Journal of Oceanology and Limnology
Vol. 36 No. 4, P. 1315-1328, 2018
https://doi.org/10.1007/s00343-018-7088-7

The first complete organellar genomes of an Antarctic red alga, *Pyropia endiviifolia*: insights into its genome architecture and phylogenetic position within genus *Pyropia* (Bangiales, Rhodophyta)*

XU Kuipeng (徐奎鹏)[1], TANG Xianghai (唐祥海)[1, **], BI Guiqi (毕桂芪)[1], CAO Min (曹敏)[1], WANG Lu (王璐)[1], MAO Yunxiang (茅云翔)[1, 2, **]

[1] *Key Laboratory of Marine Genetics and Breeding* (*Ocean University of China*), *Ministry of Education*, *Qingdao 266003*, *China*
[2] *Laboratory for Marine Biology and Biotechnology*, *Qingdao National Laboratory for Marine Science and Technology*, *Qingdao 266003*, *China*

Received Mar. 20, 2017; accepted in principle May 4, 2017; accepted for publication Jun. 15, 2017

Abstract *Pyropia* species grow in the intertidal zone and are cold-water adapted. To date, most of the information about the whole plastid and mitochondrial genomes (ptDNA and mtDNA) of this genus is limited to Northern Hemisphere species. Here, we report the sequencing of the ptDNA and mtDNA of the Antarctic red alga *Pyropia endiviifolia* using the Illumina platform. The plastid genome (195 784 bp, 33.28% GC content) contains 210 protein-coding genes, 37 tRNA genes and 6 rRNA genes. The mitochondrial genome (34 603 bp, 30.5% GC content) contains 26 protein-coding genes, 25 tRNA genes and 2 rRNA genes. Our results suggest that the organellar genomes of *Py*. *endiviifolia* have a compact organization. Although the collinearity of these genomes is conserved compared with other *Pyropia* species, the genome sizes show significant differences, mainly because of the different copy numbers of rDNA operons in the ptDNA and group II introns in the mtDNA. The other *Pyropia* species have 2–3 distinct intronic ORFs in their *cox*1 genes, but *Py*. *endiviifolia* has no introns in its *cox*1 gene. This has led to a smaller mtDNA than in other *Pyropia* species. The phylogenetic relationships within *Pyropia* were examined using concatenated gene sets from most of the available organellar genomes with both the maximum likelihood and Bayesian methods. The analysis revealed a sister taxa affiliation between the Antarctic species *Py*. *endiviifolia* and the North American species *Py*. *kanakaensis*.

Keyword: Antarctic; *Pyropia endiviifolia*; plastid and mitochondrial genomes; genome structure; phylogenetic

1 INTRODUCTION

The evolution of plastids and mitochondria by endosymbiosis is a central dogma of modern eukaryotic cell biology. Both plastids and mitochondria possess their own genomes. Plastids are the light-gathering organelles of algae and plants responsible for photosynthesis, whose origin can be traced back to cyanobacteria (Reyes-Prieto et al., 2007). This photosynthetic organelle is commonly believed to have a single origin in the common ancestor of the Archaeplastida, which comprises glaucophytes, red algae (Rhodophyta), and Viridiplantae (Rodríguez-Ezpeleta et al., 2005). Red algae have the most gene-rich and cyanobacteria-like plastid genomes, followed by glaucophytes and green algae. The mitochondrial genome is a remnant of a eubacterial genome, derived specifically from within the α-Proteobacteria (Gray et al., 2001). Mitochondria play a crucial role in providing cellular energy (Ogihara et al., 2005). During the course of evolution, this endosymbiont has transferred

* Supported by the National Natural Science Foundation of China (No. 31372517), the Scientific and Technological Innovation Project Financially Supported by Qingdao National Laboratory for Marine Science and Technology (No. 2015ASKJ02), and the National Infrastructure of Fishery Germplasm Resources (No. 2016DKA30470)
** Corresponding authors: txianghai@ouc.edu.cn; yxmao@ouc.edu.cn

many of its important genes to the nuclear genome (Taanman, 1999).

The Bangiales order of red algae consists of more than 190 species (Guiry and Guiry, 2017), which are distributed worldwide from tropical seas to polar seas. It is divided into at least fifteen genus-level taxa (*Bangia*, '*Bangia*' *1*, '*Bangia*' *2*, '*Bangia*' *3*, *Boreophyllum*, *Clymene*, *Dione*, *Fuscifolium*, *Lysithea*, *Minerva*, *Miuraea*, *Porphyra*, *Pseudobangia*, *Pyropia*, and *Wildemania*) (Sutherland et al., 2011). The cold-water seaweed genus *Pyropia* includes the most economically important marine crops grown in intertidal habitats, among which *Pyropia haitanensis* and *Py*. *yezoensis* are widely harvested and traded in East Asian countries, such as China, Korea and Japan (Mumford and Miura, 1988). *Pyropia endiviifolia* (A. Gepp & E. Gepp) H. G. Choi & M. S. Hwang grows on the Antarctic islands and has been recorded on the Antarctic Peninsula, the South Orkney Islands, the South Shetland Islands and South Georgia Island (Wiencke and Clayton, 1998). This species is olive-green in color, which led to its first specific name, *Monostroma endiviifolium* A & E Gepp (Chamberlain, 1963); this was later revised to *Pyropia* (Sutherland et al., 2011).

Previous studies have revealed that a limited number of available DNA sequences results in relatively little genetic variation, which can present difficulties in phylogenetic resolution or species identification (Dutcher and Kapraun, 1994; Niwa et al., 2004; Xie et al., 2010; Sutherland et al., 2011). In recent decades, with the rapid development of next-generation DNA sequencing technologies, it has become convenient to assemble complete organelle genomes from total genomic DNA sequences at relatively low cost, especially for *Pyropia* species, which have a high proportion of organellar DNA relative to nuclear DNA (Wang et al., 2013). Complete organellar genome sequence information is not only important for genetic breeding but also for evolutionary studies. Phylogenomics is a useful tool for providing evolutionary information for species identification, taxonomy and phylogenetic analysis (Henry, 2005; Verbruggen et al., 2010; Janouškovec et al., 2013; Yang et al., 2015; Lee et al., 2016). However, most of the available information about *Pyropia* ptDNAs and mtDNAs is limited to Northern Hemisphere species, and surprisingly little is known about the organellar genomes of Southern Hemisphere species.

Here, we present the complete organellar genomes of the Antarctic species *Py*. *endiviifolia*, which were obtained using the Illumina sequencing technology, and examine its genomic features. Through comparative genomics and phylogenomic analyses, we sought to explore the genome structure and reconstruct the phylogenetic relationships among representative species.

2 MATERIAL AND METHOD

2.1 Collection of samples and morphological observations

Fresh thalli of *Py*. *endiviifolia* were collected on February 22, 2014 from intertidal transects along a rocky coastline at Fildes Peninsula, King George Island, Antarctica (62°12′S, 58°57′W). Morphological characters including thallus shape, color, texture and reproductive tissues of the specimens were examined and photographed using an Olympus BX51 microscope (OLYMPUS, Tokyo, Japan).

2.2 DNA extraction, sequencing and genome assembly

Total DNA was extracted from 10 g of frozen thallus material according to the CTAB method (Porebski et al., 1997). Purified DNA (5 μg) was fragmented and used to construct short-insert PCR-free libraries following the instructions of the Illumina Truseq™ DNA Sample Preparation Kit (Illumina, San Diego, CA, USA) and was sequenced on an Illumina Genome Analyzer. Adapters and low-quality reads (with ambiguous bases, N; length < 100 bp) were removed using the NGS QC Toolkit (Patel and Jain, 2012). The pre-processed sequences were first assembled into non-redundant contigs using Edena with default settings (Hernandez et al., 2008). Then, all contigs were mapped to the reference genomes of *Py*. *haitanensis* (NC_007932.1 and NC_017751) using the BLAST program (http://blast.ncbi.nlm.nih.gov/) with an e-value of 1e-5 and the order of the aligned contigs was verified. Finally, gaps between the contigs were filled by iterative contig extension using the PRICE software (Ruby et al., 2013). To evaluate the quality of the organelle genome sequences, especially the junctions, validation through intensive PCR-based sequencing was carried out on ABI 3730 instrument by randomly designing 20 pairs of primers (Table S1). The PCR sequences and assembled genomes were aligned using MEGA 6.0 to determine the accuracy of the assembly (Tamura et al., 2013). The complete *Py*. *endiviifolia* plastid and mitochondria genomes are available for download via GenBank with accession numbers KT716756 and KU356193.

