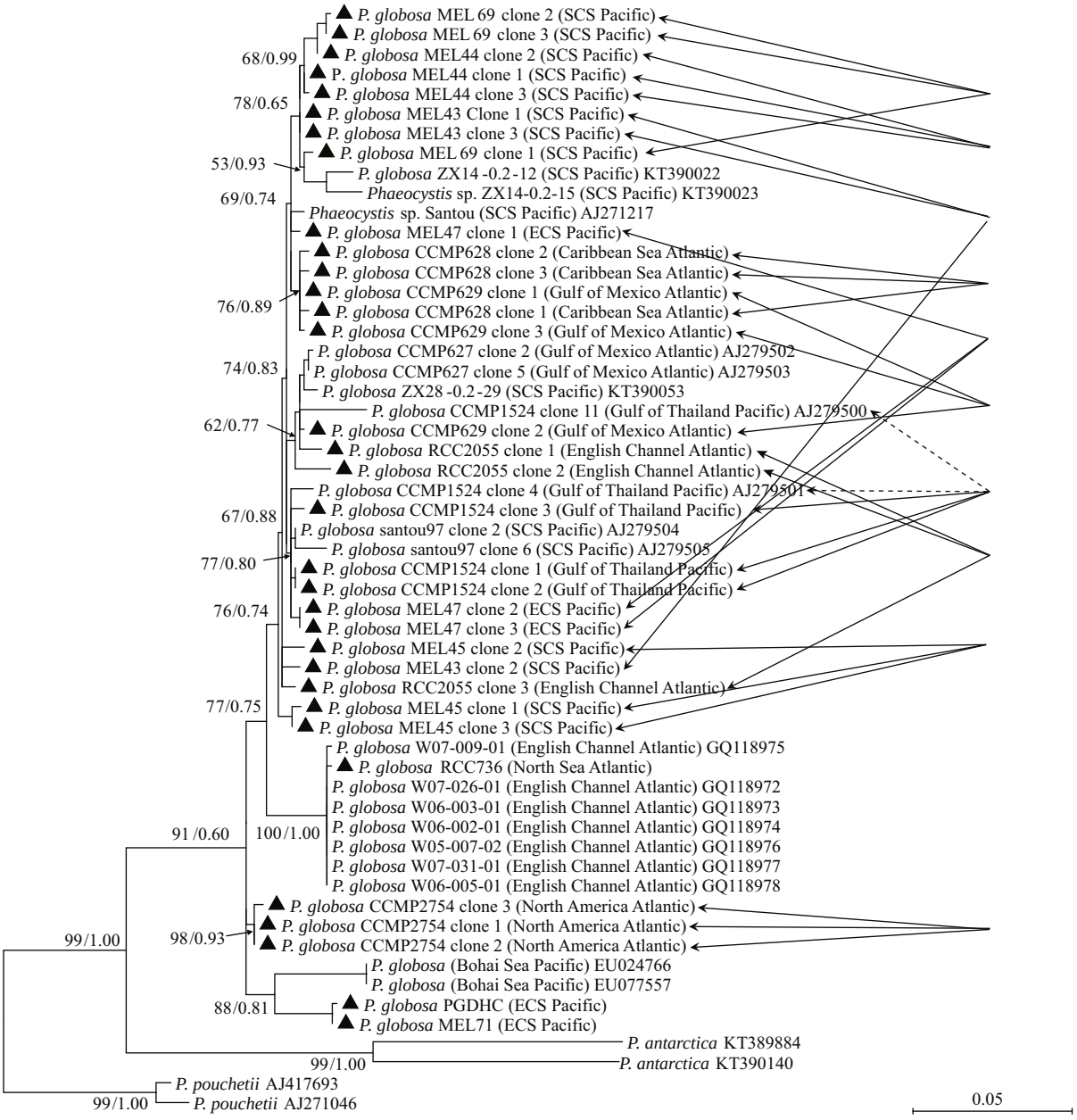



**Fig.5 Phylogenetic tree of ITS region of *Phaeocystis globosa***

Numbers indicate the bootstrap values from maximum-likelihood (left) and Bayesian (right) analyses. The sequences with a triangle were obtained by this study, and the haplotypes from one strain of *Phaeocystis* are connected by the arrows.

information indeed reflect the existence of the multiple populations within the *P. globosa* taxon, and their diversities among the different geographic populations.

### 4.1 Evaluation of the nuclear ribosomal and chloroplast DNA markers in phylogenetic analysis of geographic populations of *P. globosa*

At present, the phylogenetic relationship of *P. globosa* populations distributing in these diverse habitats has not been well understood. In general, the strains within a geographic population are easy to acquire similar evolutionary rate, and have small genetic distances under the same environmental pressure. However, there is no obvious correlation between phylogenetic relationship and geographic distribution among *P. globosa* strains based on the molecular data of nuclear ribosomal genes and intergenic spacers (Lange et al., 2002; Qu et al., 2008; Liu et al., 2010; Hu et al., 2019a). For example, it was found that a strain of *P. globosa* from the Bohai Sea is closely related to the strain isolated from the Atlantic Ocean, but far away from the strains from SCS base on

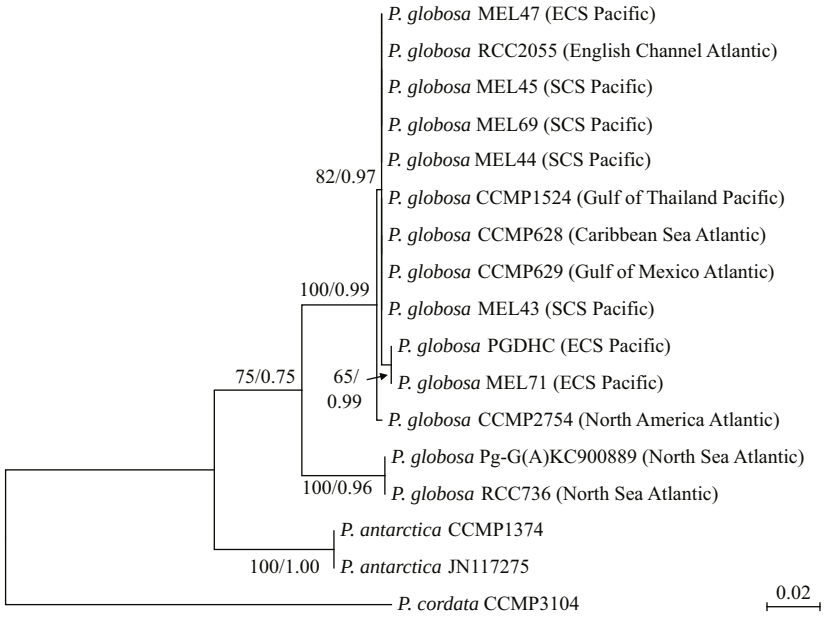



**Fig.6 Phylogenetic tree of *pet*N-*trn*S1 intergenic spacer of *P. globosa***

Numbers indicate the bootstrap values from maximum-likelihood (left) and Bayesian (right) analyses.

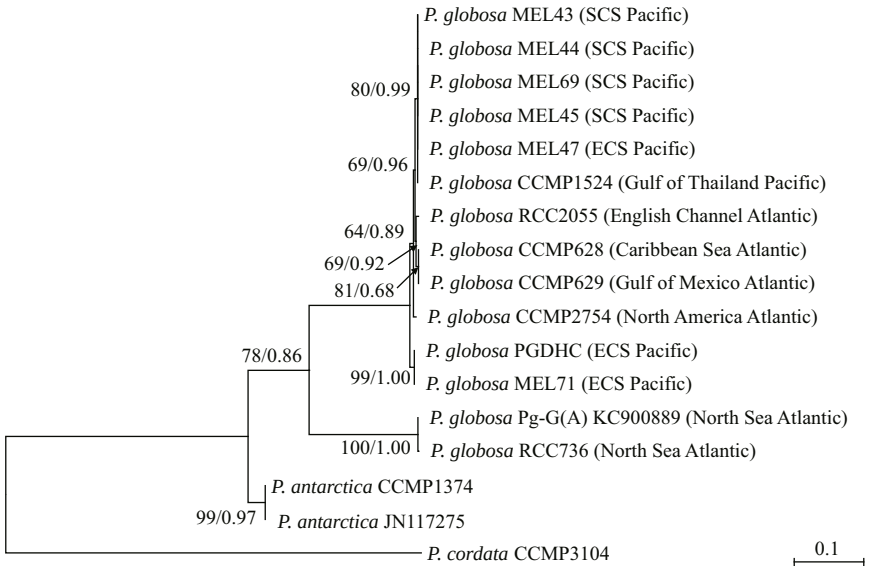



**Fig.7 Phylogenetic tree of *trn*M1-*psb*A intergenic spacer of *P. globosa***

Numbers indicate the bootstrap values from maximum-likelihood (left) and Bayesian (right) analyses.

the 18S rDNA sequences (Qu et al., 2008). Furthermore, our study found that there are nucleotide variations in 28S rDNA D1–D2 region and ITS region within a single strain cultured from either a colony or a single cell (Suppl. Figs.S1–S4), and similar results are also supported by the previous studies (Lange et al., 2002; Chen et al., 2003; Hu et al., 2019a). The reasons for the non-homogeneous copies of the nuclear ribosomal genes in *P. globosa* are somehow enigmatic.

