Resolving taxonomy and phylogenetics of benthic diatoms from single cell sequencing

Keely Lefebvre

BSc. McGill University, 2012

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

University of Ottawa

In partial fulfillment of the requirements

for the Masters of Science degree in Biology

with Specialization in Bioinformatics

Department of Biology

Faculty of Science

University of Ottawa

Abstract

Benthic diatoms are often used as indicators of water quality and past environmental conditions. This depends entirely on a reliable taxonomic system. With the advent of DNA techniques, genetic analyses can now be used in tandem with traditional microscopy in order to improve taxonomy and determine evolutionary relationships. This thesis examined a speciose genus of diatoms *Neidium* (> 300 species) and, using sequence data from molecular markers as well as traditional morphological analyses, investigated phylogenetic relationships. Fresh benthic samples from aquatic ecosystems in Eastern North America were collected; *Neidium* taxa were examined using light and scanning electron microscopy then compared to the original specimen types. A total of 124 individual cells were retrieved, amplified, and sequenced for four molecular markers (*rbcL*, 18S, *psbA*, and *psbC*). Phylogenetic reconstructions were completed using Maximum likelihood and Bayesian analyses; when compared with morphological analyses this led to the delineation of several novel *Neidium* species.

Résumé

Les diatomées benthiques sont souvent utilisées comme indicatrices de la qualité de l'eau et des conditions environnementales passées. Cela dépend entièrement d'un système taxonomique fiable. Avec l'avènement des techniques de l'ADN, les analyses génétiques peuvent désormais être utilisées de concert avec la microscopie traditionnelle afin d'améliorer la taxonomie et de déterminer les relations évolutives. Cette thèse a examiné un genre très diversifié de diatomées *Neidium* (> 300 espèces) et, en utilisant les données de séquençage de marqueurs moléculaires ainsi que des analyses morphologiques classiques, a étudié les relations phylogénétiques de ce genre. Des échantillons benthiques des écosystèmes aquatiques de l'Est de l'Amérique du Nord ont été recueillis et les taxons de *Neidium* ont été examinés à l'aide de microscopie électronique à balayage et optique, puis comparés aux types d'échantillons originaux. Un total de 124 cellules individuelles ont été isolées, amplifiées et séquencées pour quatre marqueurs moléculaires (*rbcL*, 18S, *psbA* et *psbC*). Des reconstructions phylogénétiques ont été effectuées à l'aide des approches du maximum de vraisemblance et de l'inférence bayésienne et comparées aux analyses morphologiques, ayant pour résultat la délimitation de nouvelles espèces de *Neidium.*

Acknowledgements

Firstly I would like to thank my supervisors for this project Frances Pick and Paul Hamilton. It has been a wonderful learning experience for me to work with both of you on this project. Frances, thank you for your willingness to take me on as a graduate student. Your thorough help with editing and thesis design allowed me to learn a great deal. Paul, working on this project with your guidance has been an amazing opportunity for me, and I look forward to working on the ongoing projects we have in the near future.

To my committee members and examiners Dr. Julian Starr and Dr. Nicolas Rodrigue, thank you for the helpful discussion and comments you provided towards my proposal and thesis. I very much appreciate your helpful insights into both traditional taxonomy and phylogenetic analyses. Thank you also to my examiner Dr. Jesse Vermaire for your valuable comments and discussion that you provided for my thesis.

A sincere thanks to Roger Bull for your tireless patience with many questions at the Canadian Museum of Nature's DNA lab. This project would not have been possible without you. As well, a big thank you to all of the students and volunteers at the Natural Heritage Campus who helped with lab processing and single cell picking, Nancy Hampton, Katie Baillie-David, Frédérick Caron, Danielle Vinson, and Kimberley Duff.

A big shout out goes to the Pick lab members past and current: Katy Alambo, MaryAnn Perron, Marianne Racine, Marie-Pierre Varin, and Shinjini Pal. All of your friendship, conversations, and helpful encouragement allowed me to wholly enjoy my graduate student experience.

I want to thank my solid support team of family and friends including my inspiring grandfather (96!) who has always shown me how fascinating lakes and freshwater can be from a young age. I also want to thank my sisters, Lindsay and Carmen, for always listening to my science lectures and putting up with me as an older sister. Finally, I would not be where I am today without my parents, Marc and Jeanne Lefebvre. I have learned so much from the both of you and it has allowed me to reach where I am today. Thank you for always being supportive and encouraging about my work and always pushing me to go further.

And last but definitely not least I want to thank my husband Nick. Thank you for always believing in me and allowing me to go for my goals. It is because of your support and love that I have been able to be successful in my work, and I deeply appreciate all of the sacrifices that you have made in order to let me achieve this. I am very lucky to have you as a partner.

Statement of contributions of collaborators

This thesis is my original work, which was conducted under the supervision of Dr. Frances Pick and Paul Hamilton and under the guidance of the advisory committee members Dr. Julian Starr and Dr. Nicolas Rodrigue. Chapters 1 and 4 are comprised of the General Introduction and Conclusions respectively. Chapter 2 and Appendix A have been published in peer reviewed journals. Chapter 3 will be submitted to a peer reviewed journal.

Chapter 2

P.B. Hamilton was responsible for the some of the sampling, the taxonomic observations, Figs 2.1-2.21, B.1, B.2, and some of the written manuscript. I was responsible for the molecular and morphological laboratory work, all data analysis, the bulk of the written manuscript and the remaining figures.

Chapter 3

P.B. Hamilton was responsible for some of the field sampling, some of the SEM images and editing the taxonomic observations. I was responsible for the molecular and morphological laboratory work, all data analysis, the written manuscript and all figures and tables.

Table of Contents

List of Tables

List of Figures

Abbreviations

AIC	Akaike information criterion
Au	gold
Auct. nonnull [sic]	the species description as identified by many authors
BA	Bayesian analysis
BAPP	Bayesian analysis posterior probabilities
BF	brightfield
BHUPM	Institute of Paleontology, Museum für Naturkunde, Humboldt University of Berlin
BI	Bayesian inference
BLAST	Basic Local Alignment Search Tool
BS	bootstrap
CANA	Canadian Museum of Nature Phycology Collection
DIC	differential interference contrast
DNA	deoxyribonucleic acid
ECD	Ehrenberg Collection of Drawings
ECO	Ehrenberg Collection
ESEM	environmental scanning electron microscope
ESS	effective sample size
GTR+I+G	General Time Reversible nucleotide substitution model
LM	light microscope
MAFFT	Multiple Alignment using Fast Fourier Transform
MCMC	Markov chain Monte Carlo
ML	maximum likelihood
MLBS	maximum likelihood bootstrap
NA	numerical aperture
NNI	nearest neighbour interchanges
PC	phase contrast

PCR	polymerase chain reaction
Pd	palladium
PP	posterior probabilities
rDNA	ribosomal deoxyribonucleic acid
rRNA	ribosomal ribonucleic acid
SEM	scanning electron microscope
Sensu auct.	according to some authors
Sensu auct. Nonnull.	according to many more authors
s.l.	sensu lato (in the broad sense)
SH-like	Shidmodaira–Hasegawa-like
SPR	subtree pruning regrafting
TIM3+I+G	transitional nucleotide substitution model

Genes

5.8S	5.8S ribosomal gene
18S/SSU	small ribosomal subunit
28S/LSU	large ribosomal subunit
CO1/cox1	cytochrome c oxidase I
ITS1	internal transcribed spacer 1
ITS2	internal transcribed spacer 2
psbA	photosystem II protein D1
psbC	photosystem II CP43 protein
rbcL	ribulose-1, 5-bisphosphate carboxylase/oxygenase

Chapter 1: General Introduction

1.1 Introduction

Unicellular diatoms are one of the most common forms of phototrophic microbes in aquatic systems, representing an estimated 200,000 species globally (Mann and Droop, 1996). Diatoms are also responsible for fixing at least 20% of all carbon globally per annum (Mann, 1999) and are proven bioindicators of water quality (Wu, 1999). As well, paleoecological reconstructions often use diatom assemblages to give insights into the past, including into fish communities (Gregory-Eaves *et al.,* 2003) and past climates (Chipman *et al*., 2008). These types of studies depend on a robust phylogenetic resolution of diatom fossil remains. Despite both the ecological and practical importance of diatoms, their classification and systematics remain unresolved (Mann, 1999; McManus and Katz, 2009)

1.2 Diatom Species Concepts

Diatom taxonomy has traditionally been based on the morphospecies concept that defines species by their variation in morphological characteristics (Fig. 1.1). For example, Ehrenberg (1843) first described species from the genus *Neidium* Pfitzer (1871: 39) (though they were originally in the genus *Navicula*), and work since has used the same principles: analyzing morphometric data, describing locality and environment to describe new species within the genus (Siver *et al.,* 2003; Hamilton and Jahn 2005; Cantonati *et al.,* 2010; Hamilton *et al.,* 2014). While advances in microscopy have increased our ability to study finer morphological features under increasingly higher magnification, it has also resulted in the proliferation of novel species which often lack other methods of characterization (*i.e*. genetics and life cycle data).

The biological species concept introduced by Mayr (1942), states that a species is defined by a group of individuals that have the ability to reproduce viable fertile offspring. This is the concept most usually applied to organisms. In the genus *Neidium*, mating experiments conducted by Mann (1984), suggested that reproductive isolation may occur, and recent studies which also included mating experiments further demonstrated this (Amato *et al.*. 2007; Vanormelingen *et al.,* 2008; Quijano-Scheggia *et al.,* 2009). However, the biological species concept is difficult to apply to eukaryotic protists including diatoms. This is due to the changing reproductive behaviour of diatoms (sometimes asexual, sometimes sexual), and the laboratory intensive experiments needed to explore sexual reproduction in diatoms, which may not be representative of natural populations (Mann, 1999).

The phylogenetic species concept uses molecular markers to separate species, and has been used in many taxa including green algae (Leliaert *et al*., 2009) and diatoms (Evans *et al.*, 2009; Souffreau *et al.*, 2011; Stepanek and Kociolek, 2014). Owing to the increasing ability to sequence diatoms and the proliferation of available primers, more studies are turning to methods such as barcoding to delineate diatom systematics. While this innovation has allowed for further exploration of species and biodiversity (Hebert, 2003), many traditional taxonomists caution against barcoding, insisting that it will come at a loss of traditional taxonomic expertise (DeSalle *et al*., 2005; Will and Rubinoff, 2004). As well many issues with barcoding techniques, including gene selection bias (Moritz and Cicero, 2004) show that barcoding cannot solve all issues with taxonomy and should not replace it (Alverson, 2008). Despite these shortcomings, molecular sequences analyses from DNA material are invaluable for phylogenetics.

1.3 DNA Barcoding and Molecular Markers

DNA barcoding uses relatively short fragments of DNA to identify species (Hebert *et al*., 2003) and this method has been successful for species identification in a wide variety of animal, plant and fungal species (Hebert *et al.*, 2003; Savolainen *et al*., 2005; Summerbell *et al*., 2005). Molecular markers allow for species level classifications and several studies with diatoms have been conducted at the genus (Evans *et al.*, 2007; Urbánková Veselá, 2013) and class level (Moniz and Kaczmarska 2009; Moniz and Kaczmarska 2010; Zimmermann *et al.,* 2011).

Determining the level of variation which is representative of intraspecific versus interspecific variability is a problem with DNA barcoding. Hebert *et al.* (2004) proposed a standard sequence threshold, whereby 10 times the intra-specific variation indicates a group as a separate species for the *CO1* mitochondrial marker. But this rather arbitrary value needs more anlaysis within the specific groups of study (Moritz and Cicero, 2004). While DNA barcoding will not give a definitive answer to the species concept problem currently faced in diatoms (Mann, 1999), it should help define diatom populations, and shed light on genetic variability which can help sort out phylogenies and patterns in evolution.

