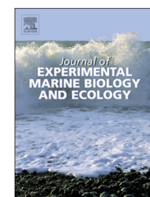
Journal of Experimental Marine Biology and Ecology 477 (2016) 31–39

Contents lists available at ScienceDirect

Journal of Experimental Marine Biology and Ecology

journal homepage: www.elsevier.com/locate/jembe



Genetic analysis of *Noctiluca scintillans* populations indicates low latitudinal differentiation in China but high China–America differences

Yongbo Pan [a,b], Lu Wang [a,b], Wenjing Zhang [a,b,c,]*, Guangxing Liu [d], Senjie Lin [a,b,c,]*

[a] State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen 361005, China
[b] Marine Biodiversity and Global Change Research Center, Xiamen University, Xiamen 361005, China
[c] Department of Marine Sciences, University of Connecticut, Groton, CT 06340, USA
[d] Department of Environmental Science, Ocean University of China, Qingdao, Shandong 266100, China

ARTICLE INFO

Article history:
Received 31 December 2015
Received in revised form 4 January 2016
Accepted 4 January 2016
Available online xxxx

Keywords:
Noctiluca scintillans
ITS rDNA
Polymorphism
Genetic distance
Phylogeny
Biogeography

ABSTRACT

Noctiluca scintillans is a globally distributed dinoflagellate that often forms a "Red Tide" bloom, but its geographically genetic differentiation is not well understood. In this study, three loci of the rRNA cistron were analyzed for populations from China and a population from Port Aransas, Texas, USA. Both phylogenetic and haplotypes network analyses based on ITS, SSU and LSU DNA markers revealed low genetic variation within and between populations along the Chinese coast, but substantial divergence based on ITS marker between the Chinese populations and the American population. Twelve single-cells were examined for intra-cell genetic variation and high levels of polymorphism were detected. The large genetic distance among the intra-cell ITS sequences (0–0.043) can exceed the species boundary (0.02 or 0.04) previously reported for dinoflagellates. High genetic diversity of intra-genomic, with the mean genetic distance being 5.1 folds greater than mean inter-genomic distance in the China Sea region, suggests high levels of gene flow and high frequency of recombination, while the divergence between the Chinese and the American populations indicates limited gene flow between the two continents. The present study suggests that the red *N. scintillans* is globally ubiquitous but genetically showing inter-region differentiation.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The large-sized heterotrophic dinoflagellate *Noctiluca scintillans* is an important bloom forming species that been reported from coastal areas worldwide (Baek et al., 2009; Gomes et al., 2014; Huang and Qi, 1997; Kopuz et al., 2014; Mikaelyan et al., 2014; Padmakumar et al., 2010; Quevedo et al., 1999). Although this species does not produce toxins, it is cataloged as a harmful algal species (HAB) as it can cause fish and marine invertebrate kills through oxygen depletion, gill clogging or the release of high levels of ammonia in the surrounding waters (Elbrächter and Qi, 1998; Okaichi and Nishio, 1976). Furthermore, the food vacuoles of this species have been found to contain toxigenic microalgae of the genera *Dinophysis* and *Pseudo-nitzschia*, suggesting that *N. scintillans* may act as a vector of phycotoxins to higher trophic levels (Escalera et al., 2007).

Only one species (*N. scintillans*) has been recognized in the genus *Noctiluca*, but two forms including the red *Noctiluca* and the green *Noctiluca* have been reported (Basu et al., 2013; Elbrächter and Qi, 1998; Hansen et al., 2004; Thompson et al., 2009). Red *Noctiluca* is heterotrophic, while the green *Noctiluca* can survive for a period of time without an external food supply for it contains a photosynthetic symbiotic *Pedinomonas noctilucae* (Harrison et al., 2011). Both of these two forms can form blooms. Green *Noctiluca* has been documented mainly in Southeastern Asia, and in recent years increasingly in the Arabian Sea (Gomes et al., 2014), while blooms of the red *Noctiluca* are more geographically widespread. In the China Seas, red *Noctiluca* blooms have been commonly found from the north to the south, including Bohai Bay, the Yellow Sea, the East China Sea, and the South China Sea (Huang and Qi, 1997; Jingzhong et al., 1985; Liu et al., 2013; Sun et al., 2010). Overall, *Noctiluca* has been found to occur over a wide temperature range, normally 10 °C to 25 °C in temperate regions, but as low as 3 °C off Alaska (Tibbs, 1967), and up to 33 °C off Mangalore in the Arabian Sea (Sulochanan et al., 2014). Such a wide temperature range for this species suggests that several genotypes with different temperature preferences may have evolved in different geographic regions (Harrison et al., 2011), however, information on genetic differentiation among different geographic populations is limited.

To examine whether there is genetic differentiation among *N. scintillans* populations living in different climes (temperature or geographic strains), a molecular analysis for *N. scintillans* from a wide latitudinal range in China was conducted. Currently, the most commonly used molecular markers for resolving species in

* Corresponding authors at: State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen 361005, China and Department of Marine Sciences, University of Connecticut, Groton, CT 06340, USA.
E-mail addresses: zhangwenjing@xmu.edu.cn (W. Zhang), senjie.lin@uconn.edu (S. Lin).

http://dx.doi.org/10.1016/j.jembe.2016.01.002
0022-0981/© 2016 Elsevier B.V. All rights reserved.

dinoflagellates have included the small subunit (SSU rDNA) (Gomez et al., 2010; Scholin et al., 1993), the large subunit (LSU rDNA) (Ki, 2010; Scholin et al., 1994), and the internal transcribed spacer (ITS) (Adachi et al., 1996; Wang et al., 2014) of the ribosomal RNA gene (rDNA), mitochondrial Cytochrome Oxidase I gene (COI) (Stern et al., 2010), and Cytochrome Oxidase B gene (COB) (Lin et al., 2009). Among these markers, ITS has the greatest sequence variability and is potentially most suitable for detecting genetic differentiation between geographic populations. Litaker and colleagues demonstrated the suitability of ITS for separating species in major dinoflagellate lineages while showing some level of sequence variation among intra-specific strains (Litaker et al., 2007). A more recent study surveyed a wider range of dinoflagellate species based on ITS and suggested that this gene could be a useful species marker for most dinoflagellate genera (Stern et al., 2012). Applications of the ITS marker as a species DNA barcode to *Alexandrium tamarense* species complex have led to the recognition of five species in this complex (Wang et al., 2014). In all of these studies, inter-strain variability in ITS was evident even though not as well defined as inter-species boundaries. Thus, the variation within the ITS region may provide useful insights about the differentiation of *Noctiluca* populations. Therefore, in this study the ITS marker was used to detect whether intra-specific variation or differentiation had occurred within red *Noctiluca* populations along the Chinese coast and between the Chinese coast (western North Pacific) and the American coast (western North Atlantic) as a result of geographical isolation. Complete ITS sequences were obtained and analyzed. For a comparison, partial SSU and LSU rDNA sequences were also analyzed.

2. Materials and methods

2.1. Sample collection

Seven locations were chosen to collect samples in this study, three from the Yellow Sea of China, two from the South China Sea, one from the East China Sea, and another one from Port Aransas, USA (Table 1). At the Chinese coast sampling locations, samples were collected using a plankton net with a filtering cod end by vertical net tows from 10-m depth to the surface, then concentrated and immediately fixed in absolute ethanol. The samples were transported to the laboratory and stored at 4 °C until analysis. For the American location, a culture (strain NS3) was isolated from a sample collected from Port Aransas, Texas (provided by Dr. Edward Buskey), which was grown with *Dunaliella tertiolecta* supplied as food (Zhang et al., 2007).