A possible explanation is that *P. globosa* is indeed a multi-species complex as indicated by the variation in DNA content among several strains (Lange et al., 2002). The other explanation is that *P. globosa* has a complex polymorphic life cycle with both flagellated and colonial cells. A haploid-diploid life cycle with both ploidies alternately existing or/and co-existing has been evidenced for *P. globosa* (Whipple et al., 2005b; Rousseau et al., 2007; Peperzak and Gäbler-Schwarz, 2012). Two main prominent features of this cycle are that sexuality is prevalent in colony bloom

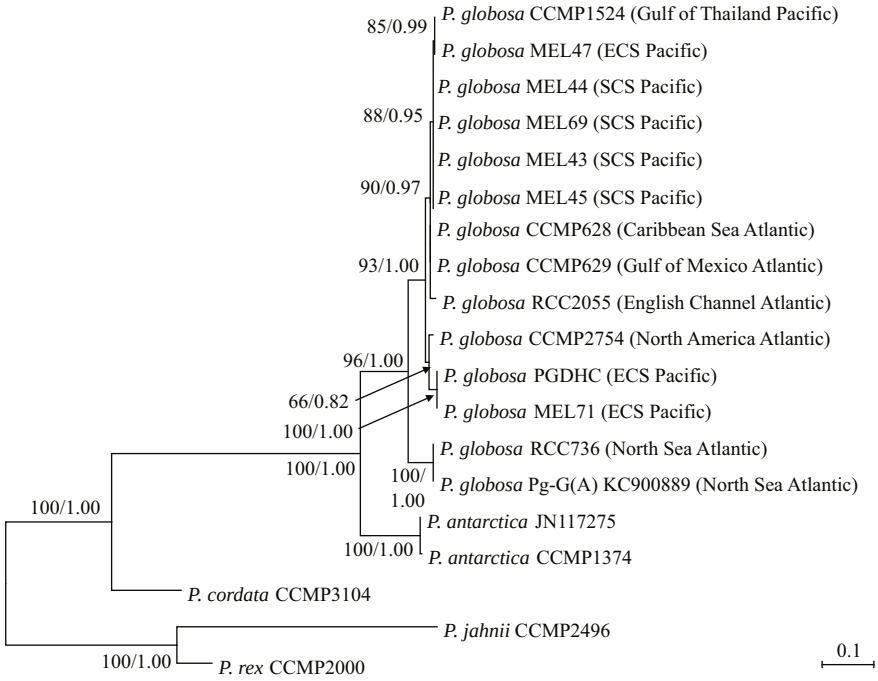



**Fig.8 Phylogenetic tree of *rbc*S-*rpl*27 intergenic spacer of *P. globosa***

Numbers indicate the bootstrap values from maximum-likelihood (left) and Bayesian (right) analyses.

formation and termination and two types of vegetative reproduction may exist at same time (Rousseau et al., 2007). Frequently sexual processes tend to cause genetic fusion between different populations. For example, a pervious study found that four haplotypes were contained in the 28S rDNA genes of two *P. globosa* strains isolated from the Beibu Gulf in SCS in 2015 and 2017, that covered a unique haplotype of another *P. globosa* strain isolated from the same region in 2016 (Hu et al., 2019a). Additionally, the habitats in the temperate and tropical sea areas are widely affected by human shipping activities, which accelerate gene flow between different populations contributing to form rDNA genetic polymorphism characteristics of *P. globosa* (Hu et al., 2019a). In any event, the non-homogeneous copies of the nuclear ribosomal genes are impossible to make interpretation of phylogenetic relationship within the *P. globosa* taxon.

Chloroplast is an important organelle involved in photosynthesis, supplying the indispensable energy for photosynthetic organism growth and development. The haploid chloroplast genome shows maternal inheritance in the genus *Phaeocystis*, which has a red algae-derived secondary plastid (Smith et al., 2014a). The chloroplast DNA has less recombination, and is substantially different from the nuclear genome (Wolfe et al., 1987). As displayed in this study, the phenomenon of non-homogeneous copies of genes is not present in the chloroplast intergenic spacers, which makes them a potential molecular marker for phylogenetic analysis among *P. globosa* geographic populations. Although it is confirmed that some conservative chloroplast genes and intergenic spacers are not suitable to act as intraspecific DNA markers of *P. globosa* (Lange et al., 2002; Yang et al., 2004; Yang et al., 2015), all of three chloroplast intergenic spacers selected in this study have high genetic resolution across different geographic strains. In sum, the chloroplast intergenic spacers are better molecular markers for intraspecific phylogeographic relationships on *P. globosa* than the nuclear ribosomal genes and intergenic spacers.

Among the three chloroplastal intergenic spacers, the *rbc*S-*rpl*27 intergenic spacer is a reasonable indicator of the genetic distance between Pacific and Atlantic strains, especially for the genetic difference between the Pacific strains, which can not be shown in the other two chloroplastal intergenic spacers. Moreover, the resolution ability of the *rbc*S-*rpl*27 intergenic spacer to *P. globosa* strains is similar with that of the concatenation of three chloroplast genetic spacers (Suppl. Fig.S5). The results imply that the genetic variation of *rbc*S-*rpl*27 intergenic spacer is the highest, and by far, it is most suitable to be served as the DNA marker for the intraspecific phylogeographic relationships of *P. globosa*.

### 4.2 Phylogeographic relationships of *P. globosa* in the Pacific and Atlantic coastal areas

In the genus *Phaeocystis*, the genetic relationships of three species *P. antarctica*, *P. globosa* and *P. pouchetii* are close, and they are speculated to be evolved from a common ancestor. When the Drake Passage opened and the Antarctic Circumpolar Current (ACC) formed, sufficiently allowed ancestral populations in the Antarctic to separate from their warm-water ancestors (Lange et al., 2002; Medlin and Zingone, 2007). While a short but deep glacial period happened around 23 MYA (Paul et al., 2000), water temperature was cool enough to allow the ancestral Antarctic populations cross the equator from the south to the north (Darling et al., 2000, 2004; Lange et al., 2002; Medlin and Zingone, 2007). Then, a major warming event in the world's oceans at approximately 15 MYA separated the two polar populations to allow them to diverge into two species (*P. antarctica* and *P. pouchetii*), and the relict populations in the temperate and tropical regions evolved into today's *P. globosa* (Lange et al., 2002; Medlin and Zingone, 2007).

In this study, the legible phylogeographic relationships of *P. globosa* from the Pacific and Atlantic coastal waters are shown in the tree of *rbc*S-*rpl*27 intergenic spacer. Two strains (RCC736 and Pg-G(A)) from the North Sea in the Atlantic Ocean form an independent clade, and detach from the *P. globosa* taxon firstly; however, other Atlantic strains stay with the Pacific strains. The population has unique pigment characteristics in the North Sea. Hex-fuco content is lower than But-fuco content in cell of RCC736 culture, which is different from other 12 strains of *P. globosa* in this study (our unpublished data). Moreover, the pervious study reported that the pigment composition and DNA content of the North Sea strains were similar, but different with the other geographic strains (Vaulot et al., 1994). These data support evidences for the uniqueness of the geographic population in the North Sea. As a result, it is deduced that there are at least two genetically distant populations of *P. globosa* in the Atlantic coastal regions, and these populations may have different geographic ancestries.