An ideal molecular marker for diatoms which would display sufficient inter- and intra-specific variation has yet to be determined. Several molecular markers have been tested and used for analysis (Evans *et al*., 2007; Moniz and Kaczmarska, 2010; Moniz and Kaczmarska, 2009; Urbánková and Veselá, 2013; Zimmermann *et al*., 2011). Cytochrome c oxidase1 (*CO1*), a mitochondrial gene has been chosen as the barcoding gene for animals (Hebert *et al*., 2003, 2004). It had also shown promising results with red algae (Saunders, 2005). Evans *et al*. (2007) evaluated its ability to effectively separate species of a genus of diatom (*Sellaphora*) and found that it was able to resolve the species. However, in Moniz and Kaczmarska's (2009) assessment

of potential barcode regions, *CO1* was very difficult to successfully amplify and generally group specific primers were needed. Several studies have examined Ribulose-1, 5-bisphosphate carboxylase/oxygenase (*rbcL*), a chloroplast gene present in diatoms and other eukaryotic phototrophs as a possible barcoding gene (Evans *et al*., 2007; MacGillivary and Kaczmarska, 2011). Hamilton *et al.* (2015) (Appendix A) used *rbcL* across many genera of diatoms and found that it was able to effectively separate the species tested. The nuclear ribosomal gene complex (18S (SSU), ITS1, 5.8S, ITS2, 28S (LSU)) has been tested in a number of studies (Evans *et al*., 2007; Zimmermann *et al*., 2011). Following suggestions from the literature, my pre-thesis research Hamilton *et al.,* (2015) (Appendix A), explored the use of the 18S rDNA gene and found that while it was more conserved than *rbcL*, it was still able to separate species of diatoms. Moniz and Kaczmarska (2009) suggested that the 5.8S + ITS2 region might serve as ideal for barcoding because of its ease of amplification. For this region the secondary structures of ITS2 were used to aid in the alignment, and variability.

1.4 Sequencing Techniques for Diatoms

Several techniques have been described in order to obtain genetic data needed for phylogenetic studies on diatoms. Auinger *et al.* (2008) experimented with levels of sodium thiosulfate in order to effectively amplify DNA from cells that were preserved with Lugol's solution, a common preservative for environmental samples of algae (and used for long-term storage and archiving). Henrichs *et al.* (2008) successfully amplified nuclear microsatellites from single cells that were preserved in Lugol's solution. Lang and Kaczmarska (2011), harvested single celled diatoms from cultures and completed a nested PCR strategy on them for the *rbcL* and *ITS* regions. Unused DNA extraction product was pooled and spread on a scanning electron microscopy (SEM) stub, and the valves were then examined under an SEM for identification.

These protocols, while useful, are limited in their scope of study because of the need for cultures. Not all field collected samples can be successfully cultured in the laboratory, and those strains which can may not be representative of *in situ* populations. Furthermore, it has long been known that long-term culturing of diatoms leads to abnormalities in diatom cells (e.g. Estes and Dute, 1994). These abnormalities in phenotype have roots in genotypic expression, thus cultured diatoms may show genetic variations that are not found in natural populations.

In my pre-thesis work, we introduced a technique which can be used to examine cryptic species within natural diatom populations (Hamilton, Lefebvre and Bull (2015), Appendix A). The technique allows for a single diatom cell of interest, collected from the field or from a collection, to be isolated and genes of interest to be amplified and sequenced. Both live cells and RNAlater preserved cells can be used to amplify multiple molecular markers for phylogenetic study of an individual (Hamilton *et al.,* 2015). This method allows the use of field populations, and is thus able to determine the inherent genetic variation of natural populations without the need for culturing of samples from the field.

1.5 Phylogeny of the diatom genus *Neidium*

Historically, diatoms have been separated into two broad groups based on the radial vs. bilateral symmetry of their characteristic silica frustules: Centrics and Pennates. The latter encompass the largest number of species including many speciose genera. For example, the genus *Neidium* in the family Neidiaceae is a group of widely distributed freshwater benthic pennate diatoms. *Neidium* taxa occur in mostly acidic and low conductance brackish ponds, wetland areas and seepage pools, and are generally a rarer species in a system, often comprising less than 2% of the total diatom assemblage (Round *et al.,* 1990). In practice, the genus *Neidium* is useful in assessing water quality near mining sites (Herlory *et al*., 2013), and heavy metal

pollution (Ivorra *et al*., 1999) due to their ability to tolerate high levels of copper (Schroeder, 1939) and acidophilic environments (Potapova and Charles, 2002).

Pfitzer (1871) first described the genus based on chloroplast structure and general valve outline. Recently, Hamilton and Jahn (2005) defined and emended the genus based on the generic type species of *N. affine* Ehrenberg (1843) with an individual from the Ehrenberg collection material from in Newfoundland. Key morphological features of the genus *Neidium* include: longitudinal canals, apical flaps, and bilaterally deflected proximal raphe ends (a longitudinal groove in the valve) (Fig. 1.2). Over the last 80 years, 300 species have been described or transferred to this genus (Hamilton *et al*., 2014). Despite the long taxonomic history of this genus, there has been considerable confusion in the identification of many of the *Neidium* taxa including common species such as *N. amphigomphus*, *N. dilatatum* and the species complex of *N. affine* (Hamilton and Jahn, 2005). The large (>100 µm) *Neidium* taxa studied within this thesis were among the first to be described by Ehrenberg and Recently two new genera were proposed based on morphometric data alone (Cantonati *et al.*, 2010; Lange-Bertalot and Genkal, 1999). However, there has been little genetic information gathered on this genus and all species descriptions have been based on morphology alone. Owing to the importance of *Neidium* as a bioindicator, correct species identifications are critical. Thus using a single cell molecular approach coupled with morphological data could likely improve the phylogeny of *Neidium* taxa.

1.6 Thesis rationale and objectives

In order to best typify microscopic organisms, it has been proposed to use a multidisciplinary approach that uses data from molecular, morphological and micromorphological sources to delimit species and determine their relationships (Alverson, 2008; De Clerck *et al*., 2013). With respect to diatoms by adding molecular sequence data along

with morphometric analysis and SEM imaging to verify and correct existing taxonomical work, species can be more accurately defined and described. While DNA barcoding techniques face a number of theoretical considerations, investigating current diatom phylogeny with these techniques may allow for greater insight. They also allow the opportunity to explore cryptic species, and tease out current phylogenetic uncertainties.

The goal of my thesis research was to use a multidisciplinary approach of traditional morphological methods and multi-locus phylogenetic methods to define and delimit species in the genus *Neidium*. This benthic diatom genus was chosen for two reasons: 1) its large size suitable for the single cell isolation technique, and 2) the current phylogenetic confusion existing in the literature. The samples sites within this study were chosen because 1) Ehrenberg used these sites in his original descriptions (Newfoundland and New York), and 2) include habitats which *Neidium* taxa are associated with (Ontario, Québec, and Nova Scotia). To accomplish this, in Chapter 2 I first isolated large (>100 μm) *Neidium* taxa and using the single cell molecular methodology (Appendix) began to sequence these taxa for two molecular markers (18S and *rbcL*, Chapter 2 and Appendix). These larger (>100 μm) *Neidium* were the first to be described by Ehrenberg and their validity as phylogenetically distinct taxa was assessed. The objective of Chapter 3 was to: increase the number of molecular markers used in the phylogenetic analyses, delimit the smaller (<100 μm), more morphologically similar species, and to increase the sampling size and geographic range of the *Neidium* taxa examined. From these markers I was able to delimit several species within the genus *Neidium* from their phylogenetic relationships (Chapter 2 and Chapter 3). I was then able to compare the phylogenetic results to the morphological data and original species descriptions, and in some cases describe new species

(Chapter 2 and Chapter 3). Chapters 1 and 4 consist of the general introduction and thesis conclusions respectively.

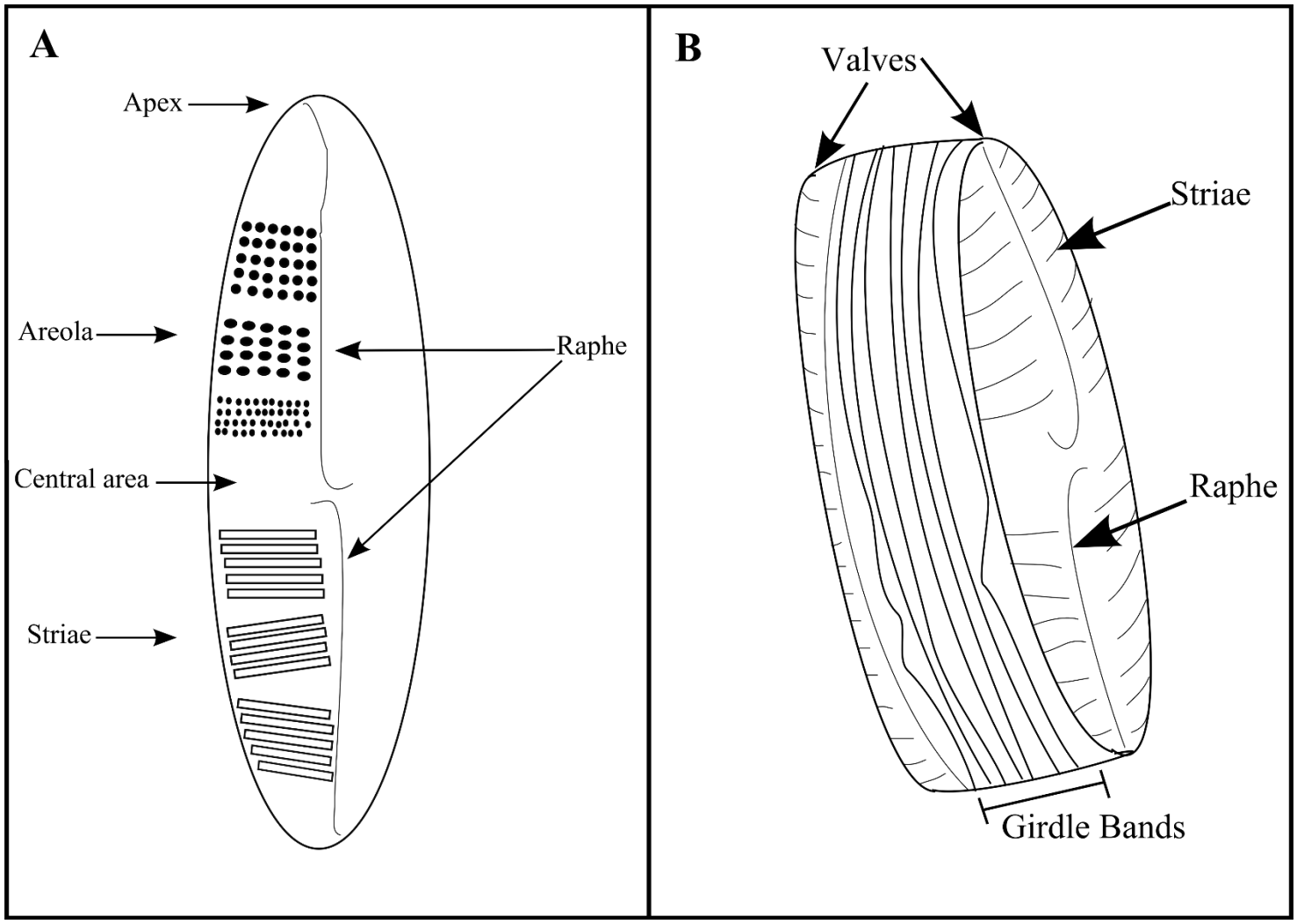


Figure 1.1 Line drawing illustrating common features present on pennate diatom frustules. A. Valve view of pennate diatom. B. Girdle view of pennate diatom.