2.2. Single-cell isolation and DNA extraction

From the samples collected in China, single cells were isolated with a pipette under the microscope for PCR amplification. Each single cell was washed three times in absolute ethanol and pipetted into a 0.2-mL PCR tube. A total of 114 cells were isolated, among them, 86 were used for ITS (ITS1/5.8S/ITS2 and partial LSU rDNA) amplification and 28 were used for SSU rDNA amplification (Table 1). PCR amplification was conducted after ethanol evaporation. For the American sample (strain NS3), cells were harvested when the culture had depleted the food alga *D. tertiolecta*, and DNA was extracted following reported CTAB protocol (Zhang et al., 2007). The DNA extract was used as the template for PCR.

2.3. PCR, cloning and sequencing

PCR was performed using universal SSU rDNA eukaryotic primers 18ScomF-3end and com28SR1 (Table 2) designed from the conserved 3′ region of SSU and the 5′ region of LSU. This primer set was expected to amplify a ~ 1.5 kb DNA fragment containing partial SSU, the entire ITS (ITS1/5.8S/ITS2) and partial LSU rDNA. PCR amplification in a total volume of 50 μL was under the following conditions: 94 °C for 5 min, 35 cycles of 15 s at 95 °C, 30 s at 56 °C and 80 s at 72 °C, followed by an additional step of extension at 72 °C for 10 min. The secondary (semi-nested) PCR was performed using the product (after a 200-fold dilution) as the template using 18ScomF-3end and com28SR2 as the primers to achieve a sufficient amount of PCR product for directly sequencing, using similar PCR conditions as described above. Additionally, a species-specific primer set, 18SnoctiF and 28SnoctiR, was designed to amplify only the *N. scintillans* ITS gene. To obtain multi-gene phylogenies for comparison, SSU rDNA was amplified using the primer set 18ScomF1-18SnoctiR1 in the first PCR, and 18ScomF1-18SnoctiR2 in the second PCR reaction (semi-nested). In all the PCR reactions, high pru EX Taq (TAKARA) was used to ensure high fidelity of DNA replication. PCR products were then purified by TIAN gel Midi Purification kit (TIANGEN), and directly sequenced from both ends with corresponding primers at Shanghai Sangon Biological Engineering Technology & Services Co, Ltd., China. For each sampling location in China, one or two purified PCR products, each from a single cell, were randomly selected to clone into a T-vector (p-GEM-T easy vector cloning kit, Promega) and 5 – 17 clones were sequenced for each cell. For the American strain, the PCR product was cloned and 8 clones were sequenced.

2.4. Examination of chimeras and pseudogenes

PCR amplification can sometimes generate chimeric DNA, which would lead to overestimation of the degree of intra-genomic polymorphism or biodiversity. Using the online program Ballerophon (http://comp-bio.anu.edu.au/bellerophon/bellerophon.pl) with the default settings (300 bp window), the potential of each sequence being a chimera was examined. The results were further manually checked using previously reported criteria (Miranda et al., 2012). Another possible confounding factor is the existence of rRNA pseudogenes, which are well-known within dinoflagellates genomes (Thornhill et al., 2007). Mutations or significant indels in conserved regions are indicative of pseudogenes (Scholin et al., 1993; Thornhill et al., 2007), were used as criteria to define pseudogenes.

2.5. Sequence quality check and polymorphism analysis

Cloned sequences of each single-cell were aligned using ClustalX software 2.0 (Larkin et al., 2007) using default settings. Based on the alignment, sequences were inspected against directly sequenced

Table 1
Information of sampling locations in this study.

Sampling location	Origin	Collect time	Longitude	Latitude	Temperature (°C)	Salinity	Number of cells amplified (ITS)	Number of cells amplified (18S)
SN3	Port Aransas of USA	2004.11	−97.05	27.8383	23.0	23.0	–	–
ZY	Yellow Sea of China	2013.12	120.2605	36.0333	14.0	30.3	11	8
ZH	Yellow Sea of China	2012.10	120.2352	36.0245	13.8	29.9	6	6
OF	Yellow Sea of China	2013.12	120.2993	36.0235	12.9	29.8	16	2
XM	East China Sea	2013.8	118.0992	24.4089	28.2	29.5	19	8
BO	South China Sea	2011.6	108.2821	21.2338	31.5	30.5	17	2
BT	South China Sea	2011.6	109.0724	21.1206	30.3	30.7	17	2

Table 2
List of primers used in this study.

Primer name	Primer sequence 5′ → 3′	Source	Annealing temperature
18ScomF-3end	GTCGTAACAAGGTTTCCGTAGGTG	Bai et al. (2002)	56 °C
Com28SR1	TCACGCATAGTTCACCATCTTTCG	Wang et al. (2014)	56 °C
Com28SR2	TTAGACTCCTTGGTCCGTGTTT	Wang et al. (2014)	56 °C
18ScomF1	GCTTGTCTCAAAGATTAAGCCATGC	Zhang et al. (2005)	56 °C
18SnoctiR1	GAATGATCCTTCCGCAGGTT	This study	56 °C
18SnoctiR2	TTCTCCTTCCTCTAAGTGATGAG	This study	56 °C
18SnoctiF	GTCGTAACAAGGTTTCCGTAGGTG	This study	58 °C
28SnoctiR	GTTTCTTTTCCTCCGCTTACTTAT	This study	58 °C

chromatographs to correct possible PCR and sequencing errors, and to identify the intra-genomic polymorphic sites, following a previously reported procedure (Wang et al., 2014) with a small modification. The modification essentially identified a mutation that occurred more than twice at the same site within one genome as true intra-genomic polymorphism. The number of polymorphic sites were determined using DnaSP 4.10.7 software (Rozas et al., 2003). The quantity and rate (%) of intra-genomic polymorphisms were presented graphically using Weblogo (http://weblogo.threeplusone.com/). As the clones analyzed were from a culture of *Noctiluca* from America, rather than single cells, the polymorphism detected was classified as intra-strain polymorphism which could exist either within a single cell or between cells in the culture.

2.6. Phylogenetic analysis and haplotype network calculated

Phylogenetic trees were performed to examine the biogeographical distribution of *Noctiluca*. Only three sequences in GenBank were downloaded and analyzed, including one SSU + ITS + LSU sequence (GenBank accession no. GQ380592.1, South China Sea), and two SSU sequences (GenBank accession no. AF022200.1 from Canada and DQ388461.1 from Long Island Sound, USA). Two SSU sequences (GenBank accession no. AB297469.1, AB297470.1) were not included in the analyses because they were too short and contained a series of gaps compared to the sequences obtained by the present study. MrModeltest 2 (Nylander, 2004) was used to find the optimal model, which indicated that the HKY + I model was suitable for ITS and SSU, while GTR + I model was suitable for LSU. These models were used for all methods of phylogenetic analyses unless indicated otherwise.

