Running head
Luciferin binding protein in dinoflagellates

Diversity of the Luciferin Binding Protein Gene in Bioluminescent Dinoflagellates - Insights from a New Gene in *Noctiluca scintillans* and Sequences from Gonyaulacoid Genera

Martha Valiadi [a,1] and M. Debora Iglesias-Rodriguez [a,b]

a University of Southampton, Ocean and Earth Science, Waterfront Campus, Southampton, U.K.

b University of California Santa Barbara, Department for Ecology, Evolution and Marine Biology, Santa Barbara, California, U.S.A.

1 Present address: Max Planck Institute for Evolutionary Biology, Department of Evolutionary Ecology, August-Thienemann-Strasse 2, 24306 Pl□n, Germany.

Correspondence
M. Valiadi, Max Planck Institute for Evolutionary Biology, August-Thienemann-Strasse 2, 24306 Pl□n, Germany, Tel: +49-4522-763277; Fax: +49-4522-763310; email: Martha.valiadi@gmail.com and valiadi@evolbio.mpg.de

ABSTRACT

Dinoflagellate bioluminescence systems operate with or without a luciferin binding protein, representing two distinct modes of light production. However, the distribution, diversity and evolution of the luciferin binding protein gene within bioluminescent dinoflagellates are not well known. We used PCR to detect and partially sequence this gene from the heterotrophic dinoflagellate *Noctiluca scintillans* and a group of ecologically important gonyaulacoid species. We report an additional luciferin binding protein gene in *N. scintillans* which is not attached to luciferase, further to its typical combined bioluminescence gene. This supports the hypothesis that a profound re-organization of the bioluminescence system has taken place in this organism. We also show that the luciferin binding protein gene is present in the genera *Ceratocorys*, *Gonyaulax* and *Protoceratium,* and is prevalent in bioluminescent species of *Alexandrium*. Therefore, this gene is an integral component of the standard molecular bioluminescence machinery in dinoflagellates. Nucleotide sequences showed high within-strain variation among gene copies, revealing a highly diverse gene family comprising multiple gene types in some organisms. Phylogenetic analyses showed that, in some species, the evolution of the luciferin binding protein gene was different from the organism's general phylogenies, highlighting the complex evolutionary history of dinoflagellate bioluminescence systems.

Keywords
Bioluminescence; gene evolution; luciferase; luciferin; PCR

Dinoflagellates are the most ubiquitous protists that are capable of bioluminescence. The ecological function of dinoflagellate bioluminescence is thought to be defence against predators (Abrahams and Townsend 1993, Buskey et al. 1983, Fleisher and Case 1995) and could therefore play a significant role in ecosystem structure. Bioluminescence in dinoflagellates is brought about by the reaction of a luciferin substrate and the enzyme luciferase (LCF) within the scintillons, which are the organelles that contain the components for the bioluminescence reaction (Desa and Hastings 1968, Johnson et al. 1985, Nicolas et al. 1987). In some species, the reaction also consists of a luciferin-binding protein (LBP) (Knaust et al. 1998, Schmitter et al. 1976). The function of LBP is thought to be to protect luciferin from autoxidation because it is naturally highly reactive with oxygen (Morse et al. 1989, Nakamura et al. 1989). A flash of light is triggered by an action potential along the vacuole membrane, induced by mechanical agitation of the cell, with the likely involvement of a voltage gated proton channel such as that recently discovered in a dinoflagellate (Smith et al. 2011). This leads to a reduction in pH within the scintillons, which activates LCF and triggers the LBP to release the luciferin, making it available for oxidation by LCF (Fogel and Hastings 1971, Hastings 2009).

The composition of bioluminescence systems in diverse dinoflagellate taxa is not well studied. This is especially true for luciferin binding protein genes (*lbp*), whose distribution and evolution has received a lot less attention than luciferase genes (*lcf*). However, the potential significance of *lbp* to the bioluminescence system is highlighted in transcriptional studies on several *Alexandrium* species and *Lingulodinium polyedrum*, where *lbp* has been found to either dominate or be very abundant in the transcript pool (Erdner and Anderson 2006, Hackett et al. 2005, Jaeckisch et al. 2011, Tanikawa et al. 2004, Toulza et al. 2010, Uribe et al. 2008). These studies have also provided the first sequences of *lbp* from *Alexandrium* species.

The dinoflagellate species *L. polyedrum* (formerly *Gonyaulax polyedra*) has been the main model organism for studies of LBP, representing the first organism from which this protein was first isolated, and *lbp* was first cloned and fully sequenced (Lee et al. 1993, Machabée et al. 1994, Morse et al. 1989). The *lbp* of *L. polyedrum* is taken as representative for the Gonyaulacales, the taxonomic order in which the ability to produce light is predominant (Poupin et al. 1999, Valiadi et al. 2012) and which includes several ecologically important bloom forming and/or toxic species. The *lbp* gene family of *L. polyedrum* is composed of two gene types, *lbp*α and *lbp*β, which share 86% sequence identity and encode for two protein isoforms that are expressed in equal amounts (Machabée et al. 1994, Morse et al. 1989). Additionally, each gene type follows the typical dinoflagellate genomic organization and is present in several non-identical copies arranged in tandem within the genome (Machabée et al. 1994, Tanikawa et al. 2004). Strikingly, LBP can make up 1% of the total protein of *L. polyedrum* (Morse et al. 1989).

Sequencing of a bioluminescence gene in *Noctiluca scintillans* revealed that *lbp* in dinoflagellates comes in diverse forms as a result of significant evolutionary events such as gene fission or fusion (Liu and Hastings 2007). Discovery of *lbp* in this organism also showed that bioluminescent systems utilising LBP are common in dinoflagellates. The bioluminescence gene of *N. scintillans* is unusual as *lcf* and *lbp* are found as two domains in one gene, while they are normally separate genes in all photosynthetic species studied to date (Liu and Hastings 2007). The origin of the hybrid structure of this *lcf/lbp*, has been suggested to be either by fusion of the two genes of photosynthetic species in *N. scintillans* or, conversely, fission of the *N. scintillans* gene into the photosynthetic species (Liu and Hastings 2007). Evidence suggests that the bioluminescence system of *N. scintillans* indeed represents the ancestral type, as its lack of typical dinoflagellate mitochondrial mRNA editing (Zhang and Lin 2008), a dinokaryotic nucleus restricted to the gamete stage (Taylor 2004)

and phylogenetic analyses of ribosomal genes (Fukuda and Endoh 2008, Gomez et al. 2010a, Ki 2010) point towards it being a primitive dinoflagellate species. Phylogenetic analyses of protein coding genes have contradicted these conclusions (Orr et al. 2012), yet still place *N. scintillans* in an ancestral position relative to other known bioluminescent dinoflagellates. Therefore, the assumption that its bioluminescence genes represent the ancestral type is still supported.

An emerging body of work shows substantial evidence that *lbp,* with its complex evolutionary history, is an important part of dinoflagellate bioluminescence systems. In fact, the only genus known to lack this protein is *Pyrocystis* (Knaust et al. 1998, Schmitter et al. 1976). Since bioluminescence is possible with and without LBP, this could represent two distinct mechanisms for the production of light by dinoflagellates, for which the origin is unknown. Therefore, our aim was to explore the composition of dinoflagellate bioluminescence systems, by detecting which organisms contain *lbp*. We amplified *lbp* from *N. scintillans* and several ecologically important, including toxic bloom-forming, gonyaulacoid dinoflagellates species. We report a new *lbp* in *N. scintillans* that is not attached to *lcf* and new *lbp* sequences obtained from four gonyaulacoid genera. Our results provide new significant insight into the distribution and diversity of *lbp* in bioluminescent dinoflagellates as well as the evolution of the structure of *lbp* and *lcf* as separate or combined genes, thus, contributing to our understanding of light production systems in marine plankton.

