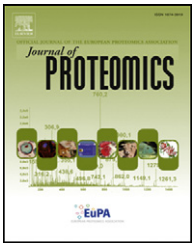
JOURNAL OF PROTEOMICS 86 (2013) 85–96

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot



# Quantitative proteomic analysis reveals evolutionary divergence and species-specific peptides in the *Alexandrium tamarense* complex (Dinophyceae)

*Cheng Li, Yong Zhang, Zhang-Xian Xie, Zhi-Ping He, Lin Lin, Da-Zhi Wang**

*State Key Laboratory of Marine Environmental Science/College of the Environment and Ecology, Xiamen University, Xiamen, 361005, China*

ARTICLE INFO

*Article history:*
Received 22 January 2013
Accepted 1 May 2013
Available online 15 May 2013

*Keywords:*
*Alexandrium tamarense* complex
Quantitative proteomics
Two dimensional fluorescence difference gel electrophoresis
Peptide mass fingerprints
Peptide amino acid sequences
RuBisCO

ABSTRACT

The *Alexandrium tamarense/catenella/fundyense* complex is the major causative agent responsible for harmful algal blooms and paralytic shellfish poisoning around the world. However, taxonomy of the *A. tamarense* complex is contentious and the evolutionary relationships within the complex are unclear. This study compared protein profiles of the *A. tamarense* complex collected from different geographic regions using the two dimensional fluorescence difference gel electrophoresis (2-D DIGE) approach, and identified species-specific peptides using MALDI-TOF/TOF mass spectrometry. The results showed that three *Alexandrium* morphotypes presented significantly different protein expression patterns with about 30–40% shared proteins. However, ecotypes from different geographic regions within a species exhibited the same expression patterns, although a few proteins were altered in abundance. Several proteins, i.e. ribulose-1,5-bisphosphate carboxylase oxygenase form II, plastid protein NAP50, methionine S-adenosyltransferase, and peridinin-chlorophyll a-binding protein, were identified and presented different shift patterns in isoelectric point and/or molecular weight in the 2-D DIGE gels, indicating that amino acid mutation and/or posttranslational modification of these proteins had occurred. The species-specific peptide mass fingerprint and amino acid sequence of ribulose-1,5-bisphosphate carboxylase oxygenase were characterized in the *A. tamarense* complex, and amino acid substitution occurred among them. This study indicated that evolutionary divergence had occurred at the proteomic level in the *A. tamarense* complex, and that the species-specific peptides could be used as potential biomarkers to distinguish the three morphotypes.

Biological significance
Scientific question: The *Alexandrium tamarense/catenella/fundyense* complex is the major causative agent responsible for harmful algal blooms and paralytic shellfish poisoning around the world. However, taxonomy of the *A. tamarense* complex is contentious and the evolutionary relationships within the complex are unclear, which has seriously impeded our understanding of *Alexandrium*-causing HABs and, consequently, the monitoring, mitigation and prevention.
Technical significance: This study, for the first time, compared the global protein expression patterns of eight ecotypes from the *A. tamarense* complex and identified species-specific peptides using a quantitative proteomic approach combining 2-D DIGE and MALDI-TOF/TOF MS.

* *Corresponding author*. Tel.: +86 592 2186016; fax: +86 592 2180655.
E-mail address: dzwang@xmu.edu.cn (D.-Z. Wang).


http://dx.doi.org/10.1016/j.jprot.2013.05.007

Scientific significance: This study demonstrated that the evolutionary divergence had occurred in the *A. tamarense* complex at the proteomic level, and the complex should be classified into three species, i.e. *A. tamarense*, *A. catenella*, and *A. fundyense*. Moreover, the species-specific peptide mass fingerprints could be used as potential biomarkers to distinguish the three morphotypes.



## 1. Introduction

*Alexandrium* is a dinoflagellate genus which is widely spread around the world, and which not only causes harmful algal blooms but also results in paralytic shellfish poisoning (PSP) [1]. It is estimated that PSP toxins result in more than 2000 intoxication incidents per year on a global basis, with an overall mortality rate of 15% [2]. In addition to human intoxications, PSP has been implicated in the deaths of fish, birds and marine mammals [3]. Over the past few decades, the blooms formed by *Alexandrium* species have been significantly increasing in frequency, intensity and distribution, and this has resulted in serious economic and public concern [4].

Traditionally, the *Alexandrium tamarense* species complex is classified into three morphospecies, *A. tamarense*, *Alexandrium catenella* and *Alexandrium fundyense*, based on differences in the shape of their characteristic plates, the presence or absence of a ventral pore between plates 1′ and 4′, and their chain-forming ability [5]. However, these conventional morphological features are somewhat inconsistent [6–8]. For example, the ventral pore, the most reliable character to separate *A. tamarense* from *A. catenella*, is also present in the first apical plate of *A. catenella* from which it is supposed to be absent [9]. So morphology is not a good indicator of evolutionary relationships within the *A. tamarense* complex. Studies show that members of the *A. tamarense* complex from the same region cannot be separated from each other based on 18S rDNA or the D1/D2 region of 28S rDNA [10,11]. The phylogenetic analysis of the LSU rDNA gene of *Alexandrium* shows that the *A. tamarense* complex can be divided into distinct geographic clades, such as North America (NA), Temperate Asia (TA), West Europe (WE) and Mediterranean (ME) clades, but it is not possible to differentiate between the three morphotypes [12,13]. A global biogeographic study of the *A. tamarense* complex indicates that five phylogenetic groups are more likely cryptic species, and geographically based species names are no longer indicative of the range occupied by members of each group [5]. Intragenomic SSU rDNA polymorphism analysis suggests that the three original *Alexandrium* morphospecies designations are invalid, and the strains do not group based on geographic locations, although some subclades are predominated by part of a morphotype from a region [14]. Overall, no consensus has been reached regarding the delineation of species within the *A. tamarense* complex.

Proteomics is a valuable tool in the measurement of natural variation within and between populations in evolutionary ecology, which links the genotype to the phenotype [15]. The 2D gel-based proteomic approach has been applied to study natural variations among *Arabidopsis* ecotypes [16] (Chevalier et al.) and marine mussels [17]. It has also been applied to identify and distinguish marine *Mytilus* species [18], hake [19], fish [20] and shrimp [21]. Comparison of proteomic reference maps of thecate (armored) and athecate (naked or unarmored) dinoflagellate species reveals species-specific protein profiles which could be used to distinguish closely related species within the same family, and even the geographically distinct isolates [22]. Recently, the peptide mass fingerprint (PMF) technique has been used to classify dinoflagellate species. Specific PMFs of ten dinoflagellate species including *Karenia*, *Alexandrium*, *Prorocentrum* and *Scrippsiella* are identified, and they can be used as potential biomarkers for species discrimination [23]. Moreover, the protein/peptide expression profiles from MALDI-TOF-MS are evaluated for species identification of field samples, and several species-specific peaks of *Karenia digitata* are identified [24]. These studies indicate that the judicious application of proteomics could identify a large suite of specific protein targets related to the organisms of interest, e.g. *Alexandrium*, and thus provide a protein phenotype. This snapshot of the protein expression of *Alexandrium* could then serve as a foundation on which to test specific assumptions and further characterize the specific proteins using proteomic techniques.

In this study, we compared the protein profiles of eight *Alexandrium* ecotypes, including four *A. catenella*, three *A. tamarense* and one *A. fundyense*, which are phylogenetically close, and analyzed the PMF and MS/MS spectrum of several proteins presented in the three *Alexandrium* species using a combination of the two-dimensional difference gel electrophoresis (2-D DIGE) method and the MALDI-TOF/TOF MS approach. Species-specific protein expression patterns were observed among members of the *A. tamarense* complex. The peptide amino acid sequences of several abundant proteins were deduced using *de novo* sequencing software and manual interpretation [25]. A comparison of the peptide amino acid sequences of ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) form II in the three *Alexandrium* species revealed species-specific peptides. This study indicated that evolutionary divergence had occurred at the proteomic level in the *A. tamarense* complex, and that the species-specific peptides could be used as potential biomarkers to discriminate among the three morphotypes.