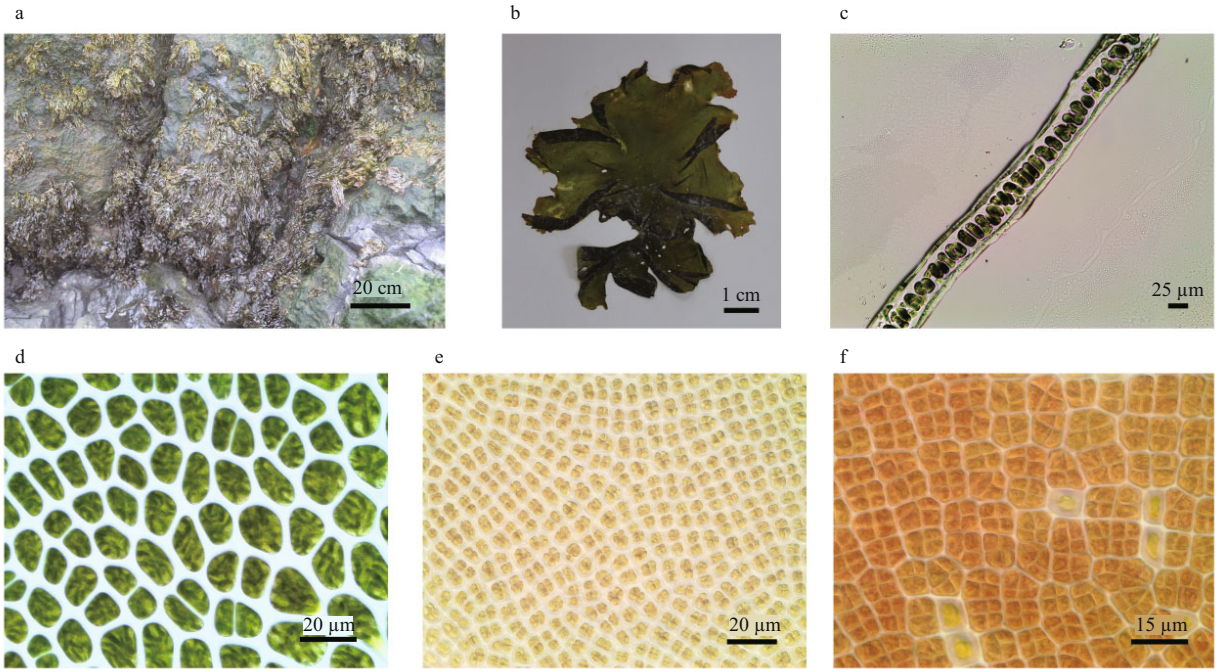


Fig.1 *Pyropia endiviifolia*

a. *Pyropia endiviifolia* on intertidal rocks on King George Island; b. leafy gametophytes; c. cross-section of the vegetative region of the thallus; d. surface view of basal rhizoidal cells; e. surface view of spermatangia; f. surface view of zygotosporangia.

2.3 Genome annotation and analysis

The organellar genomes were annotated using ORF-finder (http://www.ncbi.nlm.nih.gov/projects/gorf/) and aligned via BLASTX and BLASTN searches at the NCBI website (http://blast.ncbi.nlm.nih.gov/). tRNAs were identified using the tRNAscan-SE 1.21 web server (http://lowelab.ucsc.edu/tRNAscan-SE/) and rRNAs were identified using the RNAmmer 1.2 server (http://www.cbs.dtu.dk/services/RNAmmer/). Genome maps were drawn with OGDraw (Lohse et al., 2007). Multiple genomes were aligned using MAFFT version 5 and visualized using the mVISTA tool (Mayor et al., 2000; Katoh et al., 2005). A structure comparison was generated by Mauve with the 'Use seed families' option (Darling et al., 2004).

2.4 Phylogenetic analyses

To elucidate the phylogenetic position of *Py. endiviifolia*, the concatenated protein-coding amino acid sequences from both the plastid and mitochondrial genomes were used to construct a phylogenetic tree (Table S2). The genome sequences were aligned using the program MAFFT version 5 and were adjusted manually (Katoh et al., 2005). The aligned sequences were trimmed using trimAl with the option 'automated1' (Capella-Gutiérrez et al., 2009). Maximum likelihood (ML) analysis was conducted using RaxML-8.2.4 (Stamatakis, 2014). The best model and parameter settings were chosen according to the Akaike information criterion by ProtTest 3.0 for ML analysis (Abascal et al., 2005). The ML searches used the cpREV+G+I substitution model for plastid sequences and JTT+G+I substitution model for mitochondrial sequences (-f a, 1 000 bootstrap replicates). Bayesian analyses were carried out using MrBayes3.2 with the best ProtTest model noted above (Huelsenbeck and Ronquist, 2001). Four independent Markov Chain Monte Carlo chains were run simultaneously and sampled every 100 generations for a total of 1 000 000 generations. The first 10% of the trees were discarded as a "burn-in."

3 RESULT

3.1 Morphological analysis

On the basis of morphology and life history observations (Wang et al., 2008; Guiry, 2015), the specimen was identified as *Py. endiviifolia* (Chamberlain, 1963; Wiencke and Clayton, 1998). This species was very distinctive in the local region because of its position high in the inter-tidal zone, its rough texture compared with other *Pyropia* species and its dark greenish color, which became blackish on drying (Fig.1a). The gametophyte blades measured