MATERIALS AND METHODS

Dinoflagellate strains, bioluminescence detection and nucleic acid extractions

We used a collection of dinoflagellate gDNA that was generated over several years from our own cultures and from donations by colleagues, following the methods described in Valiadi *et al.* (2012). Details of the strains used here are provided in Table 1. Our analyses were only based on genomic DNA (gDNA) from these strains because some had lost their bioluminescence at the time of sampling and thus cDNA would be unsuitable for presence/absence tests of *lbp*. In addition to our DNA collection, we obtained a culture of *N. scintillans* for bioluminescence testing and nucleic acid extractions (see Table 1). For nucleic acid extractions, cells were harvested at stationary phase when cultures were dense and the abundance of prey cells (*Dunaliella* sp.) was minimal. Cells were collected by gently filtering approximately 400 mL of culture onto 25 mm diameter, 5 µm pore size Nuclepore polycarbonate membrane filters (Whatman, U.K.). They were disrupted using a micropestle in the presence of liquid nitrogen. Extraction of gDNA was carried out as previously described in Valiadi *et al.* (2012). As we could not efficiently amplify *lbp* from the gDNA of *N. scintillans* with one of our specific primer pairs (see Results), we amplified from cDNA instead. Extraction of RNA was carried out using the Nucleospin RNA II kit (Macherey-Nagel, Germany). The quantity and purity of the DNA and RNA were assessed using a Nanodrop spectrophotometer (ND-3000, Nanodrop, USA). The RNA was reverse transcribed to cDNA using the High Capacity RNA-to-cDNA kit (Applied Biosystems, USA).

Primer design

Six complete sequences of *lbp* were obtained from GenBank: four from *L. polyedrum* (L06908, L19071-3), one from *A. catenella* (EU236684) and one from *N. scintillans* (JF838193). Translated amino acid sequences were aligned using ClustalW2 (www.ebi.ac.uk/ClustalW) and primers were predicted by the CODEHOP program in the most conserved gene regions. Primers for photosynthetic species *lbp* were designed using the Consensus Degenerate Hybrid Oligonucleotide Primer (CODEHOP) program (http://bioinformatics.weizmann.ac.il/blocks/codehop.html) (Rose et al. 1998).The consensus

clamp was based on a codon usage table for *L. polyedrum*. Primers that were selected based on least core degeneracy and highest score of the consensus clamp, were checked for sequence similarity to other genes within dinoflagellates or other organisms using the Basic Local Alignment Search Tool (BLAST) of the National Centre for Biotechnology Information (NCBI) database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and their compatibility was assessed using the program PrimerList (http://primerdigital.com/tools/PrimerList.html). Several CODEHOP primer pairs were tested, the details of which are given in the supplementary materials (Figure S1 and Table S1). The primer set that amplified *lbp* from the highest number of species was DinoLbpF2 (CGAGCGCGTGATCGCNGTNAARGA based on consensus AA: DERVIAVKE) and DinoLbpR1 (CCACGATGCACATGGAGCKYTGCATRTA based on consensus AA: YMQRSMCLLD). These primers amplified ~1,180bp of the *lbp* which represents ~54% of the *lbp* open reading frame in *L. polyedrum*.

As the *lbp* of *N. scintillans* is highly divergent from photosynthetic species and does not contain suitable sites to design common PCR primers, specific primers were designed to amplify and sequence nearly the whole open reading frame (ORF) of the *lcf/lbp* gene, using the program Primer3 (http://frodo.wi.mit.edu/primer3/) and the published full *lcf/lbp* gene of *N. scintillans* (GenBank accession no. JF838193). The resulting PCR primers were NocORF23F (ACCGAAGCTGTTTTGGATTG) and NocORF 2418R (TGGCCGAGAGAGAAAAGAAA), where the numbers refer to the target nucleotide positions. These were in the N-terminal region and near the 3' of the *lbp*, respectively, producing a 2395-bp fragment which corresponds to 87% of the ORF. These primers were only effective on cDNA, producing only very faint bands when gDNA was used as the PCR template. To amplify from the gDNA, a second set of primers, NocORF178F (GGCGATGAGGTTCAGTCAAT) and NocORF2165R (ACGTCATGTCCTTCCTCCAC) was used which amplified a smaller fragment (72%) of the *lcf/lbp* gene.

Polymerase chain reactions

The PCR protocol was initially optimized on *L. polyedrum* and subsequently modified during further testing of other organisms. Reactions were carried out in 25-µl volumes containing: 0.12 µM each primer, 250 µM each dNTP 1x PCR buffer, 0.5 U GoTaq polymerase (Promega, Southampton, UK) and 50-100 ng of DNA. PCR reactions commenced with 5 minutes at 95 ºC for initial template denaturation, followed by 35 cycles at 95 ºC for 45 seconds, 61 ºC for 30 seconds, 68 ºC for 1.5 minutes, and a final extension at 68 ºC for 5 minutes. All PCR reactions were also tested with a lower annealing temperature of 58 ºC in addition to the 61ºC, with all other parameters remaining the same, in order to improve the yield of some faint bands and to confirm that lack of amplification in some species was not due to high PCR stringency.

For *N. scintillans* DNA and cDNA, the PCR reaction components were as above but with 0.1 µM each primer. The PCR cycle for both primer pairs was as follows: 5 minutes at 95 ºC for initial denaturation, followed by 35 cycles of 95 ºC for 45 seconds, 49 ºC for 30 seconds, 68 ºC for 3 minutes and a final extension at 68 ºC for 10 minutes. Additionally, gDNA amplification was achieved using 1 U HotStarTaq polymerase (Qiagen, Manchester, U.K.) with a 15 min initial denaturation of the DNA, instead of the GoTaq polymerase. The reaction mixture was also supplemented with 5% v/v DMSO.

Cloning and sequencing

Sequencing of *lbp* was done selectively in organisms where *lbp* was found for the first time and organisms whose *lbp* had only been sequenced from a short *locus* in previous studies (e.g. some *Alexandrium* ESTs). The PCR products of *lbp* were excised from agarose gels and

cloned into the pCR4 vector in the TOPO TA cloning kit (Invitrogen, Paisley, UK) or using the pGEM-T Easy kit (Promega, UK). Two clones were sequenced for each PCR product using primers M13F and M13R and internal primers for *N. scintillans*. The resulting sequences were trimmed to the length of highest quality base identification (at least 650 bp with > 99.9% accuracy as determined by quality scores) and were reconstructed to the full length PCR product using the CAP3 sequence assembly program (http://pbil.univ-lyon1.fr/cap3.php). A further seven clones were sequenced from *A. tamarense* CCMP1598 and PLY173 to further investigate within strain sequence variation. Sequencing was done by Eurofins MWG Operon (Ebersberg, Germany).

Sequence analyses

Sequences were analysed using the BLASTn tool of the NCBI database, to ensure that the correct gene had been amplified. Sequence alignments were constructed in MEGA v. 5 (Tamura et al. 2011) using the integrated ClustalW (Thompson et al. 1994) and were subsequently manually improved. Where sequences of clones from the same PCR product showed a difference at only one nucleotide position, this was considered likely to be a sequencing error and a consensus sequence was derived. Pseudogenes were identified based on stop codons, partly due to insertions and deletions (indels) in the sequences that caused shifts in the open reading frame. Indels were considered to be true (i.e. not sequencing error) when they were comprised of more than one nucleotide or, if a single base indel, when it was present in both sequenced clones.

Bayesian phylogenetic analyses were performed in MrBayes 3.2 (Ronquist and Huelsenbeck 2003). The best models for our data were selected using the model test function of MEGA v.5. According to the Akaike information criterion, the models were the Generalised Time Reversible with a proportion of invariant sites gamma distributed rates among sites (GTR+I+G) for nucleotide data and the WAG+G model for the amino acid data. Consensus trees were based on 10 independent runs which were run until the average standard deviation of split frequencies was below 0.02 and the Potential Scale Reduction Factor (PSRF) had reached 1.00 indicating convergence of the analyses. Maximum likelihood analyses were conducted in MEGA v.5 and the reliability of the trees was statistically assessed by 100 bootstrap replications.

For the analyses of nucleotide sequences both cDNA and DNA sequences were used in the same tree so to include the whole dataset. We assume that mRNA editing did not bias the phylogenetic analyses as it has only been reported to occur in mitochondrial (Lin et al. 2002, Zhang and Lin 2005) and minicircle genes (Zauner et al. 2004) of dinoflagellates. In contrast, no significant differences between gDNA and cDNA have ever been reported in dinoflagellate bioluminescence genes (Liu and Hastings 2007, Liu et al. 2004, Machabée et al. 1994, Okamoto et al. 2001) or other well studied nuclear genes (e.g. Bachvaroff and Place 2008, Kim et al. 2011). We also confirmed that including cDNA in our gDNA tree did not cause any bias by excluding cDNA from our phylogenetic analysis (Fig. S2).

