environmental microbiology

Environmental Microbiology (2014) **16**(2), 467–485 doi:10.1111/1462-2920.12275

Cob gene pyrosequencing enables characterization of benthic dinoflagellate diversity and biogeography

Gurjeet S. Kohli,[1,5] Brett A. Neilan,[1,3]* Mark V. Brown,[1,3] Mona Hoppenrath[4] and Shauna A. Murray[2,5]

[1]*School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, New South Wales, 2052, Australia.*

[2]*Sydney Institute of Marine Sciences, Sydney, New South Wales 2088, Australia.*

[3]*Australian Centre for Astrobiology, University of New South Wales, Sydney, New South Wales 2052, Australia.*

[4]*Senckenberg am Meer, Deutsches Zentrum für Marine Biodiversitätsforschung (DZMB), Wilhelmshaven 32689, Germany.*

[5]*Plant Functional Biology and Climate Change Cluster (C3), University of Technology Sydney, Sydney, New South Wales 2007, Australia.*

Summary

Dinoflagellates in marine benthic habitats living epiphytically on macroalgae are an important but highly understudied group of protists. Many produce toxins that can have severe economic impacts on marine-based economies, and improved monitoring tools are required to enhance the management of toxin-related hazards. We analysed the distribution and diversity of epibenthic dinoflagellates inhabiting eight sites in Cocos (Keeling) Islands, Papua New Guinea, and Broome and Exmouth, Western Australia. We used pyrosequencing approaches based on two DNA barcoding marker genes – 18S ribosomal RNA (rRNA) and mitochondrial cytochrome b (*cob*) – and compared these to an approach based on clone libraries (197 sequences) using the *cob* gene. Dinoflagellate sequences accounted for 133 [64 unique operational taxonomic units (OTU)] out of 10 529 18S rRNA gene sequences obtained from all samples. However, using the dinoflagellate specific assay targeting the *cob* gene marker, we obtained 9748 (1217 unique OTU) dinoflagellate sequences from the same environmental samples, providing the largest, to date, set of dinoflagellate *cob* gene sequences and reliable estimates of total dinoflagellate richness within the samples and biogeographic comparisons between samples. This study also reports the presence of potentially toxic species of the genera *Gambierdiscus*, *Ostreopsis*, *Coolia*, *Prorocentrum* and *Amphidinium* from the above-mentioned geographical regions.

Received 7 January, 2013; accepted 1 September, 2013. *For correspondence. E-mail b.neilan@unsw.edu.au; Tel. (+61) 9385 3235; Fax (+61) 9385 1591.

Introduction

More than 2000 extant species of dinoflagellates, a group of microbial eukaryotes, are known (Taylor *et al*., 2008). However, our knowledge of their diversity and global distribution in specific environments is far from comprehensive. Dinoflagellates in marine benthic habitats are an important but understudied group of protists. They are widely distributed in both tropical and temperate environments, living epibenthically on dead corals, sediments and numerous macroalgae (Murray, 2010). Some benthic taxa, such as the species of *Gambierdiscus*, *Prorocentrum* and *Ostreopsis* produce secondary metabolites (toxins) that are linked to seafood-related toxin diseases worldwide (Murakami *et al*., 1982; Usami *et al*., 1995; Chinain *et al*., 2010; Kalaitzis *et al*., 2010). For example, ciguatera fish poisoning affects between 50 000 and 500 000 people per year (Fleming *et al*., 1998), and is caused by the ingestion of fish that have accumulated toxins produced by *Gambierdiscus* spp. Similarly, *Ostreopsis* spp. produces the secondary metabolite palytoxin, which is highly toxic and is found throughout certain reef-associated food webs, and causes human illness and death (Usami *et al*., 1995; Deeds and Schwartz, 2010). *Prorocentrum lima* and related species of *Prorocentrum* are producers of okadaic acid (Murakami *et al*., 1982; MacKenzie *et al*., 2005), a toxin that bioaccumulates in shellfish and causes diarrhetic shellfish poisoning following consumption (Quilliam and Wright, 1995). Recently, a thecate peridinioid dinoflagellate was identified as a potent producer of pinnatoxin, a toxin that accumulates in oysters (Smith *et al*., 2011). With reports of a 60% increase in ciguatera fish poisoning in the Pacific islands over the last decade, despite being highly underreported (Skinner *et al*., 2011), it is important to understand the biogeography and distribution of interstitial (sand-dwelling) and epiphytic benthic dinoflagellates worldwide. An increased knowledge of their diversity,

ecology and distribution would enable improved monitoring and management of toxin-related hazards, and would allow the development of a baseline in order to determine the effect of future changes to ocean temperature, reef condition and currents.

Most of the studies that describe benthic dinoflagellate diversity have investigated it using microscopic observation of live or fixed field samples, or have characterized single cells and cultures using molecular genetic techniques, and more recently a combination of genetics and microscopy has been used (Taylor, 1979; Fukuyo, 1981; Carlson and Trindall, 1985; Ballantine *et al*., 1988; Faust and Balech, 1993; Morton and Faust, 1997; Pearse *et al*., 2001; Okolodkov *et al*., 2007; Rhodes *et al*., 2010). While these methods are accurate, they do not generally provide information about the total dinoflagellate diversity in a particular area. For example, one of the most comprehensive taxonomic studies of epibenthic dinoflagellates to date revealed 12 genera and 39 species, including unidentified taxa (Turquet *et al*., 1998); however, there are indications that many cryptic species of dinoflagellates are present in field samples (Stern *et al*., 2010). Also, microscopic observations are time-consuming and require taxonomic expertise.

In the last decade, numerous attempts have been made to describe the diversity of microbial eukaryotes directly from environmental samples based on sequencing of ribosomal RNA (rRNA) genes via cloning (Moon-van der Staay *et al*., 2001; López-García *et al*., 2003), metagenomics (Cuvelier *et al*., 2010) and pyrosequencing (Brown *et al*., 2009; Edgcomb *et al*., 2011; Steven *et al*., 2012). However, these general approaches are unlikely to retrieve the full range of benthic dinoflagellate species diversity due to (i) the use of primers, which amplify all eukaryotes in environmental samples and are biased towards certain taxonomic groups (Potvin and Lovejoy, 2009) or highly abundant organisms, and can 'drown out' the signal of low-abundance organisms, such as dinoflagellates, if they constitute a very small proportion of the community; and (ii) the use of generic statistical analysis methods that have been primarily developed for estimating prokaryotic diversity and do not take into account the huge range of rRNA operons in dinoflagellate genomes.

These deficiencies highlight the need to develop molecular tools for the specific purpose of describing dinoflagellate diversity, such as polymerase chain reaction (PCR) primer sets, that can selectively amplify only dinoflagellate sequences directly from environmental samples. In 2006, a dinoflagellate-specific 18S rRNA primer set was developed for describing dinoflagellate community structure from environmental samples (Lin *et al*., 2006). This method was used to amplify most of the dinoflagellate genera as tested in that study; however, its product size (1.6 kb) is not suitable for developing pyrosequencing assays. Recently, an internal transcribed spacer (ITS) region based on a barcoding technique has revealed hidden diversity in cultured dinoflagellates (Stern *et al*., 2012). However, the potential of a method based on a region of ITS rRNA to be used as an environmental barcoding marker is limited due to the presence of many gene paralogues, the potential for unidentifiable chimaeras and priming across taxa (Stern *et al*., 2012). Other studies used the mitochondrial genes *cob* (Lin *et al*., 2009) and cytochrome c oxidase (*cox*) (Stern *et al*., 2010) to determine the diversity of dinoflagellates directly from environmental DNA samples. Although these studies revealed a reasonable amount of dinoflagellate diversity, they used conventional PCR-based cloning and Sanger sequencing, with results limited to < 200–800 clone sequences (Lin *et al*., 2009; Stern *et al*., 2010). However, with the availability of next-generation sequencing, sequencing depth has been increased exponentially and has yet to be used to reveal dinoflagellate diversity in benthic systems.

In this study, we describe the diversity of epibenthic dinoflagellates in several different geographical locations in the Australasian region. Conventional barcoding molecular techniques targeting the *cob* and 18S rRNA genes were compared with more advanced Tag-Encoded FLX 454-Pyrosequencing. The new pyrosequencing assay used a *cob*-based dinoflagellate-specific primer set (Lin *et al*., 2009), which produces short PCR products that provide good taxon resolution to identify species. Using this assay, we isolated the largest, to date, set of epibenthic dinoflagellate *cob* gene sequences that provide reliable estimates of total dinoflagellate richness within our samples.

Results

Diversity of dinoflagellates via PCR, cob *gene cloning and sequencing*

Three samples from Broome, two from Papua New Guinea (PNG), and one each from Cocos (Keeling) Island (CKI) and Exmouth were analysed via PCR and generating *cob* gene clone libraries. Sequences from all 197 clones were identified as *cob* genes representing various genera belonging to Dinophyceae, based on a high level of similarity in a BLAST search (Table 1). Phylogenetic analysis was performed to facilitate accurate identification of sequences. Depending on the level of resolution of the *cob* barcoding region, they were identified at genus or species level. Between 5 and 14 distinct dinoflagellate sequences were identified from each site, belonging to 3–8 different genera of dinoflagellates. Some of the sequences could be identified at species level, such as *Prorocentrum lima*, *Prorocentrum micans* and *Ostreopsis siamensis* (Fig. 1, Table 1).

Table 1. Details of diversity of dinoflagellates at each site based on PCR-based *cob* gene cloning and sequencing.

Sample ID	Total number of clones sequenced	Closest identifiable sister group in the phylogeny (number of different unique sequences identified phylogenetically)
B1	26	*Polarella* sp. (1), *Symbiodinium* sp. (1), *Akashiwo* sp. (1), *Gymnodinium* sp. (1), *Scrippsiella* sp. (1), *Pyrocystis* sp. (2), *Coolia sp.* (1), *Ostreopsis sp.* (1)
B2	28	*Prorocentrum lima* (1), *Prorocentrum micans* (1), *Prorocentrum* sp. (1), *Gymnodinium catenatum* (1), *Akashiwo* sp. (1), *Pfiesteria* sp. (1), *Coolia* sp. (2), *Ostreopsis* sp. (1), *Symbiodinium* sp. (1)
B3	25	*Symbiodinium* sp. (2), *Gymnodinium* sp. (2), *Pyrocystis* sp. (1), *Coolia* sp. (1), *Ostreopsis* sp. (1)
C1	32	*Symbiodinium* sp. (1), *Prorocentrum lima* (1), *Amphidinium* sp. (1), *Gambierdiscus* sp. (1), *Coolia* sp. (2), *Ostreopsis sp.* (2)
E1	29	*Symbiodinium* sp. (1), *Polerella* sp. (1), *Prorocentrum lima* (1), *Prorocentrum* sp. (2), *Gambierdiscus* sp. (1)
P1	26	*Symbiodinium* sp. (1), *Prorocentrum* sp. (1)[a], *Ostreopsis siamensis* (1)
P2	31	*Polerella* sp. (1), *Prorocentrum lima* (1), *Prorocentrum* sp. (1), *Amphidinium* sp. (1), *Pfiesteria* sp. (1)

a. Not shown in Fig. 1.