## 2. Materials and methods

### 2.1. Organisms and culture conditions

Unialgal cultures of *A. tamarense* CI01, *A. tamarense* Polar, *A. catenella* GX02 and *A. catenella* DH03 were provided by the Collection Center of Marine Bacteria and Algae, Xiamen University, and *A. tamarense* CCAP1119-1, *A. tamarense* CCAP1119-20,

*A. tamarense* CCAP1119-29 and *A. tamarense* CCAP1119-22 were bought from the Culture Collection of Algae and Protozoa, Scottish Association for Marine Science. Detailed information on the *A. tamarense* complex is given in Table 1. The isolates were grown at 20 °C in K medium [26]. Illumination was provided at a light intensity of approximately 100 μmol photons $m^{-2}$ $s^{-1}$ with fluorescent lamps under a 14:10 h light: dark photoperiod.

When the cells entered the exponential growth phase, approximately 2 × $10^6$ vegetative cells from each *Alexandrium* ecotype were collected using centrifugation at 5000 ×*g* for 10 min at 2 °C. The pellets were subsequently transferred to 1.5 mL microcentrifuge tubes, rinsed twice with sterile seawater, and then stored at −80 °C for subsequent analysis.

### 2.2. DNA extraction and molecular characterization

DNA was extracted using the CTAB method [27]. Each cell pellet was incubated in 0.4 mL DNA lysis buffer at 55 °C for at least 12 h. 66 μL of 5 M NaCl and the 66 μL of preheated 10% (w/v) CTAB was added after incubation. The mixture was vortexed for 1 min and then incubated for 10 min at 56 °C. About 600 μL of chloroform was added and the mixture was centrifuged at 12,000 rpm for 10 min at room temperature. The supernatant was transferred to a fresh 2 mL tube and DNA binding buffer was added. Subsequently, the mixture was purified using a DNA purification kit (Zymo, USA). With the extracted DNA, a dinoflagellate specific primer set 18S comF1 (5′-GCTTGTCTCAAAGATTAAGCCATGC-3′)-18S comR1 (5′-CACCTACGGAAACC TTGTT- ACGAC 3′) was used to amplify the 18S rDNA using PCR (35 cycles of 15 s at 95 °C, 30 s at 56 °C and 45 s at 72 °C, followed by 5 min at 72 °C). The PCR product was directly sequenced, and distance matrix and phylogenetic analyses were performed for each of these datasets using ClustalW following the protocol of Miranda et al. [14].

### 2.3. Protein extraction and quantification

Proteins were extracted using Trizol following the manufacturer's protocol but with a few modifications. Briefly, the pellets were sonicated in 1 mL Trizol reagent on ice. After shaking for 15 s and adding 200 μL chloroform, the cell lysate was incubated for 5 min at room temperature. The top colorless or pale-yellow layer was removed after centrifuging at 12,000 ×*g* for 15 min at 4 °C. Next, about 300 μL ethanol was added to resuspend the bottom layer before centrifuging at 2000 ×*g* for 5 min at 4 °C, and then 2 mL isopropanol was added to the supernatant. The mixture was incubated for 1 h at −20 °C, and then, after centrifuging at 14,000 ×*g* for 30 min at 4 °C, the pellet was washed with 95% ethanol before air drying. The protein pellet was solubilized in 100 μL of lysis buffer containing 30 mM Tris, 7 M urea, 2 M thiourea and 4% (w/v) CHAPS (pH 8.5, based on the DIGE protocol). Protein quantification was performed using a 2D Quant kit (GE Healthcare, USA).

### 2.4. 2-D DIGE analysis

The pH of each sample was adjusted to pH 8.5 with the addition of labeling buffer (pH 9.5, based on the DIGE protocol) if required. 50 μg of protein from each *Alexandrium* ecotype was then labeled using CyDye DIGE Fluor minimal dye (GE Healthcare, USA) following the manufacturer's instructions. Samples were run in duplicate with a label switch in order to avoid introducing label bias. A Cy2-labeled internal standard for all gels was prepared from a mixture of all samples. Reaction was carried out for 30 min and stopped by the addition of an excess of 10 mM lysine. IPG strips (24 cm, pH range 3-10 NL, BioRad) were rehydrated overnight and focused for 100,000 Vhrs using an Ettan IPGphor III apparatus (GE Healthcare) set at 20 °C. All strips were equilibrated for 15 min with 1% DTT and proteins were subsequently alkylated for 15 min in 2.5% iodoacetamide. Both equilibrations were dissolved in equilibration buffer containing 6 M urea, 2% SDS, 50 mM Tris–HCl (pH 8.8), 30% glycerol, bromophenol blue buffer. Then the equilibrated strips were carefully placed on top of 12% acrylamide gels. The separation was performed at 10 °C in an EttanDalt system (GE Healthcare) at 1 w/gel for 30 min and then at 15 w/gel until the dye front reached the bottom of the gel.

### 2.5. Image capture and analysis

Labeled proteins in each gel were visualized using a Typhoon 9400™ (GE Healthcare) Imager with excitation of gels at 488, 532 and 633 nm, and emission at 520, 590 and 680 nm. Gel images were analyzed using the DeCyder™ 2D Differential Analysis Software v6.5 (GE Healthcare). For each spot, the protein volume from the fluorescence intensity for Cy2, Cy3 and Cy5 were determined and the Cy3/Cy2 and Cy5/Cy2 ratios calculated. Values were then normalized based on the assumption that the amount of protein per image was the same. Comparison between the different experimental groups and the control group was tested using Student's t-test. The gel chosen for picking was fixed in a solution containing 40% ethanol and 10% acetic acid for at least 2 h, and stained with silver staining [28].

**Table 1 – Origin of selected *Alexandrium tamarense* complex ecotypes.**

| No. | Ecotype (based on morphology) | Ecotype (based on molecular study) | Location | Abbreviation |
|---|---|---|---|---|
| 1 | *A. tamarense* CI01 | *A. catenella* CI01 | South China Sea, China | ACCI01 |
| 2 | *A. tamarense* Polar | *A. catenella* Polar | South Polar Region | ACP |
| 3 | *A. catenella* GX02 | *A. catenella* GX02 | South China Sea, China | ACGX02 |
| 4 | *A. catenella* DH03 | *A. catenella* DH03 | East China Sea | ACDH03 |
| 5 | *A. tamarense* CCAP1119-1 | *A. tamarense* CCAP1119-1 | Tamar Estuary, Cornwall, England | CCAP-1 |
| 6 | *A. tamarense* CCAP1119-20 | *A. tamarense* CCAP1119-20 | Loch Maddy, Scotland | CCAP-20 |
| 7 | *A. tamarense* CCAP1119-29 | *A. tamarense* CCAP1119-29 | Scalloway,Shetland, Scotland | CCAP-29 |
| 8 | *A. tamarense* CCAP1119-22 | *A. fundyense* CCAP1119-22 | Loch Ewe, west coast, Scotland | CCAP-22 |

### 2.6. Protein digestion and mass spectrometric analysis

In-gel tryptic digestion of silver-stained 2D proteins was carried out with 5 ng/μL Promega Sequencing Grade Modified Trypsin in 10 mM $NH_4HCO_3$ following the protocol of Wang et al. [29].