In the other clade, there is no nucleotide difference in the *rbc*S-*rpl*27 intergenic spacer among four SCS strains. Among them, MEL43, MEL44, and MEL69 were isolated from the Beibu Gulf in the western waters of SCS in the years of 2017 and 2018, respectively, and MEL45 was isolated from the Zhujiang (Pearl) River estuary in the eastern waters of SCS. The eastern and western waters of SCS are connected thought narrow Qiongzhou Strait. There is a strong year-round westward flow in the Qiongzhou Strait, which is a source of water transport to the Beibu Gulf existing cyclonic circulation (Shi et al., 2002). This kind of hydrological characteristics is well supported that four SCS strains are belong to a same geographic population. The other two Pacific strains (MEL47 and CCMP1524) have closely phylogenetic relationship with the SCS population. They were respectively isolated from the Quanzhou coastal area in ECS and the Gulf of Thailand, adjacent to SCS, which help to explain reasonably the close genetic relationship with the SCS population. It seems unreasonable that the two ECS strains (MEL71 and PGDHC) keep far away the other Pacific strains, and gather with the CCMP2754 strain from the North American coastal area. It is pluckily suspected these ECS strains may be invasive species from the Northeastern Atlantic region.

## 5 CONCLUSION

In this study, it was confirmed that the region of chloroplast intergenic spacer was a good target to be used for the intraspecific phylogeographic study of *P. globosa*, and a clear phylogenetic relationship of *P. globosa* strains was firstly revealed by the high-resolution molecular marker, the chloroplast *rbc*S-*rpl*27 intergenic spacer. The results displayed that strains from the North Sea Atlantic separated from the species *P. globosa* firstly, and kept far away from the other Atlantic strains and all of Pacific strains, which implied that there were at least two genetically distant populations of *P. globosa* in the Atlantic coastal regions.

## 6 DATA AVAILABILITY STATEMENT

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

## 7 ACKNOWLEDGMENT

We deeply appreciate valuable comments of Dr. QIU Limei from the Institute of Oceanology, Chinese Academy of Sciences (IOCAS), and free provision of two strains of *Phaeocystis globosa* RCC736 and RCC2055 by Prof. Daniel Vaulot and Ian Probert from the Roscoff Biological Station.

**References**


Antajan E, Chrétiennot-Dinet M J, Leblanc C, Daro M H, Lancelot C. 2004. 19'-hexanoyloxyfucoxanthin may not be the appropriate pigment to trace occurrence and fate of *Phaeocystis*: the case of *P. globosa* in Belgian coastal waters. *Journal of Sea Research*, **52**(3): 165-177.

Bakker F T, Culham A, Daugherty L C, Gibby M. 1999. A trnL-F based phylogeny for species of *Pelargonium* (*Geraniaceae*) with small chromosomes. *Plant Systematics and Evolutio*n, **216**(3-4): 309-324.

Blomster J, Maggs C A, Stanhope M J. 1998. Molecular and morphological analysis of *Enteromorpha intestinalis* and *E. compressa* (Chlorophyta) in the British Isles. *Journal of Phycology*, **34**(2): 319-340.

Cadée G C. 1996. Accumulation and sedimentation of *Phaeocystis globosa* in the Dutch Wadden Sea. *Journal of Sea Research*, **36**(3-4): 321-327.

Carbonell-Caballero J, Alonso R, Ibañez V, Terol J, Talon M, Dopazo J. 2015. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. *Molecular Biology and Evolution*, **32**(8): 2 015-2 035.

Charlson R J, Lovelock J E, Andreae M O, Warren S G. 1987. Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature*, **326**(6114): 655-661.

Chen Y Q, Shao P, Wang N, Zhou H, Qu L H, Medlin L K. 2003. Molecular identification of bloom-forming species *Phaeocystis globosa* (Prymnesiophyta) and its dispersal based on rDNA ITS sequence analysis. *Acta Oceanologica Sinica*, **22**(2): 243-253.

Chen Y Q, Wang N, Zhang P, Zhou H, Qu L H. 2002. Molecular evidence identifies bloom-forming *Phaeocystis* (Prymnesiophyta) from coastal waters of southeast China as *Phaeocystis globosa*. *Biochemical Systematics and Ecology*, **30**(1): 15-22.

Crooks G E, Hon G, Chandonia J M, Brenner S E. 2004. WebLogo: a sequence logo generator. *Genome Research*, **14**: 1 188-1 190.

Darling K F, Kucera M, Pudsey C J, Wade C M. 2004. Molecular evidence links cryptic diversification in polar planktonic protists to quaternary climate dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(20): 7 657-7 662.

Darling K F, Wade C M, Stewart I A, Kroon D, Dingle R, Brown A J L. 2000. Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. *Natur*e, **405**(6782): 43-47.

Davidson A T, Marchant H I. 1992. The biology and ecology of *Phaeocystis* (Prymnesiophyceae). *In*: Round F E eds. Progress in Phycological Research. Biopress, Bristol. p.1-45.

Decelle J, Probert I, Bittner L, Desdevises Y, Colin S, de Vargas C, Gali M, Simó R, Not F. 2012. An original mode of symbiosis in open ocean plankton. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(44): 18 000-18 005.

Hai D N, Lam N N, Dippner J W. 2010. Development of *Phaeocystis globosa* blooms in the upwelling waters of the south central coast of Viet Nam. *Journal of Marine Systems*, **83**(3-4): 253-261.

Hu X K, Zhang Q C, Chen Z F, Kong F Z, Wang J X, Yu R C. 2019a. Genetic diversity of *Phaeocystis globosa* strains isolated from the Beihu Gulf, the South China Sea. *Oceanologia et Limnologia Sinica*, **50**(3): 601-609. (in Chinese with English abstract)

Hu Z M, Uwai S, Yu S H, Komatsu T, Ajisaka T, Duan D L. 2011. Phylogeographic heterogeneity of the brown macroalga *Sargassum horneri* (Fucaceae) in the northwestern Pacific in relation to late Pleistocene glaciation and tectonic configurations. *Molecular Ecology*, **20**(18): 3 894-3 909.

Hu Z X, Deng Y Y, Tang Y Z. 2019b. Scanning and transmission electron microscopy observation on morphology and ultrastructure of *Phaeocystis globosa* from Beibu Gulf, China. *Oceanologia et Limnologia Sinica*, **50**(3): 621-629. (in Chinese with English abstract)

Jacobsen A. 2002. Morphology, relative DNA content and hypothetical life cycle of *Phaeocystis pouchetii* (Prymnesiophyceae); with special emphasis on the flagellated cell type. *Sarsia*, **87**(5): 338-349.

Kalyaanamoorthy S, Minh B Q, Wong T K F, von Haeseler A, Jermiin L S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**(6): 587-589.

Katayama H, Tachibana M, Iketani H, Zhang S L, Uematsu C. 2012. Phylogenetic utility of structural alterations found in the chloroplast genome of pear: hypervariable regions in a highly conserved genome. *Tree Genetics & Genomes*, **8**(2): 313-326.

Kholina A B, Kozyrenko M M, Artyukova E V, Sandanov D V, Andrianova E A. 2016. Phylogenetic relationships of the species of *Oxytropis* DC. subg. *Oxytropis* and *Phacoxytropis* (Fabaceae) from Asian Russia inferred from the nucleotide sequence analysis of the intergenic spacers of the chloroplast genome. *Russian Journal of Genetics*, **52**(8): 780-793.

Kimura T, Iketani H, Kotobuki K, Matsuta N, Ban Y, Hayashi T, Yamamoto T. 2003. Genetic characterization of pear varieties revealed by chloroplast DNA sequences. *Journal of Horticultural Science and Biotechnology*, **78**(2): 241-247.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, **35**(6): 1 547-1 549.

Lancelot C, Billen G, Sournia A, Weisse T, Colijn F, Veldhuis M J W, Davies A, Wassman P. 1987. *Phaeocystis* blooms and nutrient enrichment in the continental coastal zones of the North Sea. *Ambio*, **16**(1): 38-46.

Lange M, Chen Y Q, Medlin L K. 2002. Molecular genetic delineation of *Phaeocystis* species (Prymnesiophyceae) using coding and non-coding regions of nuclear and plastid genomes. *European Journal of Phycology*, **37**(1):

77-92.

Li P R, Zhang S J, Li F, Zhang S F, Zhang H, Wang X W, Sun R F, Bonnema G, Borm T J A. 2017. A phylogenetic analysis of chloroplast genomes elucidates the relationships of the six economically important *Brassica* species comprising the triangle of U. F*rontiers in Plant Science*, **8**: 111, https://doi.org/10.3389/fpls.2017.00111.