Bayesian phylogenetic analysis was implemented in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). Two simultaneous Markov chain Monte Carlo (MCMC) chains were carried out for 6,000,000 generations, by the end of which the standard deviation of split frequencies was below 0.01. Sampling frequency was one every 100 generations and the burn-in value was set at 25%. Maximum likelihood (ML) tree was constructed in MEGA 6.0 (Tamura et al., 2013) with 1000 bootstrap replicates. A genetic distance file was constructed using the F84 model (Felsenstein and Churchill, 1996) in the DNADIST software in PHYLIP package 3.6a. (Plotree and Plotgram, 1989). Neighbor-joining (NJ) trees were inferred with 1000 bootstrap resamplings. After that, the CONSENSE software was used to generate a consensus tree. All the phylogenetic trees were visualized with TreeView 1.6.6 (Page, 2001).

Haplotype networks for directly sequenced sequence were calculated in TCS v1.21 (Clement et al., 2000) using statistical parsimony by the 95% plausible connection limit, and gaps were treated as the fifth state of the character.

2.7. Calculation of intra- and inter-genomic genetic distances

To investigate the divergence of *Noctiluca*, intra- and inter-genomic genetic distances were calculated based on ITS sequences. A previously reported distance cutoff values (0.02 or 0.04) as species boundary (Litaker et al., 2007; Stern et al., 2012) were used as reference. To ensure comparability, the same software and parameters were used in sequence alignment. The resulting alignments were saved as NEXUS files and imported into PAUP*4b10 software (Swofford, 2002) to calculate genetic distances using simple uncorrected pair-wise (p) distance matrices as previously reported (Wang et al., 2014).

3. Results

3.1. Sequencing results

With all the five primer sets used in this study (Table 2), all the first step PCR amplification did not work well, likely due to low DNA concentration within single-cells, while all of the primer combinations gave good results in the second (semi-nested) PCR. After gel recovery and direct sequencing, 28 partial SSU (1714 bp), 86 entire ITS (534 bp) and 16 partial LSU (716 bp) sequences were obtained. Additionally, three SSU and nine ITS (eight single cells and strain SN3) PCR products were randomly selected for cloning, and in total 36 SSU, 81 ITS and 94 LSU clones were sequenced, respectively (Table 3).

3.2. Identification of chimeras and pseudogenes

No chimeras were found from the direct sequencing results, but one, three and four of the SSU, ITS and LSU clones, respectively were identified as chimera candidates. The frequency of chimeric DNA was very low (< 1.6%, < 3.7% and <4.3%). Besides, three cloned sequences of ITS were suspected to be potential pseudogenes, for these sequences contained more than a 30-bp deletion within the ITS1 region. One of the SSU cloned sequences also contained a series of mutations and small indels, and was thus classified as a pseudogene. All of these chimeric sequences and pseudogene were excluded from further analyses.

3.3. Characterization of genomic polymorphism

From each of the 12 clone libraries 5 to 17 sequences were recovered (Table 3). The result showed that each of the clone libraries were usually comprised of one highly represented sequence and several unique (i.e. single-clone) sequences. For ITS, the eight single-cell genomes from the Chinese populations and the strain NS3 showed 1 to 25 polymorphic sites, which were scattered across the ITS1 and ITS2 region, with only three sites in the 5.8S region (Fig. 1). When all these sequences were compared, the total number of polymorphic sites increased to 50. Most of these polymorphic sites were nucleotide substitutions, with only four sites of deletion occurring at the end of ITS2. Among the deletions, three continuous deletions (positions 522 to 524) occurred in the NS3 strain from Port Aransas of American, and one (position 525) shared by several sequences of the cell OF_1, which originated from the Yellow Sea of China. Within any given genome, differences between sequences were characterized primarily by substitution of several bases. When inter-genomes were compared between the same or different sampling area of China, the trend was similar, however, there were conserved difference at certain positions between the Chinese and the NS3 strains. For example, at site 48, all the sequences from the Chinese populations have a C while the sequences from the NS3 strain have A. This was true also for sites 363 and 491. Other sites showed less clear

Table 3
Intragenomic or intra-strain polymorphisms in ITS and rDNA sequences based on 12 single cells/strain.

Cell name	Origin	Partial 18S (1714 bp)				Complete ITS (534–535 bp)					Partial 28S (716 bp)			
		N	NU	V	V (%)	N	NU	V	D	V (%)	N	NU	V	V (%)
SN3*	Port Aransas of USA	–	–	–	–	8	6	0–5	3	0–0.94	–	–	–	–
ZY_1	Yellow Sea of China	–	–	–	–	10	7	0–6	0	0–1.12	10	5	0–3	0.42
ZY_2	Yellow Sea of China	–	–	–	–	12	9	0–12	0	0–2.24	16	1	0	0
ZY_3	Yellow Sea of China	13	10	0–12	0.70	–	–	–	–	–	–	–	–	–
ZH_1	Yellow Sea of China	–	–	–	–	6	2	0–1	0	0–0.19	6	1	0	0
OF_1	Yellow Sea of China	–	–	–	–	6	2	0–14	1	0–2.62	6	3	0–19	2.65
XM_1	East China Sea	–	–	–	–	11	8	0–26	0	0–4.86	16	6	0–12	1.68
XM_2	East China Sea	–	–	–	–	7	4	0–4	0	0–0.75	7	2	0–1	0.14
XM_3	East China Sea	11	4	0–3	0.18	–	–	–	–	–	–	–	–	–
BO_1	South China Sea	–	–	–	–	8	4	0–15	0	0–2.80	17	3	0–14	1.96
BT_1	South China Sea	–	–	–	–	14	10	0–24	0	0–4.49	16	8	0–19	2.65
BT_2	South China Sea	12	9	0–13	0.70	–	–	–	–	–	–	–	–	–

Abbreviations: N, number of sequences; NU, number of unique sequences; V, variable sites; D, deletion sites; V (%), variation rate (%).
* This is a strain.

geographic differentiation; for instance, at site 145, the NS3 strain have a T while the sequences of cells ZY_1, ZH_1, OF_1 and XM_2 from China have a C, and cells ZY_2, XM_1, BO_1 and BT_1 from China had either a T or a C nucleotide. Such transitional substitution occurred in many sites.

For a comparison, the degree of intra-genomic nucleotide polymorphism of SSU and partial LSU was also examined, and both of them exhibited high nucleotide polymorphism. The variation rate of these genes ranged 0 – 0.70% and 0 – 2.65%, respectively, lower than that of ITS, which ranged 0 – 4.86% (Table 3).

3.4. Phylogenetic analysis and haplotype network construction

Five major groups were identified based on 87 ITS directly sequenced sequences from single cell (86 new and 1 previously published) and 8 cloned sequences from strain NS3 (for no single cell was available for analysis for this American population). This grouping received varying bootstrap support values (BP) and posterior probability (PP) values (Fig. 2A, only major nodes are shown). All of the groups