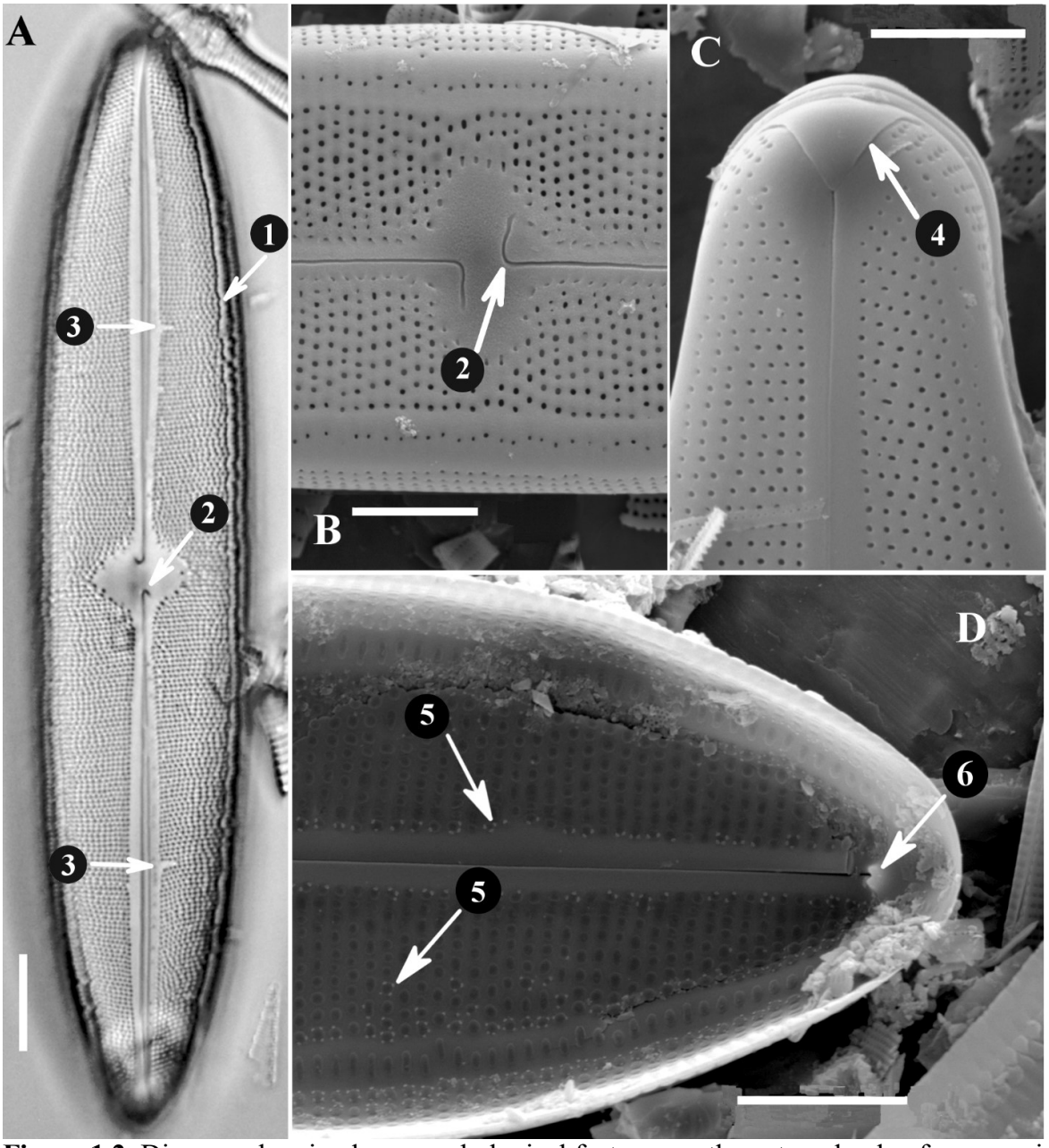


Figure 1.2. Diagram showing key morphological features on the external valve face associated with the genus *Neidium*. 1. Longitudinal canals. 2. Bilaterally deflected raphe ends. 3. Voight faults. 4. Apical flaps. 5. Renilimbia. 6. Helictoglossae. A: CANA 100072, LM of *Neidium* sp. B: CANA 108131, SEM of *Neidium* sp. C: CANA 100021, SEM of *Neidium* sp. D: CANA 93446, SEM of *Neidium* sp. Scale bar = 10 µm (A, B), 2 µm (C), 5 µm (D).

Chapter 2: Morphology and molecular studies on large *Neidium* species (Bacillariophyta) of North America, including an examination of Ehrenberg's types

Keely Lefebvre[1*] and Paul B. Hamilton[2]

[1] Ottawa-Carleton Institute of Biology, University of Ottawa

[2] Research and Collections Division, Canadian Museum of Nature

This chapter was published as:

Lefebvre, K.E., Hamilton, P.B. 2015. Morphology and molecular studies on large *Neidium* species (Bacillariophyta) of North America, including an examination of Ehrenberg's types. Phytotaxa 220 (3): 201—223.

DOI: http://dx.doi.org/10.11646/phytotaxa.220.3.1

This article is reprinted with some minor modifications with permission from the publisher, Magnolia Press.

2.1 Abstract

The genus *Neidium* contains a large array of diatoms with a wide range in structural and morphological forms. Many of the larger species (>100 μm) in this genus are old taxa dating back to the 1800s. However, there continues to be confusion over these large species including *N. iridis*, *N. dilatatum*, *N. firma*, and *N. amphigomphus*. In this study, selected *Neidium* taxa from North America were examined using LM and SEM images from both Ehrenberg's original samples and present day samples from Ontario (Canada) and New York State (USA). As well, *Neidium* individuals were isolated from Adirondack Park, New York (USA) and Ontario (Canada), amplified using a nested PCR protocol and sequenced for *rbcL* and 18S barcoding genes. The sequence data was concatenated to construct phylogenetic trees using Maximum Likelihood and Bayesian Analysis techniques. Here we present emended species descriptions and sequence data of four previously named *Neidium* taxa: *N. tumescens*, *N. hitchcockii*, *N. dilatatum* and *N. amphigomphus*. In addition, we designate isolectotypes for *N. hitchcockii*, *N. dilatatum* and *N. amphigomphus*. A new species is also formally described—*N. fossum*, *sp. nov*.—with a designated holotype and sequence data. *Neidium fossum* is distinguished by its size, longitudinal canal structure, central area and proximal raphe ends. Future work combining traditional morphological methods and phylogenetic methods will allow for further delineation of *Neidium* species and other diatom taxa.

Keywords: *Neidium*; *Neidium fossum*; Phylogenetics; diatoms; *rbcL*; 18S; algae; North America

2.2 Introduction

The diatom genus *Neidium* currently contains over 300 taxa and has a history dating back to the late 1800's when Pfitzer (1871) separated a small group of taxa from *Navicula* Bory de Saint-Vincent (1822: 128) based on cellular structure, especially chloroplast formation and orientation. Pfitzer identified four species—*N. affine* (Ehrenberg 1843: pl.2/5, fig. 4) Pfitzer (1871: 39), *N. amphigomphus* (Ehrenberg 1843: 417) Pfitzer (1871: 39), *N. amphirhynchus* (Ehrenberg (1843: 417)) Pfitzer (1871: 39), and *N. firmum* (Kützing 1844: 92) Pfitzer (1871: 39) —which still belong within *Neidium*. Later Cleve (1894) studied the valve morphology and made interesting observations on the proximal and distal raphe fissures, combined with unique areolae along the valve margins. Cleve added 10 more taxa to this genus. From these early studies, six species were described by Ehrenberg (1843) based on specimens from North and South America. Although Ehrenberg (1843, 1854) was thorough in his early work, using lower brightfield microscopic studies (*ca*. 400×), he either changed his concept of described species through time or merged morphological forms into single species. To this day, confusion in the identification of taxa like *N. iridis* (Ehrenberg 1843: 418) Cleve (1894: 69), *N*, *amphigomphus*, *N. amphirhynchus*, *N. dilatatum* (Ehrenberg 1843: 418) Cleve (1894: 70), *N. maximum* (Cleve 1894: 69) Meister (1912: 109) and *N. ampliatum* (Ehrenberg 1854: 16) Krammer in Krammer and Lange-Bertalot (1985: 101) exists, while other taxa like *N. hitchcockii* (Ehrenberg 1843: 418) Cleve (1984: 69) are well defined (Hamilton *et al*., 1995). This confusion was further linked to the similarity of valve morphologies between many of the species. The identification problem faced by Ehrenberg was illustrated by Hamilton and Jahn (2005) with the typification of the generic type *N. affine* using Ehrenberg's type material from Newfoundland, Canada. In addition, later researchers including Cleve, Mayer, Hustedt, Reimer and Krammer made interpretations of

N. iridis and *N. amphigomphus* which did not resolve the phylogeny of these large-sized *Neidium* species.

At present, eight individual cultures have been used to identify four genes and three taxa in the genus *Neidium*. These include *N. affine auct. nonnull*., *N. bisulcatum* (Lagerstedt 1873: 31) Cleve (1894: 68), *N. productum* (Smith 1853: 51) Cleve (1894: 69) (Bruder and Medlin 2007; Theriot *et al*., 2010) and *Neidium* sp. (Witkowski *et al*., 2014). The genes currently used in *Neidium* gene sequencing include *rbcL*, 18S, *psbC* and 28S. The two prominent genes *rbcL* and 18S used in these works have been identified as relatively good discriminators of diatom species (Evans *et al*., 2007; Hamilton *et al*., 2015; Mann *et al*., 2010; Moniz and Kaczmarska, 2009). In addition, several studies have used barcoding techniques with these genes to study cryptic and cosmopolitan diatom species (Souffreau *et al*., 2011, 2013; Urbánková and Veselá, 2013; Witkowski *et al*., 2014). Recent developments in single-cell gene amplifications and sequencing across all the diatom orders have opened the possibility of sequencing rarer taxa in genera which are difficult to culture and sequence (Hamilton *et al*., 2015; Lang and Kaczmarska, 2011; Ruck *et al*., 2014). Further, molecular information can also be obtained from single isolates in fixed samples (RNAlater, Lugol's solution, buffered formalin) which further open the possibility of studying rare taxa from isolated regions of the world (Auinger *et al*., 2008; Godhe *et al*., 2002; Hamilton *et al*., 2015).

The objective of this study was to validate several large (>100 µm), historical taxa first described by Ehrenberg from North America representing the genus *Neidium* using molecular and morphological metrics. Single cell sequencing of two genes (*rbcL* and 18S) with taxa replication was used for molecular studies. Morphology studies were completed using extensive research on original material, drawings and descriptions and connection with extant populations

from North America. In combination, this study merged historical and classical morphological studies with multi-gene sequencing spanning a research time period from 1832 to present.

2.3 Materials and Methods

Thirty-three live samples (Table 2.1) were collected from lakes, ponds and streams from south-eastern Ontario and western Quebec (Canada), and northern New York State (Adirondack Park, USA). In the lab, subsamples were taken from the live material for morphological study and DNA analysis. Samples were kept live in the lab at normal room conditions for approximately one week. Through this period single living cells were isolated and prepared for DNA extraction and amplification. In addition, after the first week, subsamples were also fixed with RNAlater® (http://www.protocol-online.org/prot/Protocols/ RNAlater-3999.html) and Lugol's iodine solution (10g I_2, 20g of KI, 200 mL ddH_2O) for storage. For long term conservation, freeze dried subsamples for morphological studies were deposited in the Canadian Museum of Nature Phycology Collection (CANA 93650, 93279–93284, 93371–93375, 100064–100073).

Micas and original samples from the Ehrenberg collection representing *Neidium* taxa from North America were examined at the Institute of Paleontology, Museum für Naturkunde, Humboldt University of Berlin and at the Canadian Museum of Nature. Finally, a fossil sample in the CANA collection (CANA 80874) from Cherryfield Maine was examined for *N. tumescens*.

