Phycologia Volume 55 (2), 134–146 Published 12 February 2016

18S rDNA phylogeny of *Pseudo-nitzschia* (Bacillariophyceae) inferred from sequence-structure information

Hong Chang Lim[1], Sing Tung Teng[2], Po Teen Lim[3], Matthias Wolf[4]*† and Chui Pin Leaw[3]*†

[1]*Tunku Abdul Rahman University College, Johor Branch, 85000 Segamat, Johor, Malaysia*
[2]*Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia*
[3]*Bachok Marine Research Station, Institute of Ocean and Earth Sciences, University of Malaya, 16310 Bachok, Kelantan, Malaysia*
[4]*Department of Bioinformatics, Biocenter, University of Wuerzburg, D-97074 Wuerzburg, Germany*

Abstract: We explored the 18S rDNA sequences of *Pseudo-nitzschia* and its close relatives, together with their individual secondary structure information to reconstruct phylogenies. Forty-eight taxa classified in the Bacillariaceae and Achnanthaceae were analyzed by character-based (maximum parsimony and maximum likelihood) and distance-based (neighbour-joining) methods. Three raphid diatoms from the Eunotiaceae were used as out-groups. The sequence-structure alignment was generated by 4SALE (i.e. a one-letter encoded alignment that contained the individual secondary structure information) and applied to the character-based phylogeny reconstruction. All analyses highly support *Pseudo-nitzschia* + *Fragilariopsis* as monophyletic. The *Pseudo-nitzschia*/*Fragilariopsis* clade is further divided into three subclades. One clade comprises members in the *P. seriata* and *P. americana* complexes and species of *Fragilariopsis*; the second and third clades constitute members in *P. pseudodelicatissima* and *P. delicatissima* complexes. Among all analyses, the distance-based method yielded a more highly resolved phylogenetic framework for the *Pseudo-nitzschia*/*Fragilariopsis* clade than did the character-based analyses. Other than those currently known genetic markers, 18S rDNA was also useful in exploring the intrageneric relationships.

Key Words: Alignment, 18S rDNA, 4SALE, *Pseudo-nitzschia*, Secondary structure

INTRODUCTION

The marine diatom *Pseudo-nitzschia* H. Peragallo had generated considerable research interest since the first report of its association with amnesic shellfish poisoning (ASP) in 1987 (Bates *et al*. 1989). At least 19 of the 45 known species of *Pseudo-nitzschia* produced the neurotoxin domoic acid (Dao *et al*. 2014, 2015; Teng *et al*. 2014), responsible for ASP (reviewed in Lelong *et al*. 2012).

The nuclear-encoded ribosomal genes (5.8S, 18S and 28S rDNAs), internal transcribed spacer regions (ITS1 and ITS2) and the mitochondrial gene, cytochrome oxidase gene (*cox*1), have been widely used for the last decade in combination (concatenated) or separately as molecular markers to resolve phylogenetic relationships of *Pseudo-nitzschia* (Lundholm *et al*. 2002a; Lim *et al*. 2014; Tan *et al*. 2015; references herein). The presence of cryptic (morphologically similar but genetically different) and pseudocryptic (minor morphological differences plus genetic differences) species were discovered within the genus, and molecular methods were used to aid the underlying taxonomy (e.g. Lundholm *et al*. 2003, 2006, 2012; Amato & Montresor 2008; Quijano-Scheggia *et al*. 2009; Lim *et al*. 2012a, 2013; Orive *et al*. 2013; Teng *et al*. 2014, 2015; Ruggiero *et al*. 2015). The genus comprised at least four species complexes, *viz*. *P. pseudodelicatissima* (Hasle) Hasle complex *sensu* Lundholm *et al*. (2003), *P. delicatissima* (Cleve) Heiden complex *sensu* Lundholm *et al*. (2006), *P. seriata* (Cleve) Peragallo complex *sensu* Hasle *et al*. (1996) and *P. americana* (Hasle) Fryxell complex *sensu* Lundholm *et al*. (2002b).

In diatoms, length-conservative regions of 18S rDNA have been used mostly to resolve phylogenetic relationships at higher taxonomic levels (Medlin *et al*. 1993, 1996, 2008; Kooistra & Medlin 1996; Sorhannus 1997, 2007; Lundholm *et al*. 2002a; Medlin & Kaczmarska 2004; Alverson *et al*. 2006; Alverson 2008). The 18S rDNA has a conserved secondary structure, and this information is useful in guiding orthologous alignments of 18S rDNA sequences. Several models and databases that consider base pairing in rRNA molecules are available for sequence-structure alignments (reviewed in Medlin 2010, 2016).

Often, the choices of different data types, taxon sampling schemes, out-group taxa, optimality criteria and alignments can impact the outcome of the phylogenetic analysis. In advancing sequence-structure alignment, Wolf and coworkers introduced a new approach. The sequences and their individual secondary structures were simultaneously aligned using a specified scoring matrix, fitted to a 12-letter alphabet originally constructed for ITS2 sequence-structure data. Instead of using a 4×4 scoring matrix, they used a 12×12 matrix for each nucleotide (see fig. 3 in Wolf *et al*. 2014), with its three structural states (paired left, paired right or unpaired), that is, the dot-bracket notation for RNA secondary structure (Seibel *et al*. 2006, 2008; Wolf *et al*. 2008; Koetschan *et al*. 2012; Merget *et al*. 2012). This approach was recently reintroduced and applied to other RNA molecules (Wolf *et al*. 2014). In general, this method increased robustness and accuracy (Keller *et al*. 2010) as

* Corresponding authors (matthias.wolf@biozentrum.uni-wuerzburg.de, cpleaw@um.edu.my).
DOI: 10.2216/15-78.1