Table 1 Plastid gene content for *Pyropia endiviifolia*

Gene group	Gene name
Photosystem I	*psa*M, *psa*L, *psa*K, *psa*J, *psa*I, *psa*F, *psa*E, *psa*D, *psa*C, *psa*B, *psa*A
Photosystem II	*psb*Z, *psb*Y, *psb*X, *psb*W, *psb*V, *psb*T, *psb*N, *psb*L, *psb*K, *psb*J, *psb*I, *psb*H, *psb*F, *psb*E, *psb*D, *psb*C, *psb*B, *psb*A
Protochlorophyllide reductase	*chl*N, *chl*L, *chl*I, *chl*B
Phycobiliproteins	*apc*F, *apc*E, *apc*D, *apc*B, *apc*A, *apc*P, *cpe*B, *cpe*A, *cpc*G, *cpc*B, *cpc*A
Cytochrome b/f complex	*pet*N, *pet*M, *pet*J, *pet*G, *pet*F, *pet*D, *pet*B, *pet*A
ATP synthase	*atp*I, *atp*H, *atp*G, *atp*F, *atp*E, *atp*D, *atp*B, *atp*A
RNA polymerase	*rpo*C2, *rpo*C1, *rpo*B, *rpo*A
Ribosomal proteins (SSU)	*rps*9, *rps*8, *rps*7, *rps*6, *rps*5, *rps*4, *rps*3, *rps*20, *rps*2, *rps*19, *rps*18, *rps*17, *rps*16, *rps*14, *rps*13, *rps*12, *rps*11, *rps*10, *rps*1
Ribosomal proteins (LSU)	*rpl*9, *rpl*6, *rpl*5, *rpl*4, *rpl*36, *rpl*35, *rpl*34, *rpl*33, *rpl*32, *rpl*31, *rpl*3, *rpl*29, *rpl*28, *rpl*27, *rpl*24, *rpl*23, *rpl*22, *rpl*21, *rpl*20, *rpl*2, *rpl*19, *rpl*18, *rpl*16, *rpl*14, *rpl*13, *rpl*12, *rpl*11, *rpl*1
Hypothetical chloroplast *orf*s	*ycf*7, *ycf*65, *ycf*63, *ycf*61, *ycf*59, *ycf*46, *ycf*4, *ycf*39, *ycf*38, *ycf*37, *ycf*36, *ycf*35, *ycf*34, *ycf*33, *ycf*3, *ycf*29, *ycf*23, *ycf*22, *ycf*21, *ycf*20, *ycf*19, *ycf*17, *ycf*12
Transfer RNAs	*trn*Y(GTA), *trn*W(CCA), *trn*V(TAC), *trn*V(GAC), *trn*T(TGT), *trn*T(GGT), *trn*S(TGA), *trn*S(GGA), *trn*S(GCT), *trn*S(CGA), *trn*R(TCT), *trn*R(CCT), *trn*R(CCG), *trn*R(ACG), *trn*Q(TTG), *trn*P(TGG), *trn*N(GTT), *trn*M(CAT), *trn*M(CAT), *trn*M(CAT), *trn*L(TAG), *trn*L(TAA), *trn*L(GAG), *trn*L(CAA), *trn*K(TTT), *trn*I(GAT), *trn*I(GAT), *trn*H(GTG), *trn*G(TCC), *trn*G(GCC), *trn*F(GAA), *trn*E(TTC), *trn*D(GTC), *trn*C(GCA), *trn*A(TGC), *trn*A(TGC), *trn*A(GGC)
Ribosomal RNAs	*rrs*B, *rrs*A, *rrl*B, *rrl*A, *rrf*B, *rrf*A
Other genes	*tuf*A, *tsf*, *trx*A, *trp*G, *trp*A, *thi*S, *thi*G, *tat*C, *syh*, *syf*B, *suf*C, *suf*B, *sec*Y, *sec*G, *sec*A, *rne*, *rbc*S, *rbc*R, *rbc*L, *pre*A, *pgm*A, *pbs*A, *omp*R, *odp*B, *odp*A, *ntc*A, *ndh*I, *nbl*A, *moe*B, *inf*C, *inf*B, *ilv*H, *ilv*B, *gro*EL, *glt*B, *gln*B, *fts*H, *ftr*B, *fab*H, *dsb*D, *dna*K, *dna*B, *dfr*, *clp*C, *cem*A, *ccs*A, *ccs*1, *cbb*X, *car*A, *bas*1, *acc*D, *acc*B, *acc*A , *arg*B
Open reading frames	*orf*68, *orf*621, *orf*62, *orf*58, *orf*565, *orf*382, *orf*327, *orf*320, *orf*287, *orf*238, *orf*231, *orf*203, *orf*198, *orf*174, *orf*149, *orf*148, *orf*122, *orf*121, *orf*114, *orf*111, *orf*107, *orf*108

5–20 cm in length and 4–15 cm in width and were monostromatic (Fig.1b, c). They attached to the rocks via abundant rhizoidal cells at the base of the thallus (Fig.1d). The almost colorless spermatangia around the margins of the gametophytes were formed by repeated division of vegetative cells (Fig.1e). The red carposporangia were formed by direct transformation of the vegetative cells while the fertilized zygotosporangium divided mitotically (Fig.1f).

3.2 Organellar genome features of *Py. endiviifolia*

The plastid genome of *Py*. *endiviifolia* was 195 784 base pairs (bp) long and contained two direct non-identical repeat (DR) regions encoding 16S, 23S, 5S rRNA and two tRNA genes (trnI, trnA). These two repeats divided the circular molecule into a 150.6-kb large single copy (LSC) region and a 35.6-kb small single copy (SSC) region (Fig.2a). The overall GC content was 33.28%. The plastid genome encoded a total of 253 genes, consisting of 210 protein-coding genes (including 23 hypothetical protein genes (*ycf*s) and 22 function-unknown open reading frames (*orf*s)), 37 tRNA genes and 6 rRNA genes, which comprised 75.83%, 1.44% and 4.58% of the total sequence, respectively (Table 1). Similar to other Bangiales species (*Py*. *haitanensis*, 254 genes; *Py*. *yezoensis*, 256 genes; *Bangia fuscopurpurea*, 250 genes), all the genes in the LSC and SSC regions were single copy without introns, and 14 genes overlapped (*psb*C–*psb*D, *atp*D–*atp*F, *ycf*24–*ycf*16, *rps*19–*rpl*2, *rpl*23–*rpl*4, *car*A–*orf*238 and *rpl*24–*rpl*14).

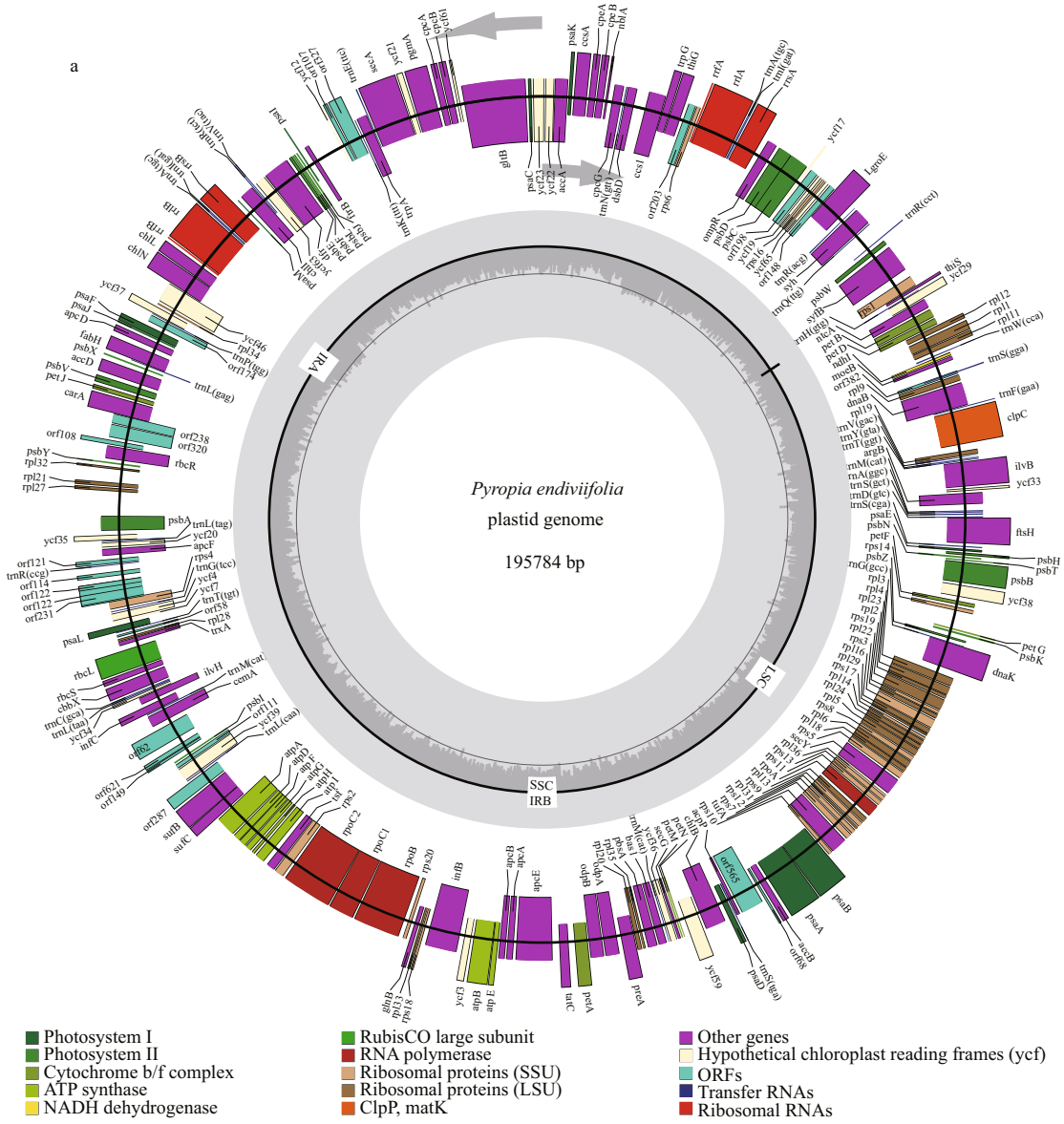
The mitochondrial genome of *Py*. *endiviifolia* contained 53 genes, was 34 603 bp in length and had 30.46% GC content (Fig.2b). The genome contained 2 ribosomal RNA genes (1 large subunit and 1 small subunit), 25 transfer RNAs, 3 *orf*s, 2 secY-independent transporter proteins (*ymf*s), 4 ribosomal proteins, and 17 genes related to electron transport and oxidative phosphorylation (Table 2). The protein-coding, tRNA and rRNA genes comprised 58.41%, 5.44% and 18.65% of the whole sequence, respectively.