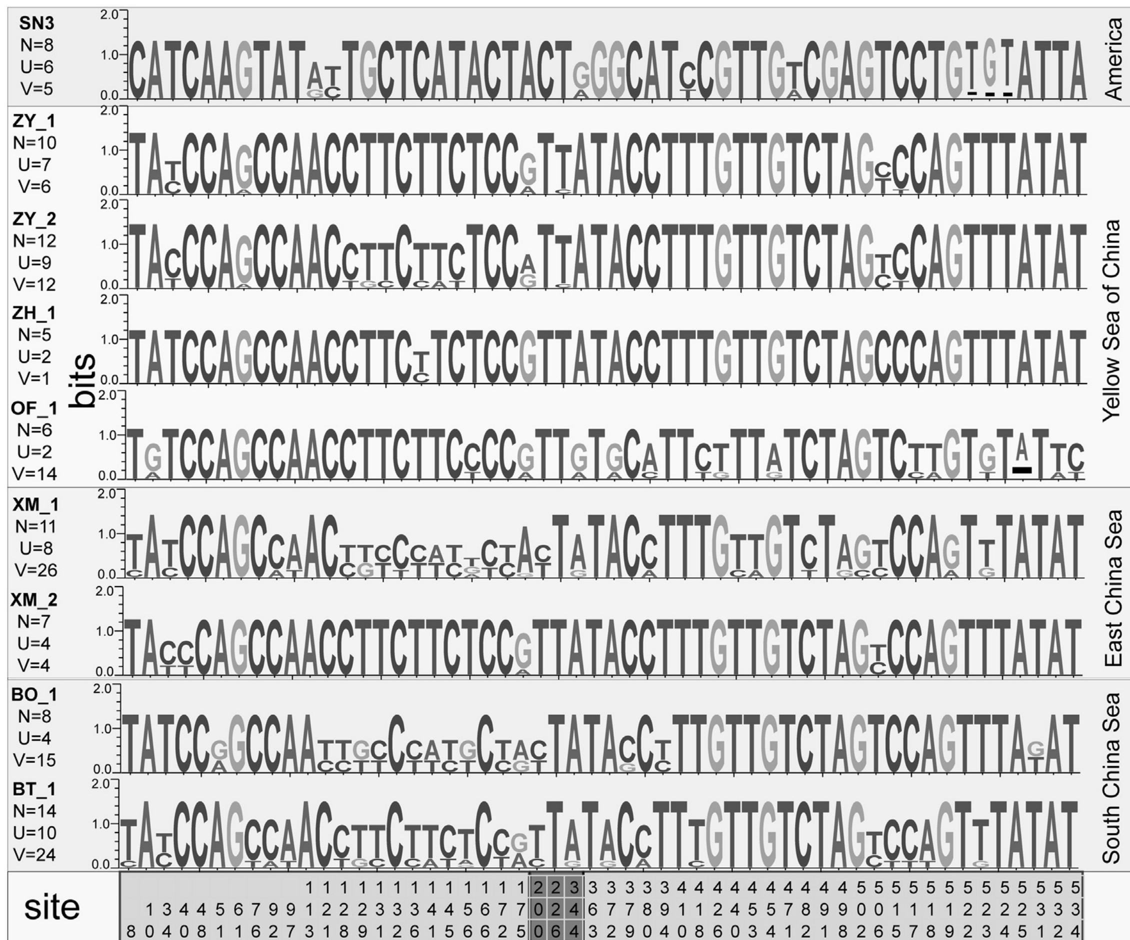


Fig. 1. ITS intra/inter-genomic polymorphisms in eight single-cells and strain SN3. Shown on the left are the cell/strain label, the total number of cloned sequence (N), the total number of unique cloned sequence (U) and the number of variation site (V) they contain; on the right is the sampling area; while on the bottom are the sites of this gene (total 534–535 bp). The overall height of the stack indicates the sequence conservation at that site, while the height of symbols within the stack indicates the relative frequency of the nucleic acid at that site. Two bits correspond to 100% conservation. "–"depicts a deletion.

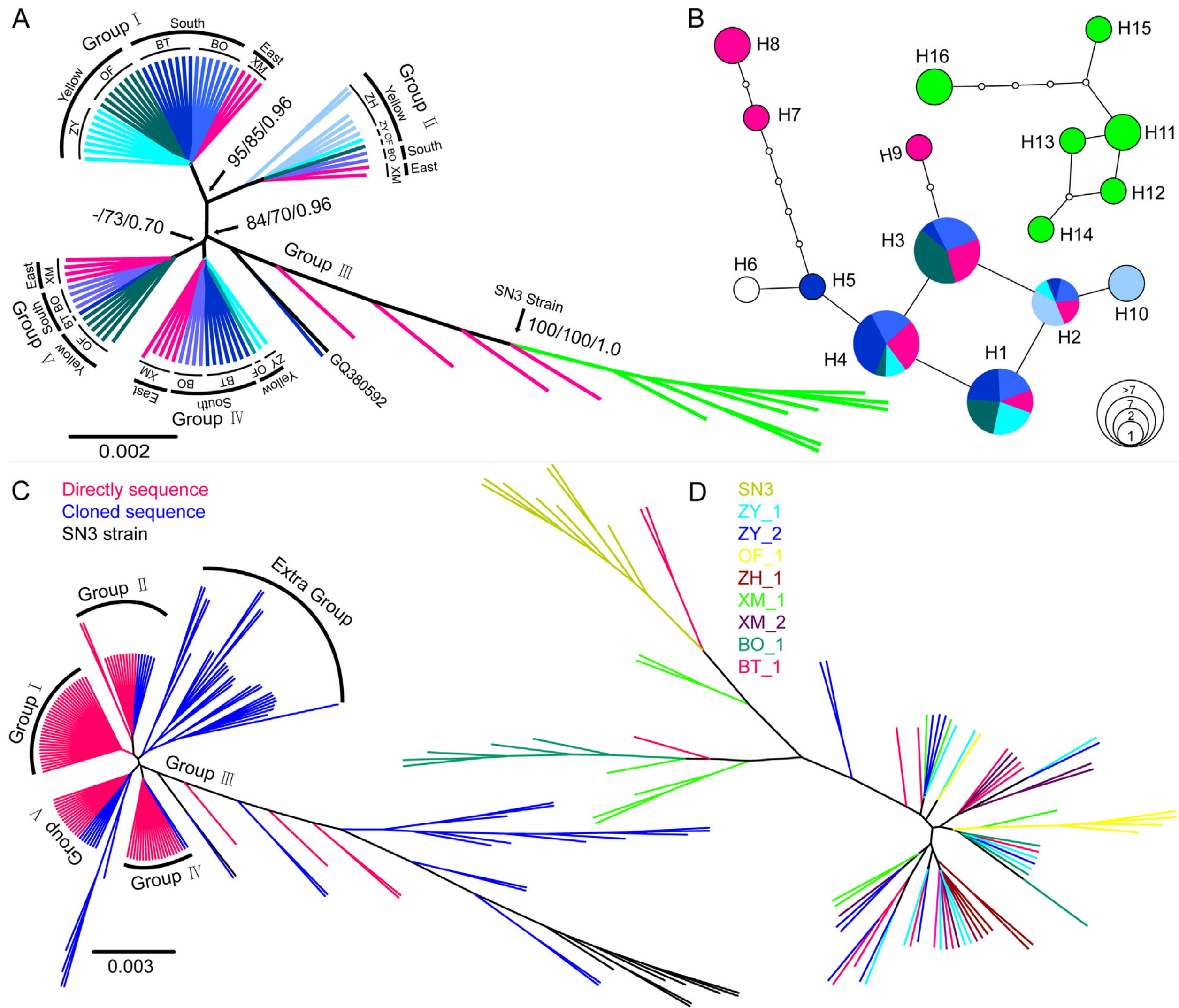


Fig. 2. Phylogenetic tree (A, C, D) and haplotypes network (B) constructed from ITS sequences. Fig. A based on directly sequenced sequences, the numbers on the nodes represent the bootstrap support values (BP)/posterior probabilities (PP) produced by the Neighbor-joining, Maximum-likelihood analyses and Bayesian inference methods respectively. Values are shown only on the major nodes, and lower than 50% or 0.50 are omitted. Clades and haplotypes are colored according to the respective sampling location. The scale bar indicates two substitutions per 1000 sites. The sizes of circles representing haplotypes reflect the number of sequences that share a haplotype. Fig. C represents the relationships between directly sequences and cloned sequences, while D represents the relationships within 8 single-cell plus strain SN3. Clades are colored according to the respective single-cell. Fig. C and D are drawn to the same scale bar denotes three substitutions per 1000 sites.