Liss P S, Malin G, Turner S M, Holligan P M. 1994. Dimethyl sulphide and *Phaeocystis*: a review. *Journal of Marine Systems*, **5**(1): 41-53.

Liu H L, Zhang Q, Jiang Q M, Ma B. 2010. Analyses of sequence variation and secondary structure on rDNA ITS regions of the strains of redtide-caused harmful algae *Phaeocystis globosa*. *Ecological Science*, **29**(5): 432-437. (in Chinese with English abstract)

Liu J S, Peng X C, Yang W D. 2007. AFM observation of hemolytic extracts from *Phaeocystis globosa* on rabbit erythrocytes. *Journal of Tropical Oceanography*, 2**6**(2): 55-58. (in Chinese with English abstract)

Medlin L, Zingone A. 2007. A taxonomic review of the genus *Phaeocystis*. *Biogeochemistry*, **83**(1-3): 3-18.

Paul H A, Zachos J C, Flower B P, Tripati A. 2000. Orbitally induced climate and geochemical variability across the Oligocene/Miocene boundary. *Paleoceanography*, **15**(5): 471-485.

Peperzak L, Colijn F, Vrieling E G, Gieskes W W C, Peeters J C H. 2000. Observations of flagellates in colonies of *Phaeocystis globosa* (Prymnesiophyceae); a hypothesis for their position in the life cycle. *Journal of Plankton Research*, **22**(12): 2 181-2 203.

Peperzak L, Gäbler-Schwarz S. 2012. Current knowledge of the life cycles of *Phaeocystis globosa* and *Phaeocystis antarctica* (Prymnesiophyceae). *Journal of Phycology*, **48**(3): 514-517.

Qi Y Z, Chen J F, Wang Z H, Xu N, Wang Y, Shen P P, Lu S H, Hodgkiss I J. 2004. Some observations on harmful algal bloom (HAB) events along the coast of Guangdong, southern China in 1998. *Hydrobiologia*, **512**(1-3): 209-214.

Qu L Y, Lv S H, Gao C L, Li Y, Sun P, Sun X Q. 2008. Structure and sequence analysis of 18s rDNA and ITS gene of *Phaeocystis* isolate from the Bohai Sea. *Advances in Marine Science*, **26**(2): 200-206. (in Chinese with English abstract)

Raman G, Choi K S, Park S. 2016. Phylogenetic relationships of the fern *Cyrtomium falcatum* (Dryopteridaceae) from Dokdo Island based on chloroplast genome sequencing. *Genes*, **7**(12): 115.

Ronquist F, Teslenko M, van der Mark P, Ayres D L, Darling A, Höhna S, Larget B, Liu L, Suchard M A, Huelsenbeck J P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *System Biology*, **61**(3): 539-542.

Rousseau V, Chrétiennot-Dinet M J, Jacobsen A, Verity P, Whipple S. 2007. The life cycle of *Phaeocystis*: state of knowledge and presumptive role in ecology. *Biogeochemistry*, **83**(1-3): 29-47.

Rousseau V, Lantoine F, Rodriguez F, LeGall F, Chrétiennot-Dinet M J, Lancelot C. 2013. Characterization of *Phaeocystis globosa* (Prymnesiophyceae), the blooming species in the Southern North Sea. *Journal of Sea Research*, **76**: 105-113.

Rousseau V, Vaulot D, Casotti R, Cariou V, Lenz J, Gunkel J, Baumann M. 1994. The life cycle of *Phaeocystis* (Prymnesiophycaea): evidence and hypotheses. *Journal of Marine Systems*, **5**(1): 23-39.

Schoemann V, Becquevort S, Stefels J, Rousseau V, Lancelot C. 2005. *Phaeocystis* blooms in the global ocean and their controlling mechanisms: a review. *Journal of Sea Research*, **53**(1-2): 43-66.

Seoane S, Zapata M, Orive E. 2009. Growth rates and pigment patterns of haptophytes isolated from estuarine waters. *Journal of Sea Research*, **62**(4): 286-294.

Shen P, Van Rijssel M, Wang Y, Lu S H, Chen J F, Qi Y Z. 2004. Toxic *Phaeocystis globosa* strains from China grow at remarkably high temperatures. *In*: Steidinger K A, Landsberg J H, Tomas C R, Vargo G A eds. Harmful Algae 2002. Florida Fish and Wildlife Conservation Commission, Florida Institute of Oceanography and Intergovernmental Oceanographic Commission of UNESCO, St. Petersburg. p.396-398.

Shi M C, Chen C S, Xu Q C, Lin H C, Liu G M, Wang H, Wang F, Yan J H. 2002. The role of Qiongzhou Strait in the seasonal variation of the South China Sea circulation. *Journal of Physical Oceanography*, **32**(1): 103-121.

Smith D R, Arrigo K R, Alderkamp A C, Allen A E. 2014a. Massive difference in synonymous substitution rates among mitochondrial, plastid, and nuclear genes of *Phaeocystis* algae. *Molecular Phylogenetics and Evolution*, **71**: 36-40.

Smith W O Jr, Liu X, Tang K W, DeLizo L M, Doan N H, Nguyen N L, Wang X D. 2014b. Giantism and its role in the harmful algal bloom species *Phaeocystis globosa*. *Deep Sea Research Part II*: *Topical Studies in Oceanography*, **101**: 95-106.

Stefels J, Dijkhuizen L, Gieskes W W C. 1995. DMSP-lyase activity in a spring phytoplankton bloom off the Dutch coast, related to *Phaeocystis* sp. abundance. *Marine Ecology Progress Series*, **123**: 235-243.

Stefels J, Van Boekel W H M. 1993. Production of DMS from dissolved DMSP in axenic cultures of the marine phytoplankton species *Phaeocystis* sp. *Marine Ecology Progress Series*, **97**: 11-18.

Vaulot D, Birrien J L, Marie D, Casotti R, Veldhuis M J W, Kraay G W, Chrétiennot-Dinet M J. 1994. Morphology, ploidy, pigment composition, and genome size of cultured strains of *Phaeocystis* (Prymnesiophyceae). *Journal of Phycology*, **30**(6): 1 022-1 035.

Vitales D, Aragay J, Garnatje T, Garreta A G, Lluch J R. 2019. Phylogeography of *Dictyota fasciola* and *Dictyota mediterranea* (Dictyotales, Phaeophyceae): unexpected patterns on the Atlantic-Mediterranean marine transition and taxonomic implications. *PeerJ*, **7**(6): e6916, https://doi.org/10.7717/peerj.6916.

Wang J X, Kong F Z, Chen Z F, Zhang Q C, Yu R C, Zhou M J. 2019. Characterization of pigment composition of six strains of *Phaeocystis globosa*. *Oceanologia et Limnologia Sinica*, **50**(3): 611-620. (in Chinese with English abstract)

Wassmann P, Ratkova T, Reigstad M. 2005. The contribution of single and colonial cells of *Phaeocystis pouchetii* to spring and summer blooms in the north-eastern North Atlantic. *Harmful Algae*, **4**(5): 823-840.

Whipple S J, Patten B C, Verity P G. 2005a. Colony growth and evidence for colony multiplication in *Phaeocystis pouchetii* (Prymnesiophyceae) isolated from mesocosm blooms. *Journal of Plankton Research*, **27**(5): 495-501.

Whipple S J, Patten B C, Verity P G. 2005b. Life cycle of the marine alga *Phaeocystis*: a conceptual model to summarize literature and guide research. *Journal of Marine Systems*, **57**(1-2): 83-110.

Winnepenninckx B, Backeljau T, De Wachter R. 1993. Extraction of high molecular weight DNA from mollusks. *Trends in Genetics*, **9**(12): 407.

Wolfe K H, Li W H, Sharp P M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America*, **84**(24): 9 054-9 058.

Yan H D, Zhang X Q, Fu C, Huang L K, Yin G H, Nie G, Xu W Z, Liu X C, Chen T, Lee S, Ma X, Peng Y, Yan Y H, Ling Y, Liu W. 2015. Chloroplast DNA variation and genetic structure of *Miscanthus sinensis* in southwest China. *Biochemical Systematics and Ecology*, **58**: 132-138.

Yang Q C, Chen L N, Hu X L, Zhao L, Yin P H, Li Q. 2015. Toxic effect of a marine bacterium on aquatic organisms and its algicidal substances against *Phaeocystis globosa*. *PLoS One*, **10**(2): e0114933, https://doi.org/10.1371/journal.pone.0114933.