3.3 Genome conservation and dissimilarities

Multiple alignment of 10 plastid sequences of Bangiales was conducted to further understand the structure and sequence similarity of the *Py*. *endiviifolia* plastid genome. Using *Py*. *yezoensis* as a reference, the sequence identity alignment results were plotted (Fig.3a). The results revealed high similarity across the *Pyropia* ptDNAs. The majority of variations

Table 2 Mitochondrial gene content for *Pyropia endiviifolia*

Gene group	Gene name
Complex I (NADH dehydrogenase)	*nad*1, *nad*2, *nad*3, *nad*4, *nad*4L, *nad*5, *nad*6
Complex II (succinate dehydrogenase)	*sdh*2, *sdh*3, *sdh*4
Complex III (ubiquinol cytochrome c reductase)	*cob*
Complex IV (cytochrome c oxidase)	*cox*1, *cox*2, *cox*3
ATP synthase	*atp*6, *atp*8, *atp*9
Ribosomal proteins	*rps*3, *rps*11, *rps*12, *rpl*16
Open reading frames	*orf*169, *orf*238, *orf*546, *ymf*16, *ymf*39
Transfer RNAs	*trn*L(UAA), *trn*S(GCU), *trn*L(UAG), *trn*D(GUC), *trn*H(GUG), *trn*M(CAU), *trn*S(UGS), *trn*Q(UUG), *trn*U(UCA), *trn*G(UCC), *trn*I(GAU), *trn*E(UUC), *trn*M(CAU), *trn*V(UAC), *trn*R(ACG), *trn*V(TAC), *trn*F(GAA), *trn*A(UGC), *trn*Y(GUA), *trn*K(UUU), *trn*R(UCU), *trn*C(GCA), *trn*G(GCC), *trn*N(GUU), *trnP*(UGG)
Ribosomal RNAs	*rnl*, *rns*



To be continued

Fig.2 Continued

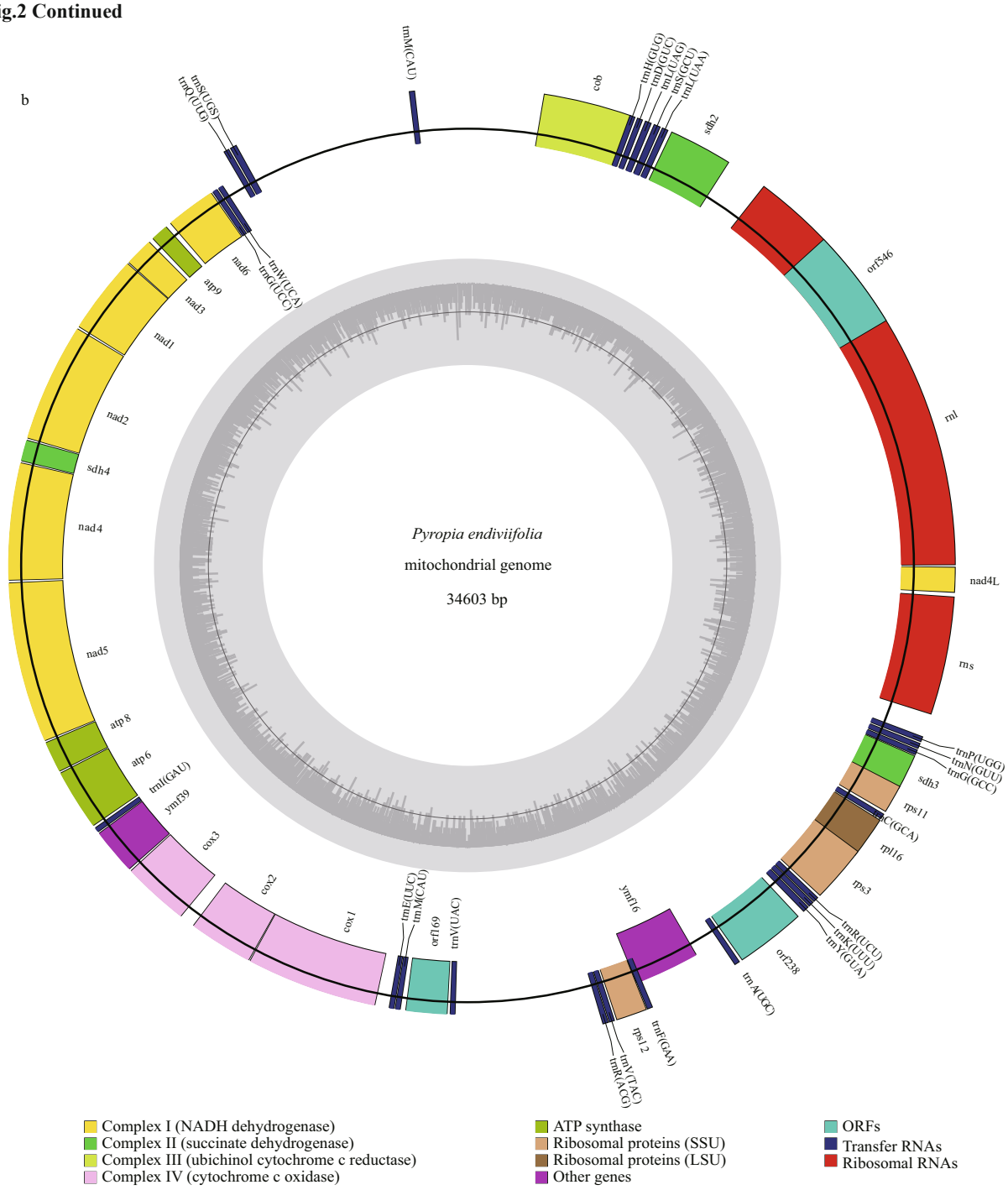
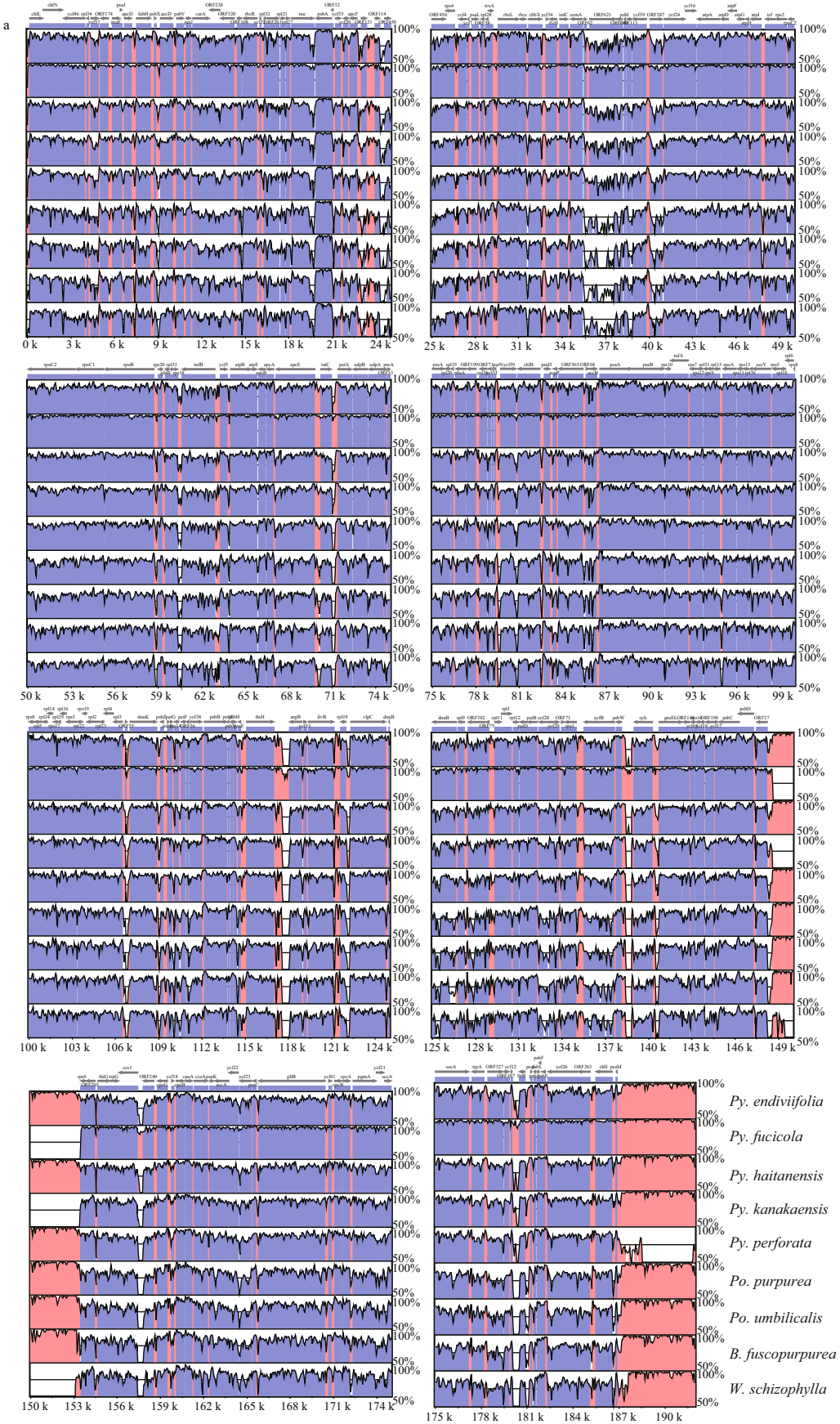