Mass Spectrometric analysis was performed using an AB SCIEX MALDI TOF–TOF™ 5800 Analyzer (AB SCIEX, Foster City, CA) in reflection positive-ion mode as previously described by Wang et al. [29]. For MS mode the 850–4000 m/z mass range was used with 1000 laser shots per spectrum. The PMF peak detection criteria used were a minimum S/N of 10, a local noise window width mass/charge (m/z) of 250, and minimum full-width half-maximum (bins) of 2.9. For MS/MS analysis, an energy of 1 KV was used for collision-induced dissociation, and 2000 acquisitions were accumulated for each MS/MS spectrum. The peak detection criteria used were a minimum S/N of 3, a local noise window width (m/z) of 200 and minimum full-width half-maximum (bins) of 2.9. Database searching involved GPS Explorer™ software (version 3.6, AB SCIEX) running a MASCOT search algorithm (v2.2, Matrix Science, London, UK) for protein identification.

### 2.7. De novo sequencing

*De novo* sequencing was conducted using DeNovo Explorer™ software (AB SCIEX) followed by manual confirmation of most sequences. The *de novo* sequencing parameters were set as follows: trypsin as protease; one maximum missed cleavage; the error tolerance of a parent- and fragment-mass: 0.08 μg; fixed modification: carbamidomethylation of cysteine and variable modification: methionine oxidation. The *de novo* deduced peptide sequences were submitted to MS-BLAST following the methods of Wang et al. [25].

### 2.8. Immunoblotting analysis of RuBisCO

In order to validate the protein identification made using the mass spectrometric data, and to prove that protein spots shift was due to natural differences between *Alexandrium* strains rather than the protein separation technique, western blotting was performed using the RuBisCO II antibody (donated by Prof. Senjie Lin from the University of Connecticut, USA). The 2DE gel was electrotransferred onto a PVDF membrane (0.45 μm; Millipore) at a constant current of 100 mA overnight at 4 °C in a transfer buffer (pH 8.3) containing 0.025 mM Tris–HCl, 0.192 mM glycine, 20% methanol. After blotting, the membrane was blocked with PBS containing 5% (w/v) non-fat dry milk powder at room temperature for 2 h. After five washes in PBS buffer with 0.05% (v/v) Tween20, the membrane was probed with anti-RuBisCO II (1:5000). Then the blots were incubated with a secondary antibody, horseradish peroxidase-goat anti-rat immunoglobulin G (IgG) (H + L) (Santa Cruz Biotechnology, Santa Cruz, CA, USA) (1:5000 dilutions) for 1 h at room temperature. The protein spots then interacted with ECL solution (Beyotime, China) and were exposed to imaging films (Kodak, USA) for detection.

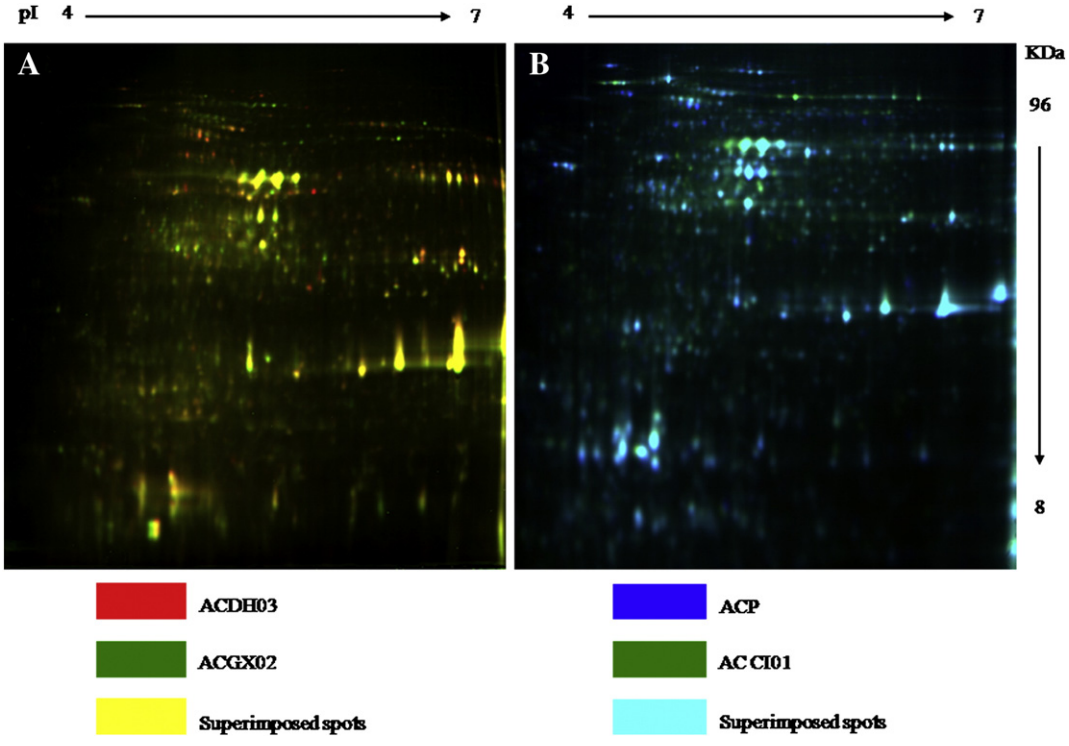



**Fig. 1 – Representative 2D-DIGE gel images of *A. catenella*. The proteins were separated using isoelectric focusing (IPG: 24 cm, non-linearly covering a pH of 4-7) in the first dimension, and 12.5% SDS-PAGE gel in the second dimension. Gels were scanned using a Typhoon™ Imager. Pseudocolors were used for each DIGE Fluor dye. A, superimposed images from equal amounts of CyDye labeled proteins of ACDH03 (red) and ACGX02 (green); B, superimposed images from equal amounts of CyDye labeled proteins of ACP (blue) and ACCI01 (green).**

## 3. Results

### 3.1. Molecular analysis of different ecotypes of the A. tamarense complex

The original information concerning eight *Alexandrium* ecotypes from different geographic regions is shown in Table 1. Three of them were isolated from the South and East China Seas, one from the South Polar region, and four from the English and Scottish coasts. Among them, six ecotypes were *A. tamarense* and two were *A. catenella* based on their traditional morphological features. However, further analysis of ITS1-5.8S-ITS2 and 18S of eight ecotypes using ClustalW software indicated that the eight ecotypes were divided into three groups: four ecotypes, ACCI01, ACP, ACGX02 and ACDH03 belonged to *A. catenella*; three ecotypes CCAP-1, CCAP-20 and CCAP-29 belonged to *A. tamarense*, and one ecotype CCAP-22 belonged to *A. fundyense* (Supplementary Fig. 1). The phylogenetic tree also showed that the eight ecotypes were classified into three groups. These results suggested that the three morphospecies of the *A. tamarense* complex did not conform to molecular phylogenetics or morphological species definition.