Diversity of dinoflagellates via the Cob *tag-encoded FLX 454-pyrosequencing (cTEFP)*

Four samples from Broome, two from CKI and one from Exmouth were analysed via cTEFP. A total of 30 809 high-quality sequences were recovered, of which 9748 were greater than or equal to 380 bp in length. More than 98.46% of the 9748 sequences were identified as dinoflagellate *cob* sequences, with closest match to those of known dinoflagellates in the GenBank (nr/nt) database. This demonstrated that the primer pair dinocob4F and dinocob6R (Lin *et al*., 2009) is highly dinoflagellate-specific, and that they can be used to amplify dinoflagellate sequences directly from environmental samples.

The composition of the dinoflagellate populations varied greatly within each sample. Sequences belonging to the order Peridiniales were the most abundant in the whole dataset, with a total of 3245 sequences across all the seven samples. At 99% sequence similarity, 400 operational taxonomic units (OTUs) were obtained, of which 243 occurred more than once (Figs S1–S3, supporting information data). More than 90% of the sequences could not be taxonomically assigned at the genus level; however, sequences belonging to the genera *Scrippsiella*, *Peridinium* and *Pfiesteria* were identified via phylogenetic analysis (Figs S1–S3, supporting information data). Collectively, a total of 1967 sequences were identified as belonging to the order Gonyaulacales in all the seven environmental samples. At 99% sequence similarity, 279 OTUs were obtained, out of which 156 were represented more than once in the whole dataset. Sequences similar to *Gambierdiscus*, *Ostreopsis*, *Coolia*, *Alexandrium*, *Protoceratium* and *Gonyaulax* were found (Fig. 2). However, more than 50% of the sequences remain unidentified due to the lack of reference species. Sample coverage ranged between 87% and 99% for both Peridiniales and Gonyaulacales, indicating that most of the diversity within these orders was covered successfully (Table 2).

A total of 1838 sequences were obtained from the Gymnodiniales representing 101 OTUs at a level of 97% sequence similarity. Of these, 82 were found more than once in the whole dataset. Samples from Broome and CKI were highly diverse. Five to ten per cent of these sequences were similar to previously documented species. However, the majority of sequences could not be identified due to the lack of reference species, including all sequences from Exmouth (Fig. 3). Sample coverage ranged between 90% and 99%, indicating that most of the Gymnodiniales diversity was covered (Table 2).

A total of 889 sequences were identified as belonging to the order Prorocentrales. In phylogenetic analyses, the order often shows two distinct clades (Grzebyk *et al*., 1998; Hoppenrath and Leander, 2008; Murray *et al*., 2009). While acknowledging that habitat is not a consistent character for distinguishing clades or genera (Murray *et al*., 2005; Hoppenrath and Leander, 2008) because the majority of species in these clades are from a particular habitat, the clades are here referred to for simplicity as 'clade 1' (mostly benthic) and 'clade 2' (mostly planktonic). At 99% similarity, 126 OTUs were identified, of which 97 were represented by more than one sequence. Phylogenetic analysis revealed 69 OTUs in clade 1 and 28 in clade 2 (Fig. 4). Sequences similar to *Prorocentrum lima* and *Prorocentrum cassubicum* were found in the benthic clade from Broome, Exmouth and CKI (Fig. 4); however, most of the sequences in the planktonic clade could not be identified. Sample coverage was found to be more than 96%, with the exception of one sample from CKI (Table 2).

For the Suessiales, a total of 508 sequences were identified, which represented 23 OTUs at a level of 94% sequence similarity. Phylogenetic analysis revealed sequences similar to *Symbiodinium* sp. and *Polarella* sp. were present in the various samples (Fig. 4). There was a minimum of 12 OTUs that were not similar to any

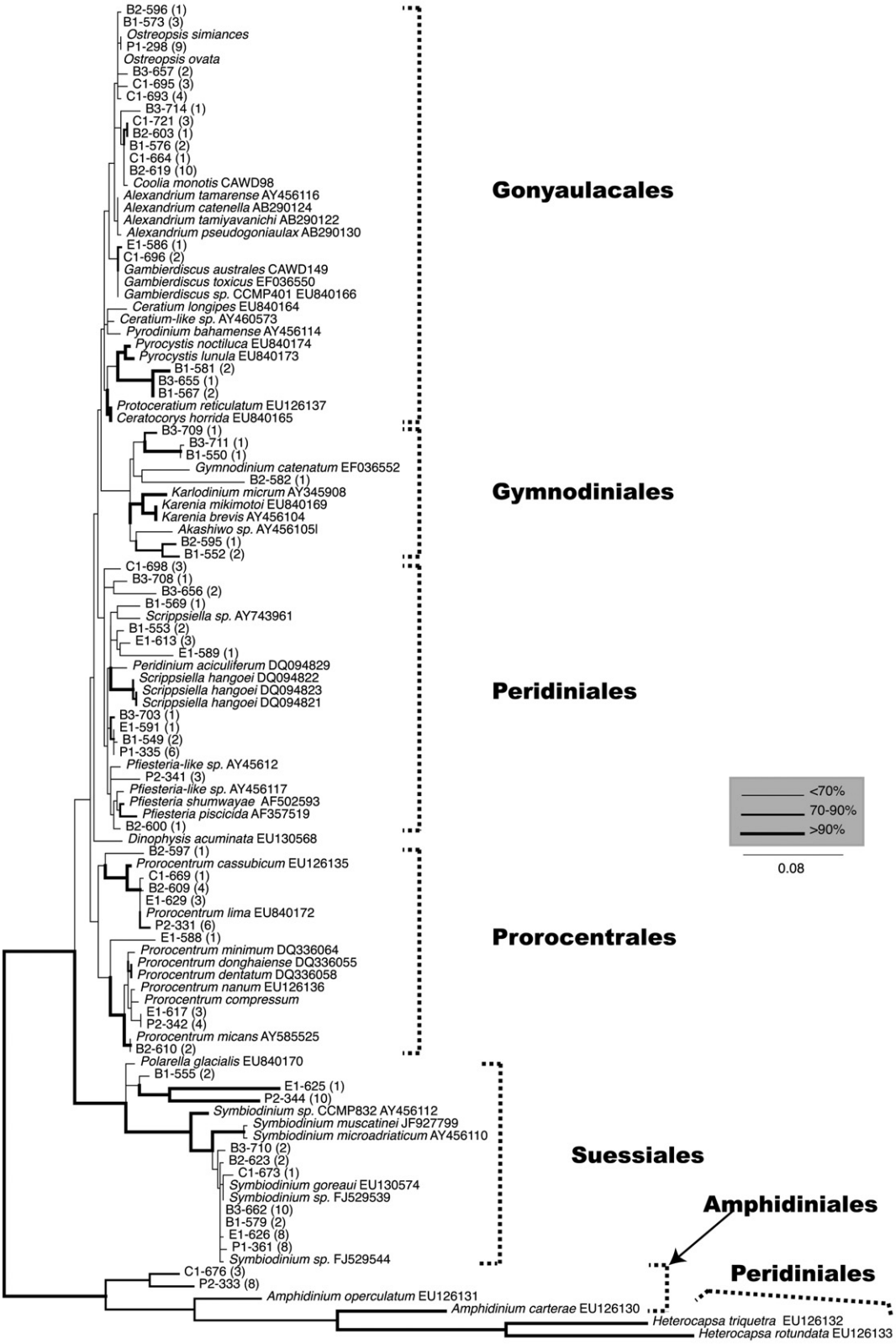

Fig. 1. Phylogenetic analysis using maximum likelihood of *cob* sequences obtained from PCR-based cob gene cloning and sequencing. Major orders of Dinophyceae were shown on the right. Bootstrap values were based on 500 replicates; the thickest branches denote bootstrap value of > 90%, medium-thick branches values of 70–90% and thin branches values of < 70%. Sequence names are followed by the number of clones obtained for each sequence in brackets. B1-B4 are samples from Broome, C1-C2 from CKI, E1 from Exmouth and P1-P2 from PNG. Reference sequence names are followed by their accession numbers.