RESULTS

A second *lbp* in *N. scintillans*

Primers designed to amplify nearly the whole of the *N. scintillans lcf*/*lbp* open reading frame amplified two distinct cDNA fragments. One fragment was ~2.4-kb long, which was the expected size, and another unexpected smaller fragment of ~1.5 kb-length (Fig. 1). The same amplification pattern was present in both cDNA and gDNA, although in the latter amplification was weak despite several modifications of the PCR cycle and trials with PCR additives, precluding sequencing of these fragments from gDNA. As the primers were

sequence-specific and the unexpected product appeared as a clean bright band that was unlikely to be a non-specific product, we investigated its identity further. We sequenced two clones from both fragments amplified from the cDNA (i.e. the ~2.4-kb and ~1.5-kb fragments) and sequences from both clones of each sequenced fragment were identical. To obtain a partial *lcf/lbp* sequence from the gDNA we used a second primer pair. As this primer pair did not target the N-terminal region, only the combined *lcf/lbp* was amplified.

The 2396-bp cDNA fragment was the combined *lcf*/*lbp* gene previously reported for this organism (Liu and Hastings 2007), consisting of part of the N-terminal region and the *lcf* domain followed by the *lbp* domain. The nucleotide sequence identity between this cDNA sequence and gDNA from this strain as well as that previously reported for *N. scintillans* (Liu and Hastings 2007) was ~ 94%. The 1,497-bp cDNA fragment was also an *lbp* with high nucleotide sequence identity of to the *lbp* domain of the combined *lcf/lbp*; 95.7% to the cDNA and 95.4-96.1% to the gDNA sequences amplified here from the same strain, and 94% to that amplified by Liu and Hastings (2007). Additionally, both fragments contained an identical N-terminal gene region which is 108-bp long. An alignment of the two translated cDNA sequences (AA) (Fig. 1) revealed that, in the shorter fragment, the N-terminal region led directly into the *lbp* domain in the absence of an *lcf* domain. Therefore these sequences, which share the N-terminal region and *lbp* domain, but not the *lcf* domain, correspond to two different genes: the combined *lcf*/*lbp* and a single separate *lbp*. By using cDNA for this analysis we also confirmed that *N. scintillans* does indeed express both of these bioluminescence-related genes.

Identification of *lbp* in gonyaulacoid species

The DinoLbp primers yielded PCR products of approximately 1,180 bp from 11 strains out of the 18 tested (Table 1), revealing that *lbp* was present in all members of the Gonyaulacales with the exception of the genera *Ceratium* and *Fragilidium*. Also, *lbp* was not detected in the genera *Protoperidinium* and *Pyrocystis*. It is important to note that, due to the large differences of *lbp* among organisms, the efficient amplification of *lbp* from *A. monilatum*, *C. horrida* and *G. spinifera* only became possible after a reduction in the PCR stringency (i.e. annealing temperature), as described in the Materials and Methods. Finally, the DinoLbp primres amplified *lbp* in all species known to contain it except *N. scintillans*, which contains a very divergent *lbp*-like sequence (Figure 2; Liu and Hastings 2007).

Diversity and phylogeny of *lbp* in dinoflagellates

Sequences of different clones obtained from each strain were never identical, but rather had at least 3 nucleotide differences. The variation in sequences amplified by our primers was further investigated by additional sequencing in two strains of *Alexandrium tamarense*. This revealed high variability within both strains with sequence identities as low as 90% in *A. tamarense* PLY173A.

The phylogenetic analyses of both *lbp*s of *N. scintillans* based on nucleotide and amino acid sequences (Fig. 2A and 2B), showed that genomic sequences, which all originated from the combined *lcf*/*lbp*, were more similar to each other (at least 98.9% identity) than to their cDNA counterpart (up to 96.1% identity) despite originating from different strains. The cDNA sequence of the single *lbp* also preferentially grouped with the cDNA rather than the gDNA of the combined *lcf*/*lbp*. These differences received high statistical support in all analyses.

Phylogenetic analyses of Gonyaulacales *lbp* sequences based both on nucleotide and amino acid sequences (Fig. 2A and 2B) generally agreed well and showed that sequences of closely related species were intermingled because of the variability within strains. For example, sequences of *Ceratocorys horrida* and *Protoceratium reticulatum* sequences were

as similar to each other as within each organism. Accordingly, the sequences of *Gonyaulax spinifera* were grouped with *C. horrida* and *P. reticulatum* (88-90% nucleotide identity to both species) in a single well-supported clade (C/G/P) because their sequences showed less divergence (9% differences) than that known to exist among gene copies within each strain (e.g. up to 15% in *L. polyedrum*, (Machabée et al. 1994)). Similarly, within *Alexandrium*, sequences did not cluster according to morphospecies or ribotype but rather randomly in three distinct and well-supported groups, with the most diverse clade A3 consisting of three morphospecies or two phylotypes. Sequences from *Alexandrium monilatum* (Am) and *L. polyedrum* (Lp) each formed monospecific clades in the nucleotide analyses and the sequence from *Pyrodinium bahamense* (Pb) was also divergent from the other gonyaulacoid clades (Fig. 2A); all received high statistical support. *Pyrodinium bahamense lbp* was most similar to *A. tamarense* at the nucleotide level (78% identity). At the amino acid level, *A. monilatum* (Am) and *P. bahamense* clustered more tightly with the C/P/G clade. Additionally, *P. bahamense* showed highest similarity to *Alexandrium catenella* at the amino acid level (75%), instead of *A. tamarense*. These differences were also reflected on the position of the *P. bahamense* sequence which differed between the phylogenetic analyses (Fig. 2A, B), placing nearer to the C/G/P and Am clades in the amino acid analysis. The *lbp* of *L. polyedrum* is distinct within the gonyaulacoids forming a basal clade within this group. Sequences from *N. scintillans* formed the most distinct clade (Ns) in our analyses (Fig. 2A, B), having the least identity to any of the gonyaulacoid strains (maximum identity 60% to *L. polyedrum* at both nucleotide and amino acid levels).

DISCUSSION

A second *lbp* in *N. scintillans* highlights the complex evolution of dinoflagellate bioluminescence systems

Our finding of a second *lbp* in *N. scintillans* that occurs as a separate gene, in addition to the combined *lcf*/*lbp*, provides significant new insight into the evolution of dinoflagellate bioluminescence systems. Our analyses also allowed us to better define the border of the N-terminal region. By comparing our data to the full sequence on GenBank, the N-terminal region is 36-AA long, not 29 AA as stated by Liu and Hastings (2007).

A second LBP was hypothesised by Liu and Hastings (2007) to be responsible for a ~65.3 kDa protein being labelled in *N. scintillans* protein extracts by an anti-LBP antibody, although a corresponding gene was not found in their study. Based on the full *lcf*/*lbp* translated AA sequence, we estimate the molecular weight of the single *lbp* (i.e. *lbp* domain and N-terminal) to be 67.5 kDa, and is therefore likely to be the gene for this previously reported protein band. This is indirect evidence that the single *lbp* sequenced in our study does indeed produce a distinct protein. Liu and Hastings (2007) had speculated that if there is a second *lbp* in *N. scintillans*, it could have a role in binding luciferin in the scintillons or it could act to store luciferin in the cytoplasm. Partitioning of members of a gene/protein family to different mechanistic roles within the cell is an interesting possibility, especially with regard to the link of bioluminescence with cell physiology, and warrants further investigation.

Our data suggest that the hypothesized fission or fusion of *lcf* and *lbp* that has led to the current organization of the respective genes in photosynthetic species and *N. scintillans* (Liu and Hastings 2007), is likely to have occurred within *N. scintillans.* If the *lcf/lbp* of *N. scintillans* is ancestral to the *lcf* and *lbp* of photosynthetic species, as suggested by their general phylogeny (Orr et al. 2012), the single *lbp* of *N. scintillans* could have been formed by mRNA splicing of *lcf*/*lbp* and subsequent retrotransposition into the genome, that is extensive in dinoflagellates (Bachvaroff and Place 2008, Slamovits and Keeling 2008). The identical N-terminal gene region in both genes supports this scenario. Having separate genes

for *lcf* and *lbp* allows differential regulation of each gene. This is advantageous if, for example, *lbp* needs to be stoichiometrically proportional to luciferin (Fogel and Hastings 1971), but *lcf*, which already has triple catalytic capacity in photosynthetic species, can be re-used. Thus, a potential gene fission event in *N. scintillans* could have played a decisive role in the subsequent evolution of dinoflagellate bioluminescence systems.