### 3.2. Comparison of 2-D DIGE protein reference maps of the A. tamarense complex

The representative 2-D DIGE maps, corresponding to the eight *Alexandrium* ecotypes are shown in Figs. 1 to 3. The intra-species comparisons of four *A. catenella* ecotypes and three *A. tamarense* ecotypes are shown in Figs. 1 and 2, respectively, while the inter-species comparison of different *Alexandrium* species are shown in Fig. 3. After analysis using DeCyder software, the inter- and intra-species matching results are shown in Tables 2 and 3. An average of ca. 2000 spots for each ecotype was detected. For intra-species comparison, the ecotypes within the same species exhibited identical protein expression patterns and no exclusive protein spots were found for each ecotype, but 5–10% of the protein spots varied in abundance in each ecotype (Table 2). For inter-species comparison, three *Alexandrium* species presented

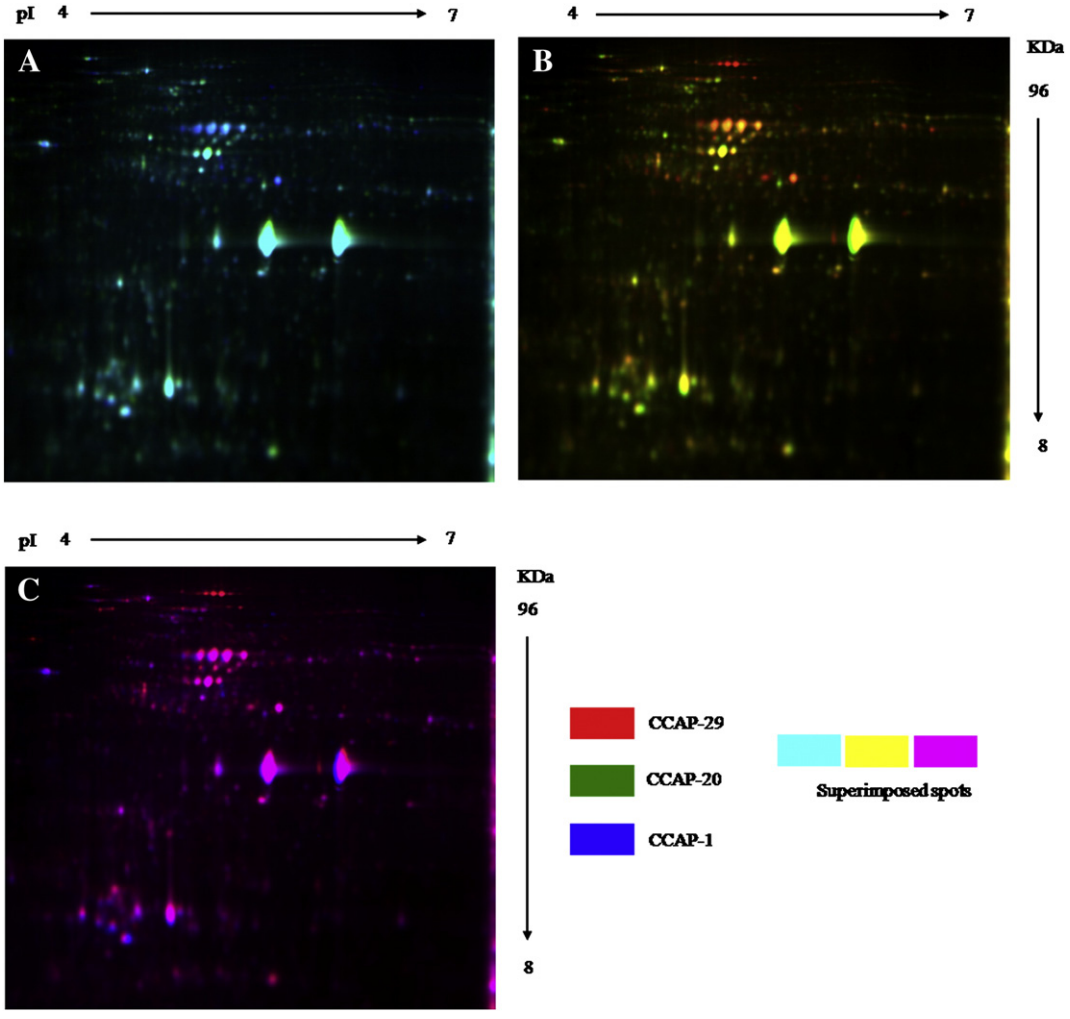



**Fig. 2 – Representative 2D-DIGE gel images of *A. tamarense*. Pseudocolors were used for each DIGE Fluor dye. A, Superimposed images from equal amounts of CyDye labeled proteins of CCAP-1 (blue) and CCAP-20 (green); B, superimposed images from equal amounts of CyDye labeled proteins of CCAP-20(green) and CCAP-29 (red). C, superimposed images from equal amounts of CyDye labeled proteins of CCAP-29 (red) and CCAP-1 (blue).**

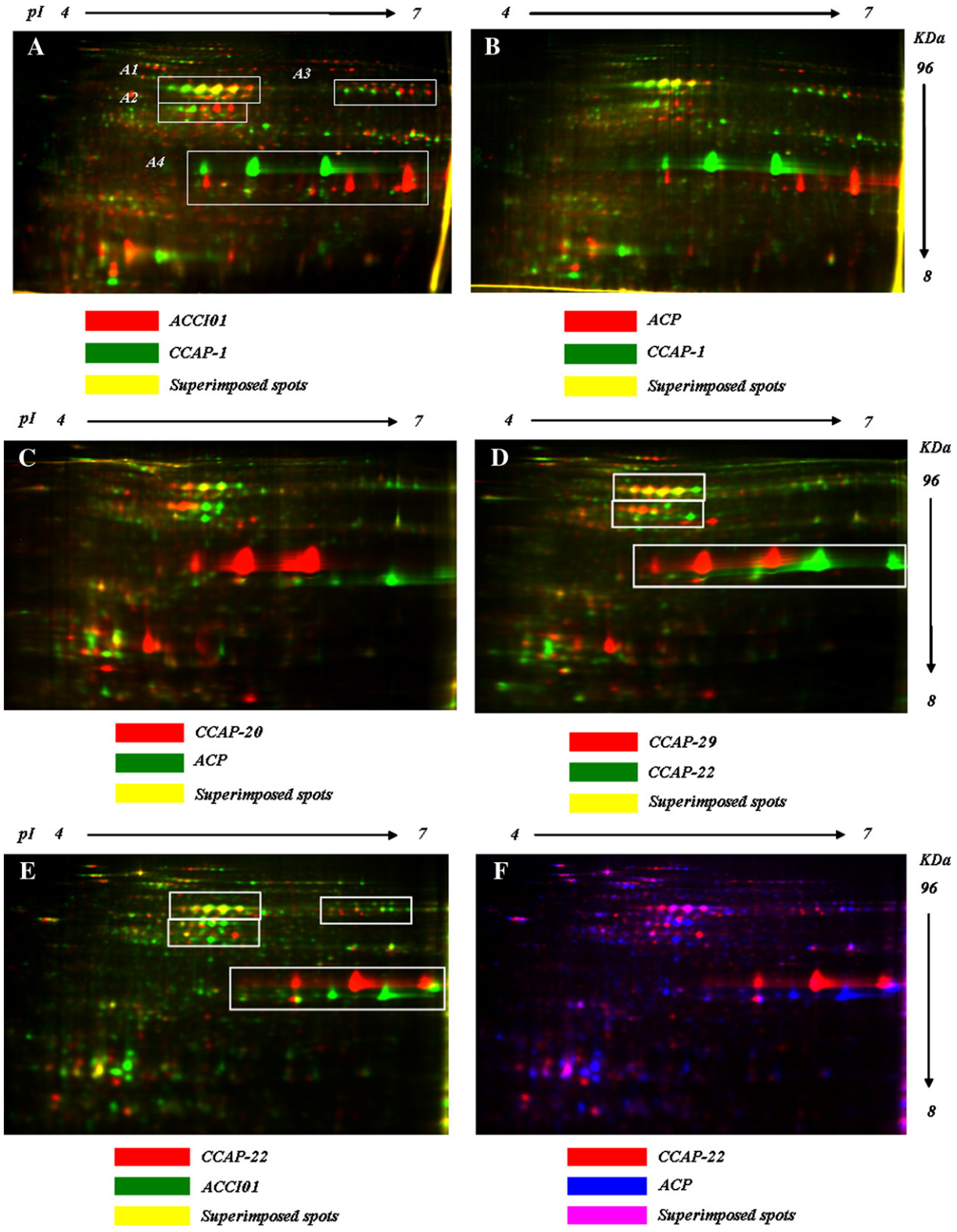



**Fig. 3 – Comparison of 2D-DIGE images among different *Alexandrium* species. (A)–(C): *A. catenella* and *A. tamarense*; (D): *A. tamarense* and *A. fundyense*; (E)–(F): *A. catenella* and *A. fundyense*. A white rectangle shows the protein shifts in the maps.**

significantly different protein expression patterns: approximately 30–40% of protein spots were shared among *A. catenella*, *A. fundyense* and *A. tamarense* 2-D DIGE gels. However, about 60–70% of protein spots differed in molecular weight (MW) or/and isoelectric point (pI) among the different ecotypes (Table 3).