Yang Z M, Zhang Q, Xie S T, Han B P, Lv S H, Hodgkiss I J. 2004. Sequence analyses of chloroplastic *psa*A gene fragment from *Phaeocystis globosa*. *Journal of Tropical and Subtropical Botany*, **12**(5): 435-439. (in Chinese with English abstract)

Zapata M, Jeffrey S W, Wright S W, Rodríguez F, Garrido J L, Clementson L. 2004. Photosynthetic pigments in 37 species (65 strains) of Haptophyta: implications for oceanography and chemotaxonomy. *Marine Ecology Progress Series*, **270**: 83-102.

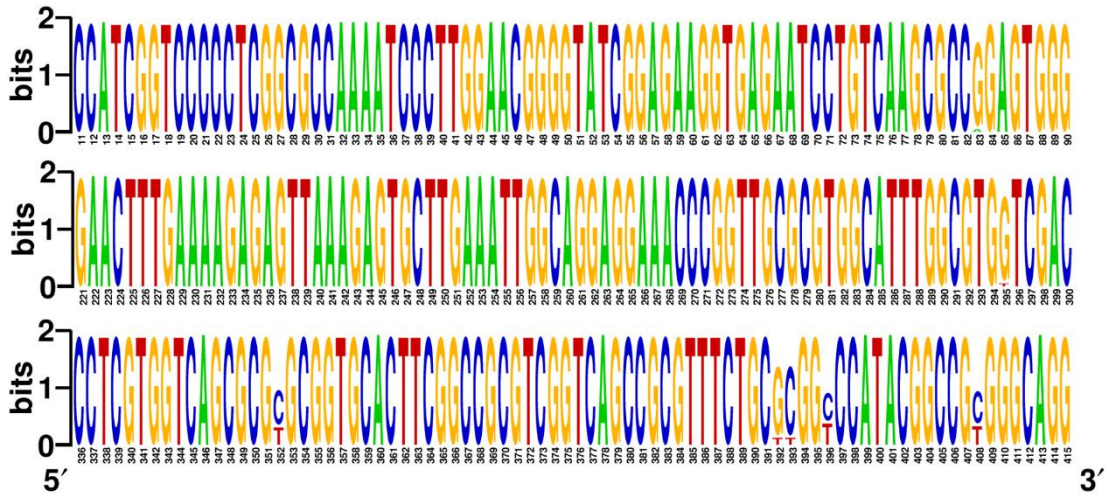
Zhang J, Huang X, Huang D J, Zhang Y, Huang L K, Lu L, Yan H D. 2016. Studies on genetic diversity and phylogenetic relationships of limpograss (*Hemarthria altissima*) and related species based on combined chloroplast DNA intergenic spacer data. *Biochemical Systematics and Ecology*, **69**: 91-100.

Zhou T, Ruhsam M, Wang J, Zhu H H, Li W L, Zhang X, Xu Y C, Xu F S, Wang X M. 2019. The complete chloroplast genome of *Euphrasia regelii*, pseudogenization of *ndh* genes and the phylogenetic relationships within Orobanchaceae. *Frontiers in Genetics*, **10**: 444, https://doi.org/10.3389/fgene.2019.00444.

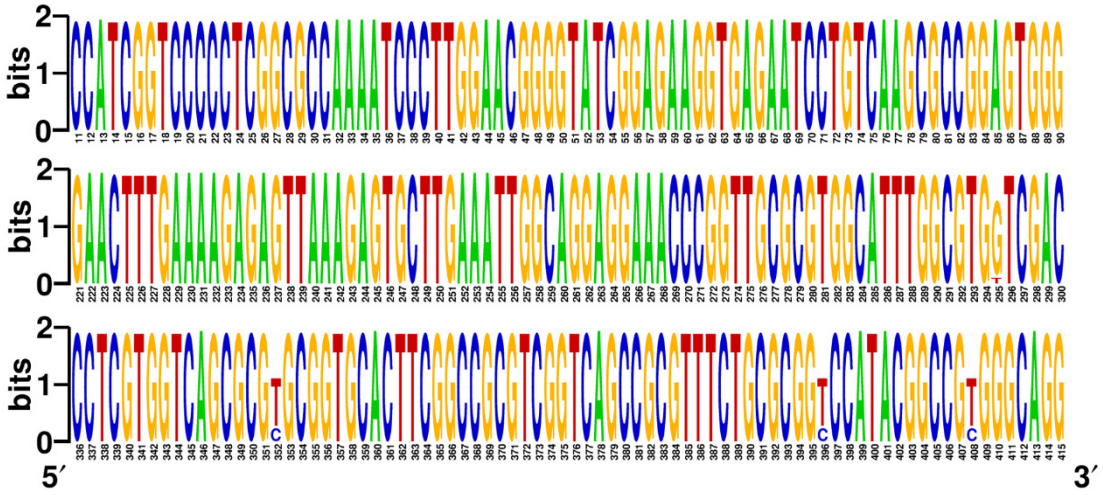
Zingone A, Chrétiennot-Dinet M J, Lange M, Medlin L. 1999. Morphological and genetic characterization of *Phaeocystis cordat*a and *P. jahnii* (Prymnesiophyceae), two new species from the Mediterranean Sea. *Journal of Phycology*, **35**(6): 1 322-1 337.

## Electronic supplementary material

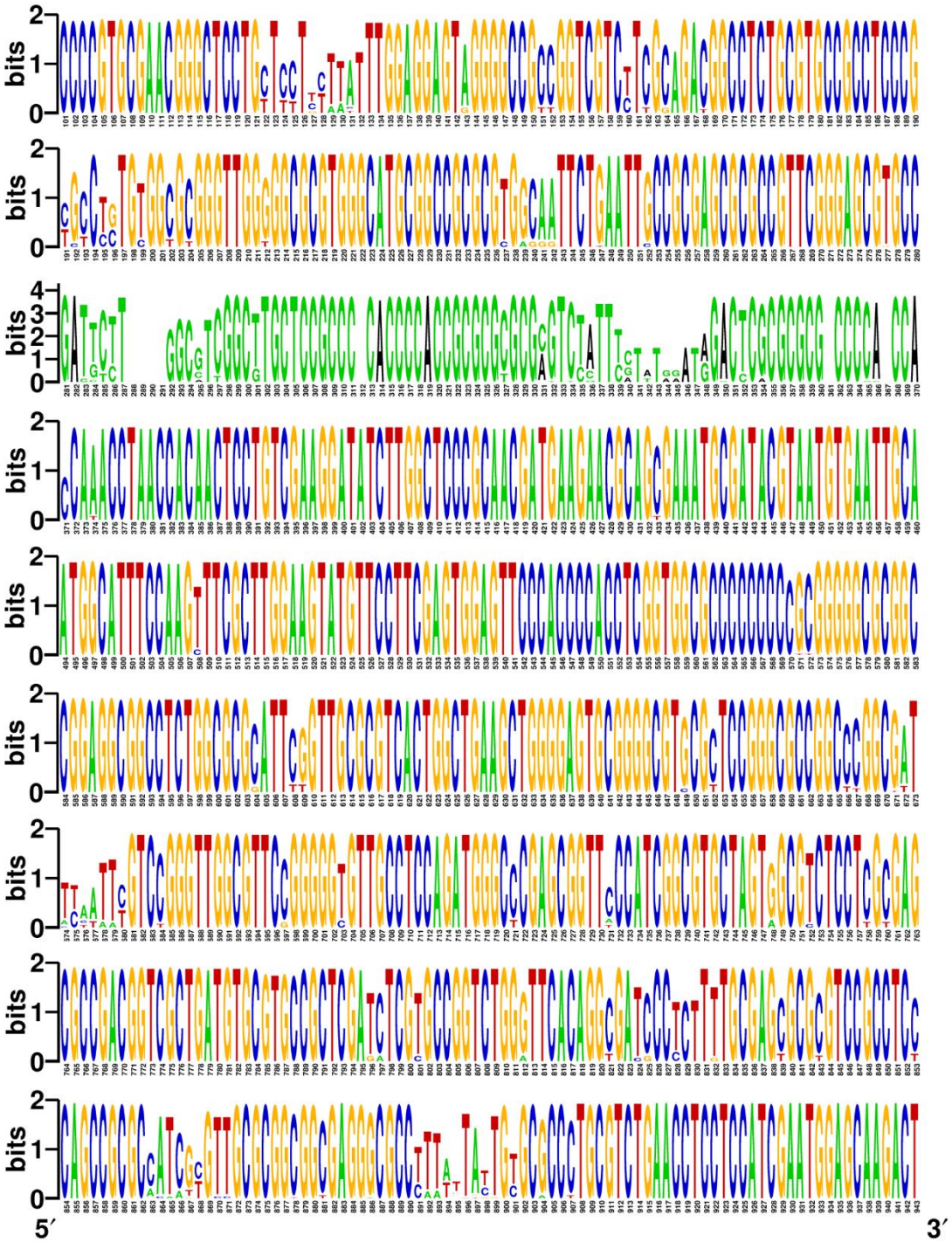
Supplementary material (Supplementary Tables S1–S6, Figs.S1–S5) is available in the online version of this article at https://doi.org/10.1007/s00343-020-9304-5.