Fig.2 Genome maps of the *Pyropia endiviifolia* plastid (a) and mitochondrion (b)

Genes shown outside the outer circle are transcribed clockwise and those inside are transcribed counterclockwise. Genes belonging to different functional groups are color coded. The dashed area in the inner circle indicates the GC content of the organellar genome.

resulted from small insertions or deletions in intergenic regions. As expected, the rDNA regions were more conserved than the single-copy regions, and the coding regions were more conserved than the intergenic regions. *Pyropia* showed some differences when compared with other genera of Bangiales. For example, the intergenic regions between the *pet*G-*rps*14 genes were longer in the *Pyropia* ptDNAs. Notably, the similarity of *orf*621 was very low between *Pyropia* and other groups. Collinearity analysis showed that the architecture of the ptDNAs was highly conserved without any large rearrangements, despite their evolutionary distance (Fig.4a). The only apparent distinction was a single



To be continued

Fig.3 Continued

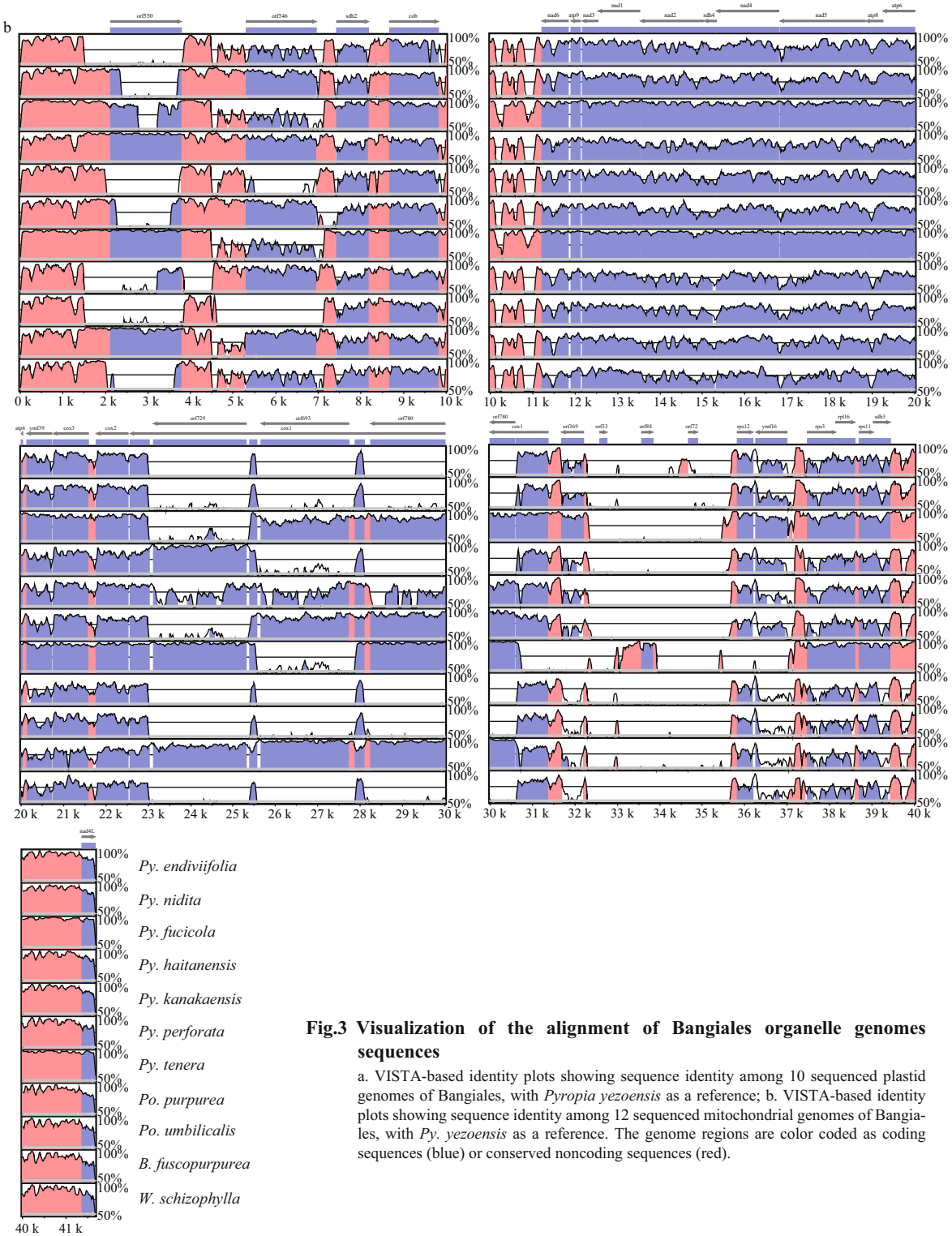


Fig.3 Visualization of the alignment of Bangiales organelle genomes sequences

a. VISTA-based identity plots showing sequence identity among 10 sequenced plastid genomes of Bangiales, with *Pyropia yezoensis* as a reference; b. VISTA-based identity plots showing sequence identity among 12 sequenced mitochondrial genomes of Bangiales, with *Py. yezoensis* as a reference. The genome regions are color coded as coding sequences (blue) or conserved noncoding sequences (red).

copy rDNA region in the species *Py. perforate* and *Wildemania schizophylla*. By contrast, *Py. endiviifolia*, *Py. yezoensis*, *Py. haitanensis*, *Porphyra* and *Bangia* possessed two direct non-identical rDNA repeats (*Py. fucicola* and *Py. kanakaensis* had partial genomes).