2.3.1 Morphological studies

Samples for morphological analysis were digested with hot H_2SO_4:HNO_3. After the removal of organic matter, the samples were washed through a series of dilutions (at least 5×) to remove

oxidative byproducts and residual acid. Subsample slurries were dried onto microscope coverglass and mounted onto microscope slides using Naphrax®. Microscopic studies were conducted using a Leica DMR microscope equipped with a Pixelink digital camera or a Nikon Microphot-FX microscope with the same camera system. Specimens were observed at magnifications ranging from 640–1000× using NPlan or Plan Apo objectives (NA 1.25–1.35) with bright field (BF), differential interference contrast optics (DIC), and phase contrast (PC) optics. Scanning electron microscopy preparations were made by drying a subsample of cleaned material onto aluminum foil and mounted onto aluminum stubs using double sided carbon tape. In some cases, samples for SEM study were filtered on either 8.0 or 0.8 µm filter papers and mounted on aluminum stubs using double sided carbon tape. The mounted stubs were coated with ~500 Å of Au or AU-Pd and examined with an FEI XL30 (FEI, Hillsboro, USA) tungsten filament environmental SEM (ESEM) with accelerating voltages between 10–20 kV and a working distance of 5–10 mm. Morphological metrics included valve length, width, stria density, areola density, and number of canals. Fine structural details like termination of raphe ends, areolae structural formation, orientation of renilimbia and formation of helictoglossae were recorded. All terminology follows Stosch (1975), Krammer and Lange-Bertalot (1986), and Siver *et al*. (2003). Type specimens of each species were selected to best represent the general morphology of the population. Type specimens represent a median form recognizing the variability of valve across the size diminution gradient.

At the Institute of Paleontology, Museum für Naturkunde, Humboldt University of Berlin (BHUPM), selected micas and samples (if available) were studied. Ten samples cited for *Neidium* (*Navicula*) taxa from Ehrenberg (1843) were examined:

1. Ehrenberg 1843, p. 315, pl. II: fig. V. 4: “Real del Monte, 8556 Fufs über der Meeresflache, leg. C. Ehrenberg”. Sample no. 1228, drawing no. 2068 (dry material available).

2. Ehrenberg 1843, p. 337: “Westpoint, New York, 40° N 74° W, lebende Probe aus Torf-Wasser, leg. Prof. Bailey”. Sample no. 1755a, drawing no. 2073 (dry material available).

3. Ehrenberg 1843, p. 344: “Stratford, Connecticut, Fossile, leg. Prof. Bailey”. Sample no. 1763, drawing no. 2234 (dry material available).

4. Ehrenberg 1843, p. 342: “Andover (Andower), Massachusetts, fossile, leg. Prof. Bailey”. Sample no. 1771, drawing no. 2239 (dry material available)

5. Ehrenberg 1843, p. 349: “Boston, Massachusetts, lebende , peat meadow near Boston, leg. Prof. Bailey”, sample no. 1696, drawing no. 2238 (no material available)

6. Ehrenberg 1843, p. 350: “Bridgewater, Massachusetts, Torferde, leg. Prof. Hitchcock”, Sample no. 1769, drawing no. 2241 (dry material available)

7. Ehrenberg 1843, p. 352: “Pelham, Massachusetts, Torf, leg. Prof. Hitchcock”, Sample no. 1767, drawing no. 2237 (dry material available)

8. Ehrenberg 1843, p. 347: “Smithfield, Rhode Island, Fossile, , leg. Prof. O. Mason”, no sample, drawing no. 2236 (no material available (?))

9. Ehrenberg 1843, p. 356: “Blue Hill, Maine I, fossile, leg. C.T. Jackson”, Sample no.1775, drawing no. 2246 (no material available (?))

10. Ehrenberg 1843, p. 357: “Blue Hill, Maine II, fossile, leg. C.T. Jackson”, Sample no.1776, drawing no. 2247 (no material available (?))

Specimens labelled *Navicula amphigomphus*, *Navicula dilatata*, and *Navicula hitchcockii* were searched for on the drawing sheets; if there were no labels, the published drawings were consulted and compared. At the top of each individual drawing Ehrenberg wrote the number of the corresponding mica-preparation (Lazarus and Jahn, 1998). The significant micas along with other mica replicates from the same sample were examined and photographed at 200× and 400× with a Zeiss microscope without oil immersion.

2.3.2. Phylogenetic studies

All single cell diatom isolations, nested PCR and sequencing methods followed the protocol from Appendix A: Hamilton, Lefebvre and Bull (2015). Nested PCR amplifications were performed on partial coding regions of the chloroplast gene ribulose1, 5-biphosphate carboxylase/oxygenase large subunit and the nuclear gene 18S rRNA using primers listed in Table 2. The sequences were edited in Geneious v. 6.1.5, and the consensus sequences were aligned using the MAFFT alignment tool. For the phylogenetic analyses, *Luticola goeppertiana*, *Biremis* sp., and *Scoliopleura peisonis* were used as outgroups (Table 2.1). The *rbcL* dataset was trimmed to 1281 characters, and the 18S dataset was trimmed to 1067 characters.

2.3.3 Phylogenetic analyses

The nucleotide substitution model used in phylogenetic testing was determined using likelihood scores and AIC values of each gene in JModel Test v.2.1.4 (Darriba *et al*. 2012). All phylogenetic analyses were carried out using the transitional nucleotide substitution model (TIM3+I+G). Bayesian phylogenetic analysis (BA) was carried out in BEAST v.3.1.2 (Drummond *et al*. 2012), where a Monte Carlo Markov Chain (MCMC) sampler was run for 10 million generations for both the individual gene and the concatenated gene datasets with runs sampled every 1000th generation. The first 10% were discarded as burn-in. The convergence and

stationarity of the BA results were analysed in Tracer v1.6 (Rambaut *et al*. 2013). Maximum likelihood analyses (ML) were carried out on PhyML v3.1 (Guidon and Gascuel 2003), with 1000 bootstrapped replicates.

2.4 Results

2.4.1 Phylogenetic Analysis

Thirty-three *Neidium* specimens were sequenced and analysed both morphometrically and phylogenetically. Thirty-three sequences were determined for 18S, and twenty-nine sequences were determined for *rbcL*, with four *rbcL* sequences published from Hamilton *et al*. 2015. LM images of each of the diatoms isolated and sequenced can be seen in Appendix B, Fig. B.3. The dataset included 1281 characters and 1067 characters, for *rbcL* and 18S sequences respectively. Phylogenetic trees of each gene showed the same topologies, thus only the concatenated phylogenetic analysis is shown. Both the BA tree with posterior probabilities (PP) and the ML tree with bootstrap values (BS) are shown, Figs 2.48 and 2.49 respectively. The phylogenetic trees produced by BA and ML (-lnL = 6314.30382) had similar topologies, with many of the same major divisions (Figs 2.48–2.49).

Within the datasets, there was robust support for four different species, *Neidium hitchcockii* (1.00, 99%), *N. amphigomphus* (0.99, 100%), *N. tumescens* (1.00, 100%), and *N. dilatatum* (1.00, 100%), PP and BS respectively. There was also one specimen which is described as a novel species, *N. fossum*. *Neidium tumescens* and *N. dilatatum* are seen to be most closely related to one another, whereas *N. amphigomphus* is far removed from the other specimens (Figs 2.48–2.49). *Neidium fossum* appears as a distinct entity within several different *N.* sp's. *Neidium*

hitchcockii appears to be a sister to all other Neidium species in both trees (1.00, 71%), MLBS and BAPP respectively.

2.4.2 Species Descriptions

Division Bacillariophyta

Subdivision Bacillariophytina

Class Bacillariophyceae

Order Naviculales

Sub-order Neidiineae

Family Neidiaceae

Neidium tumescens (Grunow in Schmidt *et al*. 1877: pl.49, fig. 10) Cleve (1894: 70) s. str. (emend. K.Lefebvre and P.B.Hamilton) (Fig. 2.22)

Basionym: *Navicula* (*firma* var.) *tumescens* Grunow in Schmidt, Commissions–Verlag Von Ludwig Siever's Buchandlung (1877: fig. 49: 10)

Emended species description (individuals examined for morphological analyses: 15, examined for molecular analyses: 2): Valves are broadly lanceolate to elliptical with narrowly rostrate to attenuated ends (Fig. 2.22). Valve length: 181–216 µm, width: 55–83 µm. Striae are weakly radiate at the centre changing to parallel and slightly convergent at the apices. The striae appear densely spaced, although individual stria and areola can be discerned in LM, 15–17 present in 10 µm. Voigt faults are present approximately ½ way from the margin of the central area to the apex on the secondary side of the valve. Areolae are more densely

spaced closer to the axial area, 14–18 in 10 μm becoming less dense towards the valve margins 11–14 in 10 μm. Axial area is linear-elliptical from center to apex with the central area transapically expanded and rounded. Central area extends over 1/4 to 1/5 the valve width. Ten or more longitudinal canals are present along the sides of the valve. The first canal closest to the axial area is wider than other canals. In LM, the raphe appears filiform and linear-elliptical, in SEM the raphe is linear. The proximal raphe endings are tightly hooked and the distal ends terminate in a lacinia (bifurcate in appearance). Internally, the raphe is straight, the distal raphe ending forms a small helictoglossa positioned on the edge of a small silica plate at the apex. The proximal raphe fissures from the two raphe branches form narrow linear helictoglossae which are weakly connected by a fine silica ridge. Renilimbia are prominent on areolae along the axial area and the longitudinal canals. The areolae are chambered and interconnected along the striae.

Observations:—*Navicula tumescens*' defining features were the very large number of canals (10+), as well as its unique shape and large size (Fig. 2.22, type locality). It is much more elliptical than other large species like *N. amphigomphus* (Figs 2.26–2.30) or *N. dilatatum* (Figs 2.23–2.24), with ends that are much more pinched. All the longitudinal canals internally have linear to linear-elliptical apertures (Güttinger 1989). The first longitudinal canal is always larger and internally is raised as a small dome. The remaining longitudinal canals on the outside are flush with the valve face and internally are weakly raised. Areolae within the valve wall form a complex chambered network similar to *N. dilatatum*, *N. amphigomphus*, *N. fossum*, *sp. nov.* and *N. hitchcockii*. The central area helictoglossae are distinct separated, but united by a fine silica ridge. Examples of *N. tumescens* in the literature includly Patrick and Reimer (1966: 34, fig. 4),

Metzeltin and Lange-Bertalot (2007: 180, figs 1–2), Camburn and Charles (2000: 22, fig.4) and Güttinger (1989: 2.05.32-1). One specimen was observed from fossil material collected at the type locality Cherryfield, Maine (Fig. 2.22). This is a freshwater species found in acidic waters from eastern North America, not an Arctic species.

Neidium dilatatum (Ehrenberg 1843: 418) Cleve (1894: 70) (Figs 2.16, 2.23–2.24, 2.35–2.38)

Basionym: *Navicula dilatata* Ehrenberg in Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin (1843: 418)

Original figure: Ehrenberg in Mikrogeologie, (1854: pl. 3(II): 5, pl. 3(IV): 13, pl. 4(I): 5, pl. 4(II): 9, pl. 5(III): 1)

Emended species description (individuals examined for morphological analyses: 28, examined for molecular analyses: 6): Valves are elliptic-lanceolate with narrowing pointed apices (Figs 23–24). Valve length: 174–269 µm, and width 40–55: µm. The axial area from center to apex is linear-elliptical. The central area is transapically elliptical, covering 1/2 to 1/3 the valve width (Figs 2.23–2.24). The valve has many (≥5) small longitudinal canals of even size along each margin (Figs 2.23–2.24). In LM, the raphe appears filiform and linear-elliptical, in SEM the raphe is linear and the proximal ends tightly hook in opposite directions (Figs 2.35–2.38). Striae are evenly spaced and weakly radiate at the centre becoming weakly convergent at the apices with 16–19 in 10 µm. One Voigt fault is clearly present half way between the centre and apex on the secondary side of the valve (Figs 2.23–2.24, 2.36–2.37). The areolae are small, fine rounded, and evenly spaced 14–18 in 10 µm. Externally in SEM the areolae are small round to elliptical and weakly recessed. Fine cribra are present within the areolae (Figs

2.35–2.36). In LM the areolae appear as fine perforations (Figs 2.23–2.24). Lacinia are clearly present externally (Fig. 2.35). Internally, rounded elevated longitudinal canals are visible (Fig. 2.38), and the fine rounded areolae are in tightly packed, evenly spaced striae (Figs 2.37–2.38). In intact specimens the internal valve face is covered by a hymen. Renilimbia are present around areolae along the axial area and areolae opening into the longitudinal canals. Renilimbia are also present around randomly identified areolae on the internal valve face. The central helictoglossae are small elevated nodules close to the edge of the central area, they are weakly connected by a small fine silica ridge (Fig. 2.38 arrow). The terminal helictoglossae are small and facing inward on a thickened apical nodule. Areolae form chambers within the valve wall and the chambers are interconnected along the striae (Fig. B.2d).