The phylogenetic analyses of *N. scintillans* gDNA and cDNA *lbp* sequences, revealed that sequences from the respective categories clustered together. This was unexpected because mRNA editing is not known in any genes of *N. scintillans*, neither in bioluminescence genes (Liu and Hastings 2007), nor even in mitochondrial genes that are edited in other dinoflagellates (Zhang and Lin 2008). An alternative explanation could be that the different primer sets used to amplify gDNA and cDNA might be biased towards different groups of gene copies within the genome as we have previously found with different primer sets for *lcf* (Valiadi et al. 2012). Furthermore, the sequences obtained from the gDNA may not be those that are expressed. Nevertheless, this surprising finding requires further investigation.

Lbp is widely distributed in bioluminescent gonyaulacoid dinoflagellates

We investigated the distribution and diversity of *lbp* in several closely related gonyaulacoid dinoflagellates. Our results represent the first reports of *lbp* in *Ceratocorys horrida*, *Gonyaulax spinifera* and *Protoceratium reticulatum*, as well as the first long partial sequences of *lbp* retrieved from *Alexandrium affine*, *A. fundyense*, *A. monilatum* and *A. tamarense*. Our results are also in agreement with a confirmed absence of LBP in *Pyrocystis* spp. (Knaust et al. 1998, Schmitter et al. 1976). However, the lack of *lbp* detection in *N. scintillans*, which is known to contain this sequence, means that the negative PCR results in *Ceratium longipes*, *Fragilidium* cf. *subglobosum* and *Protoperidinium crassipes* are unlikely to be conclusive for the absence of this gene. Despite several efforts, the large differences in *lbp* sequences among organisms precluded a "universal" primer pair for *lbp* (see supplementary materials), as we have previously done for *lcf* (Valiadi et al. 2012). We propose that an antibody for *L. polyedrum* LBP, which has been successfully used to detect LBP in *N. scintillans* (Liu and Hastings 2007), may be more suitable than PCR to initially detect LBP in distantly related bioluminescent dinoflagellates. However, at the genomic and transcriptomic level, our PCR primers are useful to investigate the distribution and evolution of *lbp* in several ecologically important gonyaulacoid dinoflagellate genera.

Our finding that *lbp* is present in several gonyaulacoid dinoflagellate species indicates that bioluminescence systems utilizing LBP are more common than those that do not. This means that *Pyrocystis lunula*, which has been used as the model organism to study the chemistry of dinoflagellate bioluminescence, may not be representative of typical dinoflagellate bioluminescence systems, since *Pyrocystis* is the only genus that appears to lack LBP (Knaust et al. 1998, Schmitter et al. 1976). *Pyrocystis lunula* luciferin is an open chain tetrapyrrole, hypothesized to be a photo-oxidation breakdown product of chlorophyll *a* (Topalov and Kishi 2001) and is thought to be universal in dinoflagellates because LCF from any dinoflagellate can use it as a substrate to produce light (Nakamura et al. 1989). According to this paradigm, Liu and Hastings (2007) hypothesized that heterotrophic dinoflagellates acquire luciferin nutritionally from ingested prey, either directly or from the degradation of chlorophyll from the prey. However, the heterotrophic dinoflagellate *Protoperidinium crassipes* can maintain its bioluminescence intensity for one year in the absence of chlorophyll or luciferin containing food (Yamaguchi and Horiguchi 2008) and must, therefore, contain a luciferin originating from a different precursor molecule. Also, *L. polyedrum* only contains luciferin at night when photo-oxidation is not possible. The presence of *lbp* in several photosynthetic species suggests that even within these species there

might be an alternative luciferin molecule that requires LBP for stabilization (Morse et al. 1989) in contrast to *P. lunula* luciferin (Knaust et al. 1998). Therefore, more studies on the structure of dinoflagellate luciferin, particularly of LBP containing species, are necessary to fully understand the biochemistry of dinoflagellate bioluminescence and its link with other physiological pathways in the cell.

High variability of *lbp* in bioluminescent dinoflagellates

Dinoflagellate genes occur in large gene families with multiple non identical gene copies that also include numerous pseudogenes (Bachvaroff and Place 2008, Kim et al. 2011, Lin et al. 2009, Moustafa et al. 2010, Reichman et al. 2003, Rowan et al. 1996, Zhang and Lin 2003). Even highly conserved ribosomal genes require deep sequencing of gene copies to successfully resolve their phylogenetic differences within a species complex (Miranda et al. 2012). Dinoflagellate *lbp* is no exception and the sequences amplified in this study showed that *lbp* is a very large and highly diverse gene family, in agreement with previous work (Machabée et al. 1994, Tanikawa et al. 2004). Additionally, phylogenetic analyses of *Alexandrium* sequences suggest the presence of more than one *lbp* type, similar to previous observations with their *lcf* sequences (Valiadi et al. 2012). Multiple gene types in genes involved in bioluminescence seem to be common in both *lbp* and *lcf* of several Gonyaulacales species (present study, Machabée et al. 1994, Okamoto et al. 2001, Valiadi et al. 2012) creating further divergence among *lbp* gene copies in the genome. This sequence variation among gene copies in dinoflagellates poses a limitation on phylogenetic analyses based on protein coding genes, since sequences can phylogenetically overlap not only between different species but also across genera (Kim et al. 2011, Valiadi et al. 2012).

In this study, intermingling of sequences from more than one species was most common for closely related genera, such as *Ceratocorys*, *Protoceratium* and *Gonyaulax*, *lbp* sequences of which formed a monophyletic clade, consistent with other phylogenetic analyses using *lcf* , ribosomal and other protein coding genes (Gomez et al. 2010b, Orr et al. 2012, Valiadi et al. 2012). Intermingling of closely related species was also observed within *Alexandrium* which largely formed one large monophyletic clade. The phylogenetic analyses shown in Fig. 2 suggest that groups A1-A3 most likely correspond to different groups of gene copies. For example, *A. affine* sequences grouped with *A. tamarense* Group 3 and 4 sequences, in agreement with a previous *lcf* phylogeny (Valiadi et al. 2012), even though it does not belong to the *A. tamarense* species complex. However, our data do suggest that, even within *Alexandrium*, the evolution of *lbp* is consistent with the general phylogenetic position of species when more pronounced differences are present, such as the close clustering of *Alexandrium* Group 1 *lbp* sequences. Also, *A. monilatum lbp* formed its own clade in both nucleotide and amino acid analyses, being divergent and possibly paraphyletic to other *Alexandrium* species, which is in accordance with its *lcf* and rDNA phylogeny as well as its unique toxin profile (Hsia et al. 2006, Rogers et al. 2006, Valiadi et al. 2012). The *lbp* of *Pyrodinium bahamense* was divergent from other clades within the gonyaulacoids and showed highest similarity to *Alexandrium*, which is consistent with phylogenies based on ribosomal and protein coding genes (Orr et al. 2012) as well as with saxitoxin production in both genera. *Lingulodinium polyedrum* sequences formed a monospecific clade which was basal within the Gonyaulacales. This is in agreement with the phylogeny of *lcf* (Valiadi et al. 2012) and SSU rDNA (Howard et al. 2009, Saldarriaga et al 2001), but not with its phylogenetic position based on concatenated analysis of ribosomal and protein coding genes, where it is rather embedded deeply within the Gonyaulacales (Orr et al. 2012). The contradiction of *lbp* evolution in *L. polyedrum* and *A. affine* to their general phylogeny, as well as the intriguing phylogenetic affiliations of *A. monilatum* and *P.bahamense lbp*, make

these organisms interesting to investigate potential horizontal gene transfers and changes in the composition of genes during the evolution of dinoflagellate bioluminescence systems.

This study has provided new information on the composition of dinoflagellate bioluminescence systems. A new *lbp* in *N. scintillans* has revealed that separate bioluminescence genes were formed early on in the evolution of bioluminescent dinoflagellates. Also, based on our finding that *lbp* is common in several ecologically important dinoflagellate species, we propose that bioluminescent systems with *lbp* should be viewed as the standard in bioluminescent dinoflagellates and, thus, the biochemistry of such systems warrants further investigation. While *lbp* might be too diverse within each organism to easily dissect its evolution among closely related species without conducting extensive sequencing, it is valuable for tracking more ancestral events in the evolution of dinoflagellate bioluminescence systems. This is especially relevant for the birth of new gene variants and their relation to the concurrent evolution of *lcf*. We anticipate that further sequencing of *lbp* using our primers in key species such as *A. affine*, *A. monilatum*, *P. bahamense* and *L.polyedrum*, combined with *de novo* sequences arising from transcriptomics studies in a diverse range of species will be key to unraveling the diversity and evolution of bioluminescence in dinoflagellates.