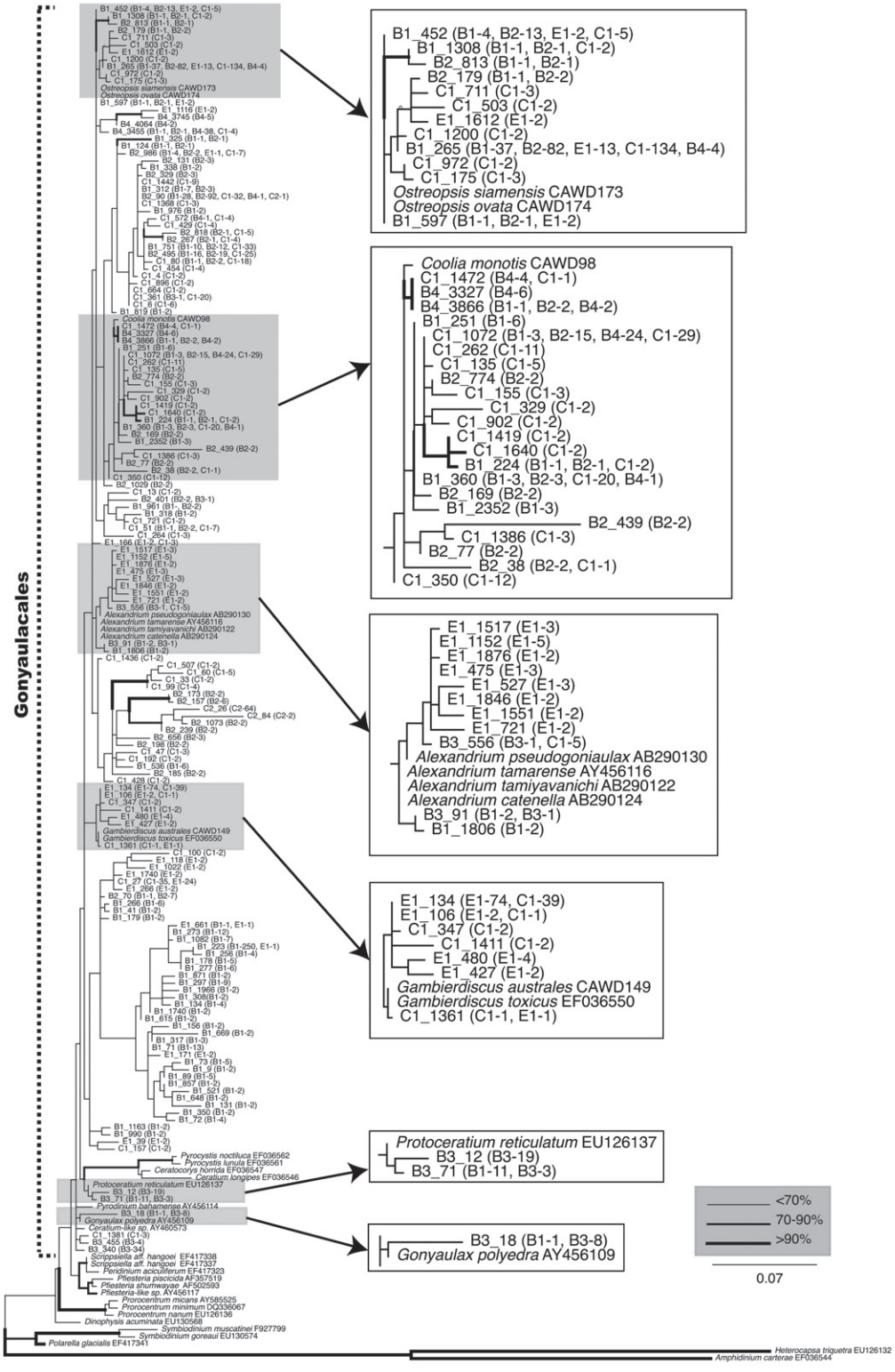

Fig. 2. Phylogenetic analysis using maximum likelihood of *cob* sequences identified as Gonyaulacales during cTEFP analysis. Bootstrap values were based on 500 replicates, and they are indicated by branch thickness; the thickest branches > 90%, medium-thick branches 70–90% and thin branches < 70%. Each unique OTU is represented by one representative sequence, for example B3_71 (B1-11, B3-3); B3_71 is the name of the representative sequence, and this OTU/sequence was found 11 times in sample B1 and 3 times in sample B3. B1-B4 are samples from Broome, C1-C2 from CKI and E1 from Exmouth. Reference sequence names are followed by their accession numbers.

Table 2. Data obtained during cTEFP analysis.

Sample ID	B1	B2	B3	B4	C1	C2	E1	Total
No. of SEQs obtained	5627	3264	2578	4540	4368	3730	6702	30809
No. of SEQs greater than 380 bp	2381	1181	761	402	1713	1227	2083	9748
No. of *cob* SEQs	2362	1176	753	391	1705	1210	2001	9598
SEQs identified as Peridiniales	430	238	287	17	260	846	1167	3245
No. of unique SEQs at 99% SEQ similarity	97	48	52	11	58	156	112	400
Coverage calculated at 99% SEQ similarity (%)	91.39	93.69	91.98	64.70	89.23	92.31	95.88	–
No. of unique SEQs for phylogenetic analysis[a]	58	42	33	8	42	82	78	243
Species identified (no. of genotypes)	*Pfiesteria* like sp. (3) Unknown genotypes (55)	*Pfiesteria* like sp. (4) Unknown genotypes (38)	*Scrippsiella* like sp. (1) *Pfiesteria* like sp. (3) Unknown genotypes (29)	Unknown genotypes (8)	*Pfiesteria* like sp. (3) *Scrippsiella* like sp. (7) Unknown genotypes (32)	*Scrippsiella* like sp. (3) *Pfiesteria* like sp. (3) Unknown genotypes (76)	*Scrippsiella* like sp. (3) Unknown genotypes (75)	
SEQs identified as Gonyaulacales	561	365	80	97	616	68	180	1967
No. of unique SEQs at 99% SEQ similarity	170	80	17	22	121	4	44	279
Coverage calculated at 99% SEQ similarity (%)	92.86	89.86	88.75	87.62	92.37	97.05	90.05	–
No. of unique SEQs for phylogenetic analysis[a]	65	40	9	11	63	3	29	156
Species identified (No. of genotypes)	*Ostreopsis* like sp. (1) *Coolia* like sp. (6) *Alexandrium* like sp. (2) *Protoceratium* like sp. (1) *Gonyaulax* like sp. (1) Unknown genotypes (54)	*Ostreopsis* like sp. (1) *Coolia* like sp. (9) Unknown genotypes (30)	*Alexandrium* like sp. (2) *Protoceratium* like sp. (2) *Gonyaulax* like sp. (1) Unknown genotypes (4)	*Coolia* like sp. (5) Unknown genotypes (6)	*Coolia* like sp. (13) *Alexandrium* like sp. (1) *Gambierdiscus* sp. (1) *Gambierdiscus* like sp. (4) Unknown genotypes (44)	Unknown genotypes (3)	*Ostreopsis* like sp. (1) *Alexandrium* like sp. (8) *Gambierdiscus* sp. (1) *Gambierdiscus* like sp. (4) Unknown genotypes (15)	
SEQs identified as Gymnodiniales	911	238	213	10	373	79	14	1838
No. of unique SEQs at 97% SEQ similarity	44	22	23	4	13	16	2	101
Coverage calculated at 97% SEQ similarity (%)	99.01	99.98	97.65	.90	99.19	97.46	99.99	–
No. of unique SEQs for phylogenetic analysis[a]	40	26	21	7	11	14	2	82
Species identified (no. of genotypes)	*Gymnodinium* like sp. (2) *Karlodinium* like sp. (1) *Akashiwo* like sp. (1) *Amphidinium carterae* like sp. (1) *Amphidinium* like sp. (1) Unknown genotypes (34)	*Gymnodinium catenatum* like sp. (1) *Gymnodinium* like sp. (3) *Karlodinium* like sp. (1) *Akashiwo* like sp. (1) *Amphidinium* like sp. (1) Unknown genotypes (19)	*Gymnodinium* like sp. (1) *Karlodinium* like sp. (3) *Karlodinium micrum* like sp. (1) *Akashiwo* like sp. (2) Unknown genotypes (14)	*Gymnodinium* like sp. (2) Unknown genotypes (5)	*Gymnodinium* like sp. (1) *Akashiwo* like sp. (1) *Amphidinium* like sp. (3) Unknown genotypes (6)	*Gymnodinium* like sp. (2) *Amphidinium* like sp. (5) Unknown genotypes (7)	*Amphidinium* like sp. (2)	
SEQs identified as Prorocentrales	89	186	7	28	165	3	411	889
No. of unique SEQs at 99% SEQ similarity	23	43	2	5	34	2	57	126
Coverage calculated at 99% SEQ similarity (%)	89.88	91.39	–	92.85	94.54	66.66	94.64	–
No. of unique SEQs for phylogenetic analysis[a]	20	41	2	5	34	2	48	97

Table 2. *cont.*

Sample ID	B1	B2	B3	B4	C1	C2	E1	Total
Species identified (no. of genotypes)	*Prorocentrum cassubicum* (1) Unknown genotypes (19)	*Prorocentrum lima* (2) *Prorocentrum* like sp. (5) Unknown genotypes (34)	Unknown genotypes (2)	*Prorocentrum lima* (1) *Prorocentrum* like sp. (2) Unknown genotypes (2)	*Prorocentrum* like sp. (5) *Prorocentrum cassubicum* (1) Unknown genotypes (28)	Unknown genotypes (2)	*Prorocentrum lima* (1) *Prorocentrum* like sp. (4) Unknown genotypes (43)	
SEQs identified as Suessiales	111	22	143	96	66	16	54	508
No. of unique SEQs at 94% SEQ similarity	6	4	4	5	12	6	5	23
Coverage calculated at 94% SEQ similarity (%)	99.09	99.99	99.99	97.91	96.96	93.75	99.97	–
No. of unique SEQs for phylogenetic analysis[a]	6	4	4	4	12	6	5	23
Species identified (no. of genotypes)	*Symbiodinium* like sp. (3) *Polarella* like sp. (1) Unknown genotype (2)	*Symbiodinium* like sp. (1) *Polarella* like sp. (1) Unknown genotype (2)	*Symbiodinium* like sp. (1) *Polarella* like sp. (1) Unknown genotype (2)	*Symbiodinium* like sp. (3) Unknown genotype (1)	*Symbiodinium* like sp. (4) Unknown genotype (8)	*Symbiodinium* like sp. (3) Unknown genotype (3)	*Symbiodinium* like sp. (4) *Polarella* like sp. (1)	
SEQs identified as Dinophysiales	5	3	–	–	4	–	–	12
No. of unique SEQs at 98% SEQ similarity	1	1	–	–	1	–	–	3
SEQs that could not be identified	255	124	23	143	221	198	175	1139
No. of unique SEQs at 99% SEQ similarity	87	38	12	30	71	39	49	284
Coverage calculated at 99% SEQ similarity (%)	85.09	82.25	78.26	93.00	87.33	96.96	92.57	–

a. The number of unique sequences for phylogenetic analysis were calculated by pooling the sequences from all the seven samples belonging to a particular order together, and sequences that occurred only once in this whole dataset, across seven samples at specific percentage similarity, were removed. For example, after pooling all the gonyaulacalean sequences together from all the samples, if a sequence was present once in multiple samples like B1 and B2, it was not deleted; only if the sequence was present in sample B1 once and not present in any other sample was it deleted.

documented/sequenced species (Fig. 4). Sample coverage was high, indicating that more than 94% of Suessiales diversity was covered for each site (Table 2). Although most of the genera in Dinophysiales are planktonic (Hernández-Becerril *et al*., 2008) surprisingly, a total of 12 sequences were identified as belonging to this order, originating from the two Broome samples and one CKI sample (Table 2). At 98% sequence identity, 3 OTUs were obtained.

Diversity of dinoflagellates via ribosomal tag-encoded FLX 454-pyrosequencing (rTEFP)

Since all the samples analysed via cTEFP showed a large diversity of dinoflagellates, an attempt was made to analyse the samples via rTEFP to determine and identify more species in all the seven marine samples. A total of 10 529 18S rRNA gene sequences were obtained from all the samples. The 18S rRNA gene primers amplify all eukaryotes; therefore, only 133 sequences were identified as being highly similar to those of dinoflagellates via phylogenetic analysis (Fig. 5), revealing 18 distinct genera and 21 species from Broome, CKI and Exmouth (Table 3).