were composed of mixed strains. Group I consisted of 37 cells — 16 were from the Yellow Sea of China, 15 were from the South China Sea, and 6 were from the East China Sea. Group II was sister to Group I with strong support values (95, 85, 0.96), containing 12 cells — 8 were from Yellow Sea of China, 2 were from the East China Sea and 2 were from the South China Sea. Group III was composed of many highly divergent clusters and branched with Groups IV and V with poor support values (84, 70, 0.94). This group contained a strain from Hong Kong (GenBank accession no. GQ380592), 4 clades from the East China Sea, and 1 clade from the South China Sea. Additionally, the American strain SN3 formed an independent clade in this group showing higher divergence relative to other clades, with strong support values (100, 100, 1.0). While Groups IV and V were sister groups with poor support values (–, 73, 0.70) both composed vary cells from the three sampling areas in the China Seas.

Within these 95 ITS sequences, a total of 16 haplotypes were identified (Fig. 2B). Among them, four major haplotypes were shared by cells within the China Sea regions (H1–H4), while other haplotypes were unique but occurred only once to twice (H5–H9). The American strain sequences belonged to six different haplotypes (H10–H16) and formed an independent haplotype network, which was not associated with the China haplotype network, indicating high divergence (Fig. 2B).

Whether the 81 ITS clone sequences from nine clone libraries could cluster together to form a coherent single-cell/strain clade was also examined. The result showed that all the sequences were randomly distributed in the phylogenetic tree except cell ZH_1 (N = 6) and the strain SN3 (N = 8) (Fig. 2D). When the directly sequenced sequences were combined with these cloned sequences to infer the phylogenetic tree, an extra group composed of cloned sequences was separated (Fig. 2C). Compared to the directly sequenced sequences, cloned sequences formed varying length blanches, indicating high genetic diversity within single-cell/strain rather than between cells.

In comparison, only six and four haplotypes were identified based on the 31 SSU (28 new and 3 previously published) and 17 LSU (16 new and 1 previously published) directly sequenced sequences (Fig. 3A and D), respectively, however, similar to ITS, cloned sequences of both SSU and LSU exhibited high genetic divergence than directly sequenced sequence. In other words, the genetic variations within single-cells were higher than between cells (Fig. 3B, C, E, and F).

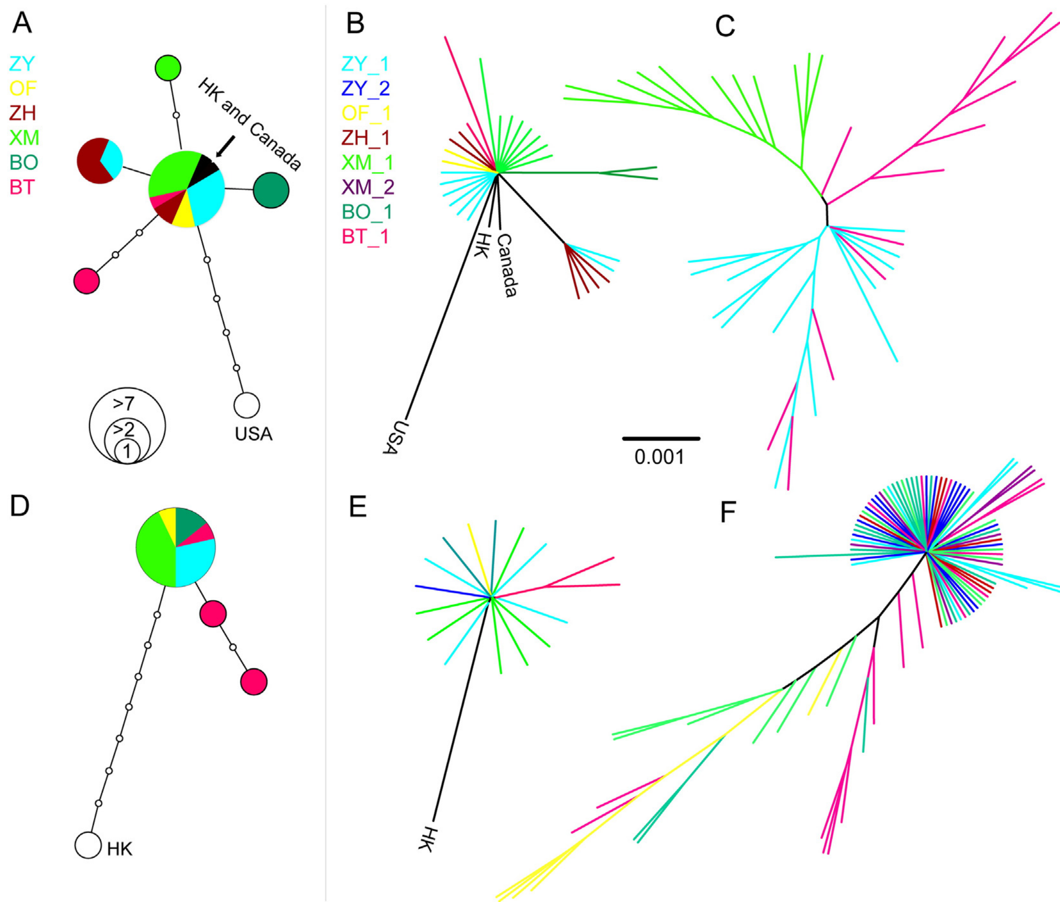


Fig. 3. Haplotypes network and phylogenetic trees constructed from SSU and LSU rDNA genes. Fig. A and B based on 31 SSU directly sequenced sequences; Fig. D and E based on 17 LSU directly sequenced sequences; Fig. C based on 36 SSU cloned sequences from three single-cells; Fig. F based on 94 LSU cloned sequences from eight single-cells, both of these trees and haplotypes were colored according to the respective sampling location/cells. The sizes of circles representing haplotypes reflect the number of sequences that share a haplotype. All of these four trees are drawn to the same scale to facilitate comparison of the rates of evolution between genes. The scale bar indicates one substitution per 1000 sites.

3.5. Intra- and inter-genomic genetic distances

These results showed that the intra-genomic genetic distance (p value) based on ITS in the eight cells isolated from the Chinese coast ranged from 0 to 0.043 (Table 4). The distances in four of these cells exceeded the 0.02 species boundary proposed previously (Stern et al., 2012), including OF_1, XM_1, BO_1 and BT_1, among which distances in XM_1 and BT_1 exceeded the more stringent species boundary 0.04 proposed earlier (Litaker et al., 2007), however, the mean values of these single-cells only ranged from 0.001 to 0.020. Similarly, the p values among all the Chinese populations (86 cells), within or between sampling locations, only ranged from 0 to 0.019, all less than 0.02 (Table 4).

In contrast, the genetic distances within the NS3 strain (i.e. including intra- and inter-genomic distances) were very low, ranging from 0 to 0.009, however, large distances were found between the Chinese populations and the NS3 strain, ranging from 0.032 to 0.056. In addition, when the divergence of cloned sequences between directly sequenced sequences within China was compared, the mean p value among cloned sequences (totally 73 sequences from 8 single-cells of China) was 0.015 (intra-genomic), while that among the 86 directly sequenced sequences was 0.003 (inter-genomic), and the ratio between them was around 5.1.