ACKNOWLEDGEMENTS

We deeply thank Michael I. Latz, Aika Yamaguchi and Ana Amorim who kindly donated cultures or DNA samples that for the present study. This research was financially supported by the Luminescence and Marine Plankton project (Defence Science and Technology Laboratory and Natural Environment Research Council joint grant scheme proposal. ref 1166, awarded to John T. Allen and DIR), and the Office of Naval Research (ONR award number N000140410180, awarded to DIR).

LITERATURE CITED

Abrahams, M. V. & Townsend, L. D. 1993. Bioluminescence in dinoflagellates: A test of the burgular alarm hypothesis. *Ecology,* 74:258-260.

Bachvaroff, T. R. & Place, A. R. 2008. From stop to start: Tandem gene arrangement, copy number and *Trans*-Splicing Sites in the Dinoflagellate *Amphidinium carterae*. *PLoS ONE,* 3:e2929.

Buskey, E., Mills, L. & Swift, E. 1983. The effects of dinoflagellate bioluminescence on the swimming behavior of a marine copepod. *Limnol. Oceanogr.,* 28:575-579.

Desa, R. & Hastings, J. W. 1968. The characterization of scintillons. Bioluminescent particles from the marine dinoflagellate, *Gonyaulax polyedra*. *J. Gen. Physiol.,* 51:105-122.

Erdner, D. L. & Anderson, D. M. 2006. Global transcriptional profiling of the toxic dinoflagellate *Alexandrium fundyense* using Massively Parallel Signature Sequencing. *BMC Genomics,* 7:88.

Fleisher, K. J. & Case, J. F. 1995. Cephalopod predation facilitated by dinoflagellate luminescence. *Biol. Bull.,* 189:263-271.

Fogel, M. & Hastings, J. W. 1971. A substrate-binding protein in the *Gonyaulax* bioluminescence reaction. *Arch. Biochem. Biophys.,* 142:310-321.

Fukuda, Y. & Endoh, H. 2008. Phylogenetic analyses of the dinoflagellate *Noctiluca scintillans* based on beta-tubulin and Hsp90 genes. *Eur. J. Protistol.,* 44:27-33.

Gomez, F., Moreira, D. & Lopez-Garcia, P. 2010a. Molecular phylogeny of noctilucoid dinoflagellates (Noctilucales, Dinophyceae). *Protist,* 161:466-478.

Gomez, F., Moreira, D. & Lopez-Garcia, P. 2010b. *Neoceratium* gen. nov., a new genus for all marine species currently assigned to *Ceratium* (Dinophyceae). *Protist,* 161:35-54.
Hackett, J. D., Scheetz, T. E., Yoon, H. S., Soares, M. B., Bonaldo, M. F., Casavant, T. L. & Bhattacharya, D. 2005. Insights into a dinoflagellate genome through expressed sequence tag analysis. *BMC Genomics,* 6:80.
Hastings, J. W. 2009. *Dinoflagellate luminescence and its circadian regulation* [Online]. Available: http://photobiology.info/Hastings.html [Accessed 06 June 2013].
Hsia, M. H., Morton, S. L., Smith, L. L., Beauchesne, K. R., Huncik, K. M. & Moeller, P. D. R. 2006. Production of goniodomin A by the planktonic, chain-forming dinoflagellate *Alexandrium monilatum* (Howell) Balech isolated from the Gulf coast of the United States. *Harmful Algae,* 5:290-299.
Jaeckisch, N., Yang, I., Wohlrab, S., Glöckner, G., Kroymann, J., Vogel, H., Cembella, A. & John, U. 2011. Comparative genomic and transcriptomic characterization of the toxigenic marine dinoflagellate *Alexandrium ostenfeldii*. *PLoS ONE,* 6:e28012.
Johnson, C. H., Inoue, S., Flint, A. & Hastings, J. W. 1985. Compartmentalization of algal bioluminescence - autofluorescence of bioluminescent particles in the dinoflagellate *Gonyaulax* as studied with image-intensified video microscopy and flow cytometry. *J. Cell Biol.,* 100:1435-1446.
Ki, J. S. 2010. Nuclear 28S rDNA phylogeny supports the basal placement of *Noctiluca scintillans* (Dinophyceae; Noctilucales) in dinoflagellates. *Eur. J. Protistol.,* 46:111-120.
Kim, S., Bachvaroff, T. R., Handy, S. M. & Delwiche, C. F. 2011. Dynamics of actin evolution in dinoflagellates. *Mol. Biol. Evol.,* 28:1469-1480.
Knaust, R., Urbig, T., Li, L. M., Taylor, W. & Hastings, J. W. 1998. The circadian rhythm of bioluminescence in *Pyrocystis* is not due to differences in the amount of luciferase: A comparative study of three bioluminescent marine dinoflagellates. *J. Phycol.,* 34:167-172.
Lee, D. H., Mittag, M., Sczekan, S., Morse, D. & Hastings, J. W. 1993. Molecular cloning and genomic organization of a gene for luciferin-binding protein from the dinoflagellate *Gonyaulax polyedra*. *J. Biol. Chem.,* 268:8842-8850.
Lin, S., Zhang, H., Hou, Y., Zhuang, Y. & 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.
Lin, S., Zhang, H., Spencer, D. F., Norman, J. E. & Gray, M. W. 2002. Widespread and extensive editing of mitochondrial mRNAs in dinoflagellates. *J. Mol. Biol.,* 320:727-740.
Liu, L. Y. & Hastings, J. W. 2007. Two different domains of the luciferase gene in the heterotrophic dinoflagellate *Noctiluca scintillans* occur as two separate genes in photosynthetic species. *PNAS,* 104:696-701.
Liu, L. Y., Wilson, T. & Hastings, J. W. 2004. Molecular evolution of dinoflagellate luciferases, enzymes with three catalytic domains in a single polypeptide. *PNAS,* 101:16555-16560.
Machabée, S., Wall, L. & Morse, D. 1994. Expression and genomic organization of a dinoflagellate gene family. *Plant Mol. Biol.,* 25:23-31.
Miranda, L. N., Zhuang, Y., Zhang, H. & Lin, S. 2012. Phylogenetic analysis guided by intragenomic SSU rDNA polymorphism refines classification of "*Alexandrium tamarense*" species complex. *Harmful Algae,* 16:35-48.
Morse, D., Pappenheimer, A. M. & Hastings, J. W. 1989. Role of a luciferin-binding protein in the circadian bioluminescent reaction of *Gonyaulax polyedra*. *J. Biol. Chem.,* 264:11822-11826.

Moustafa, A., Evans, A. N., Kulis, D. M., Hackett, J. D., Erdner, D. L., Anderson, D. M. & Bhattacharya, D. 2010. Transcriptome profiling of a toxic dinoflagellate reveals a gene-rich protist and a potential impact on gene expression due to bacterial presence. *PLoS ONE,* 5:e9688.

Nakamura, H., Kishi, Y., Shimomura, O., Morse, D. & Hastings, J. W. 1989. Structure of dinoflagellate luciferin and its enzymic and nonenzymic air-oxidation products. *J. Am. Chem. Soc.,* 111:7607-7611.

Nicolas, M. T., Sweeney, B. M. & Hastings, J. W. 1987. The ultrastructural localization of luciferase in three bioluminescent dinoflagellates, two species of *Pyrocystis*, and *Noctiluca*, using anti-luciferase and immunogold labelling. *J. Cell Sci.,* 87:189-196.

Okamoto, O. K., Liu, L., Robertson, D. L. & Hastings, J. W. 2001. Members of a dinoflagellate luciferase gene family differ in synonymous substitution rates. *Biochemistry,* 40:15862-15868.

Orr, R. J. S., Murray, S. A., Stüken, A., Rhodes, L. & Jakobsen, K. S. 2012. When naked became armored: An eight-gene phylogeny reveals monophyletic origin of theca in dinoflagellates. *PLoS ONE,* 7:e50004.

Poupin, J., Cussatlegras, A. S. & Geistdoerfer, P. 1999. Plancton Marin Bioluminescent. *Rapport scientifique du Loen.* Brest, France.

Reichman, J. R., Wilcox, T. P. & Vize, P. D. 2003. PCP gene family in *Symbiodinium* from *Hippopus hippopus*: Low levels of concerted evolution, isoform diversity and spectral tuning of chromophores. *Mol. Biol. Evol.,* 20:2143-2154.