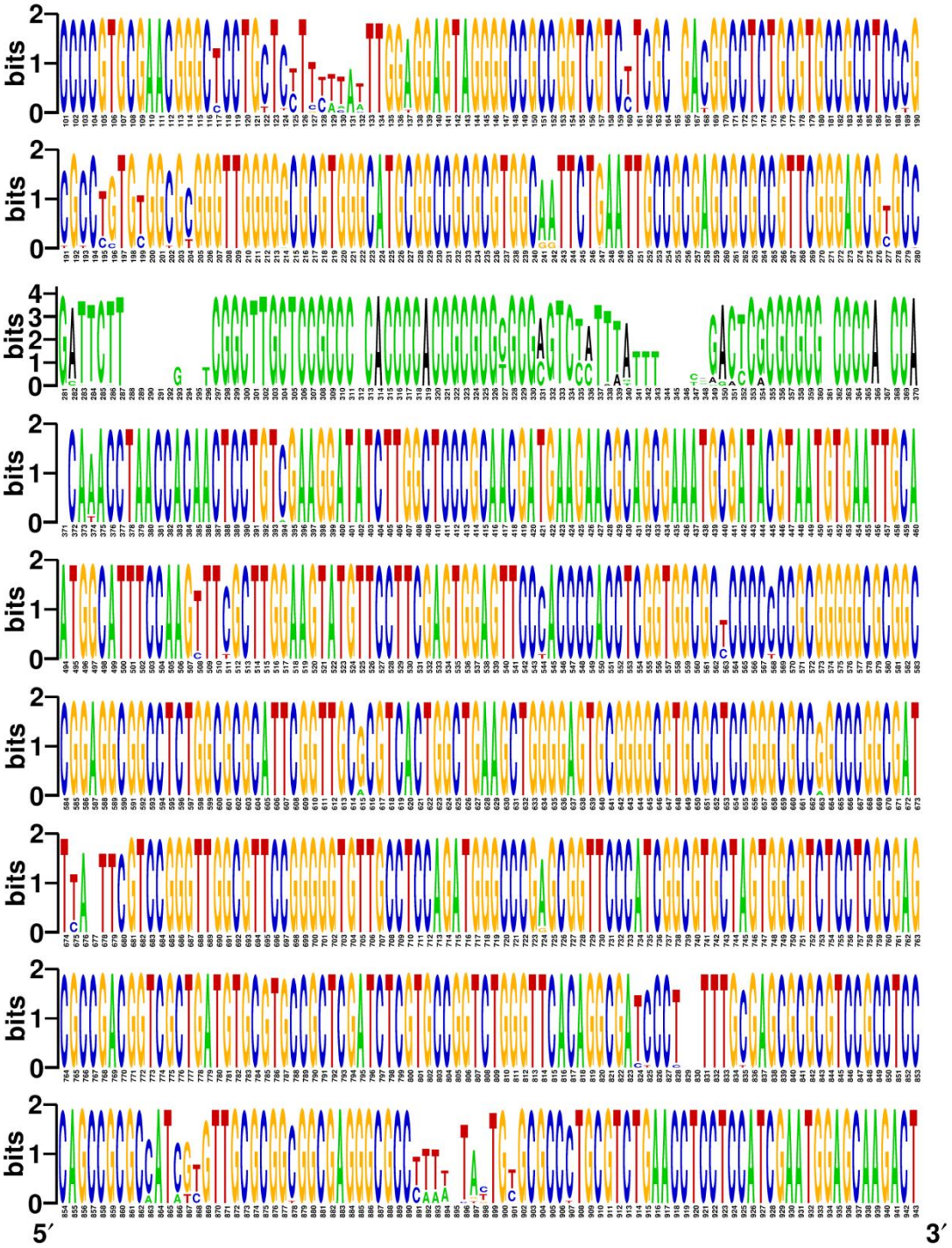

Suppl. Fig.S1 WebLogo of 28S rDNA D1-D2 region of *Phaeocystis globosa* strain MEL43



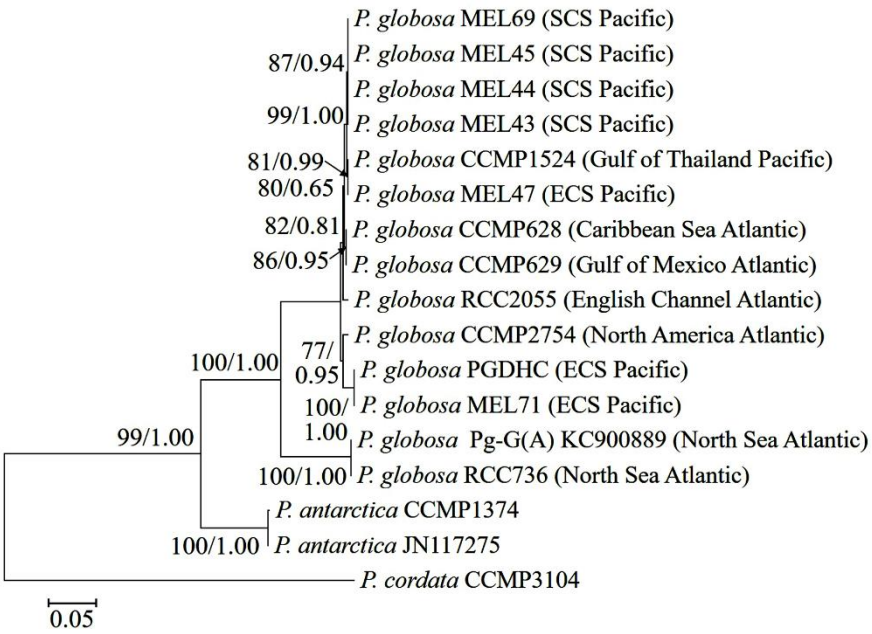

Suppl. Fig.S2 WebLogo of 28S rDNA D1-D2 region of *Phaeocystis globosa* strain MEL47




Suppl. Fig.S3 WebLogo of ITS region of *Phaeocystis globosa* strain MEL43




Suppl. Fig.S4 WebLogo of ITS region of *Phaeocystis globosa* strain MEL47




Suppl. Fig.S5 Phylogenetic tree of concatenation of *pet*N-*trn*S1, *trn*M1-*psb*A and *rbc*S-*rpl*27 intergenic spacers of *Phaeocystis globosa*. Numbers indicated the bootstrap values from maximum-likelihood (left) and bayesian (right) analyses.