Based on the above analysis, ACCI01, ACP, CCAP-1, CCAP-20, CCAP-29 and CCAP-22 were selected for principal component

**Table 2 – Intra-species protein matching results among different 2-D DIGE gels analyzed using Decyder software.**

| Species | Unchanged (protein number) | Decreased | Increased |
|---|---|---|---|
| ACDH03 vs ACGX02 | 88.0% (1753) | 6.8% | 5.2% |
| ACCI01 vs ACP | 89.1% (2140) | 5.7% | 5.1% |
| CCAP-29 vs CCAP-20 | 79.9% (1893) | 10.0% | 10.1% |
| CCAP-29 vs CCAP-1 | 80.9% (1915) | 9.8% | 9.3% |
| CCAP-20 vs CCAP-1 | 82.7% (1958) | 8.6% | 8.7% |

Note: Unchanged-protein spots with a fold change less than 1.5 in abundance; decreased and increased-protein spots with a fold change greater than 1.5 in abundance.

analysis (PCA). PCA of the global protein spots showed that the ecotypes from *A. catenella* or from *A. tamarense* were grouped together (Fig. 4), while the three species, *A. fundyense*, *A. catenella* and *A. tamarense* were separated into three distinct groups. This PCA result was consistent with the comparative results of 2D-DIGE maps among the *A. tamarense* complex ecotypes.

### 3.3. *Expressions of representative proteins in the A. tamarense complex*

Several proteins, namely, RuBisCO, methionine S-adenosyltransferase (MAT), nitrogen- associated protein 50 (NAP50) and peridinin chlorophyll a binding protein (PCBP) were identified in *A. catenella*, *A. tamarense* and *A. fundyense* (Fig. 5). However, they presented significantly different expression patterns: RuBisCO, MAT and NAP50 shifted in p*I*, while PCBP shifted in both p*I* and MW (Fig. 5 A1-A4). The western blotting result further authenticated these variations between *A. catenella* and *A. tamarense*. Four isoforms of RuBisCO were identified in both *Alexandrium* species, but they presented different expression pattern and p*I* shift. CR1, CR2, CR3 and CR4 were detected in *A. tamarense* ecotypes, inclining to the acidic end, while AR1, AR2, AR3 and AR4 were detected in *A. catenella* ecotypes, inclining to the alkaline end (Fig. 6). These results indicated that amino acid mutation and/or posttranslational modification of these proteins occurred in the *A. tamarense* complex.

**Table 3 – Inter-species protein matching results among different 2-D DIGE gels analyzed using Decyder software.**

| Species | Matched (protein number) | Unmatched |
|---|---|---|
| CCAP-20 vs ACP | 42.1% (1021) | 57.9% |
| ACCI01 vs CCAP-1 | 29.9% (578) | 70.1% |
| ACP vs CCAP-1 | 34.8% (666) | 65.3% |
| CCAP-22 vs ACCI01 | 37.2% (893) | 62.8% |
| CCAP-22 vs ACP | 34.7% (832) | 65.3% |
| CCAP-29 vs CCAP-22 | 37.8% (847) | 62.2% |

Note: Matched-protein spots with identical MW and p*I*; unmatched-protein spots differed in MW or/and p*I*.

### 3.4. *Peptide mass fingerprints and amino acid sequences of RuBisCO*

Four RuBisCO isoforms were selected to interpret the variations in PMFs among the *A. tamarense* complex species. The peptide maps obtained from the MALDI-TOF mass spectra are shown in Supplementary Fig. 2. The PMF of *A. catenella* possessed a unique peak at 1666.7 Da, which was absent in the four other RuBisCO isoforms from *A. tamarense* and *A. fundyense*. Instead, *A. tamarense* and *A. fundyense* both presented a peak of 1530.8 Da which was absent in *A. catenella*. The unique PMF peak of *A. fundyense* was 2435.2 Da. The PMFs from *A. tamarense* and *A. fundyense* possessed a peak of 1682.7 Da which was absent in the spectra of the three isoforms (AR2, AR3 and AR4) and presented a very low signal intensity in isoform AR1 of *A. catenella*.

The peptide amino acid sequence candidates deduced from MS/MS spectra using DeNovo Explorer™ software are shown in Supplementary Figs. 3 and 4. Comparison of amino acid sequences between peptides at 1530.8 and 1588.8 Da, and 1666.7 and 1682.7 Da, indicated that amino acid substitution of RuBisCO occurred in the *A. tamarense* complex. The difference of 58 Da mass between 1530.8 and 1588.8 Da was justified using a single amino acid substitution in the RuBisCO sequence (Supplementary Fig. 3), and neutral amino acid A (Alanine, 89 Da) in the peptide (1530.8 Da) was substituted by acidic amino acid E (Glutamic acid, 147 Da) in the peptide (1588.8 Da) (Supplementary Fig. 3). Meanwhile, the difference of 16 Da mass between 1666.7 and 1682.7 Da indicated that neutral amino acid A (Alanine, 89 Da) in the peptide (1666.7 Da) replaced uncharged polar residue S (Serine, 105 Da) in the peptide (1682.7 Da) (Supplementary Fig. 4). Further comparison of amino acid sequences of RuBisCO, using reference sequences in the NCBInr database, indicated that these two positions were very mutable positions among the dinoflagellates (Table 4). These results further demonstrated that evolutionary divergence had occurred in the *A. tamarense* complex.

## 4. Discussion

### 4.1. *Different Alexandrium species display contrasting protein expression patterns*

This study compared the global protein expression patterns of eight ecotypes from the *A. tamarense* complex using the 2-D DIGE approach, and found that the three *Alexandrium* species presented significantly different protein expression patterns, and more than 60% of protein spots shifted in either MW or p*I*, or both, in the 2-D DIGE gels. The PCA results also demonstrated that the eight *Alexandrium* ecotypes were classified into three distinct groups even though these ecotypes were collected from different geographic regions. Comparison of four identified proteins indicated that these proteins presented different shifting patterns: RuBisCO, MAT and NAP50 shifted in p*I* while PCBP shifted in both p*I* and MW, indicating that protein mutation or/and posttranslational modification occurred in the *A. tamarense* complex. However, the ecotypes of the same species from different geographic regions presented the same protein expression patterns: all protein spots were identical on 2-D DIGE gels and no exclusive protein spots were found in each