Discussion

Comparison of cTEFP, rTEFP and cob *cloning-based sequencing methods*

We initially generated *cob* clone libraries from our samples, which provided preliminary estimates of the level of diversity present. Between 25 and 35 clones per sample were sequenced, but this method did not show complete dinoflagellate diversity, as certain genera recorded via a microscope-based survey were not found (data not shown). Therefore, we designed the cTEFP assay to provide a more complete picture of dinoflagellate diversity in benthic environments. With an average of 1300 *cob* sequences per sample, the estimated coverage of the samples suggested that more than 90% of dinoflagellate *cob* sequence diversity was recovered via cTEFP (Table 2). All the genera that were observed using the light microscope were detected using cTEFP. This level of selective deep sequencing has never been achieved for dinoflagellates in benthic systems. We also used an SSU-based rTEFP analysis to survey eukaryotic diversity in the same samples. Only 2–3% of the sequences found using this method were of dinoflagellate origin, as compared

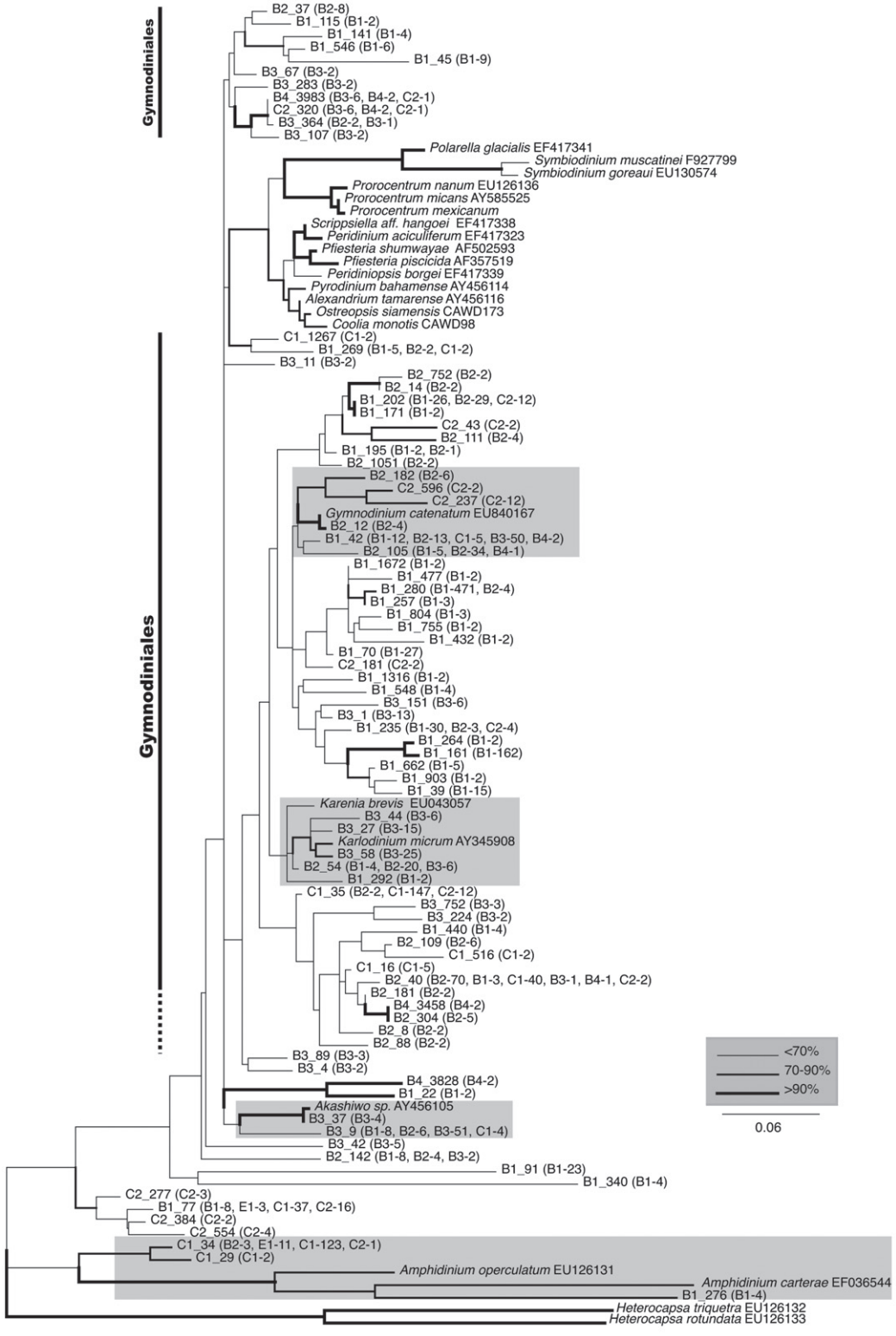

Fig. 3. Phylogenetic analysis using maximum likelihood of *cob* sequences identified as Gymnodiniales during cTEFP analysis. Bootstrap values were based on 500 replicates, and they are indicated by branch thickness; the thickest branches > 90%, medium-thick branches 70–90% and thin branches < 70%. Each unique OTU is represented by one representative sequence, for example B4_3983 (B3-6, B4-2, C2-1); B4_3983 is the name of the representative sequence, and this OTU/sequence was found six times in sample B3, 2 times in sample B4 and 1 time in sample C1. B1-B4 are samples from Broome, C1-C2 from CKI and E1 from Exmouth. Reference sequence names are followed by their accession numbers.

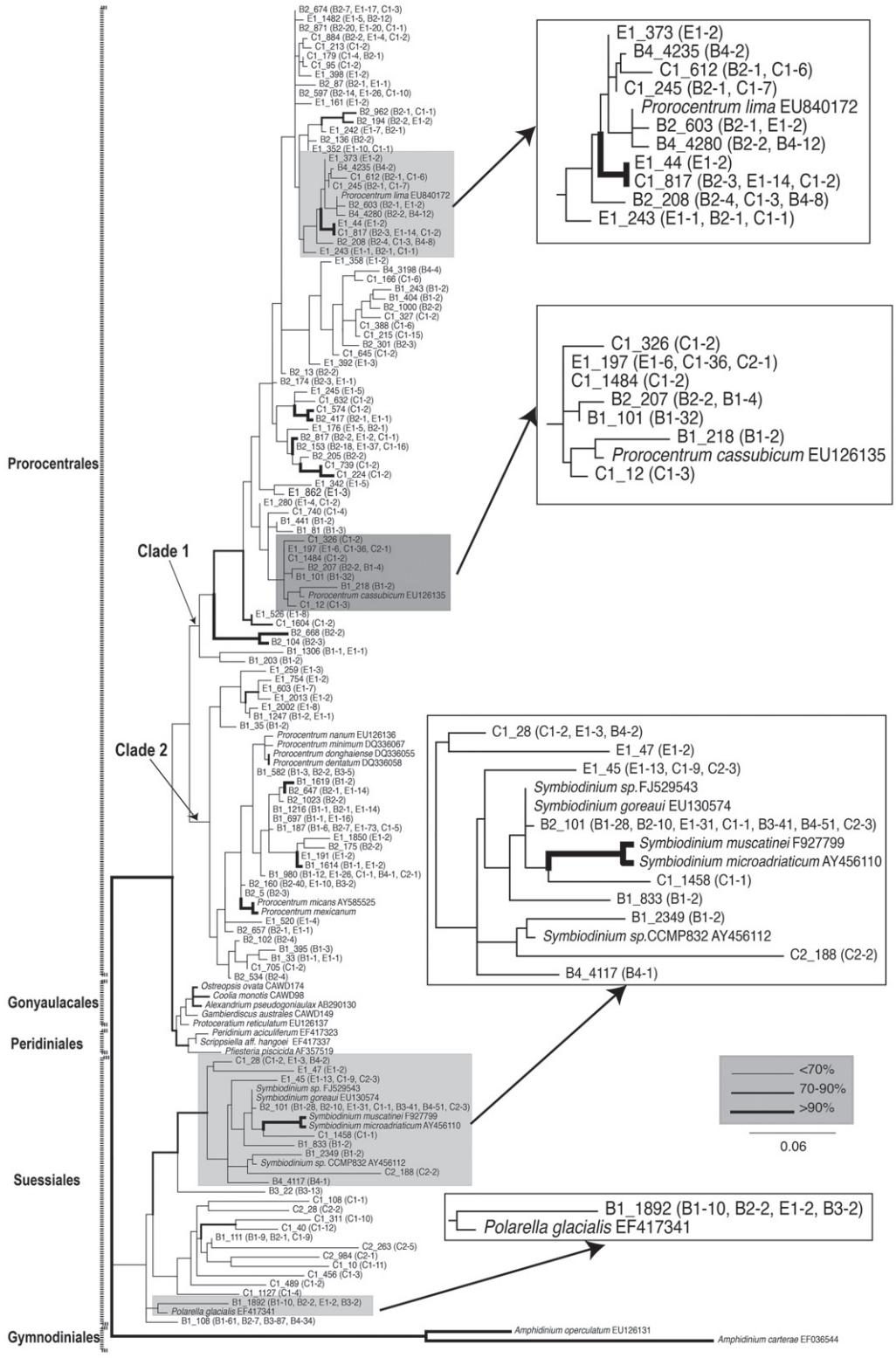

Fig. 4. Phylogenetic analysis using maximum likelihood of *cob* sequences identified as Suessiales and Prorocentrales during cTEFP analysis. Bootstrap values were based on 500 replicates, and they are indicated by branch thickness; the thickest branches > 90%, medium-thick branches 70–90% and thin branches < 70%. Each unique OTU is represented by one representative sequence, for example E1_197 (E1-6, C1-36, C2-1); E1_197 is the name of the representative sequence, and this OTU/sequence was found six times in sample E1, 36 times in sample C1 and 1 time in sample C2. B1-B4 are samples from Broome, C1-C2 from CKI and E1 from Exmouth. Reference sequence names are followed by their accession numbers.

Fig. 5. Phylogenetic analysis using maximum likelihood of SSU sequences obtained from rTEFP analysis. Bootstrap values were based on 500 replicates; the thickest branches denote bootstrap value of > 90%, medium-thick branches values of 70–90% and thin branches values of < 70%. B1-B4 are samples from Broome, C1-C2 from CKI and E1 from Exmouth. Reference sequence names are followed by their accession numbers.

Table 3. Data obtained during rTEFP analysis.

Sample ID	Total number of sequences obtained	Total number of unique sequences obtained	Total number of dinoflagellate sequences obtained	Total number of unique dinoflagellate sequences	Closest identifiable sister group in the phylogeny (number of different unique sequences identified phylogenetically)
B1	1319	220	4	3	*Bysmatrum* sp. (2), *Ostreopsis* sp. (1)
B2	1432	330	7	4	*Coolia* sp. (2), *Ostreopsis* sp. (1)
B3	1403	390	35	20	*Heterocapsa triquetra* (1), *Heterocapsa* sp. (2), *Duboscquodinium collinii* (1), *Gymnodinium* sp. (3), *Gymnodinium catenatum* (1), *Prorocentrum* sp. (2), *Polarella* sp. (3), *Protoceratium reticulatum* (1), *Protoceratium* sp. (2), *Protoperidinium pellucidum* (2), *Protoperidinium bipes* (2), *Ceratium* sp. (1), *Ceratium fusus* (1), *Amphidinium massartii* (1), *Alexandrium fraterculus* (1), *Alexandrium* sp. (1)
B4	2074	348	7	3	*Prorocentrum* sp. (2), *Ostreopsis* sp. (1)
C1	1598	302	20	12	*Prorocentrum* sp. (1), *Polarella* sp. (1), *Prorocentrum concavum* (1), *Galeidinium rugatum* (1), *Amphidinium massartii* (3), *Coolia* sp. (2), *Ostreopsis cf. ovata* (1), *Ostreopsis* sp. (1)
C2	1323	336	26	14	*Gymnodinium* sp. (1), *Prorocentrum* sp. (2), *Symbiodinium* sp. (1), *Polarella* sp. (1), *Amphidiniella sedentaria* (1), *Bysmatrum* sp. (2), *Amphidinium massartii* sp. (3), *Coolia* sp. (1)
E1	1380	319	34	8	*Symbiodinium* sp. (1), *Prorocentrum concavum* (1), *Galeidinium rugatum* (2), *Amphidinium massartii* (1), *Gambierdiscus carpenteri* (2), *Gymnodinium* sp. (1)[a]

a. Not shown in Fig. 5.