Type:—UNITED STATES OF AMERICA. Massachusetts: Andover (Andower), fossil, *Prof. Bailey*, *ca. 1842*, Ehrenberg Sample no. 1771 (lectotype designated here: BHUPM! slide ECO-101, isolectotype: circled specimen illustrated here as Fig. 2.16).

Observations:—*Neidium dilatatum* was recorded from 10 localities in Ehrenberg's original work including living and fossil materials (Figs 2.1–2.3, 2.9–2.15). All the drawings show a large linear-elliptical to elliptical valve with narrowly rounded almost apiculate apices. *Neidium dilatatum* was also shown to be larger than *N. amphigomphus* (Figs 2.1–2.3). In eight of the drawings, multiple lines were presented along the valve margins (Figs 2.1–2.3, 2.10–2.13, 2.15). In nine of the drawings round-elliptical central areas were drawn (Figs 2.1–2.3, 2.9–2.12, 2.14–2.15). Specimens examined from Ehrenberg's mica collection confirm the general morphological

features described above (Fig. B.1). However, size differences for specimens were observed between sites. The smaller valves (Figs 2.9, 2.13, 2.14) <160 μm, from Stratford, West Port and Wrentham, also appear to have one larger longitudinal canal along the margin with associated smaller canals (see Fig. B.1, m–q). LM and SEM examination of possible syntype material from West Point, and Stratford confirm smaller valves with one larger longitudinal canal. In contrast, the larger linear-elliptical valves (>150 μm) have many longitudinal canals of more or less the same size. Further, the smaller valves have the same size range as *N. amphigomphus*. Based on direct comparisons of valve size between *N. amphigomphus* and *N. dilatatum*, from Andover (Andower) (Fig. B.2), Boston and Bridgewater, and large valves observed from Pelham, Smithfield and Blue Hill, *N. dilatatum* sensu stricto is a large linear-elliptical valve with multiple longitudinal canals of even size. The size ranges for observed Ehrenberg samples were length 164–264 μm, width 39–56 μm, and 14–17 striae in 10 μm. In the past, this taxon has been not been identified and used in the common taxonomic literature.

Specimens of *N. dilatatum* were commonly identified as *N. iridis sensu auct*. or sometimes *N. amphigomphus sensu auct*. *nonnull*. An example can be found in Krammer and Lange-Bertalot (1986: 104, figs 1, 4). Specimens identified as *N. iridis sensu auct*. *nonnull*. (Camburn and Charles 2000: 21, fig. 1; Metzeltin and Lange-Bertalot 2007: 179, figs 3–5) are likely to also be *N. dilatatum*, given the large size range of the specimen. At the smaller end of the size range *N. dilatatum* can be confused with *N. amphigomphus*, but comparisons of the shapes and sizes of the areolae from centre to valve margin (becoming larger in *N. amphigomphus*) and central area (small round in *N. amphigomphus*) can assist in distinguishing the two taxa. Metzeltin and Lange-Bertalot (2007: 183, Figs 1–4) show fossil specimens identified as *N. amphigomphus* from Florida which fit some of the characteristics of *N. dilatatum* such as overall shape, and

areola density. However, the smaller size (118–171 μm) of these fossils falls under the extant size ranges we observed.

Within the samples analysed phylogenetically (Table 1), *N. dilatatum* was long and narrow with straight to mildly elliptical sides. Based on genetic analysis, there is a large range in valve size and overall morphology (Fig. B.3, l–q). The ends are narrowly rounded (Figs 2.23–2.24, 2.36–2.37), which was also seen with the valves of *N. fossum*, *sp. nov*. (Figs 2.31–2.34, 2.44, 2.46–2.47) compared to the apiculate ends of *N. amphigomphus* (Figs 2.40, 2.42). *Neidium dilatatum* was larger than *N. fossum* (Figs 2.31–2.34, 2.47), with a rounder central area, and very tightly hooked proximal raphe ends (Figs 2.35, 2.38). Five or more evenly sized longitudinal canals are present in *N. dilatatum* (Fig. 2.35) compared to *N. fossum* with a single large canal (Fig. B.2, f, arrow) and numerous small canals (Figs 2.43, 2.45, 2.47, Fig. B.2, f).

Neidium dilatatum appears to be globally distributed.

Neidium firmum (Kützing 1844: 92) P.B.Hamilton and K.Lefebvre, *comb. nov.* (Fig. 2.21)

Basionym: Navicula firma Kützing in Bacillarien oder Diatomeen, (1844: 92, fig. 21: 10)

Emended species description (individuals examined for morphological analyses: 1, examined for molecular analyses: not available): Valve is linear-elliptical with narrow rounded apices (Fig. 21). Valve length 79 μm, and width 16–17 μm. The axial area from centre to apex covering *ca*. 1/6 of the valve width. The central area is broad, transapically elliptical, covering 1/2 the valve width. The valve has 1, possibly up to 3 longitudinal canals of even size along each margin. In LM, the raphe appears filiform and linear-elliptical with broadly curved and deflected proximal raphe ends and terminal raphe end forms a lacinia. Striae are weakly

radiate at the centre becoming weakly convergent at the apices with 17 in 10 μm. Voigt faults are present on the secondary side of the valve. The areolae are large, round to elliptical, and evenly spaced 10–11 in 10 μm. In LM, the presence of a small pore in the center of the areolae (Fig. 2.21, arrow) suggests that the areolae are within surface depressions on the valve face.

Type:—ITALY. Grosseto: San Fiore, fossil, *ca. 1840*, BM! 18414 (lectotype designated here: individual at microscope coordinates from the center of the type sticker over to the microscope coverslip x = -17.49 mm, y = -2.27 mm, illustrated here as Fig. 2.21).

Observations: Only one large *Neidium* valve was found on the type slide which is here shown as *Navicula affine* var. *firma*. The slide was not the original examined by Kützing, but made later from original material. The valve does not exactly match the line drawing by Kützing (1844: 21, Fig. 10). The overall length matches, but the line drawing presents a valve with a length to width ratio of 4:1 while the observed specimen is 5:1. In addition, the central area in the line drawing is round covering over 1/3 the valve width while the central area on the valve from the type slide is elliptical and covers 1/2 the valve width. However, based on the linear lines along the margin, there appears to be only one or a couple of longitudinal canals, which agrees with the valve presented. Since this slide was the only slide from San Fiore in Kützing's slide collection (D.M. Williams, pers. comm.), we conclude that the observed *Neidium* specimen should represent the species and is here identified as the lectotype.

Neidium hitchcockii (Ehrenberg 1843: 418) Cleve (1984: 69) (Figs 2.19, 2.25)

Basionym: *Navicula hitchcockii* Ehrenberg in Abhandlungen der Königlichen Akademis der Wissenschaften zu Berlin, 1843: 418

Original figure: Ehrenberg in Mikrogeologie, (1854:pl. 5(III): 11)

Emended species description (individuals examined for morphological analyses: 96, examined for molecular analyses: 1): Valves are linear with triundulate margins, the ends narrow sharply forming rostrate apices (Fig. 2.25). Valve length: 62–85 µm, width: 14–22 µm. The axial area is broad, linear and centrally expanded. In SEM, the axial area is elevated and separated from the valve face by a thickened conopeum-like ridge (Hamilton *et al*. 1995: Figs 14, 17, 18). Small fine poroids are present in the axial area. The central area is small and elliptical, slightly expanded from the axial area. In LM, the raphe appears filiform and linear-elliptical, in SEM the raphe is linear and the proximal ends are long and deflected (not curved) into the central area. Sometimes the terminal end of the proximal raphe fissure is forked (Hamilton *et al*. 1995: Fig. 17). The striae are weakly radiate at the centre to weakly convergent at the apices, 19 – 22 in 10 um. Areolae are large and rounded, 20–24 in 10 µm. In SEM, externally the areolae opening are occluded by fine subsurface cribra (Hamilton *et al*. 1995: Fig. 17). Internally, the areolae are positioned with broad surface depressions (Supplemental Plate 2). In intact specimens, the internal valve face is covered by a hymen. One large, elevated, longitudinal canal is present along each margin. The canals have some or no poroid openings to the external valve face. In SEM, internally the longitudinal canal is at the edge of the valve face mantle junction with elliptical (not round) openings (Fig. B.2). Renilimbia are present around areolae along the axial area and areolae opening into the longitudinal canals (Hamilton *et al*. 1995: fig. 17). Renilimbia are also found around random

areolae on the internal valve face. The central helictoglossae is present on a transapically raised central nodule and form two distinct nodular thickenings weakly connected by a small ridge of silica (Fig. B.2). The terminal helictoglossae are small and facing inward on a thickened apical nodule (Fig. B.2). The valve wall is chambered, areolae interconnected primarily along the striae and into the longitudinal canal (Hamilton *et al*. 1995: Fig. 19, 21). Copulae (3–4) have one to two rows of poroids along the par exterior (Hamilton *et al.* 1995: Fig 16).

Type:—UNITED STATES OF AMERICA. Massachusetts: Bridgewater, peat material, *Prof. Hitchcock*, *ca*. *1842*, Ehrenberg sample no. 1769 (lectotype designated here: BHUPM! ECO-102, circled specimen, illustrated here as Fig. 2.19).

Observations: *Neidium hitchcockii* was first recorded by Ehrenberg from Bridgewater Massachusetts (Fig. 2.19). The distinctive morphological shape along with size can distinguish this taxon in Ehrenberg's original line drawings and the micas (Fig. B.1, a–d). A more recent LM and SEM study of *N*. *hitchcockii* (Hamilton *et al*. 1995) shows distinct deflected central raphe ending sometimes forked, an elevated axial region extending from apex to apex, renilimbia around the areolae, and the presence of lacinia. These characteristics were also seen in specimens from the *N*. *hitchcockii* sample, which was sequenced (Fig. 2.25, B.3, T). A few specimens were observed from type material and one individual has been selected as the lectotype (Fig. 2.19). Size ranges for the observed individuals were length 46–80 µm, width 14–19 µm, and 18–21 striae in 10 µm. Positive identifications of this taxon include Patrick and Reimer 1966 (pl. 36: 2), Krammer and Lange-Bertalot (1986: 107, fig.

3), Hamilton *et al*. (1995: figs 1, 8–24), and Camburn and Charles (2000: 20, fig. 26). More complete comparisons with *Neidium baicalense* Janitsky (1936: 693) and *N. sauramoi* Mölder (1937: 30) are still required. At present, *N. hitchcockii* is considered to be very widespread (North America eastern and western regions, Europe, Asia, Australia and New Zealand).