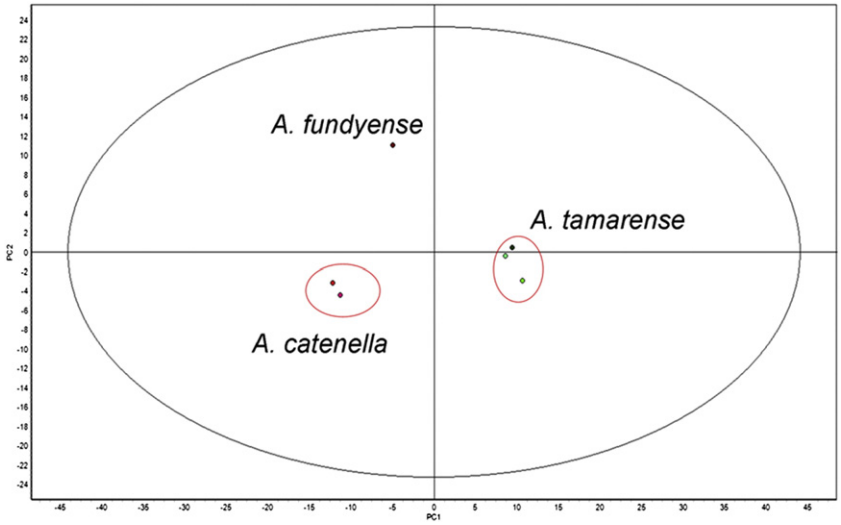



**Fig. 4 – Principal component analysis of the global protein spots of six *Alexandrium* ecotypes using DeCyder software.**

ecotype. These results suggested that evolutionary divergence had occurred in the *A. tamarense* complex, and the difference in protein expression was caused by species rather than geographical differences. A number of studies demonstrate significant variation in protein expression patterns within and between species [30]. For example in a proteomic study of natural variation between eight *Arabidopsis* ecotypes, only one-quarter of protein spots is shared by all accessions [16]. When maritime pine plants from seven origins are compared, less than 20% of the protein spots are observed simultaneously in all patterns [31]. It is postulated that the proteome level is close to the phenotypic level where natural selection is acting, affecting more directly protein variation through balancing, divergent or directional selection [30].

It should be pointed out that the abundance of several proteins altered among different ecotypes of the same species. For example, RuBisCO isoform AR1 changed 2.15 fold between ecotypes ACP and ACCI01, but the fold change was less than 1.09 among four RuBisCO isoforms of ecotypes ACGX02 and ACDH03. The same phenomenon was also observed in our *A. tamarense* intra-species comparison. The abundant alterations among different ecotypes might be caused by their adaptive ability to ambient environments. A recent study shows that two ecotypes of *Littorina saxatilis* with distinct shore microhabitats (exposed/sheltered) differ by a considerable percentage (16%) in the proteome analysis. In particular the enzyme arginine kinase, which is involved in the energetic metabolism, is altered significantly in abundance between the two ecotypes [32]. Proteomic differentiation between ecotypes is largely insensitive to drastic environmental changes experienced during growth, suggesting a mainly genetic determination of the proteomic differentiation [33].

Overall, the substantial differences in the pattern of protein expressions indicated that a rapid evolutionary divergence occurred within the *A. tamarense* complex at the proteomic level, and protein mutation and/or posttranslational modification might have resulted in the pI and/or MW shifts in global protein spots, which needs further investigations in future.

### 4.2. *Species-specific peptides of RuBisCO as signatures for species classification*

Protein/PMF profiles generated by mass spectrometry have been used for the identification and classification of various organisms including dinoflagellates [23,24,34,35], and some species-specific peptide sequences are also identified in fish [20] and shrimp [21]. These studies suggest that the biomarker peptides can be used for species authentication purposes. In our study, the species-specific peptide mass-maps of four RuBisCO isoforms were identified and characterized. Well-defined peaks with a high S/N ratio were observed for ions corresponding to two peaks, m/z 1530.8 and 1588.8 Da in *A. tamarense*; whereas no

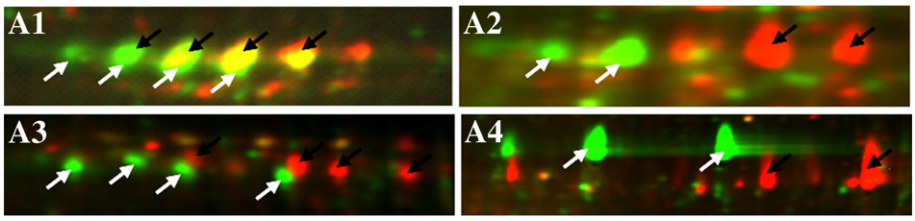



**Fig. 5 – Comparison of four identified protein expression patterns. Arrow-highlighted spots to be identified using MALDI-TOF/TOF MS as the same isoforms of the protein in respective selected areas. A1, RuBisCO II; A2, NAP50; A3, MAT; A4, PCP. White arrows indicate the proteins from *A. tamarense*, and black arrows indicate proteins from *A. catenella*.**

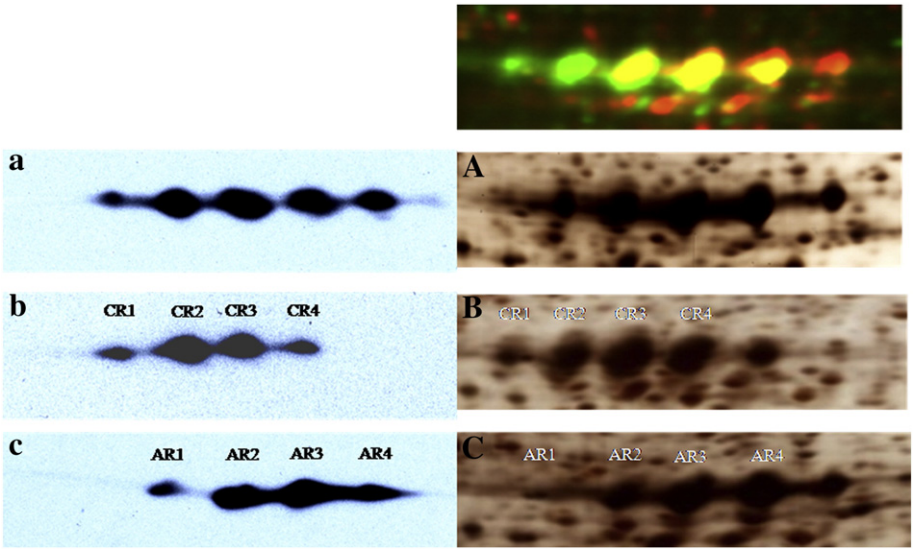



**Fig. 6 – Expressions of RuBisCO isoforms in *A. tamarense* and *A. catenella* revealed using western blotting analysis. 40 μg of soluble proteins from *A. tamarense*, *A. catenella* or mixture of the two were separated using 2-D SDS-PAGE followed by western blotting. (a) and (A). The mixture of equal amounts of CCAP-1 and ACCI01 proteins; (b) and (B) CCAP-1; (c) and (C) ACCI01. AR1, AR2, AR3 and AR4 are four RuBisCO II isoforms in *A. catenella*, and CR1, CR2, CR3 and CR4 are four isoforms of RuBisCO II in *A. tamarense*.**

significant signal could be obtained for the peak at 1530.8 Da in *A. catenella*. Interestingly, there was no significant signal of the peak at 1666.7 Da in *A. tamarense*. The peak at 2435.20 Da was unique for *A. fundyense*. Some variations in peptide intensities were expected to occur in different analyses from the four isoforms of RuBisCO within the same sample, e.g. the intensities of the peak at 1530.8 Da gradually decreased from CR1 to CR4, and the intensities of the peak at 1588.8 Da increased correspondingly in *A. tamarense* species.

*De novo* sequencing analysis indicated that the amino acid sequence of peptide at 1530.8 Da was IYDIYFPPAYLR; at 1588.8 Da was IYDIYFPPEYLR; at 1666.7 Da was LFDGPACS VIDMWR; and at 1682.7 Da was LFDGPSCSVIDMWR. It was clear that the neutral amino acid A (Alanine, 89 Da) in the peptide at 1530.8 Da was substituted by the acidic amino acid E (Glutamic acid, 147 Da) of the peptide at 1588.8 Da; meanwhile, neutral amino acid A (Alanine, 89 Da) in the peptide with 1666.7 Da replaced the uncharged polar residues S (Serine, 105 Da) of the peptide at 1682.7 Da (Supplementary Fig. 4). The neutral amino acid alanine in the sequence "IYDIYFPPAYLR" of RuBisCO was substituted by acidic glutamic acid, so the peptide at 1530.8 Da was not detected in the four isoforms of RuBisCO in *A. catenella*. Another substitution of amino acid position was neutral amino acid replacing uncharged polar residues, but which could not change the pI of the proteins.