Table 4
Estimated genetic distances between and within genomes based ITS.

Between genome					Within genome				
Location	Number of cells	Maximum	Minimum	Mean	Cell/strain name	Number of sequences	Maximum	Minimum	Mean
SN3*	–	–	–	–	SN3*	8	0.009	0	0.004
ZY	11	0.004	0	0.001	ZY_1	10	0.008	0	0.004
–	–	–	–	–	ZY_2	12	0.019	0	0.007
ZH	6	0.002	0	0.001	ZH_1	6	0.002	0	0.001
OF	16	0.004	0	0.002	OF_1	5	0.026	0	0.011
XM	19	0.019	0	0.006	XM_1	11	0.041	0	0.020
–	–	–	–	–	XM_2	7	0.006	0	0.003
BO	17	0.004	0	0.002	BO_1	8	0.026	0	0.015
BT	17	0.004	0	0.001	BT_1	14	0.043	0	0.015

* Not from a single cell, and the distance calculated was attributed to within or between genomes.

4. Discussion

4.1. Intra-genomic polymorphism

Intra-genomic polymorphism in dinoflagellates has been reported in many previous studies (Ki and Han, 2007; Miranda et al., 2012; Thornhill et al., 2007; Wang et al., 2014). For *Noctiluca*, intra-specific variation has been analyzed for SSU and LSU in Hong Kong strains only (Ki, 2010). The present study provides the first systematic profile of SSU, ITS and partial LSU rDNA intra-genomic polymorphism in *Noctiluca*. Intra-genomic sequence variation was substantial in many cases, in terms of both the number of polymorphic sites and the number of haplotypes. For example, the highest polymorphism occurred in the cell of XM_1 reaching 26 SNPs (Fig. 1). The clone sequence variation rate ranged from 0.14% to 4.86%, significantly higher than error rates usually expected of Taq polymerase (~1.21×10^{-4} errors/nt for normal Taq or ~5×10^{-5} errors/nt for High Fidelity Platinum Taq (Barnes, 1992; Cariello et al., 1991; Thornhill et al., 2007)), and also than that of the replication error of 1×10^{-9} in *Escherichia coli* (Haber, 1999). Besides, a quality control procedure in identifying polymorphic sites to prevent overestimation of polymorphism was used. Therefore, the sequence diversity reported here was most likely true genetic variations within genome.

4.2. Divergence and geography distribution of Noctiluca

The primary objective of this study was to evaluate whether intra-specific variation or differentiation had happened within red *Noctiluca* in association with geographic or latitudinal (temperature-related) isolation. Both of the phylogenetic trees and haplotypes network based on SSU, ITS and LSU revealed low variation between cells in the China Sea region, but a high divergence in comparison to the American strain based on ITS. Genetic distance based on ITS also revealed the same regularly. The high divergence between Chinese and American populations had exceeded the species cut-off value threshold (0.02 or 0.04) (Litaker et al., 2007; Stern et al., 2012). If these thresholds were used, *Noctiluca* in the two regions should be considered as different species; however, four of the eight single-cells analyzed in this study also exceeded the species boundary, indicating impact that ITS intra-genomic polymorphism can have on dinoflagellate taxonomy (Gribble and Anderson, 2007; Harrington et al., 2014; Thornhill et al., 2007). Phylogenetic analysis also showed that some of the cloned sequences from the Chinese and the American populations were affiliated within the same cluster (Fig. 2D). All these indicate that the wide geographic isolation has led to significant genetic divergence between the Chinese and the American populations, but the divergence is not sufficient to result in speciation. Meanwhile, the finding of the substantial intra-genomic variations suggests that the use of genetic distance based on ITS, SSU or LSU rDNA sequence in discriminating species and geographic populations must be done with the dominant sequence instead of a random one.

In the present study, only 16 (6 from American sequences), 6 and 4 haplotypes were identified based on ITS ($N = 95$), SSU ($N = 31$) and LSU ($N = 17$), respectively. Some studies have shown high degree of genetic diversity within other protists by using COI gene. For instance, *Carchesium polypinum* contained 29 haplotypes ($N = 100$) mainly sampled from Grand River Basin (Canada) (Gentekaki and Lynn, 2012), *Tetrahymena thermophile* 24 haplotypes ($N = 165$) isolated from ponds in the USA (Zufall et al., 2013), and *Hyalosphenia papilio* 49 haplotypes ($N = 301$) collected in 42 peatlands in North America, Europe and Asia (Heger et al., 2013). Although high diversities were observed by these studies, the biogeography and diversity of different protists were different. Actually, there have been two different opinions about this topic for the last ten years. The ubiquity theory proposes that small cell sizes, huge populations, and almost unlimited dispersal of protists in the ocean contribute to their cosmopolitan distribution (Fenchel, 1993; Finlay et al., 2004), while the opponents argue for a high degree of global diversity and endemism of protist species (Foissner, 2006; Ryšánek et al., 2014). The results from the present study showed that the number of haplotypes within China were geographically (including latitudinally) quite homogeneous, while compared to the American population these populations were generally different (Fig. 2B), suggesting a basin or continental scale endemism for *Noctiluca*.

Mutation, environmental selection, gene flow and genetic drift have been recognized as the major processes that contribute to species' diversity, but gene flow and genetic drift further influence within-species genetic diversity (Hanson et al., 2012). At a large geographic scale, gene flow may be limited due to small population sizes, which would reduce the probability of migration events or the ability of a population to establish after migration (Zufall et al., 2013). In such a case, the high divergence between the Chinese and American populations would be no surprise. While at a fine scale, along the Chinese coast, the genetic divergences within cells were 5.1 times higher than that between cells, strong evidence that extensive recombination events have happened within the cells in the species. Actually, *Noctiluca* may be an outbreeder, and up to 1024 gametes can be produced from one gametogenic cell (Fukuda and Endoh, 2006). In this scenario, the rDNA repeats within or between cells could be recombined depending on whether gametes from the same or different gametogenic cells fuse to produce a new cell, however, over time the existence of concerted evolution should homogenize the repeat variant to either correct the mutations or change all the repeats to the new sequence (Ganley and Kobayashi, 2007). The high intra-genomic diversity of ITS suggests recent recombinations within the rDNA array without sufficient time for concerted evolution to homogenize the repeats. Consequently, the inter-genomic diversity can be lower than that within the recombined single-cell. It is noteworthy that in the present study, not only the ITS but also the SSU and LSU genes showed high intra-genomic diversity (Fig. 3), suggesting high frequency of recombination in this species. On the other hand, the relatively low genetic divergence among the Chinese populations indicate strong gene flow from the north to the south China Sea.