with 98.46% in cTEFP analysis (Tables 2 and 3). Phylogenetic analysis of *cob* genes revealed the Dinophyceae to be a monophyletic lineage, with the major orders within the class (Gonyaulacales, Gymnodiniales, Peridiniales, Prorocentrales and Suessiales) forming distinct and separate clades (Figs 1–4, and Figs S1–S3, supporting information data), with the exception of *Heterocapsa* and *Amphidinium*. As compared with phylogenetic analyses based on SSU rRNA, phylogenies based on *cob* are generally more highly supported at the deeper branches (Zhang *et al*., 2005), allowing us to assign unknown sequences to a particular order. However, *cob* has been shown to evolve at different rates in different lineages of dinoflagellates (Zhang *et al*., 2005; Lin *et al*., 2009); therefore, uniform distance thresholds for genus- or species-level discrimination across the Dinophyceae are not possible to determine and should not be used (Patterson, 1999; Orr *et al*., 2012; Murray *et al*., 2012b). We overcame this issue by calculating separate cut-off values for each order, combined with phylogenetic analysis for identification purposes. However, the removal of sequences that occurred only once in the dataset (cTEFP sequence analysis, see *Experimental procedures*) may have resulted in the elimination of sequences of rare species. With the limited availability of a reference *cob* database, more than 50% of the sequences in the samples could not be identified to a particular genus, and of those assigned a genus, species-level identification often could not be achieved, for example within the *Gambierdiscus*, *Ostreopsis* and *Coolia*. This problem will be overcome by enhancing the *cob* reference database. As the reference database for 18S rRNA genes is currently much more comprehensive than the *cob* reference database, despite the low recovery of dinoflagellate sequences in rTEFP analysis, almost all the sequences could be identified at the genus or species level, therefore making it an efficient tool for monitoring toxic species in local areas.

Identification of novel dinoflagellate diversity

Epibenthic dinoflagellate diversity from PNG, CKI, Exmouth and Broome has never been studied before, as previous studies mainly focused on sand dinoflagellates (Murray *et al*., 2006a,b; 2007). The total number of unique OTUs was 77–425 per sampling site, indicating similar diversity levels in each sample. This estimate is considerably more than the diversity that has been reported in studies based on light and electron microscopic observations, which indicated that 25–40 different species may be present in these habitats (Turquet *et al*., 1998).

During cTEFP analysis of sequences of the order Peridiniales, 243 unique OTUs were repeated more than once in the *cob* dataset. Similar results were found in other groups of dinoflagellates, with a final total of 604

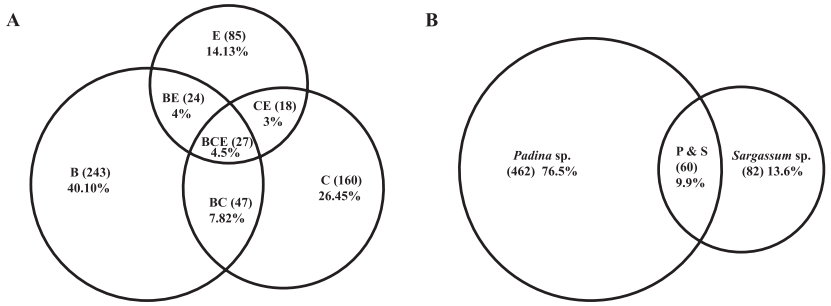

Fig. 6. (A) Venn diagram showing the distribution of unique OTUs obtained during cTEFP analysis from different geographical locations (B, Broome; C, CKI; E, Exmouth). The number in brackets represents the number of OTUs from that area. The percentage depicts the proportion of those OTUs in the pool of 604 unique OTUs from all the three regions. (B) Venn diagram showing the distribution of unique OTUs obtained during cTEFP analysis from different macroalgae where P & S stands for *Padina* sp. and *Saragassum* sp. The number in brackets represents the number of OTUs from each macroalgal species, and the percentage depicts the proportion of those OTUs in the pool of 604 unique OTUs from both types of macroalgal species.

unique OTUs across all the orders (Gonyaulacales – 156 OTUs, Peridiniales – 243 OTUs, Gymnodiniales – 82 OTUs, Prorocentrales – 97 OTUs, Suessiales – 23 OTUs, Dinophysiales – 3 OTUs, totalling to 604 OTUs all together). Such high molecular diversity may indicate the presence of morphologically highly similar but genetically diverse species (cryptic species) in benthic environments. In the past 15 years, cryptic species have been found in numerous dinoflagellate taxa, including *Scrippsiella trochoidea* (Zinssmeister *et al*., 2011), *Alexandrium minutum* (Lilly *et al*., 2005), *Amphidinium carterae*, *Amphidinium massartii* (Murray *et al*., 2012a), *Prorocentrum lima* (Nagahama *et al*., 2011) and *Peridinium limbaticum* (Kim *et al*., 2004). These results further confirm that genetic diversity of major 'barcoding' regions is generally higher than can be determined from the morphological variability of dinoflagellates.

During cTEFP data analysis, of those 604 OTUs, only 4.5% (27) were found in all the three regions, which included sequences similar to *Amphidinium*, *Ostreopsis*, *Symbiodinium* and *Prorocentrum lima* (Figs 2–4 and 6A, and Figs S1–S3, supporting information data). However, 80.7% (488) of the OTUs were unique to one region: Broome (243 OTUs), CKI (160 OTUs) and Exmouth (85 OTUs) (Figs 2–4 and 6A, and Figs S1–S3, supporting information data). A further 14.8% (89) unique OTUs were found in at least two of the three regions. This included sequences of *Gambierdiscus* from Exmouth and the CKI (Figs 2–4 and 6A, and Figs S1–S3, supporting information data). Also, as samples were collected from two different types of macroalgae, 76.5% (462 OTUs) were only found in samples collected from *Padina* sp., 13.6% (82 OTUs) were found in samples collected from *Sargassum* sp., and 9.9% (60 OTUs) were common to both macroalgal species (Fig. 6B).

In the literature, two models have been proposed to describe the biogeographic distribution of microbial eukaryotes (protists) (Finlay, 2002; Finlay and Tom, 2004; Foissner, 2009). The moderate endemism model suggests that many protists might have a restricted distribution and that endemism may be common (Foissner, 2009). The ubiquity model suggests that microbial eukaryotes are cosmopolitan and occur wherever their required habitats are realized, as a consequence of ubiquitous dispersal driven by huge population sizes, and the consequently low probability of local extinction (Finlay, 2002; Finlay and Tom, 2004). Our data revealed large differences in the diversity between sites, with a relatively small overlap, despite sampling highly similar microhabitats, associated with the epibenthic communities on *Padina* sp and *Sargassum* sp. These results are in agreement with the moderate endemism model. Further sampling at fine spatial and temporal scales is required to determine how these results relate to the wider biogeographic distribution of these organisms.

This study provides an initial report of a number of species in particular geographic regions. *Prorocentrum concavum* and *Ostreopsis* cf *ovata* were identified from Broome and Exmouth, the initial report of these species in Australia. The monotypic genus *Galeidinium* was newly described in 2005 (Tamura *et al*., 2005) and is so far only known from the western Pacific nation of Palau. This study increases its known distribution to CKI and Exmouth (Fig. 5). Our data also revealed a genetically unique lineage of *Gambierdiscus* in Exmouth, similar to *Gambierdiscus carpenteri* (Fig. 5). This site requires further investigation in order to fully characterize this likely new species. To date, the only known species of *Gambierdiscus* in Australia is *Gambierdiscus toxicus* (Murray, 2010). Recent reports have shown that each species of *Gambierdiscus* has a different toxin profile, with some being potent producers of ciguatoxins and/or maitotoxins, and others being totally nontoxic (Chinain *et al*., 2010; Litaker *et al*., 2010; Fraga *et al*., 2011). There

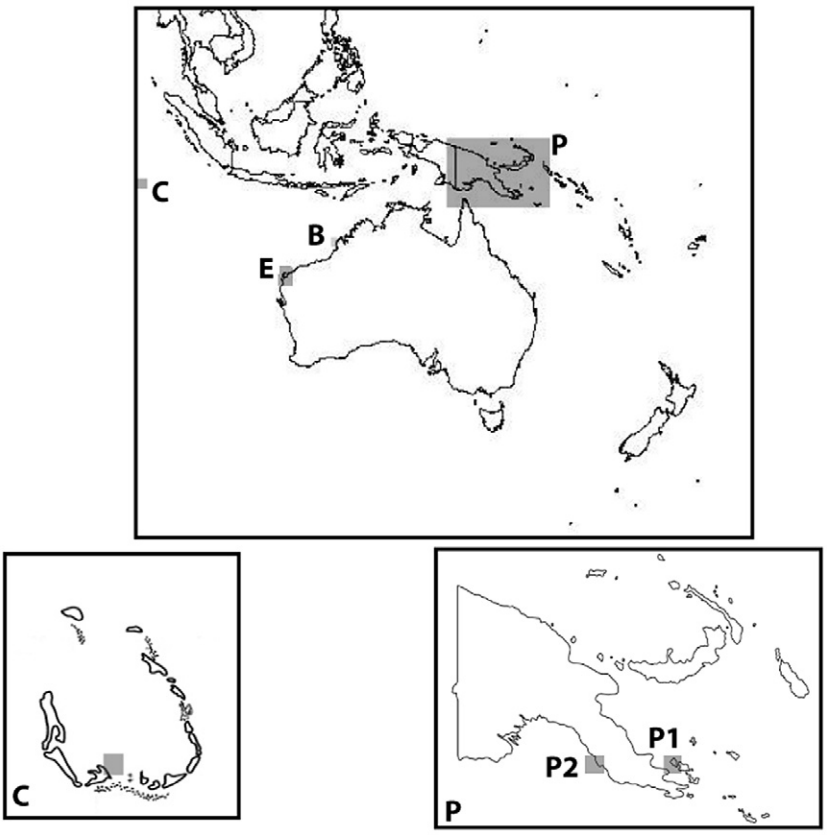

Fig. 7. Map showing location of different sampling sites: (B) Broome, Western Australia. (E) Exmouth, Western Australia. (C) Cocos (Keeling) Islands. (P1) Tawali, Milne Bay, Papua New Guinea. (P2) Lion Island, Port Moresby, Papua New Guinea. Shaded areas in the map indicate sampling sites.

are no published records of ciguatera fish poisoning from Exmouth and CKI; however, it has been reported in Queensland and numerous Pacific island nations (Lewis, 2006; Skinner *et al*., 2011). It is important to identify *Gambierdiscus* at the species level to assess the risk of ciguatera fish poisoning in these areas. Similarly, the presence of other potentially toxic species of the genera *Ostreopsis*, *Coolia*, *Amphidinium*, *Prorocentrum* and *Gymnodinium* in Exmouth and the CKI suggests the need for further investigations, for better management of toxin-related diseases. Samples from PNG revealed the presence of sequences similar to *Symbiodinium*, *Prorocentrum lima*, *O. siamensis*, *Polarella*, *Amphidinium* and *Pfiesteria* sp. (Fig. 1).