Neidium amphigomphus (Ehrenberg 1843: 417) Pfitzer (1871: 39) (Figs 2.17, 2.26–2.30, 2.39–2.42)

Basionym: *Navicula amphigomphus* Ehrenberg in Abhandlungen der Königlichen Akademis der Wissenschaften zu Berlin (1843: 417)

Original figure: Ehrenberg in Mikrogeologie, (1854: pl. 2(II): pl. 2(III): 8a, pl. 3(I): 6, pl. 3(II): 6, pl. 3(IV): 12, pl. 4(I): 6, pl. 4(III): 3, pl. 5(III): 15)

Emended species description (individuals examined for morphological analyses: 45, examined for molecular analyses: 11): Valves are linear to linear-elliptic with acute rounded apices (Figs 26–30). Valve length 73–141 µm, and width 23–38 µm. Axial area from center to apex linear to linear-elliptical. Central area is round to elliptical covering approximately 1/3 of the valve width (Figs 2.17, 2.39, 2.41). In LM, the raphe appears filiform and linear-elliptical, in SEM the raphe is linear. The proximal ends form small hooks in opposite directions and the distal raphe ends form a lacinia (Fig. 2.40). Striae are parallel to mildly oblique and loosely packed 15–19 in 10 µm. The areolae are large and regularly spaced, smaller and rounded areolae closer to the axial area, becoming larger and more elliptical towards the valve mantle with 14–20 in 10 µm (Figs 2.39–2.40). In SEM, externally the areolae form depressions with

smaller opening in the center which may or may not be occluded with cribra (Figs 2.39, 2.40, Siver *et al*. 2005, pl. 50, figs 2–3). 3–5 longitudinal canals are visible (Fig. 2.40). Internal the longitudinal canals are less evident (Fig. 2.41) and areolae are positioned within surface depressions and barely visible under the hymen in intact valves (Fig. 2.41–2.42). The terminal helictoglossae form small thickened nodules close to the apex. Renilimbia are positioned around areolae across the valve face (Fig. 2.41–2.42, Siver *et al*. 2005, pl. 50, Figs 5–6). The valve wall is chambered and the areolae are internally interconnected both apically and transapically. Each proximal helictoglossa forms a thickened nodule on top of the central nodule and are weakly connected by a small silica ridge (Fig. 2.41). In combination, the proximal helictoglossae are not aligned (Fig. 2.41, Siver *et al*. 2005, pl. 50, Fig. 4). The terminal helictoglossae form small thickened nodules close to the apex (Fig. 2.42).

Type:—UNITED STATES OF AMERICA. Massachusetts: Andover (Andower), fossil, *Prof. Bailey*, *ca*. *1842*, Ehrenberg sample no. 1771 (lectotype designated here: BHUPM! ECO-100, circled specimen illustrated here as Fig. 2.17).

Observations: *Neidium amphigomphus* was identified from eight localities in Ehrenberg's original work (Figs 2.1–2.8). All line drawings show valves with linear sides and apiculate apices. In six of the eight drawings, multiple lines are presented along the valve margins suggesting some type of unique valve formation (Figs 2.1–2.4, 2.6–2.7). The line drawings from Bridgewater, Andover (Andower) and Boston clearly show that *N*. *amphigomphus* is smaller than *N*. *dilatatum*. Specimens observed from the micas also illustrate valves with linear margins and apiculate apices. (Fig. B.1, e–h). Shadowing along the margins was also observed on some

of the mica specimens (Fig. B.1, e–f). In the current LM study, nine intact specimens from the type material showing multiple longitudinal canals were observed, and one from Bridgewater (Massachusetts) was selected as the lectotype (Fig. 2.17). The size ranges for observed specimens from Ehrenberg's collection were length 106–167 µm, width 26–42 µm, and 14–16 striae in 10 µm. This taxon has been commonly identified under many different concepts with only a few identifications representing *N. amphigomphus sensu stricto*. Examples of *N. amphigomphus sensu stricto* include Patrick and Reimer (1966: 34, fig. 2) and Siver *et al*. (2005: 49, figs 1–4, 50, figs 1–6). No specimens of *N. amphigomphus* were presented in Kramer and Lange-Bertalot (1986).

Within the specimens of *N. amphigomphus* sequenced and analysed phylogenetically, two key characteristics were evident. One, the distinct acute and apiculate apices of the valve and two, the areolae tend to become larger and more linear towards longitudinal canals (Figs 2.40, 2.42). This is in contrast to the smaller rounded areolae of *N. dilatatum* (Figs 2.23–2.24, 2.35, 2.37) and *N. fossum* (Figs 2.31–2.34, 2.44–2.45, 2.47). *Neidium amphigomphus* also has a circular central area, with 3–5 longitudinal canals (Figs 2.39–2.42) which contrasts with the broad central area and broad longitudinal canal with associated smaller canals in *N. fossum* (Figs 2.43–2.47).

***Neidium fossum* K.Lefebvre and P.B.Hamilton, *sp. nov*. (Figs 2.18, 2.20, 2.31–2.34, 2.43–2.47)**

Individuals examined for morphological analyses: 27, examined for molecular analyses: 1.

Valves are elliptic-lanceolate, with narrowing to subacute apices (Figs 2.31–2.34). Valve

length 92–142 µm and width 22.5–34.0 µm. Central area is transapically broad and elliptical, covering 1/3 to 1/2 the valve width. The axial area from centre to apex is linear-elliptical. In LM, the raphe appears filiform and linear-elliptical, in SEM the raphe is linear. Proximal ends form hooks in opposite directions and the distal raphe ends form a lacinia (Figs 2.43–2.47). Striae are parallel and moderately spaced becoming weakly convergent towards the apices, 17–21 in 10 µm. Voigt faults are present on the secondary side of the valve. Areolae are regularly spaced and irregular in shape, 17–20 in 10 µm. In SEM, externally the areolae are flat to the valve surface sometimes with cribra occlusions and not in deep depressions (Figs 2.43–2.44). Internally, the areolae are positioned within small depressions (Fig. 2.47). One large longitudinal canal is present along each valve margin (Figs 2.43–2.47). Additional canals 2–3 or more may extend along the valve mantle. Externally, the longitudinal canal is flat with the valve face (Fig. 2.43). Internally the canal forms a bulge along the valve margin (Fig. 2.46). A single row of linear to linear-elliptical areolae open externally and internally from the canal (Figs 2.45–2.46). Internally, areolae are covered by a hymen (Fig. 2.45). Renilimbia are positioned around areolae along the axial area, and the region of the longitudinal canals (Figs 2.45–2.46). Renilimbia can also be found randomly scattered around areolae on the valve face. Areolae form chambers in the valve wall that are highly interconnected transapically (Fig. 2.47, Fig. B.2, f). Each proximal helictoglossa forms a double elevated ribbed nodule that is loosely connected to the adjacent helictoglossa by a thin ridge of silica (Figs 2.45, 2.47). The terminal helictoglossae form on a small thickened nodules close to the apex (Figs 2.46, 2.47).

Type:—UNITED STATES OF AMERICA. New York: Adirondack Park, 43.90870 N 74.43944 W, *K. Lefebvre*, *5 May 2014* (holotype: CANA! 100072, circled specimen, illustrated here as Fig. 2.33).

Etymology:—The specific epithet (*fossum*), Latin meaning ditch, trench or canal, refers to the large longitudinal canal seen in *N. fossum*

Observations:—*Neidium fossum* was identified from four of Ehrenberg's samples, (Andover (Andower) MA, Bridgewater MA, Pelham MA and Stratford CT), where it had been listed under either *N. dilatatum* or *N. amphigomphus* (Figs 2.18, 2.20, Fig. B.2). From the Ehrenberg samples, the size ranges of *N. fossum* observed were length: 90–158 µm, width: 22–40 µm and 19–26 striae in 10 µm. Within the modern samples observed in the phylogenetic study, *N. fossum* (Figs 2.31–2.34) has a similar overall shape to *N. dilatatum* (Figs 2.23–2.24), however *N. fossum* is much smaller than *N. dilatatum* with a maximum length of 160 µm. This species can be differentiated from *N. amphigomphus*, by its narrowly rounded ends, and longitudinal canal structure of one large and several small canals (Figs 2.43–2.47). It can be differentiated from *N. dilatatum* (Figs 2.35–2.38) with its large longitudinal canal, smaller overall size, more transapically elliptical central area and its straight bifurcating terminal raphe ends (Figs 2.43, 2.46). There is taxonomic confusion around large linear-elliptical *Neidium* taxa with multiple longitudinal canals extending along the valve margins. One clade which is represented by one larger canal with associated smaller canals, can be linked to *N. fossum*. Specimens identified as *Neidium iridis sensu auct nonull*.—Camburn and Charles (2000: 21, fig. 2), Metzeltin and Lange-Bertalot (2009: 93, figs 1–4, 94, figs 1–9)—and as *Neidium amphigomphus*—Metzeltin and Lange-Bertalot (2002: 51, fig. 1), Metzeltin and Lange-Bertalot (2007: 183, fig. 5, 188, fig.

3, 189, fig. 2), Metzeltin and Lange-Bertalot (2009: 95, figs. 1–3)—are *N. fossum* or closely related undescribed cryptic species. Other similar specimens identified under *N. iridis* include Patrick and Reimer (1966: 34, fig. 1) and Krammer and Lange-Bertalot (1986: 104, fig. 2).

2.5 Discussion

Within our analysis, five large species of the genera *Neidium* were found to be well defined using the genetic and morphometric data, (*N. tumescens*, *N. hitchcockii*, *N. amphigomphus*, *N. fossum* and *N. dilatatum*). Four of the genetic specimen groups were matched with species described by Ehrenberg (1841), and one new species was described. The remaining specimens within the genera *Neidium* are denoted as *N.* sp. (Figs 2.48–2.49). Future work with increased sampling and more intensive morphometric analyses will allow for the description of these specimens and further determination of where they belong within the genus *Neidium*.

During the isolation and the morphometric analysis of the large *Neidium* species in our samples, we were unable to find anything that would morphologically match *N. maximum* due to its large size. The type slides for *Neidium firmum* (originally *Navicula firma* Kützing) was also examined as a possible species close to *N. amphigomphus* and *N. fossum*. After consulting the type slide (BM 18414, Fig. 2.21) we concluded that it did not have the longitudinal canal formation and fit the size range of *N. dilatatum* (Figs 2.23-2.24). While *N. firmum* also had a similar overall shape to *N. fossum*, the stria density and areola density were much lower (17 in 10 μm). The proximal raphe ends were much larger on *N. firma* as compared to *N. fossum* (Fig. 2.43).

In this study, we did not identify *Navicula amphirhynchus* from Ehrenberg's Rel del Monte material with distinct rostrate apices. In the future, we hope to clarify this taxon. Based on its

overall shape and size, we hypothesize that the specimen may not be within the *Neidium* genus, but actually part of *Navicula*, possibly associated with *Navicula rhynchocephala* Kützing (1844: pl.30, fig. 35). Further study into these species should be able to confirm whether they are still valid, and whether they are from the genus *Neidium*.

The phylogenetic analysis into some of the larger taxa within the genus *Neidium* displayed insights into this genus and clarified some taxonomic confusion. The sequenced specimens, *N. dilatatum* had a very large size range (length: 207–427 μm) and were all placed within the same branch of the tree. Future phylogenetic work with the addition of new genes may show that the current individuals grouped under *N. dilatatum* represent more than one species, possibly *N. dilatatum* and *N. iridis*. In contrast, the similarity in large size and other characteristics between *N. iridis* and *N. dilatatum*, may suggest that they are the same species. This would imply that in the original species descriptions made by Ehrenberg, they were wrongly separated because of the large size range within the species. Further work between specimens of these species will show whether these species are in fact different or the same. *Neidium dilatatum* is presently well circumscripted from Ehrenberg's material and present day collections from North America.

Several of the large, well-described *Neidium* species such as *N. tumescens* and *N. hitchcockii* were easily separated through molecular analyses into distinct species. The lower support values for *N. hitchcockii* are likely since there was a single individual in the analysis. Increasing the number of individuals sequenced (in progress) for this species will not only strengthen its phylogenetic placement as a distinct group, it may also solidify its position as the sister taxa to all other *Neidium* taxa sequenced. Uncovering that this species is the the sister taxon to all other *Neidium* species in this analysis is consistent with morphological observations of its unique and

complex surface morphology (Hamilton *et al*. 1995) which differentiates it dramatically from other *Neidium* species.

Using molecular data to describe and delimit species must be done with caution as there are several caveats associated with using molecular data alone. One such limitation is the amount of genetic distance that different studies use to delimit species, as it is often subjective and differs between studies (Meyer and Paulay, 2005; Witt *et al*., 2006). Another issue is determining which locus should be used for differentiating species. Some loci have resolution which may only be helpful to separate out taxa at the rank of Order, while others vary widely within individuals from the same species (Leliaert *et al*., 2014). The scale at which a study examines taxa is critical for picking effective loci. These are just a few of the reasons why it is critical to combine the molecular analyses of diatom taxa with morphological analyses.