Suppl. Table S1 Information of the sequences of 28S rDNA region, ITS region, *pet*N-*trn*S1, *trn*M1-*psb*A and *rbc*S-*rpl*27 of *Phaeocystis* downloaded from Genbank using for phylogenetic analysis

| Species | Strain/clone | Sampling area | Gene | Sequence No. |
|---|---|---|---|---|
| *P. globosa* | RCC575 | Atlantic | 28S rDNA | JX660933 |
| *P. globosa* | RCC576 | Atlantic | 28S rDNA | JX660932 |
| *P. globosa* | ZX14-0.2-12 | the South China SCS Pacific | 28S rDNA, ITS | KT390022 |
| *P. globosa* | ZX14-0.2-15 | the South China SCS Pacific | 28S rDNA, ITS | KT390023 |
| *P. globosa* | RCC540 | North Atlantic | 28S rDNA | JX660930 |
| *P. globosa* | RCC1398 | China Sea Pacific | 28S rDNA | JX660936 |
| *Phaeocystis* sp. | RCC1000 | Pacific | 28S rDNA | JX660940 |
| *P. globosa* | ZX28-0.2-29 | SCS Pacific | 28S rDNA, ITS | KT390053 |
| *Phaeocystis* sp. | Ind12 | Acantharia symbiont | 28S rDNA | JX660785 |
| *Phaeocystis* sp. | AC618 | - | 28S rDNA | EU502882 |
| *Phaeocystis* sp. | Ros1 | Acantharia symbiont | 28S rDNA | JX660804 |
| *Phaeocystis sp.* | JD-2012 Ros2 | Acantharia symbiont | 28S rDNA | JX660805 |
| *Phaeocystis sp.* | PCC679 | | 28S rDNA | JX660929 |
| *P. globosa* | RCC64 | English Channel Atlantic | 28S rDNA | JX660934 |
| *P. globosa* | K 1321 | Portugal Atlantic | 28S rDNA | JX660935 |
| *P. globosa* | NIES388 | Pacific | 28S rDNA | JX660937 |
| *P. globosa* | RCC678 | North Sea Atlantic | 28S rDNA | JX660939 |
| *P. antarctica* | CCMP1871 | Southern Ocean, Antarctic | 28S rDNA | KP144258 |
| *P. antarctica* | - | - | 28S rDNA | AF289040 |
| *P. cordata* | RCC1383 | Mediterranean Sea | 28S rDNA | JX660941 |
| *P. cordata* | CCMP2495 | - | 28S rDNA | EU729459 |
| *Phaeocystis sp.* | Santou | SCS Pacific | ITS | AJ271217 |
| *P. globosa* | CCMP627 clone 2 | Gulf of Mexico Atlantic | ITS | AJ279502 |
| *P. globosa* | CCMP627 clone 5 | Gulf of Mexico Atlantic | ITS | AJ279503 |
| *P. globosa* | CCMP1524 clone 11 | Thailand Sea Pacific | ITS | AJ279500 |
| *P. globosa* | CCMP1524 clone 4 | Thailand Sea Pacific | ITS | AJ279501 |
| *P. globosa* | santou97 clone 2 | SCS Pacific | ITS | AJ279504 |
| *P. globosa* | santou97 clone 6 | SCS Pacific | ITS | AJ279505 |
| *P. globosa* | W07-009-01 | English Channel Atlantic | ITS | GQ118975 |
| *P. globosa* | W07-026-0 | English Channel Atlantic | ITS | GQ118972 |
| *P. globosa* | W06-003-01 | English Channel Atlantic | ITS | GQ118973 |
| *P. globosa* | W06-002-01 | English Channel Atlantic | ITS | GQ118974 |
| *P. globosa* | W05-007-02 | English Channel Atlantic | ITS | GQ118976 |
| *P. globosa* | W07-031-01 | English Channel Atlantic | ITS | GQ118977 |
| *P. globosa* | W06-005-01 | English Channel Atlantic | ITS | GQ118978 |
| *P. globosa* | - | Bohai Sea Pacific | ITS | EU024766 |

| | | | | |
|---|---|---|---|---|
| *P. globosa* | - | Bohai Sea Pacific | ITS | EU077557 |
| *P. pouchetii* | - | - | ITS | AJ417693 |
| *P. pouchetii* | - | - | ITS | AJ271046 |
| *P. antarctica* | KJ22-0.2-54 | - | ITS | KT389884 |
| *P. antarctica* | KJ47-3-33 | - | ITS | KT390140 |
| *P. globosa* | Pg-G(A) | North Sea Atlantic | Chloroplast genome | KC900889 |
| *P. antarctica* | CCMP1374 | - | Chloroplast genome | JN117275 |

"-": the information unknown.

Suppl. Table S2 GenBank ID of the sequences of 28S rDNA D1-D2 region, ITS region, *pet*N-*trn*S1, *trn*M1-*psb*A and *rbc*S-*rpl*27 of *Phaeocystis* obtained in this study

| Species | Strain | GenBank ID | | | | |
|---|---|---|---|---|---|---|
| | | 28S rDNA | ITS | *pet*N-*trn*S1 | *trn*M1-*psb*A | *rbc*S-*rpl*27 |
| *P. globosa* | MEL43 | MN602538 | - | MN807922 | MN928971 | MN935489 |
| *P. globosa* | MEL44 | MN602539 | - | MN807923 | MN928972 | MN935490 |
| *P. globosa* | MEL45 | MN602540 | - | MN807924 | MN928973 | MN935491 |
| *P. globosa* | MEL47 | MN602541 | - | MN807925 | MN928974 | MN935492 |
| *P. globosa* | MEL69 | MN602542 | - | MN807926 | MN928975 | MN935493 |
| *P. globosa* | MEL71 | MN602543 | MN603007 | MN807927 | MN928976 | MN935494 |
| *P. globosa* | PGDHC | MN602544 | MN603008 | MN807928 | MN928977 | MN935495 |
| *P. globosa* | CCMP628 | MN602545 | - | MN807929 | MN928978 | MN935496 |
| *P. globosa* | CCMP629 | MN602546 | - | MN807930 | MN928979 | MN935497 |
| *P. globosa* | CCMP1524 | MN602547 | - | MN807931 | MN928980 | MN935498 |
| *P. globosa* | CCMP2754 | MN602548 | - | MN807932 | MN928981 | MN935499 |
| *P. globosa* | RCC736 | MN602549 | MN603009 | MN807933 | MN928982 | MN935501 |
| *P. globosa* | RCC2055 | MN602550 | - | MN807934 | MN928983 | MN935500 |
| *P. antarctica* | CCMP1374 | / | - | MN807935 | MN928984 | MN956829 |

"-": The sequences could not be amplified and sequenced directly in this study;

"/": The sequence was not amplified in this study.

Suppl. Table 3 GenBank ID of 28S rDNA D1-D2 region clone libraries sampled from strain MEL43 and MEL47

| Strain | Clone | GenBank ID | Strain | Clone | GenBank ID |
|---|---|---|---|---|---|
| MEL43 | clone1 | MN602900 | MEL47 | clone4 | MN610015 |
| MEL43 | clone2 | MN602901 | MEL47 | clone5 | MN610016 |
| MEL43 | clone3 | MN602902 | MEL47 | clone6 | MN610017 |
| MEL43 | clone4 | MN602903 | MEL47 | clone7 | MN610018 |
| MEL43 | clone5 | MN602904 | MEL47 | clone8 | MN610019 |
| MEL43 | clone6 | MN602905 | MEL47 | clone9 | MN610020 |
| MEL43 | clone7 | MN602906 | MEL47 | clone10 | MN610021 |
| MEL43 | clone8 | MN602907 | MEL47 | clone11 | MN610022 |
| MEL43 | clone9 | MN602908 | MEL47 | clone12 | MN610023 |
| MEL43 | clone10 | MN602909 | MEL47 | clone13 | MN610024 |
| MEL43 | clone11 | MN602910 | MEL47 | clone14 | MN610025 |
| MEL43 | clone12 | MN602911 | MEL47 | clone15 | MN610026 |
| MEL43 | clone13 | MN602912 | MEL47 | clone16 | MN610027 |
| MEL43 | clone14 | MN602913 | MEL47 | clone17 | MN610028 |
| MEL43 | clone15 | MN602914 | MEL47 | clone18 | MN610029 |
| MEL43 | clone16 | MN602915 | MEL47 | clone19 | MN610030 |
| MEL43 | clone17 | MN602916 | MEL47 | clone20 | MN610031 |
| MEL43 | clone18 | MN602917 | MEL47 | clone21 | MN610032 |
| MEL43 | clone19 | MN602918 | MEL47 | clone22 | MN610033 |
| MEL43 | clone20 | MN602919 | MEL47 | clone23 | MN610034 |
| MEL43 | clone21 | MN602920 | MEL47 | clone24 | MN610035 |
| MEL43 | clone22 | MN602921 | MEL47 | clone25 | MN610036 |
| MEL43 | clone23 | MN602922 | MEL47 | clone26 | MN610037 |
| MEL43 | clone24 | MN602923 | MEL47 | clone27 | MN610038 |
| MEL43 | clone25 | MN602924 | MEL47 | clone28 | MN610039 |
| MEL43 | clone26 | MN602925 | MEL47 | clone29 | MN610040 |
| MEL43 | clone27 | MN602926 | MEL47 | clone30 | MN610041 |
| MEL43 | clone28 | MN602927 | MEL47 | clone31 | MN610042 |
| MEL43 | clone29 | MN602928 | MEL47 | clone32 | MN610043 |
| MEL43 | clone30 | MN602929 | MEL47 | clone33 | MN610044 |
| MEL43 | clone31 | MN602930 | MEL47 | clone34 | MN610045 |
| MEL43 | clone32 | MN602931 | MEL47 | clone35 | MN610046 |
| MEL43 | clone33 | MN602932 | MEL47 | clone36 | MN610047 |
| MEL43 | clone34 | MN602933 | MEL47 | clone37 | MN610048 |
| MEL47 | clone1 | MN610012 | MEL47 | clone38 | MN610049 |
| MEL47 | clone2 | MN610013 | MEL47 | clone39 | MN610050 |
| MEL47 | clone3 | MN610014 | MEL47 | clone40 | MN610051 |