The retrieval of sequences related to the freshwater species *Peridinium aciculiferum* from some sites (Figs S1–S3, supporting information data) suggests that a related species may be found in marine habitats, as has been previously reported (Logares *et al*., 2007). Sequences similar to some generally planktonic genera, such as *Protoperidinium*, *Protoceratium*, *Karlodinium*, *Gonyaulax* and *Akashiwo*, were identified in Broome, CKI and Exmouth. Using the standard sampling method developed for epiphytic dinoflagellates (Bomber *et al*., 1989; Morton and Faust, 1997; Holmes, 1998; Rhodes *et al*., 2010), samples were rinsed in ambient seawater. Using filtered seawater to rinse macroalgae could eliminate the possibility that planktonic dinoflagellates are co-isolated.

In conclusion, we found that both cTEFP and rTEFP are promising tools for estimating the diversity of dinoflagellates in benthic systems, as well as for the monitoring of potentially toxic dinoflagellates. The specificity of this cTEFP technique for assessing dinoflagellates has allowed for the amplification of a much greater proportion of the total dinoflagellate population in a particular environmental sample. However, because the reference database of *cob* sequences from known species is currently small, rTEFP may be a better tool for identification and monitoring purposes. As the *cob* reference database grows, cTEFP may, however, prove to be a promising tool for future dinoflagellate diversity studies.

Experimental procedures

Sample collection

Samples were collected from eight different sites from four different tropical regions around Australasia (Fig. 7, Table 4). Except Broome, all sampling sites were coral reefs. As dinoflagellates are known to live epibenthically on numerous macroalgae, specimens of *Padina* sp. and *Sargassum* sp. were collected (Bomber *et al*., 1989; Holmes, 1998). These

Table 4. List of sampling sites, the macroalgae the samples were collected from, their water temperature at the time of sample collection and the type of analysis done with each sample.

Sample ID	Location (Latitude, Longitude)	Macroalgae	Water temperature (°C)	Diversity analysis		
				PCR-cloning of *cob*	cTEFP	rTEFP
B1	Cable Beach, Broome, WA, Australia (17°55'34.34'S, 122°12'31.00'E)	*Padina* sp.	25	Yes	Yes	Yes
B2	Gantheaume Point, Broome, WA, Australia (17°58'24.02'S, 122°10'43.46'E)	*Padina* sp.	25	Yes	Yes	Yes
B3	Town Beach, Broome, WA, Australia (17°58'23.31'S, 122°14'9.68'E)	*Padina* sp.	25	Yes	Yes	Yes
B4	Gantheaume Point, Broome, WA, Australia (17°58'24.02'S, 122°10'43.46'E)	*Sargassum* sp.	25	No	Yes	Yes
C1	South Lagoon, Cocos (Keeling) Islands (12°9'43.76'S, 96°50'59.15'E)	*Padina* sp.	–	Yes	Yes	Yes
C2	South Island, Cocos (Keeling) Islands (12°9'45.86'S, 96°53'38.36'E)	*Sargassum* sp.	–	No	Yes	Yes
E1	Town Beach, Ningaloo reef, Exmouth, WA, Australia (21°56'21.37'S, 114° 8'26.00'E)	*Padina* sp.	26	Yes	Yes	Yes
P1	Lion Island, Port Moresby, Papua New Guinea (9°32'9.11'S, 147°16'34.32'E)	*Padina* sp.	28	Yes	No	No
P2	Tawali, Milne Bay, Papua New Guinea (10°19'54.57'S, 150°30'42.83'E)	*Padina* sp.	28	Yes	No	No

genera of macroalgae were chosen as they were present at every site. The identity of the species of *Padina* and *Sargassum* were obtained following the next generation pyrosequencing. About 750 g of macroalgae was collected from approximately 1 m deep water at low tide and briefly placed in plastic bags containing 200–500 ml of ambient seawater. They were shaken vigorously for 5 min to detach the epiphytic dinoflagellates from the macroalgae. The seawater was collected in a separate container immediately. This seawater was passed through > 100 μm mesh filter to remove any remaining larger fauna and debris. From this, approximately 50 ml of each sample was filtered using a 3 μm filter (Merck Millipore®, Billerica, MA, USA). Cells were washed from the filters using about 5 ml RNAlater (Ambion®, Austin, TX, USA) for preservation and stored at 4°C.

Culturing

To enhance the *cob* reference database, the following dinoflagellate species were obtained from the Cawthron Institute Culture Collection (Nelson, New Zealand): *Gambierdiscus australes* CAWD149, isolated from the Cook Islands (21°13'S 159°46'W) (Rhodes *et al*., 2010); *O. siamensis* CAWD173, isolated from New Zealand (34°58'S 173°16'E); *O. ovata* CAWD174 isolated from Cook Islands (21°14'S 159°47'W); *Coolia monotis* CAWD98, isolated from New Zealand (41°11'S, 173°20'E); and *Prorocentrum compressum* CAWD30, isolated from New Zealand (38°10'S, 174°41'E). All cultures were grown at 25°C, in 12:12 h light : dark cycle with 60 μmol $m^{-2} s^{-1}$ light intensity in f/2 medium. Five hundred millilitres of cultures were harvested once the cells reached early stationary phase, and pellets were preserved in 2 ml RNAlater at 4°C for DNA extraction.

DNA extraction

A 50 mg pellet of the preserved samples/culture cell pellet was collected via centrifugation (12000 × *g*, 5 min) and used to extract total genomic DNA via FastDNA® Spin kit for Soil (MP Biomedicals, Solon, OH, USA). The manufacturer's protocol was followed, and samples were stored at –20°C until PCR amplification.

PCR, construction and screening of dinoflagellate-specific cob *clone libraries*

All PCR reactions were performed in 25 μl reaction volumes containing 5 μl *Taq* polymerase buffer, 2.5 μl GC Melt (Clontech, Mountain View, CA, USA), 0.2 mM deoxynucleotide triphosphates, 10 pmol each of the forward and reverse primers, between 10 and 100 ng genomic DNA, and 0.5 U of Advantage®-GC 2 *Taq* Polymerase (Clontech). PCR for *cob* was performed using primers dinocob4f and dinocob6r (Lin *et al*., 2009), and the following cycle conditions: initial denaturation at 96°C for 3 min followed by 32 cycles of denaturation at 94°C for 20 s, primer annealing at 55°C for 20 s and an extension at 68°C for 1 min followed by a final extension step at 68°C for 3 min. The PCR amplicons were cloned using TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. More than 25 clones were

sequenced for each sample. Sequencing was performed using the PRISM BigDye cycle sequencing system and a model 373 sequencer (Applied Biosystems, Foster City, CA, USA) at the Ramaciotti Centre for Gene Function Analysis (Sydney, Australia). Sequence data were trimmed and analysed using the GENEIOUS® software (Kearse *et al*., 2012), while identity/similarity values to other sequences were determined using BLAST, in conjugation with the National Center for Biotechnology Information (NIH, Bethesda, MD, USA). Finally, phylogenetic analysis combining reference sequences and sequences obtained during this study was carried out to determine the diversity of dinoflagellates in each sample.

Tag-encoded FLX 454-pyrosequencing

Cob tag-encoded FLX 454-pyrosequencing and 18S rRNA gene ribosomal tag-encoded FLX 454-pyrosequencing was performed using primers designed in previous studies. For cTEFP assay, the primer pair dinocob4f (AGCATTTATG GGTTATGTNTTACCTTT) and dinocob6r (ATTGGCATAGG AAATACCATTCAGG) (Lin *et al*., 2009), which amplifies a 441 bp bar-coded region of the *cob* gene, was tailored for pyrosequencing. This was done by adding a biotin and fusion linker sequence at the 5' end of the forward primer and a fusion linker, and a proprietary 12 bp barcode sequence at the 5' end of the reverse primer. Sequencing was performed at the 3' end of the product rather than the 5' end using the reverse primer. For rTEFP assay, the primer pair euk-A7F (AACCTGGTTGATCCTGCCAGT) (Medlin *et al*., 1988) and euk-570R (GCTATTGGAGCTGGAATTAC) (Weekers *et al*., 1994), which amplifies a 563 bp region of the 18S rRNA gene, was tailored for pyrosequencing. This was done by adding a fusion linker and the 12 bp barcode sequence at the 5' end of the euk-A7F forward primer, and a biotin and fusion linker sequence at the 5' end of the reverse primer. Sequencing was performed from the 5' end of the product using the forward primer. PCRs and FLX-titanium sequencing (Roche, Nutley, NJ, USA) were carried out at the Research and Testing Laboratory (Lubbock, TX, USA) according to their established protocols (Dowd *et al*., 2008).

cTEFP sequence analysis

Trimming and selection of high-quality reads was done using the MOTHUR software package (Schloss *et al*., 2009). Sequences with any ambiguous base calls and/or shorter than 380 bp in length were discarded. A local BLAST using the nr/nt database was performed, and any sequence with the closest BLAST match (based on sequence percentage similarity) to a non-dinoflagellate *cob* sequences was discarded. Unique sequences were obtained using the *unique.seqs* and *cluster.fragments* command in MOTHUR. Chimeras were identified using *chimera.uchime* in MOTHUR and eliminated from the unique sequences list. After removing the chimeras, all the sequences were combined with 46 reference sequences (Table S1, supporting information data), and phylogenetic analysis was carried out to identify and segregate sequences based on their taxonomic orders, which were Gonyaulacales, Peridiniales, Gymnodiniales, Prorocentrales, Suessiales and Dinophysiales. Sequences that could not be assigned to any dinoflagellate order based on phylogenetic analysis were marked as unknown sequences and not analysed any further. This divided the dataset from each sample into six-subsample datasets based on the six orders. Sequences from each subsample were aligned and trimmed using MAFFT v6.814b (Katoh *et al*., 2002) in Geneious®. Coverage and number of unique sequences, based on percentage similarity, was calculated using the following commands in MOTHUR: *unique.seqs*, *dist.seqs*, *cluster*, *summary.single (calc=nseqs-coverage-sobs)*. In MOTHUR, coverage calculator is based on Goods equation (Good, 1953).