Collinearity analysis of 12 mtDNAs showed that most sequence blocks were conserved co-linearly, but the genome contents and lengths were significantly different (Fig.3b; Fig.4b). The observed structural differences mainly arose from the number and organization of mitochondrial group II introns in the large subunit ribosomal RNA (*rnl*) gene and the *cox*1

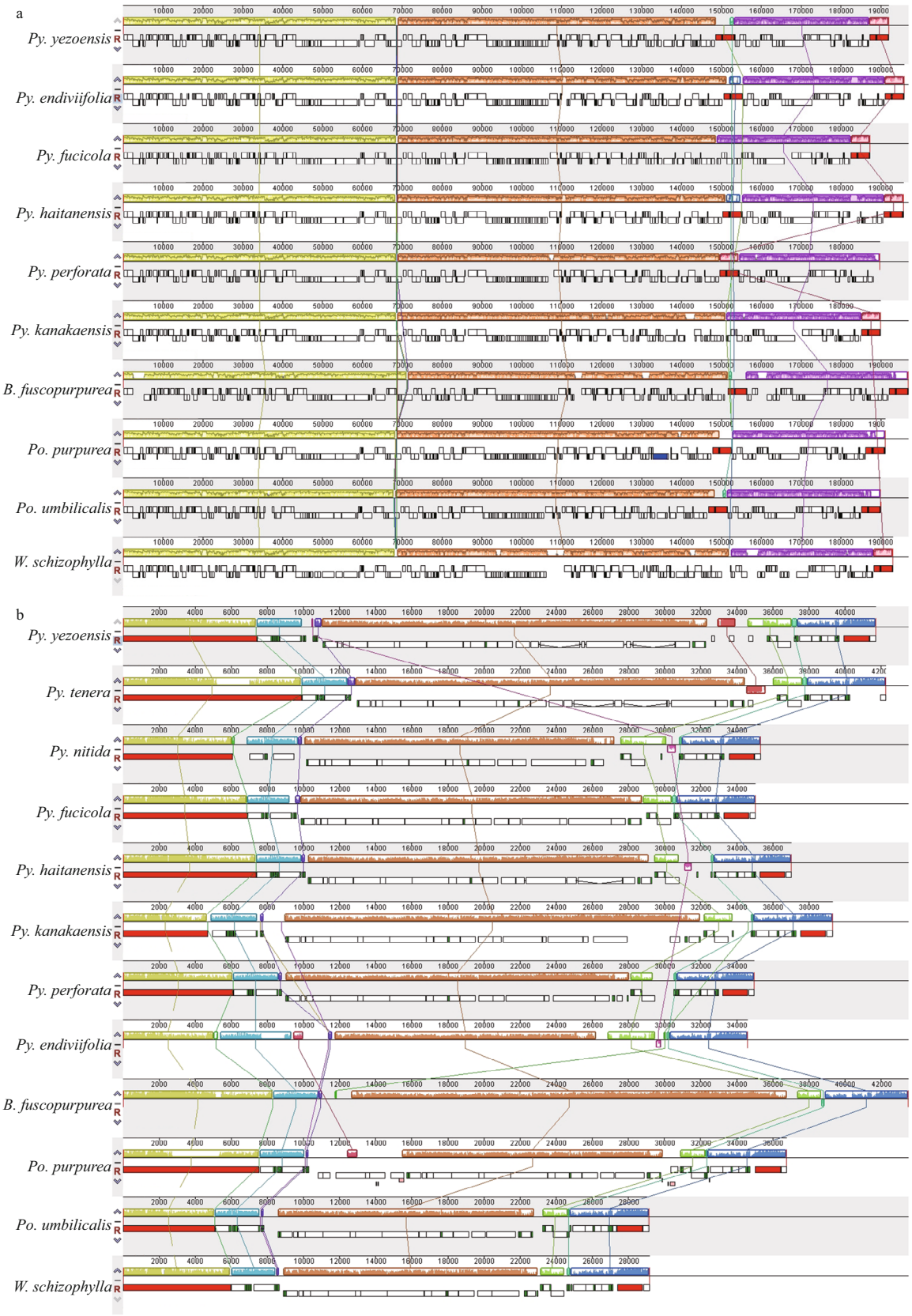


Fig.4 Local collinear block (LCB) analysis of 10 plastid genomes (a) and 12 mitochondrial genomes (b)

Each chromosome is oriented horizontally, and homologous blocks are shown as identically colored regions linked across genomes. Sequence similarities within an LCB are proportional to the heights of the interior colored bars. Large sections of white within blocks and gaps between blocks indicate lineage-specific sequences.

Table 3 General characteristics of Bangiales plastid and mitochondrial genomes

		Size (bp)	CDS	tRNA	rRNA	Introns	Intronic ORF	GC content (%)	Accession No.
Py. endiviifolia	PT	195 784	210	37	6	-	-	33.3	KT716756
	MT	34 603	26	25	2	1	1	30.5	KU356193
Py. yezoensis	PT	191 975	213	37	6	-	-	33.1	KC517072
	MT	41 688	27	27	2	5	5	32.7	NC_017837
Py. haitanensis	PT	195 597	211	37	6	-	-	33.0	KC464603
	MT	37 023	24	24	2	4	4	30.7	NC_017751
Py. fucicola	PT	187 282	203	36	3	-	-	32.7	KJ776837
	MT	35 035	31	23	2	3	3	32.5	NC_024288
Py. kanakaensis	PT	189 931	206	36	3	-	-	32.8	KJ776836
	MT	39 300	30	25	2	5	3	30.0	NC_024289
Py. perforata	PT	189 789	208	36	3	-	-	32.9	KF515973
	MT	40 042	24	23	2	4	3	31.8	KJ708761
Py. nitida	PT	-	-	-	-	-	-	-	-
	MT	35 313	28	17	1	3	3	30.4	KP890080
Py. tenera	PT	-	-	-	-	-	-	-	-
	MT	42 269	25	23	2	6	6	32.8	NC_021475
Po. purpurea	PT	191 028	209	35	6	-	-	33.0	NC_000925
	MT	36 753	29	24	2	2	2	33.5	NC_002007
Po. umbilicalis	PT	189 933	207	37	6	-	-	32.8	JQ408795
	MT	29 123	25	24	2	1	1	31.9	NC_018544
B. fuscopurpurea	PT	196 913	207	37	6	-	-	33.3	KP714733
	MT	43 517	31	23	2	5	6	33.0	NC_026905
W. schizophylla	PT	193 008	239	34	3	-	-	34.4	KP020505
	MT	29 156	26	23	2	2	2	33.2	NC_024579

gene (Table 3). The mitochondrial genomes of Bangiales possessed different numbers of introns and intronic ORFs (*orf*111, *orf*543, *orf*544, *orf*546, *orf*550) in the *rnl* gene. The mtDNA of *Py*. *endiviifolia* had only one intron, which contained one intronic ORF (*orf*546), in the *rnl* gene. The other genomes of *Pyropia* contained at least two introns with 0–3 different intronic ORFs. The *Pyropia* species had 2–3 distinct introns and intronic ORFs (*orf*693, *orf*729, *orf*789, *orf*813) in the *cox*1 gene, except *Py*. *endiviifolia*, whose *cox*1 gene had no introns. An absence of introns in the *cox*1 gene was also observed in the *Porphyra* and *Wildemania* mtDNAs. The low number of introns resulted in *Py*. *endiviifolia* having the smallest mtDNA within *Pyropia*. *B*. *fuscopurpurea* had the most intronic ORFs in the *cox*1 gene (*orf*652, *orf*693, *orf*780 and *orf*813).

3.4 Phylogenetic analyses

Trees were constructed using a dataset of 160 amino acid sequences of ptDNAs and 22 amino acid sequences of mtDNAs selected from eight representative species to examine the evolutionary position of *Py*. *endiviifolia*, and all of the nodes were inferred with strong support by the ML and BI methods (Fig.5). Within *Pyropia*, the close relationship between the Antarctic species *Py*. *endiviifolia* and the North American species *Py*. *kanakaensis* was

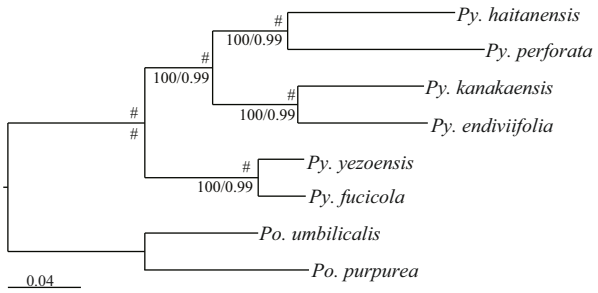


Fig.5 Phylogenetic relationships within the Bangiales clade

Numbers above the lines indicate ptDNA results; numbers below the lines indicate mtDNA results; numbers on left indicate ML bootstrap values and numbers on the right indicate Bayesian posterior probabilities. "#" indicates that the node was fully supported by both methods.