The tentative sequence "IYDIYFPPA(E)YLRLFDGPS(A)CSVI DMWR" yielded using DeNovo Explorer™ software and manual interpretation was submitted to MS-BLAST. This sequence was similar to that of the peptide in *Prorocentrum minimum*, belonging to RuBisCO form II. The amino acid pairs (A, S) which changed in *Alexandrium* were variable positions in all the homologous amino acid sequences among different dinoflagellates species (Table 4). These variable positions in the sequence of RuBisCO from any dinoflagellate species present in the databases were specific to each species. The present study concluded that an amino acid substitution occurred in the ninth position of the peptide "IYDIYFPPAYLR", which produced a peptide mass shift from m/z 1530.8 to 1588.8 Da in *Alexandrium*. Combining the substitution in the sixth and eighth positions of the peptide "LFDGPACSVIDMWR" might therefore be used as a criterion to differentiate species belonging to dinoflagellates in relation to the protein RuBisCO. For example, the changed criterional positions "E.A.S" and "S.S.N" originating from *A. catenella* and *Heterocapsa triquetra* (Table 4). For distinguishing *A. tamarense* and *A. fundyense*, it was necessary for the peptide at 2435.20 Da or 2463.20 Da to be involved in the study (Fig. 7). This interspecific variability could be used as a good biomarker for dinoflagellate species identification. Identified species-specific peptides can be used to prepare antibody-based, facilitative kits for the sensitive detection of each of the dinoflagellates. It is not necessary to obtain all the amino acid sequences of RuBisCO and, as shown in this study, the identification and characterization of specific peptides is the first step toward designing cheap and fast detection analyses of dinoflagellates.

Overall, the identification possibilities of mass spectrometry provide a fast and reliable method for the differentiation of the species which were the subject of this study. As shown in Fig. 7, a peptide at 1666.50 Da was specific for *A. catenella*, but was absent from *A. tamarense* and *A. fundyense*; while on the other hand, the peptide at 1530.80 Da was absent from *A. catenella*; and a peptide at 2435.20 Da was specific for *A. fundyense*, but was absent from *A. tamarense*. Thus, unequivocal species differentiation is clearly possible.

## 5. Conclusions

In summary, 2-D DIGE combined with MS provided us a potential useful approach to resolve the taxonomic ambiguities of closely related algal species, regardless of its disadvantages, i.e. labor

**Table 4 – Amino acid sequences of the RuBisCO fragment from different dinoflagellate species.**

| Species | Partial Amino Acid Sequence |
|---|---|
| *Symbiodinium sp.* | I Y D F Y L P P S F L R L Y D G P A V N V E D M W R |
| *Gonyaulax polyedra* | I Y D I Y F P P S Y L R F F D G P A C S I L D M W R |
| *Heterocapsa triquetra* | I Y D I Y F P P S Y L R L F D G P S C N I I D M W R |
| *Prorocentrum minimum* | I Y D I Y F P P Q Y L R L F D G P S C C V I D M W R |
| *Alexandrium tamarense** | I Y D I Y F P P A Y L R L F D G P S C S V I D M W R |
| *Alexandrium catenella** | I Y D I Y F P P E Y L R L F D G P A C S V I D M W R |

[a]*Alexandrium* species used in this study. Partial amino acid sequences of RuBisCO from six species with the highest sequence homology including *A. catenella* and *A. tamarense*. Conserved amino acids in terms of chemical homology are in the black rectangles. Tentative sequences deduced in this study using mass spectrometry analysis, and showing the changed position, are in the red rectangles.

intensive, time consumable. 2-D DIGE analysis revealed that three *Alexandrium* species presented significantly different protein expression patterns. However, no difference was found in the protein expression pattern among the ecotypes within the same species. Comparison of four proteins indicated that MW and/or pI shifts occurred between species but not within species. These results demonstrated that evolutionary divergence had occurred in the *A. tamarense* complex at the proteomic level and the complex might be classified into three species based on their species-specific protein expression patterns. The species-specific peptides of RuBisCO could be used for the systematic discrimination of the different *Alexandrium* species, even among different dinoflagellates. The deduced amino acid sequences of RuBisCO from different *Alexandrium* species as well as from other dinoflagellate species indicated that amino acid mutations had occurred in the *A. tamarense* complex. Thus, the alignment of substituted amino acids might be used to distribute the dinoflagellates into different groups. This approach appears to have the capacity to discriminate species and to generate a classification directly based on the features of their major proteins. Meanwhile, a challenge for the future is to investigate the relative contributions of these factors to the expression differences observed, and even to clarify which differences have functional consequences.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2013.05.007.

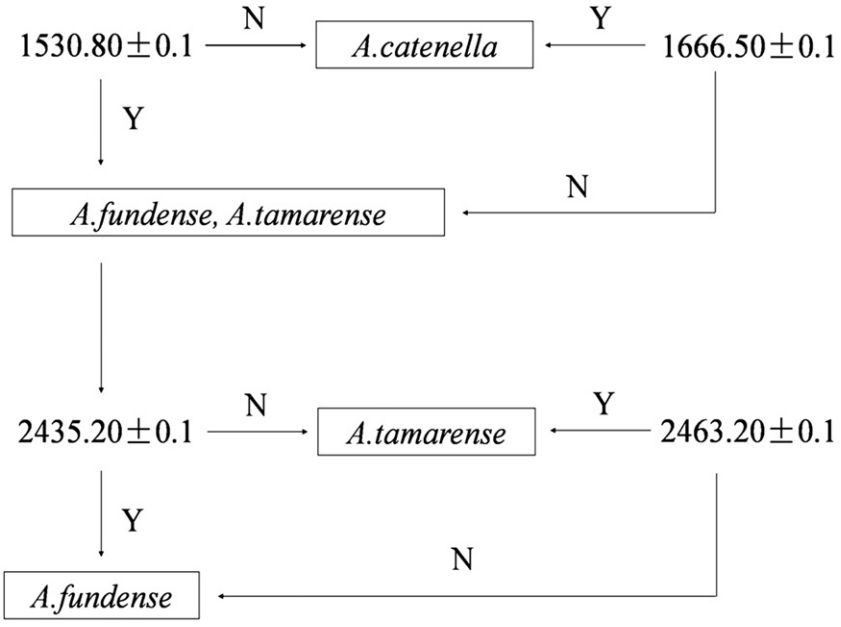



**Fig. 7 – Flow diagram for a systematic identification of three morphospecies of the *A. tamarense* complex using peptide mass fingerprinting. Specific peak masses from RuBisCO II are shown with a number (m/z). "Y" denotes the presence and "N" denotes the absence of a particular peak.**

## Acknowledgments

This work was partially supported by research grants from the National Natural Science Foundation of China (Project No. 41230961), the Ministry of Science and Technology of the People's Republic of China (Project No. 2010CB428703), the Excellent Group and the Program for New Century Excellent Talents in University to D.-Z. Wang. We thank Dr. Huan Zhang from the University of Connecticut for his assistance during DNA extraction and data analysis, and Professor John Hodgkiss from The University of Hong Kong for his help with English.