To determine the number of species/genera within each sample, an accurate identification of sequences based on percentage similarity could not be performed due to the limited availability of sequences to construct a comprehensive *cob* database. No single universal distance threshold could be established to define a distinct species or genus across the whole Dinophyceae class, as the diversity in this gene was strikingly different for some genera. To identify and estimate diversity at the genus level accurately, the minimum distance between different genera within each order was calculated, based on the percentage similarity of reference sequences. These were calculated as Gonyaulacales (99%), Peridiniales (99%), Gymnodiniales (97%), Prorocentrales (distance between different species 99%) and Suessiales (96%). Due to sequencing error, inflation of diversity estimates at a 99% cut-off has been associated with pyrosequencing in prokaryotes (Kunin *et al*., 2010). To minimize this, sequences from all the seven samples belonging to a particular order were pooled together, and sequences that occurred only once in this whole dataset, across seven samples at specific percentage similarity, were removed. For example, after pooling all the gonyaulacalean sequences together from all the samples, if a sequence was present once in multiple samples, such as B1 and B2, it was not deleted. However, if the sequence was present in sample B1 once and not present in any other sample, it was deleted. A final phylogenetic analysis was performed using unique sequences from each order and 49 reference sequences (Table S1, supporting information data) to identify the genera present in each sample.

Very limited information is available on the copy number of the *cob* gene in the mitochondrial genomes of dinoflagellates (Waller and Jackson, 2009; Jackson *et al*., 2012). Recent studies have shown that only very slight variations exist between different copies of cytochrome c oxidase 1 genes within the same species (*Scrippsiella* and *Prorocentrum* – 0.2% pairwise distances, *Symbiodinium* – 1.3% pairwise distances) (Stern *et al*., 2010). The cut-off values of different OTUs used in the current study were 96–99% of sequence similarity, as described above. This range of percentage sequence similarity is higher than the diversity observed intra-genomically in the cytochrome c oxidase 1 gene.

rTEFP sequence analysis

Sequences obtained via rTEFP were trimmed, and unique sequences were selected using the following commands in MOTHUR: *trim.seqs, unique.seqs* and *cluster.fragments*. On the selected sequences, a local BLAST using the nr/nt database was performed, and sequences with the closest BLAST match to dinoflagellate SSU sequences were selected (based on

sequence percentage similarity). These sequences were combined with 61 reference sequences (Table S2, supporting information data), and phylogenetic analysis was carried out for identification purposes.

Phylogenetic analysis

All steps were performed in Geneious® software (Kearse *et al.*, 2012). Sequences from different datasets were aligned using MAFFT v6.814b (Katoh *et al.*, 2002). Alignments were trimmed manually to ensure they spanned the same *cob/18 s* rRNA region. After aligning the sequences, the best substitution model was determined using MODELTEST (Posada and Crandall, 1998), and a maximum likelihood phylogenetic analysis was carried out using the programme PHYML (Guindon *et al.*, 2010), with 500 bootstraps.

Acknowledgements

All co-authors agree with the contents of the manuscript, and there is no conflict of interest to be declared. The authors acknowledge technical support for computing infrastructure and software development from Intersect, and in particular assistance from Joachim Mai. Financial support to conduct this study was provided by the Australian Biological Resources Study and the Australian Research Council. The authors thank the staff of the Research and Testing Laboratory, Texas, for help in designing and conducting the pyrosequencing assays, and the Ramaciotti Centre for Gene Function Analysis for conducting the Sanger sequencing. The authors also acknowledge Prof. Manfred Lenzen from the University of Sydney for collecting samples from Cocos (Keeling) Islands, and Prof. Tertulohi Matainaho and Steven Ipunesa from the University of Papua New Guinea in assisting the authors to collect samples from PNG. We thank Lesley L. Rhodes from Cawthron Institute's Culture Collection of Micro-algae for the provision of cultures. We thank the two anonymous reviewers for their constructive feedback. This is contribution number 109 from the Sydney Institute of Marine Science.

References

Ballantine, D.L., Tosteson, T.R., and Bardales, A.T. (1988) Population dynamics and toxicity of natural populations of benthic dinoflagellates in southwestern Puerto Rico. *J Exp Mar Bio Ecol* **119:** 201–212.

Bomber, J.W., Rubio, M.G., and Norris, D.R. (1989) Epiphytism of dinoflagellates associated with the disease ciguatera: substrate specificity and nutrition. *Phycologia* **28:** 360–368.

Brown, M.V., Philip, G.K., Bunge, J.A., Smith, M.C., Bissett, A., Lauro, F.M., *et al.* (2009) Microbial community structure in the North Pacific ocean. *ISME J* **3:** 1374–1386.

Carlson, R., and Trindall, D. (1985) Distribution and periodicity of toxic dinoflagellates in the Virgin Islands. In *Toxic Dinoflagellates*. Andersen, D., White, A., and Baden, D. (eds). New York, NY, USA: Elsevier Science, pp. 171–176.

Chinain, M., Darius, H.T., Ung, A., Cruchet, P., Wang, Z., Ponton, D., *et al.* (2010) Growth and toxin production in the ciguatera-causing dinoflagellate Gambierdiscus polynesiensis (Dinophyceae) in culture. *Toxicon* **56:** 739–750.

Cuvelier, M.L., Allen, A.E., Monier, A., McCrow, J.P., Messié, M., Tringe, S.G., *et al.* (2010) Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc Natl Acad Sci U S A* **107:** 14679–14684.

Deeds, J.R., and Schwartz, M.D. (2010) Human risk associated with palytoxin exposure. *Toxicon* **56:** 150–162.

Dowd, S., Sun, Y., Secor, P., Rhoads, D., Wolcott, B., James, G., *et al.* (2008) Survey of bacterial diversity in chronic wounds using Pyrosequencing, DGGE, and full ribosome shotgun sequencing. *BMC Microbiol* **8:** 43.

Edgcomb, V., Orsi, W., Bunge, J., Jeon, S., Christen, R., Leslin, C., *et al.* (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME J* **5:** 1344–1356.

Faust, M.A., and Balech, E. (1993) A further SEM study of marine benthic dinoflagellates from a mangrove island, Twin Cays, Belize including Plagiodinium belizeanum gen. et. so.nov.1. *J Phycol* **29:** 826–832.

Finlay, B.J. (2002) Global dispersal of free-living microbial eukaryote species. *Science* **296:** 1061–1063.

Finlay, B.J., and Tom, F. (2004) Cosmopolitan metapopulations of free-living microbial eukaryotes. *Protist* **155:** 237–244.

Fleming, L.E., Baden, D.G., Bean, J.A., Weisman, R., and Blythe, D.G. (1998) Seafood toxin diseases: issues in epidemiology and community outreach. In *Harmful Algae*. Reguera, B., Blanco, J., Fernandez, M.L., and Wyatt, T. (eds). Xunta de Galicia, Spain: Xunta de Galicia and Intergovernmental Oceanographic Commission of UNESCO, pp. 245–248.

Foissner, W. (2009) Protist diversity and distribution: some basic considerations. *Protist Divers Geogr Distrib* **8:** 1–8.

Fraga, S., Rodriguez, F., Caillaud, A., Diogene, J., Raho, N., and Zapata, M. (2011) Gambierdiscus excentricus sp. nov. (Dinophyceae), a benthic toxic dinoflagellate from the Canary Islands (NE Atlantic Ocean). *Harmful Algae* **11:** 10–22.

Fukuyo, Y. (1981) Taxonomical study on benthic dinoflagellates collected in coral reefs. *Bull Jpn Soc Sci Fish* **47:** 967–978.

Good, I.J. (1953) The population frequencies of species and the estimation of population parameters. *Biometrika* **40:** 237–264.

Grzebyk, D., Sako, Y., and Berland, B. (1998) Phylogenetic analysis of nine species of Prorocentrum (dinophyceae) inferred from 18s ribosomal dna sequences, morphological comparisons, and description of Prorocentrum panamensis, sp. nov. *J Phycol* **34:** 1055–1068.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59:** 307–321.

Hernández-Becerril, D.U., Ceballos-Corona, J.G.A., Esqueda-Lara, K., Tovar-Salazar, M.A., and León-Álvarez, D. (2008) Marine planktonic dinoflagellates of the order Dinophysiales (Dinophyta) from coasts of the tropical