Suppl. Table 4 GenBank ID of ITS region clone libraries sampled from strain MEL43 and MEL47

| Strain | Clone | GenBank ID | Strain | Clone | GenBank ID |
|---|---|---|---|---|---|
| MEL43 | clone1 | MN603081 | MEL47 | clone11 | MN603198 |
| MEL43 | clone2 | MN603082 | MEL47 | clone12 | MN603199 |
| MEL43 | clone3 | MN603083 | MEL47 | clone13 | MN603200 |
| MEL43 | clone4 | MN603166 | MEL47 | clone14 | MN603201 |
| MEL43 | clone5 | MN603167 | MEL47 | clone15 | MN603202 |
| MEL43 | clone6 | MN603168 | MEL47 | clone16 | MN603203 |
| MEL43 | clone7 | MN603169 | MEL47 | clone17 | MN603204 |
| MEL43 | clone8 | MN603170 | MEL47 | clone18 | MN603205 |
| MEL43 | clone9 | MN603171 | MEL47 | clone19 | MN603206 |
| MEL43 | clone10 | MN603172 | MEL47 | clone20 | MN603207 |
| MEL43 | clone11 | MN603173 | MEL47 | clone21 | MN603208 |
| MEL43 | clone12 | MN603174 | MEL47 | clone22 | MN603209 |
| MEL43 | clone13 | MN603175 | MEL47 | clone23 | MN603210 |
| MEL43 | clone14 | MN603176 | MEL47 | clone24 | MN603211 |
| MEL43 | clone15 | MN603177 | MEL47 | clone25 | MN603212 |
| MEL43 | clone16 | MN603178 | MEL47 | clone26 | MN603213 |
| MEL43 | clone17 | MN603179 | MEL47 | clone27 | MN603214 |
| MEL43 | clone18 | MN603180 | MEL47 | clone28 | MN603215 |
| MEL43 | clone19 | MN603181 | MEL47 | clone29 | MN603216 |
| MEL43 | clone20 | MN603182 | MEL47 | clone30 | MN603217 |
| MEL43 | clone21 | MN603183 | MEL47 | clone31 | MN603218 |
| MEL43 | clone22 | MN603184 | MEL47 | clone32 | MN603219 |
| MEL43 | clone23 | MN603185 | MEL47 | clone33 | MN603220 |
| MEL43 | clone24 | MN603186 | MEL47 | clone34 | MN603221 |
| MEL43 | clone25 | MN603187 | MEL47 | clone35 | MN603222 |
| MEL43 | clone26 | MN603188 | MEL69 | clone1 | MN603063 |
| MEL43 | clone27 | MN603189 | MEL69 | clone2 | MN603064 |
| MEL43 | clone28 | MN603190 | MEL69 | clone3 | MN603065 |
| MEL44 | clone1 | MN603057 | CCMP628 | clone1 | MN603066 |
| MEL44 | clone2 | MN603058 | CCMP628 | clone2 | MN603067 |
| MEL44 | clone3 | MN603059 | CCMP628 | clone3 | MN603068 |
| MEL45 | clone1 | MN603060 | CCMP629 | clone1 | MN603069 |
| MEL45 | clone2 | MN603061 | CCMP629 | clone2 | MN603070 |
| MEL45 | clone3 | MN603062 | CCMP629 | clone3 | MN603071 |
| MEL47 | clone1 | MN603084 | CCMP1524 | clone1 | MN603072 |
| MEL47 | clone2 | MN603085 | CCMP1524 | clone2 | MN603073 |
| MEL47 | clone3 | MN603086 | CCMP1524 | clone3 | MN603074 |
| MEL47 | clone4 | MN603191 | CCMP2754 | clone1 | MN603075 |
| MEL47 | clone5 | MN603192 | CCMP2754 | clone2 | MN603076 |
| MEL47 | clone6 | MN603193 | CCMP2754 | clone3 | MN603077 |

| | | | | | |
|---|---|---|---|---|---|
| MEL47 | clone7 | MN603194 | RCC2055 | clone1 | MN603078 |
| MEL47 | clone8 | MN603195 | RCC2055 | clone2 | MN603079 |
| MEL47 | clone9 | MN603196 | RCC2055 | clone3 | MN603080 |
| MEL47 | clone10 | MN603197 | | | |

Suppl. Table 6 GenBank ID of *rbc*S-*rpl*27 region clone libraries sampled from strain MEL43 and MEL47

| Strain | clone | GenBank ID | Strain | clone | GenBank ID |
|---|---|---|---|---|---|
| MEL43 | clone1 | MN990301 | MEL43 | clone37 | MN990337 |
| MEL43 | clone2 | MN990302 | MEL43 | clone38 | MN990338 |
| MEL43 | clone3 | MN990303 | MEL43 | clone39 | MN990339 |
| MEL43 | clone4 | MN990304 | MEL43 | clone40 | MN990340 |
| MEL43 | clone5 | MN990305 | MEL47 | clone1 | MN990269 |
| MEL43 | clone6 | MN990306 | MEL47 | clone2 | MN990270 |
| MEL43 | clone7 | MN990307 | MEL47 | clone3 | MN990271 |
| MEL43 | clone8 | MN990308 | MEL47 | clone4 | MN990272 |
| MEL43 | clone9 | MN990309 | MEL47 | clone5 | MN990273 |
| MEL43 | clone10 | MN990310 | MEL47 | clone6 | MN990274 |
| MEL43 | clone11 | MN990311 | MEL47 | clone7 | MN990275 |
| MEL43 | clone12 | MN990312 | MEL47 | clone8 | MN990276 |
| MEL43 | clone13 | MN990313 | MEL47 | clone9 | MN990277 |
| MEL43 | clone14 | MN990314 | MEL47 | clone10 | MN990278 |
| MEL43 | clone15 | MN990315 | MEL47 | clone11 | MN990279 |
| MEL43 | clone16 | MN990316 | MEL47 | clone12 | MN990280 |
| MEL43 | clone17 | MN990317 | MEL47 | clone13 | MN990281 |
| MEL43 | clone18 | MN990318 | MEL47 | clone14 | MN990282 |
| MEL43 | clone19 | MN990319 | MEL47 | clone15 | MN990283 |
| MEL43 | clone20 | MN990320 | MEL47 | clone16 | MN990284 |
| MEL43 | clone21 | MN990321 | MEL47 | clone17 | MN990285 |
| MEL43 | clone22 | MN990322 | MEL47 | clone18 | MN990286 |
| MEL43 | clone23 | MN990323 | MEL47 | clone19 | MN990287 |
| MEL43 | clone24 | MN990324 | MEL47 | clone20 | MN990288 |
| MEL43 | clone25 | MN990325 | MEL47 | clone21 | MN990289 |
| MEL43 | clone26 | MN990326 | MEL47 | clone22 | MN990290 |
| MEL43 | clone27 | MN990327 | MEL47 | clone23 | MN990291 |
| MEL43 | clone28 | MN990328 | MEL47 | clone24 | MN990292 |
| MEL43 | clone29 | MN990329 | MEL47 | clone25 | MN990293 |
| MEL43 | clone30 | MN990330 | MEL47 | clone26 | MN990294 |
| MEL43 | clone31 | MN990331 | MEL47 | clone27 | MN990295 |
| MEL43 | clone32 | MN990332 | MEL47 | clone28 | MN990296 |
| MEL43 | clone33 | MN990333 | MEL47 | clone29 | MN990297 |
| MEL43 | clone34 | MN990334 | MEL47 | clone30 | MN990298 |
| MEL43 | clone35 | MN990335 | MEL47 | clone31 | MN990299 |
| MEL43 | clone36 | MN990336 | MEL47 | clone32 | MN990300 |