The genetic analysis of multiple specimens from different taxa within one genus using a single-cell extraction approach is helpful in identifying potential problems with taxonomic identifications on GenBank. Our phylogenetic trees (Fig. 2.48–2.49, black arrows) show that two specimens labelled as *N. affine* are not the same species. The fact that they are not placed together in the tree suggests the presence of many *N. affine*-like forms which are poorly understood or possible cryptic species. Indeed in this study, the Genbank sequences of '*N. affine*' (Table 2.1) are shown to be more closely related to other potential species. The true *N. affine* (Hamilton and Jahn, 2005) has yet to be sequenced. One of the cultures sequenced may be similar to the recently described *N. quadratum* Kociolek (2014: 31). This highlights the strength of the single cell amplification technique, as we can definitively describe each species both morphometrically and genetically without the absolute need for cultures.

Owing to the lack of genetic sequences of diatoms in GenBank, it was difficult to find appropriate outgroups for this analysis. Our insight into the evolutionary history of the genus *Neidium* is quite limited, with few studies looking at the evolutionary relationships of this genus using phylogenetic analyses (Bruder and Medlin, 2007; Medlin and Kaczmarska, 2004; Theriot *et al*., 2010). Thus, while the outgroups used in this study are likely not the closest sister group to the genus *Neidium*, possible closer groups, like *Neidiopsis* Lange-Bertalot and Metzeltin (1999: 77) and *Neidiomorpha* Lange-Bertalot and Cantonati in Cantonati *et al.,* (2010: 196), have no sequences available in GenBank. This limitation illustrates another reason why sequencing a more diverse range of diatoms is important for both taxonomic purposes and to determine the evolutionary history for diatoms.

This study provides an example for how future taxonomic work within the diatoms can proceed. The ability to use both genetic data as well as morphometric data is described in this study. Within the genus *Neidium*, future work will focus on increasing the sample size of the smaller species sequenced in this study. With increased sample size and increased study area, more LM differences can be seen between the sequenced specimens, and using the phylogenetic relationships between species, the ability to focus on the correct grouping of specimens into species groups is possible.

2.6 Acknowledgements

The authors wish to thank Dr. Frances Pick and Dr. Julian Starr, Center for Advanced Research in Environmental Genomics, University of Ottawa, Ottawa, Canada for their insight and helpful discussions, two anonymous reviewers for their helpful suggestions and comments, Mr. Roger Bull, Canadian Museum of Nature, Ottawa, Canada for technical assistance in the laboratory, Dr.

David Williams, Natural History Museum, London, UK for access to Kützing material at the British Natural History Museum and Dr. David Lazarus, for allowing us to access to the Ehrenberg collection, Institute of Paleontology, Museum für Naturkunde, Humboldt University of Berlin, Berlin, Germany. Financial support was given to K.E.L. by the Canadian Museum of Nature under a grant by P.B.H.

Table 2.1. List of specimens sequenced in this study. Each specimen is listed with its GenBank Accession numbers, identifier, location and date of collection. Outgroups and specimens retrieved from GenBank are also listed.

GenBank Accession Number		CANA Accession Number	Identifier	Taxon	Locality and date of collection	Source
rbcL	**18S**					
AM710433.1	AM501961.1	NA	NA	*Luticola goeppertiana*	Germany	Bruder and Medlin (2007)
KM078662.1	KM078661.1	NA	NA	*Biremis sp*.	Playa Monagre (Los Santos, Panama), 20/01/2011	Witkowski, Barka and Mann, *unpublished*
HQ912473.1	HQ912609.1	NA	NA	*Scoliopleura peisonis*	UTEX FD13	Theriot *et al*. (2010)
KP325181	KP325148	CANA 100068	G4R12	*Neidium hitchcockii*	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W, 18/05/2014	
KP325182	KP325149	CANA 100068	B6R14	*Neidium sp*.	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W, 18/05/2014	
KP325183	KP325150	CANA 100068	G5R12	*Neidium sp*.	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W, 18/05/2014	
KP325184	KP325151	CANA 100068	G6R12	*Neidium amphigomphus*	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W, 18/05/2014	
KP325185	KP325152	CANA 93284	C9R13	*Neidium amphigomphus*	Big Moose Lake, Adirondack Park (New York, United States of America) 43.81723 N 74.88400 W, 10/05/2014	
KP325186	KP325153	CANA 100068	G8R12	*Neidium amphigomphus*	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W, 18/05/2014	
KP325187	KP325154	CANA 100073	F9R13	*Neidium amphigomphus*	Boat Launch, Big Moose Lake (New York, United States of America) 43.8271 N 74.83831 W, 10/05/2014	
KP325188	KP325155	CANA 100068	A3R14	*Neidium amphigomphus*	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W, 18/05/2014	
KP325189	KP325156	CANA	C3R7	*Neidium*	Roadside Pond, Hwy 56 (New York, United States of America)	

Table 2.1. Continued

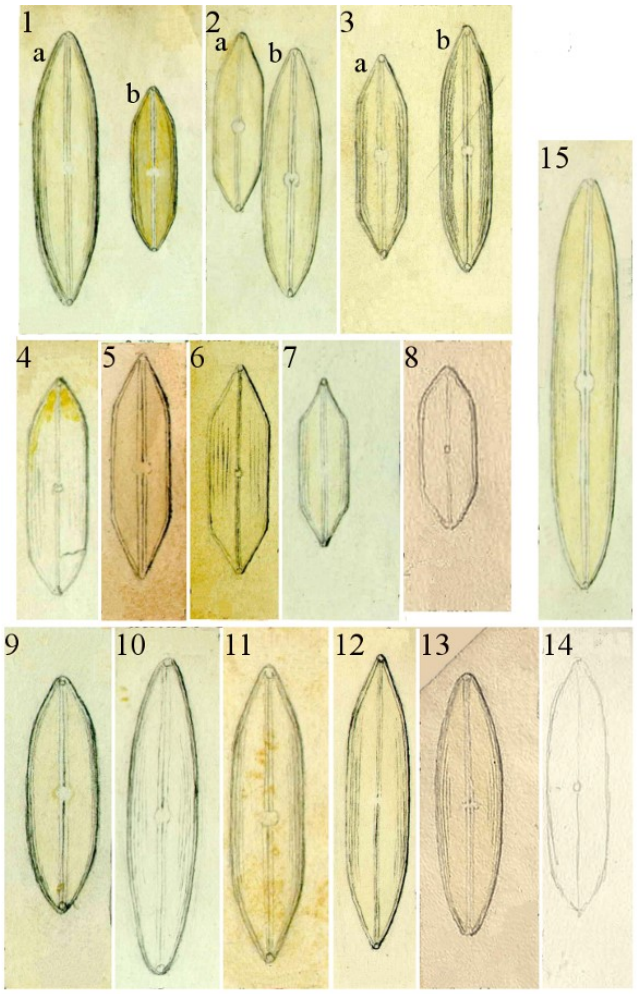
GenBank Accession Number		CANA Accession				
rbcL	**18S**	**Number**	**Identifier**	**Taxon**	**Locality and date of collection**	**Source**
		93650		*amphigomphus*	44.39273 N 74.75588 W, 26/10/2013	
KP325190	KP325157	CANA 100072	D4R12	*Neidium amphigomphus*	Wetland, Adirondack Park (New York, United States of America) 43.90870 N 74.43944 W, 11/05/2014	
KP325191	KP325158	CANA 100067	H4R13	*Neidium amphigomphus*	Owl Head Mountain Trail, Adirondack Park (New York, United States of America) 43.957394 N 74.464819 W, 11/05/2014	
KP325192	KP325159	CANA 93375	B4R12	*Neidium amphigomphus*	Wetland, Hwy 30 south of Long Lake (New York, United States of America) 43.927 N 74.44318 W, 11/05/2014	
KP325193	KP325160	CANA 100066	G7R13	*Neidium amphigomphus*	Wilson Pond, Hwy 28, Adirondack Park (New York, United States of America) 43.84326 N 74.47681 W, 10/05/2014	
KP325194	KP325161	CANA 93284	C7R13	*Neidium amphigomphus*	Big Moose Lake, Adirondack Park (New York, United States of America) 43.81723 N 74.88400 W, 10/05/2014	
KM999091	KP325162	CANA 100072	A7R12	*Neidium tumescens*	Wetland, Adirondack Park (New York, United States of America) 43.90870 N 74.43944 W, 11/05/2014	Hamilton *et al*. 2015
KM999090	KP325163	CANA 100072	C7R12	*Neidium tumescens*	Wetland, Adirondack Park (New York, United States of America) 43.90870 N 74.43944 W, 11/05/2014	Hamilton *et al*. 2015
KP325195	KP325164	CANA 93279	A4R13	*Neidium dilatatum*	Wetland, Adirondack Park (New York, United States of America) 44.44806 N 74.77439 W, 10/05/2014	
KP325196	KP325165	CANA 100072	D6R12	*Neidium dilatatum*	Wetland, Adirondack Park (New York, United States of America) 43.90870 N 74.43944 W, 11/05/2014	
KP325197	KP325166	CANA 93375	A6R12	*Neidium dilatatum*	Wetland, Hwy 30 south of Long Lake (New York, United States of America) 43.927 N 74.44318 W, 11/05/2014	
KP325198	KP325167	CANA 93281	B3R13	*Neidium dilatatum*	Creek, Adirondack Park (New York, United States of America) 44.00715 N 74.49155 W, 11/05/2014	
KP325199	KP325168	CANA 93279	A9R13	*Neidium dilatatum*	Wetland, Adirondack Park (New York, United States of America) 44.44806 N 74.77439 W, 10/05/2014	