Mexican Pacific, including two new species of the genus Amphisolenia. *J Mar Biolog Assoc U K* **88:** 1–15.
Holmes, M.J. (1998) Gambierdiscus yasumotoi sp. nov. (dinophyceae), a toxic benthic dinoflagellate from south-eastern asia. *J Phycol* **34:** 661–668.
Hoppenrath, M., and Leander, B.S. (2008) Morphology and molecular phylogeny of a new marine sand-dwelling Prorocentrum species, P. tsawwassenense (Dinophyceae, Prorocentrales), from British Columbia, Canada. *J Phycol* **44:** 451–466.
Jackson, C.J., Gornik, S.G., and Waller, R.F. (2012) The mitochondrial genome and transcriptome of the basal dinoflagellate hematodinium sp.: character evolution within the highly derived mitochondrial genomes of dinoflagellates. *Genome Biol Evol* **4:** 59–72.
Kalaitzis, J.A., Chau, R., Kohli, G.S., Murray, S.A., and Neilan, B.A. (2010) Biosynthesis of toxic naturally-occurring seafood contaminants. *Toxicon* **56:** 244–258.
Katoh, K., Misawa, K., and Kuma Ki, M.T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30:** 3059–3066.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., *et al*. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28:** 1647–1649.
Kim, E., Wilcox, L., Graham, L., and Graham, J. (2004) Genetically distinct populations of the dinoflagellate Peridinium limbatum in neighboring Northern Wisconsin lakes. *Microb Ecol* **48:** 521–527.
Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12:** 118–123.
Lewis, R.J. (2006) Ciguatera: Australian perspectives on a global problem. *Toxicon* **48:** 799–809.
Lilly, E.L., Halanych, K.M., and Anderson, D.M. (2005) Phylogeny, biogeography, and species boundaries within the Alexandrium minutum group. *Harmful Algae* **4:** 1004–1020.
Lin, S., Zhang, H., Hou, Y., Miranda, L., and Bhattacharya, D. (2006) Development of a dinoflagellate-oriented PCR primer set leads to detection of picoplanktonic dinoflagellates from Long Island sound. *Appl Environ Microbiol* **72:** 5626–5630.
Lin, S., Zhang, H., Hou, Y., Zhuang, Y., and Miranda, L. (2009) High-level diversity of dinoflagellates in the natural environment, revealed by assessment of mitochondrial cox1 and cob genes for dinoflagellate DNA barcoding. *Appl Environ Microbiol* **75:** 1279–1290.
Litaker, R.W., Vandersea, M.W., Faust, M.A., Kibler, S.R., Nau, A.W., Holland, W.C., *et al*. (2010) Global distribution of ciguatera causing dinoflagellates in the genus Gambierdiscus. *Toxicon* **56:** 711–730.
Logares, R., Rengefors, K., Kremp, A., Shalchian-Tabrizi, K., Boltovskoy, A., Tengs, T., *et al*. (2007) Phenotypically different microalgal morphospecies with identical ribosomal DNA: a case of rapid adaptive evolution? *Microb Ecol* **53:** 549–561.
López-García, P., Philippe, H., Gail, F., and Moreira, D. (2003) Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc Natl Acad Sci U S A* **100:** 697–702.
MacKenzie, L., Beuzenberg, V., Holland, P., McNabb, P., Suzuki, T., and Selwood, A. (2005) Pectenotoxin and okadaic acid-based toxin profiles in Dinophysis acuta and Dinophysis acuminata from New Zealand. *Harmful Algae* **4:** 75–85.
Medlin, L., Elwood, H.J., Stickel, S., and Sogin, M.L. (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **71:** 491–499.
Moon-van der Staay, S.Y., De Wachter, R., and Vaulot, D. (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409:** 607–610.
Morton, S., and Faust, M. (1997) Survey of toxic epiphytic dinoflagellates from the Belizean Barrier Reef Ecosystem. *Bull Mar Sci* **61:** 899–906.
Murakami, Y., Oshima, Y., and Yasumoto, T. (1982) Identification of okadaic acid as a toxic component of a marine dinoflagellate Prorocentrum lima. *Bull Jpn Soc Sci Fish* **48:** 69–72.
Murray, S. (2010) Benthic Dinoflagellates. In *Algae of Australia: Phytoplankton of Temperate Coastal Waters*. Hallegraeff, G., Bolch, C., Hill, D., Jameson, I., LeRoi, J., McMinn, A., *et al* (eds). Canberra, Australia: CSIRO Publishing, pp. 213–259.
Murray, S., Flø Jørgensen, M., Ho, S.Y.W., Patterson, D.J., and Jermiin, L.S. (2005) Improving the analysis of dinoflagellate phylogeny based on rDNA. *Protist* **156:** 269–286.
Murray, S., Hoppenrath, M., Larsen, J., and Patterson, D.J. (2006a) Bysmatrum teres sp. nov., a new sand-dwelling dinoflagellate from north-western Australia. *Phycologia* **45:** 161–167.
Murray, S., Hoppenrath, M., Preisfeld, A., Larsen, J., Yoshimatsu, S., Toriumi, S., *et al*. (2006b) Phylogenetics of Rhinodinium broomeense gen. et. nov., a peridinioid, sandodwelling dinoflagellate (Dinophyceae)1. *J Phycol* **42:** 934–942.
Murray, S., De Salas, M., Luong-Van, J., and Hallegraeff, G. (2007) Phylogenetic study of Gymnodinium dorsalisulcum comb. nov. from tropical Australian coastal waters (Dinophyceae). *Phycological Res* **55:** 176–184.
Murray, S., Ip, C.L.C., Moore, R., Nagahama, Y., and Fukuyo, Y. (2009) Are prorocentroid dinoflagellates monophyletic? A study of 25 species based on nuclear and mitochondrial genes. *Protist* **160:** 245–264.
Murray, S.A., Garby, T., Hoppenrath, M., and Neilan, B.A. (2012a) Genetic diversity, morphological uniformity and polyketide production in dinoflagellates (*Amphidinium*, dinoflagellata). *PLoS ONE* **7:** e38253.
Murray, S.A., Patterson, D.J., and Thesson, A. (2012b) Transcriptomics and microbial eukaryotic diversity: a way forward. *Trends Ecol Evol* **27:** 651–652.
Nagahama, Y., Murray, S., Tomaru, A., and Fukuyo, Y. (2011) Species boundaries in the toxic dinoflagellate Prorocentrum lima (Dinophyceae, Prorocentrales), based on morphological and phylogenetic characters. *J Phycol* **47:** 178–189.
Okolodkov, Y.B., Campos-Bautista, G., Gárate-Lizárraga, I., González-González, J.A.G., Hoppenrath, M., and Arenas, V. (2007) Seasonal changes of benthic and epiphytic

dinoflagellates in the Veracruz reef zone, Gulf of Mexico. *Aquat Microb Ecol* **47:** 223–237.
Orr, R.J.S., Murray, S.A., Stüken, A., Rhodes, L., and Jakobsen, K.S. (2012) When naked became armored: an eight-gene phylogeny reveals monophyletic origin of theca in dinoflagellates. *PLoS ONE* **7:** e50004.
Patterson, D.J. (1999) The diversity of eukaryotes. *Am Nat* **154:** S96–S124.
Pearse, I., Marshall, J., and Hallegraeff, G. (2001) Toxic epiphytic dinoflagellates from east coast Tasmania, Australia. In *Harmful Algal Blooms 2000*. Hallegraeff, G., Blackburn, S., Bolch, C., and Lewis, R.J. (eds). Hobart, Tasmania: Intergovernmental Oceanographic Commission of UNESCO, pp. 54–57.
Posada, D., and Crandall, K.A. (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.
Potvin, M., and Lovejoy, C. (2009) PCR-based diversity estimates of artificial and environmental 18S rRNA gene libraries. *J Eukaryot Microbiol* **56:** 174–181.
Quilliam, M., and Wright, J. (1995) Methods for diarrhetic shellfish poisons. In *Manual on Harmful Marine Microalgae*. Hallegraeff, G., Andersen, D., and Cembella, A. (eds). Paris, France: IOC-UNESCO, pp. 95–111.
Rhodes, L.L., Smith, K.F., Munday, R., Selwood, A.I., McNabb, P.S., Holland, P.T., *et al*. (2010) Toxic dinoflagellates (Dinophyceae) from Rarotonga, Cook Islands. *Toxicon* **56:** 751–758.
Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al*. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.
Skinner, M.P., Brewer, T.D., Johnstone, R., Fleming, L.E., and Lewis, R.J. (2011) Ciguatera fish poisoning in the Pacific Islands (1998 to 2008). *PLoS Negl Trop Dis* **5:** e1416.
Smith, K.F., Rhodes, L.L., Suda, S., and Selwood, A.I. (2011) A dinoflagellate producer of pinnatoxin G, isolated from sub-tropical Japanese waters. *Harmful Algae* **10:** 702–705.
Stern, R.F., Horak, A., Andrew, R.L., Coffroth, M.-A., Andersen, R.A., Küpper, F.C., *et al*. (2010) Environmental barcoding reveals massive dinoflagellate diversity in marine environments. *PLoS ONE* **5:** e13991.
Stern, R.F., Andersen, R.A., Jameson, I., Küpper, F.C., Coffroth, M.-A., Vaulot, D., *et al*. (2012) Evaluating the ribosomal internal transcribed spacer (ITS) as a candidate dinoflagellate barcode marker. *PLoS ONE* **7:** e42780.
Steven, B., McCann, S., and Ward, N.L. (2012) Pyrosequencing of plastid 23S rRNA genes reveals diverse and dynamic cyanobacterial and algal populations in two eutrophic lakes. *FEMS Microbiol Ecol* **82:** 607–615.
Tamura, M., Shimada, S., and Horiguchi, T. (2005) Galeidiniium rugatum gen. et sp. nov. (Dinophyceae), a new coccoid dinoflagellate with a diatom endosymbiont. *J Phycol* **41:** 658–671.
Taylor, F. (1979) The toxigenic gonyaulacoid dinoflagellates. In *Toxic Dinoflagellate Blooms*. Taylor, D., and Seliger, H. (eds). New York, NY, USA: Elsevier/North-Holland, pp. 47–56.
Taylor, F., Hoppenrath, M., and Saldarriaga, J. (2008) Dinoflagellate diversity and distribution. *Biodivers Conserv* **17:** 407–418.
Turquet, J., Quad, J., Couté, A., and Faust, M. (1998) Assemblage of benthic dinoflagellates and monitoring of harmful species in Reunion Island, SW Indian Ocean, 1993–1996. In *Harmful Algae*. Reguera, B., Blanco, J., Fernández, M., and Wyatt, T. (eds). Xunta de Galicia, Vigo, Spain: Intergovernmental Oceanographic Commission of UNESCO, pp. 44–77.
Usami, M., Satake, M., Ishida, S., Inoue, A., Kan, Y., and Yasumoto, T. (1995) Palytoxin analogs from the dinoflagellate Ostreopsis siamensis. *J Am Chem Soc* **117:** 5389–5390.
Waller, R.F., and Jackson, C.J. (2009) Dinoflagellate mitochondrial genomes: stretching the rules of molecular biology. *Bioessays* **31:** 237–245.
Weekers, P.H., Gast, R.J., Fuerst, P.A., and Byers, T.J. (1994) Sequence variations in small-subunit ribosomal RNAs of Hartmannella vermiformis and their phylogenetic implications. *Mol Biol Evol* **11:** 684–690.
Zhang, H., Bhattacharya, D., and Lin, S. (2005) Phylogeny of dinoflagellates based on mitochondrial cytochrome b and nuclear small subunit rDNA sequence comparisons. *J Phycol* **41:** 411–420.
Zinssmeister, C., Soehner, S., Facher, E., Kirsch, M., Meier, K.J.S., and Gottschling, M. (2011) Catch me if you can: the taxonomic identity of Scrippsiella trochoidea (F.Stein) A.R.Loebl. (Thoracosphaeraceae, Dinophyceae). *System Biodivers* **9:** 145–157.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Phylogenetic analysis using maximum likelihood of *cob* sequences identified as Peridiniales during cTEFP analysis. Bootstrap values were based on 500 replicates; the thickest branches denote bootstrap value of > 90%, medium-thick branches values of 70–90% and thin branches values of < 70%. Each unique OTU is represented by one representative sequence, for example B1_547 (B1-6, B2-2, B3-22, C2-2); B1_547 is the name of the representative sequence, and this OTU/sequence was found 6 times in sample B1, 2 times in sample B2, 22 times in sample B3 and 2 times in sample C2. B1-B4 are samples from Broome, C1-C2 from CKI and E1 from Exmouth. Reference sequence names are followed by their accession numbers.
Fig. S2. Phylogenetic analysis using maximum likelihood of *cob* sequences identified as Peridiniales during cTEFP analysis. Bootstrap values were based on 500 replicates; the thickest branches denote bootstrap value of > 90%, medium-thick branches values of 70–90% and thin branches values of < 70%. Each unique OTU is represented by one representative sequence, for example C1_174 (C1-6, E1-1); C1_174 is the name of the representative sequence, and this OTU/sequence was found 6 times in sample C1 and once in sample E1. B1-B4 are samples from Broome, C1-C2 from CKI and E1 from Exmouth. Reference sequence names are followed by their accession numbers.
Fig. S3. Phylogenetic analysis using maximum likelihood of *cob* sequences identified as Peridiniales during cTEFP analysis. Bootstrap values were based on 500 replicates; the thick-

est branches denote bootstrap value of > 90%, medium-thick branches values of 70–90% and thin branches values of < 70%. Each unique OTU is represented by one representative sequence, for example E1_344 (E1-46, C1-2); E1_344 is the name of the representative sequence, and this OTU/sequence was found 46 times in sample E1 and 2 times in sample C1. B1-B4 are samples from Broome, C1-C2 from CKI and E1 from Exmouth. Reference sequence names are followed by their accession numbers.

Table S1. Cytochrome b reference sequences used in this study.

Table S2. 18 s reference sequences used in this study.