Rogers, J. E., Leblond, J. E. D. & Moncreiff, C. A. 2006. Phylogenetic relationship of *Alexandrium monilatum* (Dinophyceae) to other *Alexandrium* species based on 18S ribosomal RNA gene sequences. *Harmful Algae,* 5:275-280.

Ronquist, F. & Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19:1572-1574.

Rose, T. M., Schultz, E. R., Henikoff, J. G., Pietrokovski, S., Mccallum, C. M. & Henikoff, S. 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res.,* 26:1628-1635.

Rowan, R., Whitney, S. M., Fowler, A. & Yellowlees, D. 1996. Rubisco in marine symbiotic dinoflagellates: form II enzymes in eukaryotic oxygenic phototrophs encoded by a nuclear multigene family. *The Plant Cell Online,* 8:539-553.

Schmitter, R. E., Njus, D., Sulzman, F. M., Gooch, V. D. & Hastings, J. W. 1976. Dinoflagellate bioluminescence - Comparative study of *in vitro* components. *J. Cell. Physiol.,* 87:123-134.

Slamovits, C. H. & Keeling, P. J. 2008. Widespread recycling of processed cDNAs in dinoflagellates. *Curr. Biol.,* 18:550-552.

Smith, S. M. E., Morgan, D., Musset, B., Cherny, V. V., Place, A. R., Hastings, J. W. & Decoursey, T. E. 2011. Voltage-gated proton channel in a dinoflagellate. *PNAS,* 108:18162-18167.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.,* 28:2731-2739.

Tanikawa, N., Akimoto, H., Ogoh, K., Chun, W. & Ohmiya, Y. 2004. Expressed sequence tag analysis of the dinoflagellate *Lingulodinium polyedrum* during dark phase. *Photochem. Photobiol.,* 80:31-35.

Taylor, F. J. R. M. 2004. Illumination or confusion? Dinoflagellate molecular phylogenetic data viewed from a primarily morphological standpoint. *Phycol. Res.,* 52:308-324.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994. CLUSTALW - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting,

position specific gap penalties and weight matrix choice. *Nucleic Acids Res.,* 22:4673-4680.
Topalov, G. & Kishi, Y. 2001. Chlorophyll catabolism leading to the skeleton of dinoflagellate and krill luciferins: hypothesis and model studies. *Angew. Chem.,* 40:3892-3894.
Toulza, E., Shin, M.-S., Blanc, G., Audic, S., Laabir, M., Collos, Y., Claverie, J.-M. & Grzebyk, D. 2010. Gene expression in proliferating cells of the dinoflagellate *Alexandrium catenella* (Dinophyceae). *Appl. Environ. Microbiol.,* 76:4521-4529.
Uribe, P., Fuentes, D., Valdés, J., Shmaryahu, A., Zúñiga, A., Holmes, D. & Valenzuela, P. D. T. 2008. Preparation and analysis of an expressed sequence tag library from the toxic dinoflagellate *Alexandrium catenella*. *Mar. Biotechnol.,* 10:692-700.
Valiadi, M., Iglesias-Rodriguez, M. D. & Amorim, A. 2012. Distribution and genetic diversity of the luciferase gene within marine dinoflagellates. *J. Phycol.,* 48:826-836.
Yamaguchi, A. & Horiguchi, T. 2008. Culture of the heterotrophic dinoflagellate *Protoperidinium crassipes* (Dinophyceae) with noncellular food items. *J. Phycol.,* 44:1090-1092.
Zauner, S., Greilinger, D., Laatsch, T., Kowallik, K. V. & Maier, U.-G. 2004. Substitutional editing of transcripts from genes of cyanobacterial origin in the dinoflagellate Ceratium horridum. *FEBS Lett.,* 577:535-538.
Zhang, H. & Lin, S. 2003. Complex gene structure of the form II rubisco in the dinoflagellate *Prorocentum minimum* (Dinophyceae). *J. Phycol.,* 39:1160-1171.
Zhang, H. & Lin, S. 2005. Mitochondrial Cytochrome b mRNA Editing in Dinoflagellates: Possible Ecological and Evolutionary Associations? *J. Eukaryot. Microbiol.,* 52:538-545.
Zhang, H. & Lin, S. 2008. mRNA editing and spliced-leader RNA trans-splicing groups *Oxyrrhis*, *Noctiluca*, *Heterocapsa*, and *Amphidinium* as basal lineages of dinoflagellates. *J. Phycol.,* 44:703-711.

FIGURE LEGENDS

Fig. 1. Amino acid alignment of sequences obtained from the two PCR fragments amplified from *Noctiluca scintillans* cDNA. The 2.4 kb fragment corresponds to the combined luciferase (*lcf*) and luciferin binding protein (*lbp*) genes while the 1.45 kb fragment contains only the N-terminal part and the *lbp* domain and therefore corresponds to a different gene.

Fig. 2A. Bayesian phylogenetic trees of partial *lbp* sequences based on a nucleotide alignment containing 775 positions. Values shown at the nodes are posterior probabilities followed by maximum likelihood bootstrap values based on 100 replications. The scale of branch lengths shows 2 substitutions per 10 sites. Sequences obtained from GenBank are indicated by their accession numbers which are given in brackets. Sequences from the *Alexandrium tamarense* species complex are labelled with their ribotype (G1-G4) in addition to species names. An asterisk indicates that the sequence is a pseudogene. Clades of *lbp* are shown by vertical lines. A = *Alexandrium* (with 3 sub-groups), Am = *Alexandrium monilatum*, C = *Ceratocorys horrida*, G = *Gonyaulax spinifera*, Lp = *Lingulodinium polyedrum*, Ns = *Noctiluca scintillans*, P = *Protoceratium reticulatum*, Pb = *Pyrodinium bahamense* (single divergent sequence rather than a clade)

Fig. 2B. Bayesian phylogenetic trees of partial *lbp* sequences based on an amino acid alignment containing 255 positions with presumed pseudogene sequences excluded. Values shown at the nodes are posterior probabilities followed by maximum likelihood bootstrap values based on 100 replications. The scale of branch lengths shows 2 substitutions per 10

sites. Sequences obtained from GenBank are indicated by their accession numbers which are given in brackets. Sequences from the *Alexandrium tamarense* species complex are labelled with their ribotype (G1-G4) in addition to species names. Clades of *lbp* are shown by vertical lines. A = *Alexandrium* (with 3 sub-groups), Am = *Alexandrium monilatum*, C = *Ceratocorys horrida*, G = *Gonyaulax spinifera*, Lp = *Lingulodinium polyedrum*, Ns = *Noctiluca scintillans*, P = *Protoceratium reticulatum*, Pb = *Pyrodinium bahamense* (single divergent sequence rather than a clade).

Table 1. Details of the dinoflagellate cultures tested in this study with the results of *lbp* PCR tests. For comparison, the bioluminescence (BL) and *lcf* presence data from Valiadi et al. (2012) that correspond to the DNA extracts used in this study are included in the table. Asterisks indicate PCR products that were sequenced and the corresponding GenBank accession numbers are given for new sequences obtained in this study.

Species	Source[a]	Strain	BL	LCF	LBP	Accession no.
Alexandrium affine (H. Inoue & Y. Fukuyo) Balech	CCMP	112	+	+	+*	JQ946837-8
Alexandrium fundyense Balech	CCMP	1978	+	+	+*	JQ946839-40
Alexandrium monilatum (J. F. Howell) Balech	MIL		+	+	+*	JQ946841-2
Alexandrium tamarense (Lebour) Balech	CCMP	115	+	+	+*	JQ946843-4
Alexandrium tamarense (Lebour) Balech	CCMP	1598	+	+	+*	JQ946845-53
Alexandrium tamarense (Lebour) Balech	PLY	173A	-	+	+	JQ946854-62
Ceratium longipes (Bailey) Gran	CCMP	1770	-	+	-	
Ceratocorys horrida Stein	CCMP	157	+	+	+	N/A
Ceratocorys horrida Stein	MIL		+	+	+*	JQ946863-4
Fragilidium cf. *subglobosum* (von Stosch) Loeblich III	ALISU	I097-01	-	+	-	
Gonyaulax spinifera (Clap. & Lachm.) Diesing	CCMP	409	+	+	+*	JQ946865-6
Lingulodinium polyedrum (F. Stein) J. D. Dodge	NOCS	M22	+	+	+	N/A
Noctiluca scintillans (Macartney) Kofoid et Swezy[b]	MIL		+	+	+[c]	KF601764-8
Protoceratium reticulatum (Clap. & Lachm.) Butschli	CCMP	1889	+	+	+*	JQ946867-8
Protoperidinium crassipes (Kofoid) Balech[d]	AY		+	+	-	
Pyrocystis lunula (Schütt) Schütt	CCMP	731	+	+	-	
Pyrocystis lunula (Schütt) Schütt	CCAP	1131	+	+	-	
Pyrocystis noctiluca Murray ex Haeckel	CCMP	732	+	+	-	