5. Concluding remarks

In the current study ITS, SSU and LSU rDNA markers were used to analyze the intra-specific variation or differentiation of red *N. scintillans*. Both phylogenetic trees and haplotype networks based on these genes revealed low within-region variation but high between-region (trans-Pacific Ocean) divergences. The high levels of intra-genomic polymorphism and genetic distance in these genes provide a guide to negate the possibility that the geographic isolation between the Chinese and the American populations of *Noctiluca* has led to speciation. They also suggest that the species discriminating genetic distance threshold (e.g. $p = 0.02$ or 0.04 based on ITS) should be used or modified with the intra-genomic genetic variation taken into consideration. A mean p value instead of a single p value should be used. If analysis of intra-genomic genetic distance is not feasible, the use of variation ITS, SSU or LSU rDNA sequence in discriminating species must be done with the dominant sequence (e.g. obtained from direct sequencing) instead of a random sequence (e.g. obtained by sequencing a random clone) (Miranda et al., 2012). High diversity within single-cell rather than between cells at a local geographic scale suggests high frequency of recombination, whereas the significant genetic divergence between the trans-Pacific regions suggests gene flow limitation. It is hence plausible that the biogeographic pattern of red *Noctiluca* fits the global ubiquity but showing inter-region differentiation.

Author contributions

Conceived and designed the experiments: S. L., W. Z. and L. W. Performed the experiments: Y. P., L. W. Analyzed the data: Y. P., L. W., W. Z. and S. L. Contributed reagents/ materials/ analysis tools:

W. Z., S. L. and G. L. Wrote the paper: Y. P., W. Z., S. L. and L. W. All authors reviewed the manuscript.

Acknowledgments

We thank Dr. Edward Buskey and Ms. Commie J. Hyatt from University of Texas at Port Aransas for kindly providing strain NS3 of *Noctiluca scintillans*, and thank Jun Yang from Aquatic EcoHealth, Institute of Urban Environment, Chinese Academy of Sciences and Edward A. D. Mitchell from Laboratory of Soil Biology, University of Neuchâtel for constructive comments on the manuscript. This study was supported by the National Natural Science Foundation of China (No. 41276133, No. 41129001), the Special Research Foundation for Public Welfare Marine Program (No. 201005012), the Natural Science Foundation of Fujian Province (2014J01163) and the Global Change and Air-sea Interaction Project (GASI-03-01-02-03), National Bureau of Oceanography of China. The sequences obtained in this study have been deposited in GenBank under the accession numbers KR527268-KR527331 and KR606824-KR607100 with more detailed information available in the Supplementary Material. **[SS]**

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jembe.2016.01.002.

References

Adachi, M., Sako, Y., Ishida, Y., 1996. Identification of the toxic dinoflagellates *Alexandrium catenella* and *Alexandrium tamarense* (Dinophyceae) using DNA probes and whole-cell hybridization. J. Phycol. 32 (6), 1049–1052.

Baek, S.H., Shimode, S., Kim, H.C., Han, M.S., Kikuchi, T., 2009. Strong bottom-up effects on phytoplankton community caused by a rainfall during spring and summer in Sagami Bay, Japan. J. Mar. Syst. 75 (1), 253–264.

Bai, F.-Y., Zhao, J.-H., Takashima, M., Jia, J.-H., Boekhout, T., Nakase, T., 2002. Reclassification of the *Sporobolomyces roseus* and *Sporidiobolus pararoseus* complexes, with the description of *Sporobolomyces phaffii* sp. *nov*. Int. J. Syst. Evol. Microbiol. 52 (6), 2309–2314.

Barnes, W.M., 1992. The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. Gene 112 (1), 29–35.

Basu, S., Deobagkar, D.D., Matondkar, S.G., Furtado, I., 2013. Culturable bacterial flora associated with the dinoflagellate green *Noctiluca miliaris* during active and declining bloom phases in the northern Arabian sea. Microb. Ecol. 65 (4), 934–954.

Cariello, N.F., Swenberg, J.A., Skopek, T.R., 1991. Fidelity of *Thermococcus litoralis* DNA polymerase (Vent™) in PCR determined by denaturing gradient gel electrophoresis. Nucleic Acids Res. 19 (15), 4193–4198.

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9 (10), 1657–1659.

Elbrächter, M., Qi, Z., 1998. Aspects of *Noctiluca* (Dinophyceae) population dynamics. In: Anderson, D.M., Cembella, A.D., Hallegraeff, G.M. (Eds.), Physiological Ecology of Harmful Algal Blooms. Springer, Berlin, Heidelberg, pp. 315–335.

Escalera, L., Pazos, Y., Moroño, Á., Reguera, B., 2007. *Noctiluca scintillans* may act as a vector of toxigenic microalgae. Harmful Algae 6 (3), 317–320.

Felsenstein, J., Churchill, G.A., 1996. A hidden Markov model approach to variation among sites in rate of evolution. Mol. Biol. Evol. 13 (1), 93–104.

Fenchel, T., 1993. There are more small than large species? Oikos 375–378.

Finlay, B.J., Esteban, G.F., Fenchel, T., 2004. Protist diversity is different? Protist 155 (1), 15–22.

Foissner, W., 2006. Biogeography and dispersal of micro-organisms: a review emphasizing protists. Acta Protozool. 45 (2), 111–136.

Fukuda, Y., Endoh, H., 2006. New details from the complete life cycle of the red-tide dinoflagellate *Noctiluca scintillans* (Ehrenberg) McCartney. Eur. J. Protistol. 42 (3), 209–219.

Ganley, A.R., Kobayashi, T., 2007. Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. Genome Res. 17 (2), 184–191.

Gentekaki, E., Lynn, D., 2012. Spatial genetic variation, phylogeography and barcoding of the peritrichous ciliate *Carchesium polypinum*. Eur. J. Protistol. 48 (4), 305–313.

Gomes, H., Goes, J., Matondkar, S., Buskey, E., Basu, S., Parab, S., Thoppil, P., 2014. Massive outbreaks of *Noctiluca scintillans* blooms in the Arabian sea due to spread of hypoxia. Nat. Commun. 5, 4862. http://dx.doi.org/10.1038/ncomms 5862.

Gomez, F., Moreira, D., Lopez-Garcia, P., 2010. Molecular phylogeny of noctilucoid dinoflagellates (noctilucales, dinophyceae). Protist 161 (3), 466–478.

Gribble, K.E., Anderson, D.M., 2007. High intra-individual, intra-specific, and inter-specific variability in large-subunit ribosomal DNA in the heterotrophic dinoflagellates *Protoperidinium*, *Diplopsalis*, and *Preperidinium* (dinophyceae). Phycologia 46 (3), 315–324.

Haber, J.E., 1999. DNA recombination: the replication connection. Trends Biochem. Sci. 24 (7), 271–275.

Hansen, P.J., Miranda, L., Azanza, R., 2004. Green *Noctiluca scintillans*: a dinoflagellate with its own greenhouse. Mar. Ecol. Prog. Ser. 275, 79–87.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., Martiny, J.B., 2012. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat. Rev. Microbiol. 10 (7), 497–506.

Harrington, T.C., Kazmi, M.R., Al-Sadi, A.M., Ismail, S.I., 2014. Intra-specific and intra-genomic variability of ITS rDNA sequences reveals taxonomic problems in *Ceratocystis fimbriata* sensu stricto. Mycologia 106 (2), 224–242.

Harrison, P., Furuya, K., Glibert, P., Xu, J., Liu, H., Yin, K., Lee, J., Anderson, D., Gowen, R., Al-Azri, A., 2011. Geographical distribution of red and green *Noctiluca scintillans*. Chin. J. Oceanol. Limnol. 29 (4), 807–831.

Heger, T.J., Mitchell, E.A.D., Leander, B.S., 2013. Holarctic phylogeography of the testate amoeba *Hyalosphenia papilio* (amoebozoa: arcellinida) reveals extensive genetic diversity explained more by environment than dispersal limitation. Mol. Ecol. 22 (20), 5172–5184.