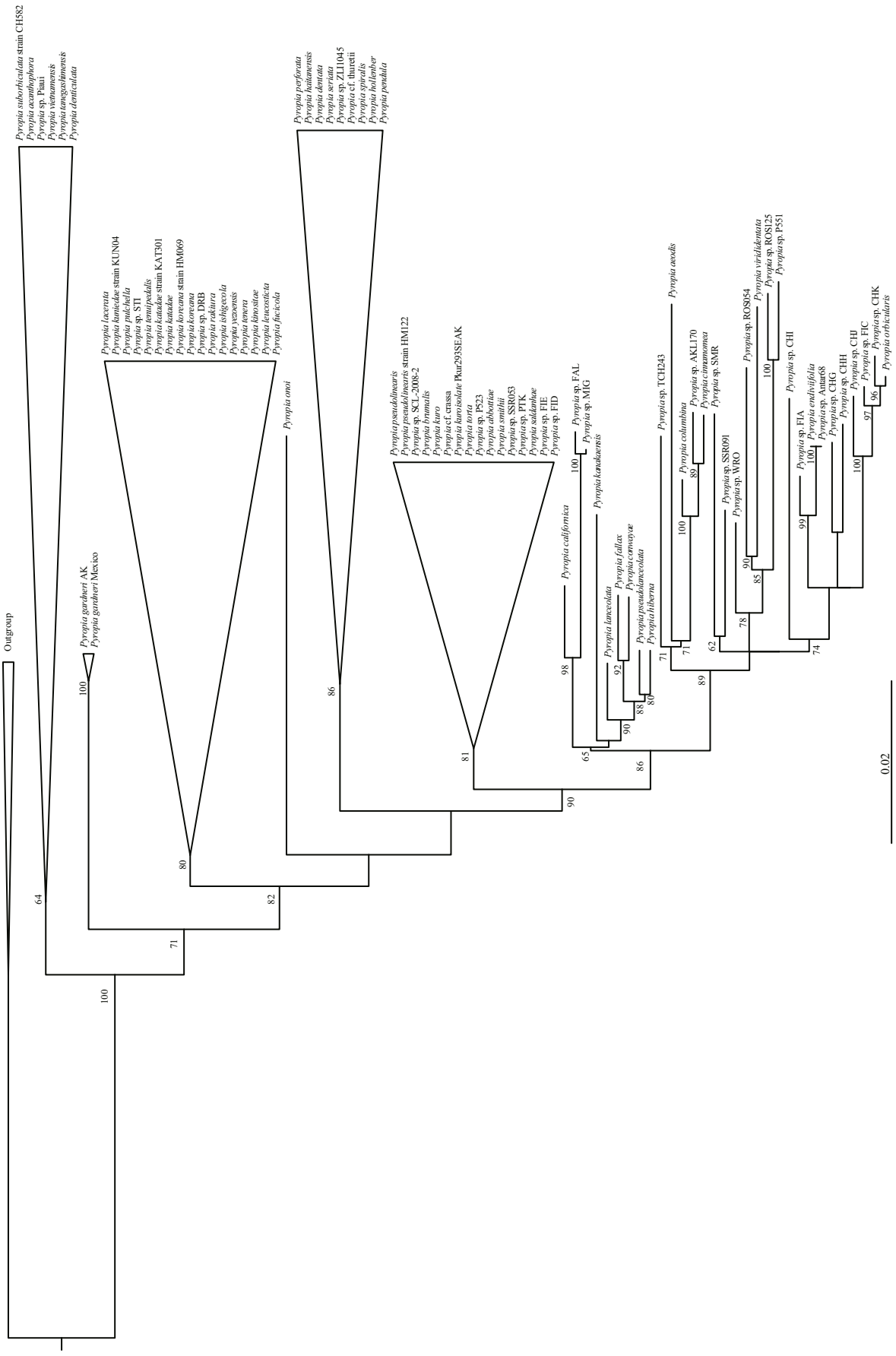


Fig.6 Maximum likelihood rooted tree for *rbcL* sequences of *Pyropia*

Porphyra purpurea and *Porphyra umbilicalis* were used as outgroups. For each node, maximum likelihood bootstrap values are indicated and only high support values (>60) are shown.

confirmed. The species *Py*. *haitanensis* and *Py*. *perforate* also formed a separate clade. These two clades clustered together with a sister relationship. The remaining species *Py*. *yezoensis* and *Py*. *fucicola* grouped together at the base of the *Pyropia* group.

To verify the phylogenetic relationships of this group, the *rbc*L genes from the *Py*. *endiviifolia* plastid genome and 81 species of *Pyropia* (Table S3) downloaded from GenBank were used for phylogenetic tree reconstruction (Fig.6). The overall topologies were consistent with the trees constructed using whole organellar genomes. The inconsistencies resulted from the low number of available organellar genomes of *Pyropia*. The phylogenetic tree topology demonstrated that *Py*. *endiviifolia* formed a well-supported clade together with the unidentified *Pyropia* sp. *Antar*68 from Admiralty Bay, King George Island, South Shetlands Archipelago, Antarctica. The sequence similarity between *Py*. *endiviifolia* and *Pyropia* sp. *Antar*68 was 100%, which indicated that they might be the same species.

4 DISCUSSION

The first plastid and mitochondrial genomes of the Antarctic red algae *Py*. *endiviifolia* were determined in this study. The organellar genomes of *Py*. *endiviifolia* have large protein-coding gene repertoires and a compact genome organization. Comparative genomic analysis revealed highly conserved collinearity across the whole organellar genomes. The differences in size among mitochondrial genomes were related to the number and organization of mitochondrial group II introns of the large subunit of the ribosomal RNA gene and the *cox*1 gene. Typically, eukaryotes possess inserted sequences termed group II introns, but these sequences are only observed in organellar genomes (Michel et al., 1982). A previous study indicated that horizontal transfers have taken place from the mitochondrial genomes of diatoms to the alga *Chattonella* (Kamikawa et al., 2009). *Pyropia endiviifolia* had no introns in its *cox*1 gene, which was unique among *Pyropia* species. The number of introns in the *rnl* gene was also lower than in other *Pyropia* species. This lack of introns could lead to a convergent and stabilized mtDNA structure. It could also be used as a basis for designing molecular markers for species identification. The structure and number of *rnl* and *cox*1 introns in *Py*. *endiviifolia* implies a specific evolutionary mechanism in this Antarctic species.

We identified two direct non-identical repeats in the *Py*. *endiviifolia* plastid genomes. By comparison, there was only one copy in *Py*. *perforate* and *W*. *schizophylla*, a phenomenon that has also been found in some Florideophyte species (*Calliarthron tuberculosum* and *Chondrus crispus*). Typically, most plastid genomes possess two large inverted repeats containing the rRNA genes. However, with the number of sequenced genomes increasing, more and more variations have been found (Hagopian et al., 2004). Analysis of the rDNA operons in *Guillardia* and *Porphyra* suggests that the directly repeated rDNA genes of the ancestral Rhodophyte were transformed into inverted repeats in *Guillardia* (Douglas, 1998). Lee et al. detected three minor structural types (R1-, R2-, and R3-type) in the Florideophyceae group, which were explained by recombination events of the duplicated rDNA operons (Lee et al., 2016). The two rDNA operons have been only partially retained or one copy has been completely lost in some red algae species. This process could lead to structural stabilization of the plastid genomes. The ancestral R1-type rDNA operon was retained in *Py*. *endiviifolia*, which implies slow evolution of the ptDNA structure.

Studies have shown that multigene phylogenies can elucidate phylogenetic relationships more exactly when the different evolutionary rates of the genes are considered (Yoon et al., 2006; Verbruggen et al., 2010). We utilized a set of ptDNA and mtDNA genes to explore the phylogenetic relationships of *Pyropia*. In this study, the Antarctic species *Py*. *endiviifolia* and the North American species *Py*. *kanakaensis* grouped together with high support in the phylogenetic analysis, rather than all the of Northern Hemisphere species clustering together first. Despite their geographical isolation, all members of this group are cold-water adapted (Brodie and Irvine, 2003), which has led to almost identical environmental selection pressure. The fixation rate in genome evolution depends on the purifying selection of the environment (Buschiazzo et al., 2012). This result implies that the selection pressure the Antarctic species has experienced was more similar to that of *Py*. *kanakaensis* than other species. However, inconsistencies in phylogenetic analysis can occur when there is sparse taxon sampling (Zhao et al., 2016). The current evidence is insufficient to interpret the origin and evolution of *Py*. *endiviifolia*, because it is the only Southern Hemisphere species with complete organellar genomes available. Therefore, more taxon information needs to be obtained and

further studies combining the nuclear, plastid and mitochondrial genomes need to be performed to better understand the relationship of this algal group.