[a] Source abbreviations: ALISU-Universidade de Lisboa culture collection, Portugal; CCMP- Culture collection of marine phytoplankton, USA; CCAP-Culture collection of algae and protozoa, UK; PLY-Plymouth culture collection, UK; NOCS- National Oceanography Centre Southampton, UK; MIL- M.I. Latz, Scripps Institution of Oceanography, USA; AY- Aika Yamaguchi, University of British Columbia, Canada.
[b] Maintained in low light (15 µmol m^{-1} s^{-2}) with algal prey *Dunaliella tertiolecta* CCMP 1320 grown in F/2 medium minus silicate.
[c] Amplified with specific primers, not COPDEHOPs
[d] Culture was grown on rice flour

```
Noc_lcflbp      SACLEFPVESCFFVMSGYTYKTEVHALHGMNKQGGFGGPLCAEPYGDEVQSIVKRLLADS 60
Noc_lbp         SACPEVPVESCFFDMSGYSYKT-------------------------------------- 22
                *** *.******* ****:***

Noc_lcflbp      DNDKSLAFNNFTDPCPELTKKQLQTMKGFDYADKKKKLPFGPLPWPTGLPTPGYVPKTNP 120
Noc_lbp         ------------------------------------------------------------

Noc_lcflbp      LNGRWVTVTGGDAEFIKKSIASGMLGSAEASKIQADVDTKKTGGMFLRITQNGEVCTVDA 180
Noc_lbp         ------------------------------------------------------------

Noc_lcflbp      SVAKFARAVRTWKSGHYFYEPLVSGSHLFGVWVLPEEYRKIGFFWEMTSGKCFRIQRNAW 240
Noc_lbp         ------------------------------------------------------------

Noc_lcflbp      FDSGYMFMRQSTEAFGRISHIFYVKVDSDPEVDSRPALQSRDFTALAGVFNAPDNLGNPY 300
Noc_lbp         ------------------------------------------------------------

Noc_lcflbp      PCQPVDLDAPVERDTWMDQNKEVVAQQAASIGKSVQELEKEKQKLVNLGWSSDNVTVHVD 360
Noc_lbp         ----------------------------------EELEKEKKQLVNLGWSSDNVTVHVD 47
                                                   :******::**************

Noc_lcflbp      ALWEALTMKARSPDKFMDVSDVKVSDEDGYLSRSMTIKANNKIVKERIWINRVASEIVFQ 420
Noc_lbp         ALWEALTMKARSPEKFMDVTDVKVSDEDGYLSRSMTIKANNKTVKERIWINRVASEIVFQ 107
                ************:*****:********************** *****************

Noc_lcflbp      PLHPDTGAPLHDECVIAVREEPNLHLEFYQRDVTDGMRSSWKLPVDVVSKSFQEVVKVAQ 480
Noc_lbp         PLHPDTGTPLNEERVIAVREEPNLHLEFYQRDVTDGMRSPWKLPVDVVSKSFQEVVNVAQ 167
                *******:**::* ************************.****************:***

Noc_lcflbp      KLESTVQPVIGLGFHSSAMENVDHDALWLAVLNEVRQPGAHAPNCSVIECDGYIERKLSS 540
Noc_lbp         KLENTVQPVIGLGFHSSALEDVNHDALWLALLNEVRQPVAHAPNCSVIECDGFIERKLSS 227
                ***.**************:*:*:*******:******* *************:*******

Noc_lcflbp      TGTTQHVYVREGEWAVVYRDVVDGKESKTECAIVLRAHPLEIEVCERNVLSGFRVHSSIP 600
Noc_lbp         IGSTQHIYVREEEWTVVYREVVEGKESQTECAIVLRAHPLEIEVCERNVLSGLRVHSSIP 287
                 *:***:**** **:****:**:****:************************:*******
```

Figure 1
126x114mm (300 x 300 DPI)

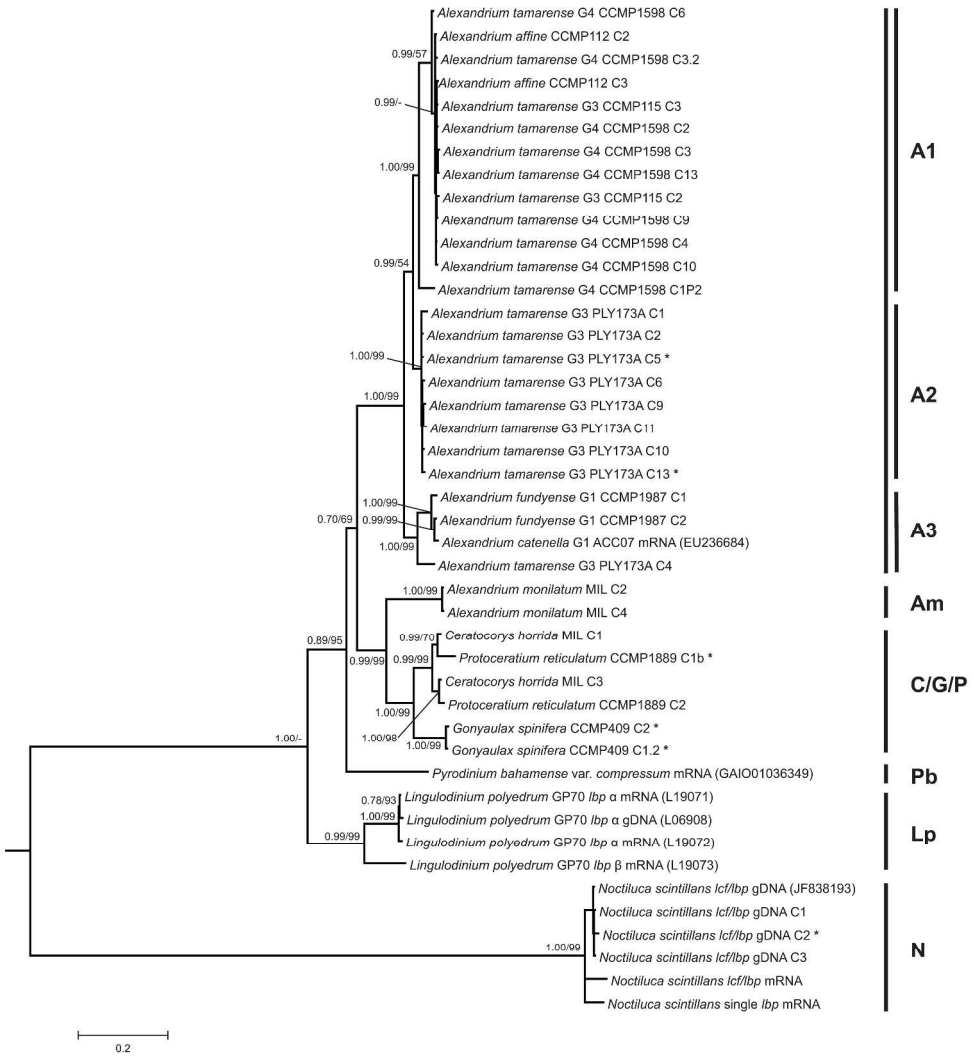

Figure 2A
384x401mm (300 x 300 DPI)