## REFERENCES

[1] Cembella A, et al. Ecophysiology and metabolism of paralytic shellfish toxins in marine microalgae. Physiological Ecology of Harmful Algal Blooms. Heidelberg: Springer-Verlag; 1998. p. 381–404.
[2] Hallegraeff GM. A review of harmful algal blooms and their apparent global increase. Phycologia 1993;32:79–99.
[3] Van Dolah FM, Leighfield TA. Diel phasing of the cell-cycle in the Florida red tide dinoflagellate, *Gymnodinium breve*. J Phycol 2002;35:1404–11.
[4] Anderson DM, Cembella AD, Hallegraeff GM. Progress in understanding harmful algal blooms: paradigm shifts and new technologies for research, monitoring, and management. Annual Review of Marine Science 2012;4: 143–76.
[5] Steidinger KA. Species of the *tamarensis/catenella* group of *Gonyaulax* and the fucoxanthin derivative-containing gymnodinioids. In: Graneli E, Sundstroem B, Edler L, Anderson DM, editors. Toxic marine phytoplankton. New York: Elsevier; 1989. p. 11–6.
[6] Penna A, Vila M, Giacobbe MG, Bertozzini E, Andreoni F, Fraga S, et al. Population diversity of the Ostreopsidaceae in the Mediterranean Sea: a preliminary study on the genetics and morphology. Harmful algae; 2002. p. 431–3.
[7] Lilly EL, Halanych KM, Anderson DM. Species boundaries and global biogeography of the *Alexandrium tamarense* complex (Dinophyceae). J Phycol 2007;43:1329–38.
[8] Genovesi B, Shin-Grzebyk MS, Grzebyk D, Laabir M, Gagnaire P-A, Vaque A, et al. Assessment of cryptic species diversity within blooms and cyst bank of the *Alexandrium tamarense* complex (Dinophyceae) in a Mediterranean lagoon facilitated by semi-multiplex PCR. J Plankton Res 2011;33: 405–14.
[9] Kim KY, Yoshida M, Fukuyo Y, Kim CH. Morphological observation of *Alexandrium tamarense* (Lebour) Balech, *A. catenella* (Whedon et Kofoid) Balech and one related morphotype (Dinophyceae) in Korea. Algae 2002;17:11–9.
[10] Scholin CA, Anderson DM. Identification of group- and strain-specific genetic markers for globally distributed *Alexandrium* (Dinophyceae). I. RFLP analysis of SSU rRNA genes. J Phycol 1994;30:744–54.
[11] Scholin CA, Herzog M, Sogin M, Anderson DM. 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 1994;30: 999–1011.
[12] Scholin CA, Hallegraeff GM, Anderson DM. Molecular evolution of the *Alexandrium tamarense* 'species complex' (Dinophyceae): dispersal in the North American and West Pacific regions. Phycologia 1995;34:472–85.
[13] Scholin CA, Anderson DM. LSU rDNA-based RFLP assays for discriminating species and strains of *Alexandrium* (Dinophyceae). J Phycol 1996;32:1022–35.
[14] Miranda LN, Zhuang Y, Zhang H, Lin S. Phylogenetic analysis guided by intragenomic SSU rDNA polymorphism refines classification of "*Alexandrium tamarense*" species complex. Harmful Algae 2012;16:35–48.
[15] Diz AP, Martínez-Fernández M, Rolán-Alvarez E. Proteomics in evolutionary ecology: linking the genotype with the phenotype. Mol Ecol 2012;21:1060–80.
[16] Chevalier F, Martin O, Rofidal V, Devauchelle AD, Barteau S, Sommerer N, et al. Proteomic investigation of natural variation between *Arabidopsis* ecotypes. Proteomics 2004;4: 1372–81.
[17] López JL, Marina A, Vázquez J, Alvarez G. A proteomic approach to the study of the marine mussels *Mytilus edulis* and *M. galloprovincialis*. Mar Biol 2002;141:217–23.
[18] López JL, Marina A, Alvarez G, Vazquez J. Application of proteomics for fast identification of species-specific peptides from marine species. Proteomics 2002;2:1658–65.
[19] Carrera M, Cañas B, Piñeiro C, Vázquez J, Gallardo JM. Identification of commercial hake and grenadier species by proteomic analysis of the parvalbumin fraction. Proteomics 2006;6:5278–87.
[20] Carrera M, Cañas B, Piñeiro C, Vázquez J, Gallardo JM. *De novo* mass spectrometry sequencing and characterization of species-specific peptides from nucleoside diphosphate kinase B for the classification of commercial fish species belonging to the family *Merlucciidae*. J Proteome Res 2007;6: 3070–80.
[21] Ortea I, Cañas B, Gallardo JM. Mass spectrometry characterization of species-specific peptides from arginine kinase for the identification of commercially relevant shrimp species. J Proteome Res 2009;8:5356–62.
[22] Chan LL, Hodgkiss IJ, Lu S, Lo SCL. Use of two-dimensional gel electrophoresis proteome reference maps of dinoflagellates for species recognition of causative agents of harmful algal blooms. Proteomics 2004;4:180–92.
[23] Lee FWF, Ho KC, Lo SCL. Rapid identification of dinoflagellates using protein profiling with matrix-assisted laser desorption/ionization mass spectrometry. Harmful Algae 2008;7:551–9.
[24] Lee FWF, Ho KC, Mak YL, Lo SCL. Authentication of the proteins expression profiles (PEPs) identification methodology in a bloom of *Karenia digitata*, the most damaging harmful algal bloom causative agent in the history of Hong Kong. Harmful Algae 2011;12:1–10.
[25] Wang DZ, Li C, Xie ZX, Dong HP, Lin L, Hong HS. Homology-driven proteomics of dinoflagellates with unsequenced genomes using MALDI-TOF/TOF and automated *de novo* sequencing. Evidence-based complementary and alternative medicine. 2011, http://dx.doi.org/10.1155/2011/471020.
[26] Keller MD, Selvin RC, Claus W, Guillard RRL. Media for the culture of oceanic ultraphytoplankton. J Phycol 1987;23: 633–8.
[27] Lin S, Corstjens PLAM. Molecular cloning and expression of the proliferating cell nuclear antigen gene from the coccolithophorid *Pleurochrysis carterae* (Haptophyceae). J Phycol 2002;38:164–73.
[28] Wang DZ, Lin L, Gu HF, Chan LL, Hong HS. Comparative studies on morphology, ITS sequence and protein profile of *Alexandrium tamarense* and *A. catenella* isolated from the China Sea. Harmful Algae 2008;7:106–13.
[29] Wang DZ, Li C, Zhang Y, Wang YY, He ZP, Lin L, et al. Quantitative proteomic analysis of differentially expressed proteins in the toxicity-lost mutant of *Alexandrium catenella* (Dinophyceae) in the exponential phase. J Proteomics 2012;75: 5564–77.

[30] Diz AP, Carvajal-Rodríguez A, Skibinski DOF. Multiple hypothesis testing in proteomics: a strategy for experimental work. Mol Cell Proteomics 2011;10:M110.
[31] Bahrman N, Zivy M, Damerval C, Baradat P. Organisation of the variability of abundant proteins in seven geographical origins of maritime pine (*Pinus pinaster Ait.*). TAG Theoretical and applied genetics 1994;88:407–11.
[32] Martínez-Fernández M, Rodríguez-Piñeiro AM, Oliveira E, Páez de la Cadena M, Rolán-Alvarez E. Proteomic comparison between two marine snail ecotypes reveals details about the biochemistry of adaptation. J Proteome Res 2008;7:4926–34.
[33] Martinez-Fernandez A, Nelson TJ, Ikeda Y, Terzic A. C-MYC-independent nuclear reprogramming favors cardiogenic potential of induced pluripotent stem cells. J Cardiovasc Transl Res 2010;3:13–23.
[34] Demirev PA, Lin JS, Pineda FJ, Fenselau C. Bioinformatics and mass spectrometry for microorganism identification: proteome-wide post-translational modifications and database search algorithms for characterization of intact *H. pylori*. Anal Chem 2001;73:4566–73.
[35] Fenselau C, Demirev PA. Characterization of intact microorganisms by MALDI mass spectrometry. Mass Spectrom Rev 2001;20:157–71.