Huang, C., Qi, Y., 1997. The abundance cycle and influence factors on red tide phenomena of *Noctiluca scintillans* (Dinophyceae) in Dapeng Bay, the South China Sea. J. Plankton Res. 19 (3), 303–318.

Jingzhong, Z., Liping, D., Baoping, Q., 1985. Preliminary studies on eutrophication and red tide problems in Bohai Bay. Hydrobiologia 127 (1), 27–30.

Ki, J.S., 2010. Nuclear 28S rDNA phylogeny supports the basal placement of *Noctiluca scintillans* (Dinophyceae; Noctilucales) in dinoflagellates. Eur. J. Protistol. 46 (2), 111–120.

Ki, J.S., Han, M.S., 2007. Informative characteristics of 12 divergent domains in complete large subunit rDNA sequences from the harmful dinoflagellate genus, *alexandrium* (Dinophyceae). J. Eukaryot. Microbiol. 54 (2), 210–219.

Kopuz, U., Feyzioglu, A.M., Valente, A., 2014. An unusual red-tide event of *Noctiluca scintillans* (Macartney) in the southeastern black sea. Turk. J. Fish. Aquat. Sci. 14, 261–268.

Larkin, M.A., Blackshields, G., Brown, N., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., 2007. Clustal W and clustal X version 2.0. Bioinformatics 23 (21), 2947–2948.

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 (5), 1279–1290.

Litaker, W.R., Vandersea, M.W., Kibler, S.R., Reece, K.S., Stokes, N.A., Lutzoni, F.M., Yonish, B.A., West, M.A., Black, M.N., Tester, P.A., 2007. Recognizing dinoflagellate species using ITS rDNA sequences. J. Phycol. 43 (2), 344–355.

Liu, L., Zhou, J., Zheng, B., Cai, W., Lin, K., Tang, J., 2013. Temporal and spatial distribution of red tide outbreaks in the Yangtze River estuary and adjacent waters, China. Mar. Pollut. Bull. 72 (1), 213–221.

Mikaelyan, A.S., Malej, A., Shiganova, T.A., Turk, V., Sivkovitch, A.E., Musaeva, E.I., Kogovšek, T., Lukasheva, T.A., 2014. Populations of the red tide forming dinoflagellate *Noctiluca scintillans* (Macartney): A comparison between the Black Sea and the northern Adriatic Sea. Harmful Algae 33, 29–40.

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.

Nylander, J., 2004. MrModeltest v2. Program Distributed by the Author. Evolutionary Biology Centre, Uppsala University.

Okaichi, T., Nishio, S., 1976. Identification of ammonia as the toxic principle of red tide of *Noctiluca miliaris*. Bull. Plankton Soc. Jpn. 23, 25–30.

Padmakumar, K., SreeRenjima, G., Fanimol, C., Menon, N., Sanjeevan, V., 2010. Preponderance of heterotrophic *Noctiluca scintillans* during a multi-species diatom bloom along the southwest coast of India. Int. J. Ocean Oceanogr. 4 (1), 45–53.

Page, R.D., 2001. TreeView. Glasgow University, Glasgow, UK.

Plotree, D., Plotgram, D., 1989. PHYLIP-phylogeny inference package (version 3.2). Cladistics 5, 163–166.

Quevedo, M., Gonzalez-Quiros, R., Anadon, R., 1999. Evidence of heavy predation by *Noctiluca scintillans* on *Acartia clausi* (copepoda) eggs off the central Cantabrian coast (NW Spain). Oceanol. Acta 22 (1), 127–131.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics 19 (12), 1572–1574.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19 (18), 2496–2497.

Ryšánek, D., Hrčková, K., Škaloud, P., 2014. Global ubiquity and local endemism of free-living terrestrial protists: phylogeographic assessment of the streptophyte alga *Klebsormidium*. Environ. Microbiol. http://dx.doi.org/10.1111/1462–2920.12501.

Scholin, C.A., Anderson, D.M., Sogin, M.L., 1993. Two distinct small-subunit ribosomal RNA genes in the north American toxic dinoflagellate *Alexandrium fundyense* (Dinophyceae). J. Phycol. 29 (2), 209–216.

Scholin, C.A., Herzog, M., Sogin, M., Anderson, D.M., 1994. Identification of group- and strain-specific genetic markers for globally distributed *Alexandrium* (Dinophyceae). II. Sequence analysis of a fragment of the LSU rRNA gene. J. Phycol. 30 (6), 999–1011.

Stern, R.F., Andersen, R.A., Jameson, I., Küpper, F.C., Coffroth, M.A., Vaulot, D., Le Gall, F., Veron, B., Brand, J.J., Skelton, H., 2012. Evaluating the ribosomal internal transcribed spacer (ITS) as a candidate dinoflagellate barcode marker. PLoS One 7 (8), e42780.

Stern, R.F., Horak, A., Andrew, R.L., Coffroth, M.A., Andersen, R.A., Küpper, F.C., Jameson, I., Hoppenrath, M., Véron, B., Kasai, F., 2010. Environmental barcoding reveals massive dinoflagellate diversity in marine environments. PLoS One 5 (11), e13991.

Sulochanan, B., Dineshbabu, A., Saravanan, R., Bhat, G.S., Lavanya, S., 2014. Occurrence of *Noctiluca scintillans* bloom off mangalore in the Arabian sea. Indian J. Fish 61 (1), 42–48.

Sun, D., Zhao, D.Z., Wen, S.Y., Xing, X.G., Wang, L., 2010. Fuzzy analysis on relationship between *Noctiluca scintillans* blooms and environmental factors. Mar. Environ. Sci. 29 (1), 70–75.

Swofford, D.L., 2002. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetic analysis version 6.0. Mol. Biol. Evol. 30 (12), 2725–2729.

Thompson, P., Baird, M., Ingleton, T., Doblin, M., 2009. Long-term changes in temperate Australian coastal waters: implications for phytoplankton. Mar. Ecol. Prog. Ser. 394, 1–19.

Thornhill, D.J., Lajeunesse, T.C., Santos, S.R., 2007. Measuring rDNA diversity in eukaryotic microbial systems: how intra-genomic variation, pseudogenes, and PCR artifacts confound biodiversity estimates. Mol. Ecol. 16 (24), 5326–5340.

Tibbs, J.F., 1967. On some planktonic protozoa taken from the track of drift station ARLIS I, 1960–61. Arctic 20, 247–254.

Wang, L., Zhuang, Y., Zhang, H., Lin, X., Lin, S., 2014. DNA barcoding species in *Alexandrium tamarense* complex using ITS and proposing designation of five species. Harmful Algae 31, 100–113.

Zhang, H., Bhattacharya, D., Lin, S., 2005. Phylogeny of dinoflagellates based on mitochondrial cytochrome B and nuclear small subunit rDNA sequence comparisons. J. Phycol. 41 (2), 411–420.

Zhang, H., Bhattacharya, D., Lin, S., 2007. A three-gene dinoflagellate phylogeny suggests monophyly of prorocentrales and a basal position for *Amphidinium* and *Heterocapsa*. J. Mol. Evol. 65 (4), 463–474.

Zufall, R.A., Dimond, K.L., Doerder, F.P., 2013. Restricted distribution and limited gene flow in the model ciliate *Tetrahymena thermophila*. Mol. Ecol. 22 (4), 1081–1091.