5 DATA AVAILABILITY STATEMENT

The authors declare that all data supporting the findings of this study are available within the methods and appendix sections.

References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*, **21**(9): 2 104-2 105.

Brodie J A, Irvine L M. 2003. Seaweeds of the British Isles. Volume 1 Rhodophyta. Part 3B Bangiophycidae. Natural History Museum, London.

Buschiazzo E, Ritland C, Bohlmann J, Ritland K. 2012. Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. *BMC Evolutionary Biology*, **12**: 8.

Capella-Gutiérrez S, Silla-Martínez J M, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**(15): 1 972-1 973.

Chamberlain Y M. 1963. The identity of *Monostroma endiviifolium* A. and E.S. Gepp. *Nova Hedwigia*, **5**: 151-155.

Darling A C E, Mau B, Blattner F R, Perna N T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, **14**(7): 1 394-1 403.

Douglas S E. 1998. Plastid evolution: origins, diversity, trends. *Current Opinion in Genetics & Development*, **8**(6): 655-661.

Dutcher J A, Kapraun D F. 1994. Random amplified polymorphic DNA (RAPD) identification of genetic variation in three species of *Porphyra* (Bangiales, Rhodophyta). *Journal of Applied Phycology*, **6**(3): 267-273.

Gray M W, Burger G, Lang B F. 2001. The origin and early evolution of mitochondria. *Genome Biology*, **2**(6): reviews1018.1.

Guiry M D, Guiry G M. 2017. AlgaeBase. World-Wide Electronic Publication, National University of Ireland, Galway, http://www.marinespecies.org/aphia.php?p=sourcedetails&id=37.

Hagopian J C, Reis M, Kitajima J P, Bhattacharya D, De Oliveira M C. 2004. Comparative analysis of the complete plastid genome sequence of the red alga *Gracilaria tenuistipitata* var. *liui* provides insights into the evolution of rhodoplasts and their relationship to other plastids. *Journal of Molecular Evolution*, **59**(4): 464-477.

Henry R J. 2005. Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants. CABI Publishing, Wallingford, Oxfordshire, UK.

Hernandez D, François P, Farinelli L, Østerås M, Schrenzel J. 2008. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Research*, **18**(5): 802-809.

Huelsenbeck J P, Ronquist F. 2001. MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics*, **17**(8): 754-755.

Janouškovec J, Liu S L, Martone P T, Carré W, Leblanc C, Collén J, Keeling P J. 2013. Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. *PLoS One*, **8**(3): e59001.

Kamikawa R, Masuda I, Demura M, Oyama K, Yoshimatsu S, Kawachi M, Sako Y. 2009. Mitochondrial group II introns in the raphidophycean flagellate *Chattonella spp*. suggest a diatom-to-Chattonella lateral group II intron transfer. *Protist*, **160**(3): 364-375.

Katoh K, Kuma K I, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, **33**(2): 511-518.

Lee J, Cho C H, Park S I, Cho J W, Song H S, West J A, Bhattacharya D, Yoon H S. 2016. Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC Biology*, **14**: 75.

Lohse M, Drechsel O, Bock R. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics*, **52**(5-6): 267-274.

Mayor C, Brudno M, Schwartz J R, Poliakov A, Rubin E M, Frazer K A, Pachter L S, Dubchak I. 2000. VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics*, **16**(11): 1 046-1 047.

Michel F, Jacquier A, Dujon B, 1982. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie*, **64**(10): 867-881.

Mumford Jr T F, Miura A. 1988. *Porphyra* as food: cultivation and economics. *In*: Lembi C A, Waaland J R eds. Algae and Human Affairs. Cambridge University Press, Cambridge.

Niwa K, Kikuchi N, Iwabuchi M, Aruga Y. 2004. Morphological and AFLP variation of *Porphyra yezoensis* Ueda form, *narawaensis* Miura (Bangiales, Rhodophyta). *Phycological Research*, **52**(2): 180-190.

Ogihara Y, Yamazaki Y, Murai K, Kanno A, Terachi T, Shiina T, Miyashita N, Nasuda S, Nakamura C, Mori N, Takumi S, Murata N, Futo S, Tsunewaki K. 2005. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Research*, **33**(19): 6 235-6 250.

Patel R K, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*, **7**(2): e30619.

Porebski S, Bailey L G, Baum B R. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter*, **15**(1): 8-15.

Reyes-Prieto A, Weber A P, Bhattacharya D. 2007. The origin

and establishment of the plastid in algae and plants. *Annual Review Genetics*, **41**: 147-168.

Rodríguez-Ezpeleta N, Brinkmann H, Burey S C, Roure B, Burger G, Löffelhardt W, Bohnert H J, Philippe H, Lang B F. 2005. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Current Biology*, **15**(4): 1 325-1 330.

Ruby J G, Bellare P, DeRisi J L. 2013. PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. *G3*: *Genes*, *Genomes*, *Genetics*, **3**(5): 865-880.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**(9): 1 312-1 313.

Sutherland J E, Lindstrom S C, Nelson W A, Brodie J, Lynch M D J, Hwang M S, Choi H G, Miyata M, Kikuchi N, Oliveira M C, Farr T, Neefus C, Mols-Mortensen A, Milstein D, Müller K M. 2011. A new look at an ancient order: generic revision of the Bangiales (Rhodophyta). *Journal of Phycology*, **47**(5): 1 131-1 151.

Taanman J W. 1999. The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta* (*BBA*)-*Bioenergetics*, **1410**(2): 103-123.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, **30**(12): 2 725-2 729.

Verbruggen H, Maggs C A, Saunders G W, Le Gall L, Yoon H S, De Clerck O. 2010. Data mining approach identifies research priorities and data requirements for resolving the red algal tree of life. *BMC Evolutionary Biology*, **10**: 16.

Wang L, Mao Y X, Kong F N, Li G Y, Ma F, Zhang B L, Sun P P, Bi G Q, Zhang F F, Xue H F, Cao M. 2013. Complete sequence and analysis of plastid genomes of two economically important red algae: *Pyropia haitanensis* and *Pyropia yezoensis*. *PLoS One*, **8**(5): e65902.

Wiencke C, Clayton M N. 1998. The life history of *Porphyra endiviifolium* from the South Shetland Islands, Antarctica. *Polar Biology*, **19**(4): 257-263.

Xie C T, Chen C S, Xu Y, Ji D H. 2010. Construction of a genetic linkage map for *Porphyra haitanensis* (Bangiales, Rhodophyta) based on sequence-related amplified polymorphism and simple sequence repeat markers. *Journal of Phycology*, **46**(4): 780-787.

Yang E C, Kim K M, Kim S Y, Lee J, Boo G H, Lee J H, Nelson W A, Yi G M, Schmidt W E, Fredericq S, Boo S M, Bhattacharya D, Yoon H S. 2015. Highly conserved mitochondrial genomes among multicellular red algae of the Florideophyceae. *Genome Biology and Evolution*, **7**(8): 2 394-2 406.

Yoon H S, Müller K M, Sheath R G, Ott F D, Bhattacharya D. 2006. Defining the major lineages of red algae (Rhodophyta). *Journal of Phycology*, **42**(2): 482-492.

Zhao L, Li X, Zhang N, Zhang S D, Yi T S, Ma H, Guo Z H, Li D Z. 2016. Phylogenomic analyses of large-scale nuclear genes provide new insights into the evolutionary relationships within the rosids. *Molecular Phylogenetics and Evolution*, **105**: 166-176.

Electronic supplementary material

Supplementary material (Supplementary Tables S1–S3) is available in the online version of this article at https://doi.org/10.1007/s00343-018-7088-7.