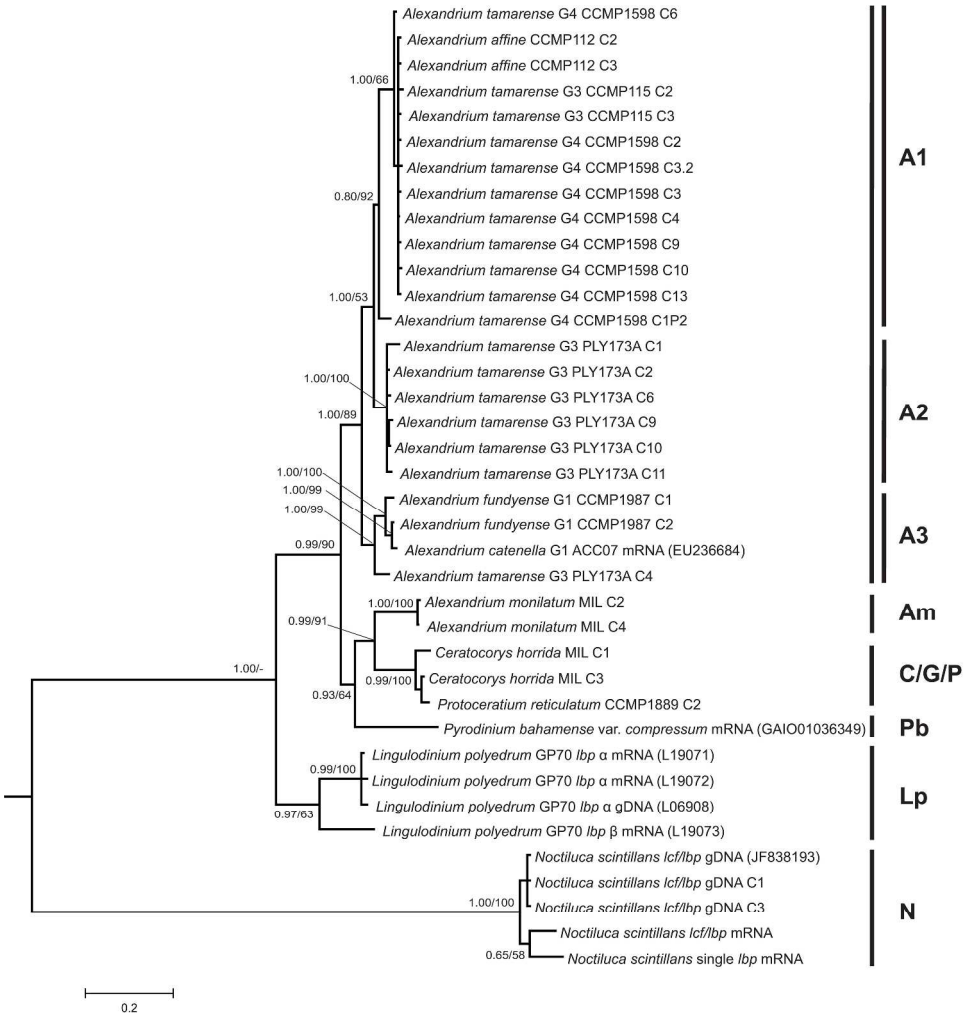

Figure 2B
334x336mm (300 x 300 DPI)

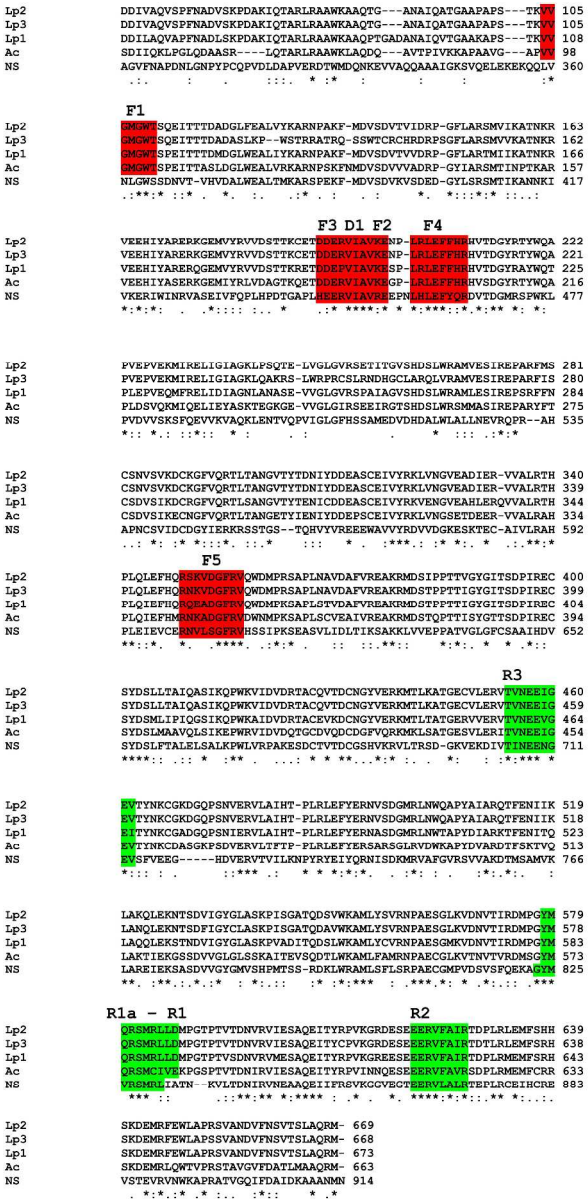

Figure S1
351x498mm (300 x 300 DPI)

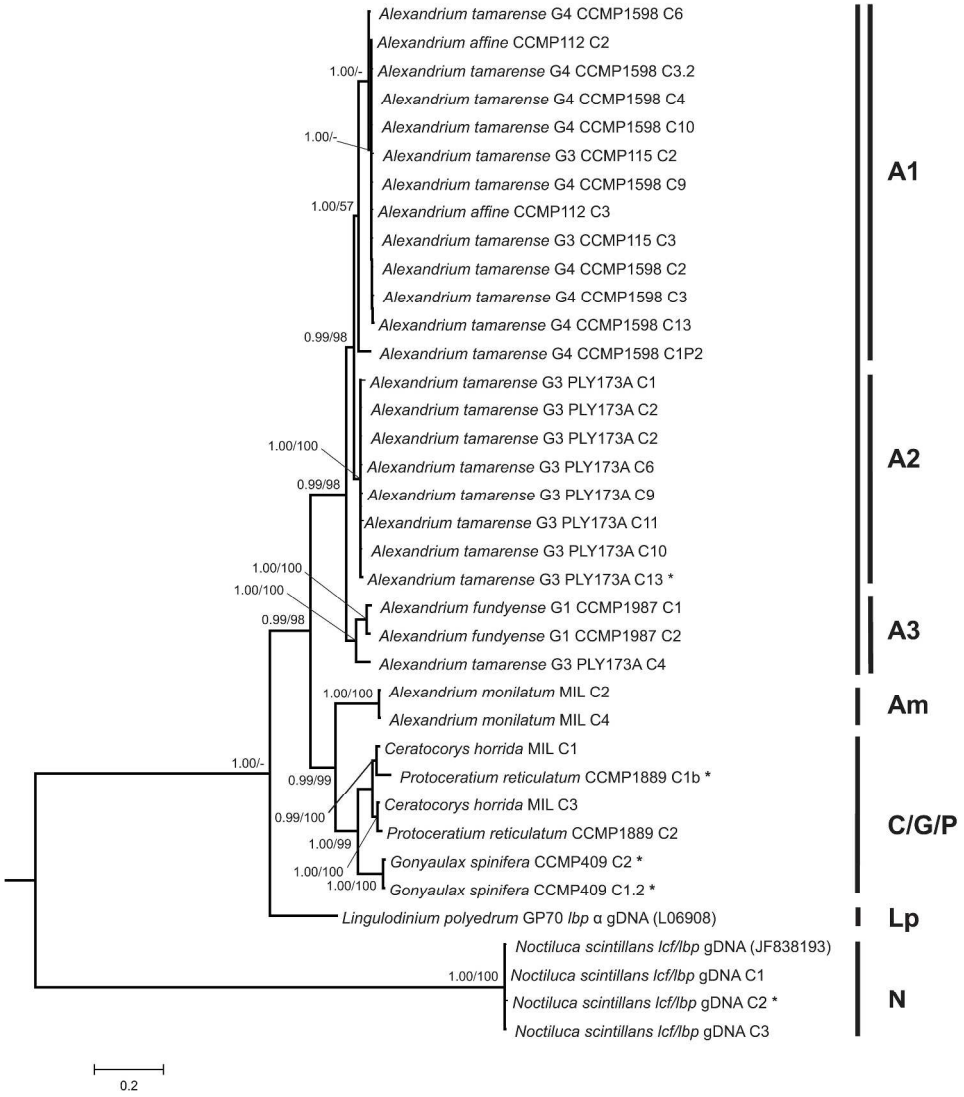

Figure S2
326x355mm (300 x 300 DPI)