Table 2.1. Continued

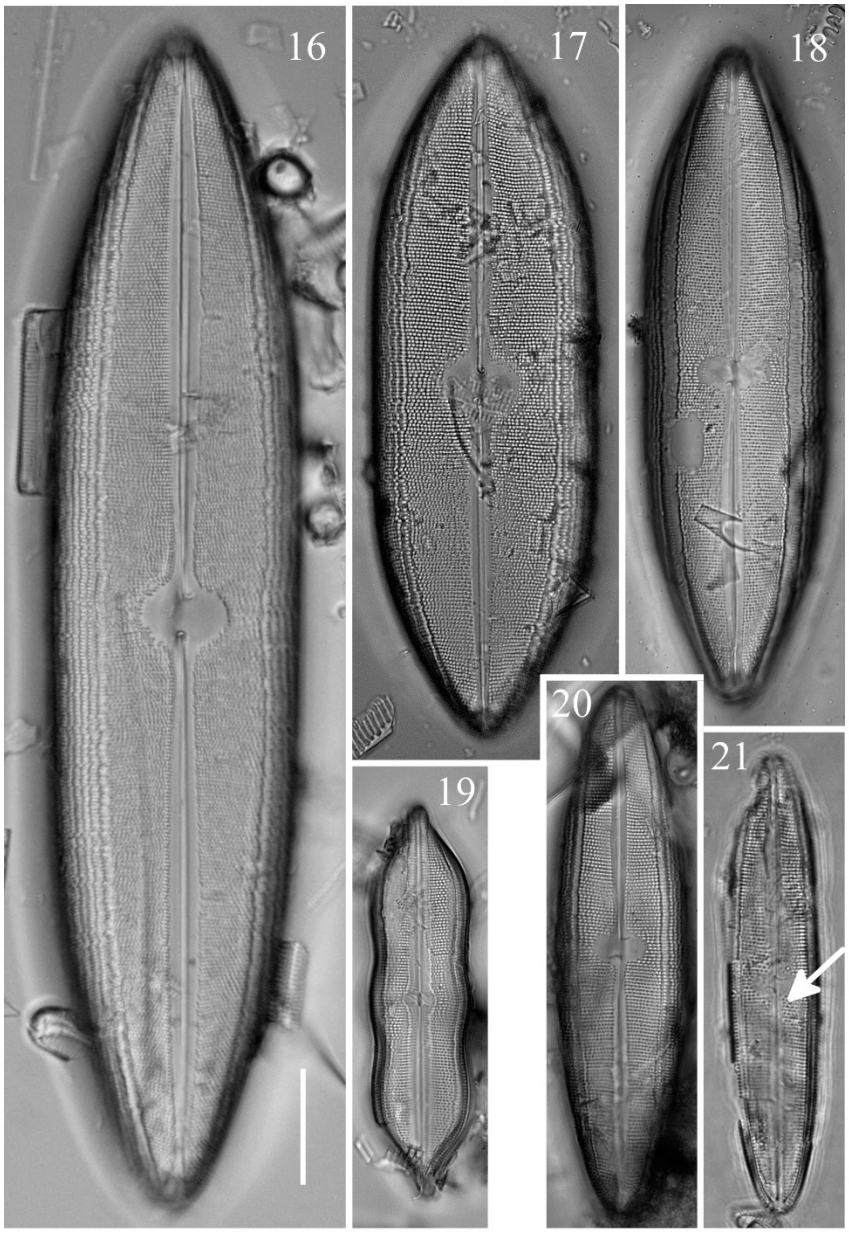
GenBank Accession Number		CANA Accession Number	Identifier	Taxon	Locality and date of collection	Source
rbcL	**18S**					
KP325200	KP325169	CANA 93372	F1R12	*Neidium dilatatum*	Pond, Hwy 56 (New York, United States of America) 44.39536 N 74.75690 W, 11/05/2014	
KM999092	KP325170	CANA 93372	D1R12	*Neidium sp*.	Pond, Hwy 56 (New York, United States of America) 44.39536 N 74.75690 W, 11/05/2014	Hamilton *et al*. 2015
KP325201	KP325171	CANA 93284	C4R13	*Neidium sp*.	Big Moose Lake, Adirondack Park (New York, United States of America) 43.81723 N 74.88400 W, 10/05/2014	
KP325202	KP325172	CANA 100072	E6R12	*Neidium sp*.	Wetland, Adirondack Park (New York, United States of America) 43.90870 N 74.43944 W, 11/05/2014	
KP325203	KP325173	CANA 93284	C3R13	*Neidium sp*.	Big Moose Lake, Adirondack Park (New York, United States of America) 43.81723 N 74.88400 W, 10/05/2014	
KM999089	KP325174	CANA 100068	G3R12	*Neidium sp*.	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W, 18/05/2014	Hamilton *et al*. 2015
KP325204	KP325175	CANA 100068	G7R12	*Neidium sp*.	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W, 18/05/2014	
KP325205	KP325176	CANA 100072	D5R12	*Neidium fossum*	Wetland, Adirondack Park (New York, United States of America) 43.90870 N 74.43944 W, 11/05/2014	
KP325206	KP325177	CANA 93375	E3R13	*Neidium sp*.	Wetland, Hwy 30 south of Long Lake (New York, United States of America) 43.92700 N 74.44318 W, 11/05/2014	
KP325207	KP325178	CANA 100072	E4R12	*Neidium sp*.	Wetland, Adirondack Park (New York, United States of America) 43.90870 N 74.43944 W, 11/05/2014	
KP325208	KP325179	CANA 100068	A1R14	*Neidium sp*.	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W, 18/05/2014	
KP325209	KP325180	CANA 100068	B1R14	*Neidium sp*.	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W, 18/05/2014	

Table 2.2. A list of the primers used in the nested PCR amplification of *rbcL* and 18S.

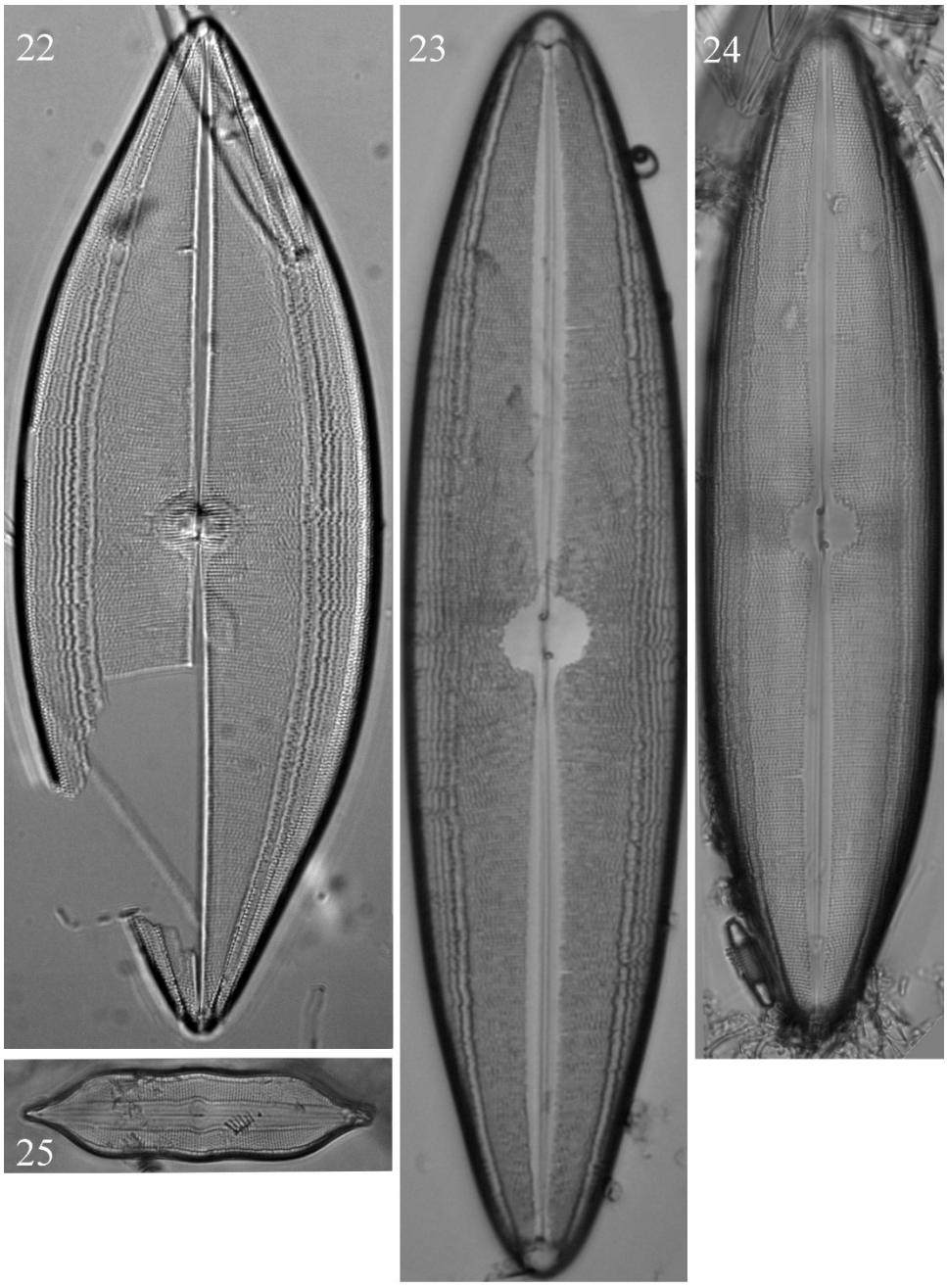
Primer Name	Primer Sequence (5'–3')	Reference
RBCL66+	TTTAAGGAGAAATAAATGTCTCAATCTG	Alverson *et al*. (2007)
RBCL dP7-	AAASHDCCTTGTGTWAGTVTC	Daugbjerg and Andersen (1997)
RBCL 1444-	GCGAAATCAGCTGTATCTGTWG	Ruck and Theriot (2011)
RBCL 40+	GGACTCGAATVAAAAGTGAACG	Ruck and Theriot (2011)
18SP2F	CTGGTTGATTCTGCCAGT	Hannen *et al.* (1999)
18SP4R	TGATCCTTCYGCAGGTTCAC	Guillou *et al.* (1999)
18S200F	YGGSRWGAYRTGGTGADTCA	Hamilton *et al*. 2015
18S1444R	GVRTRCATCAGTGTAGCGCG	Hamilton *et al*. 2015



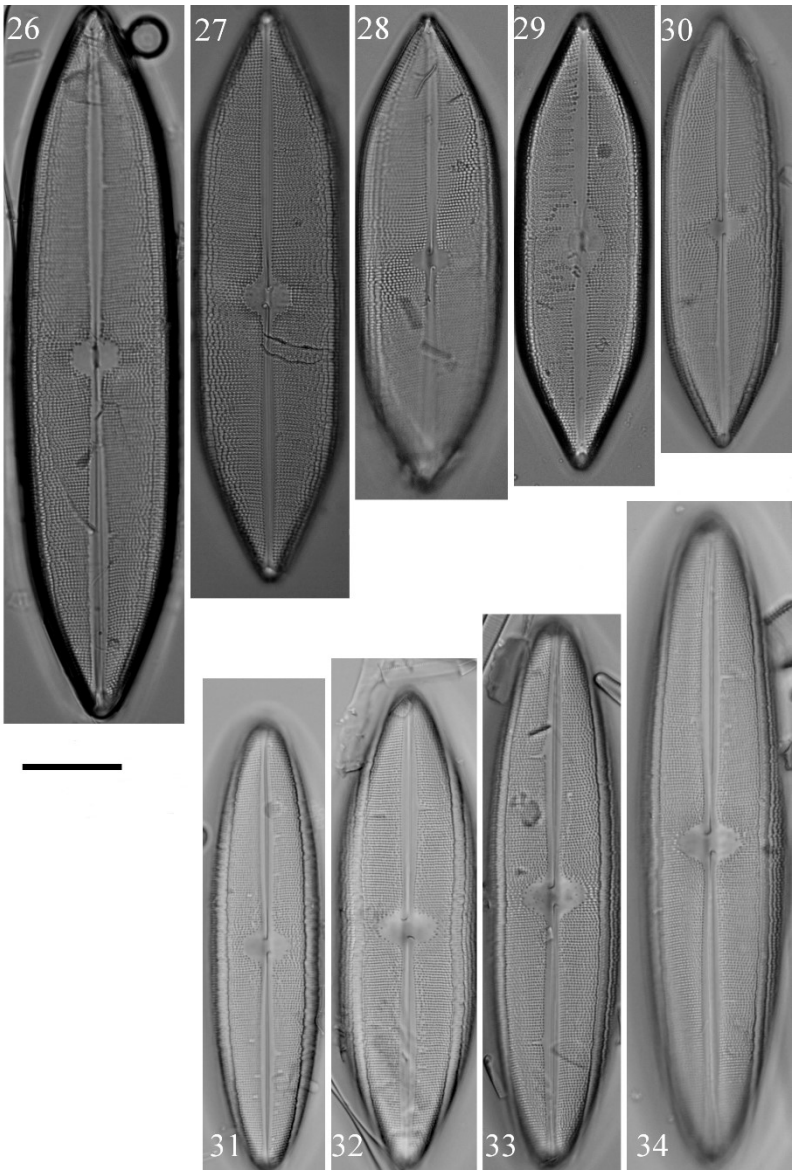
Figures 2.1–2.15. Extracted from Ehrenberg's original line drawing (ECD). 2.1b, 2.2a, 2.3a, 2.4–2.8. Identified as *Neidium amphigomphus*. 2.1b. ECD2239 Andover (Andower). 2.2a. ECD2238 Boston. 2.3a. ECDR2241 Bridgewater. 2.4. ECD2068 Rel del Monte. 2.5 ECD2233 West Point. 2.6. ECD2247 Blue Hill Pond II. 2.7 ECD2246 Blue Hill Pond I. 2.8. ECD2234 Stratford. 2.1a, 2.2b, 2.3b, 2.9–2.15. Identified as *Neidium dilatatum*. 2.1a. ECD2239 Andover (Andower). 2.2b. ECD2238 Boston. 2.3b. ECD2241 Bridgewater. 2.9. ECD2233 West Point. 2.10. ECD2236 Smithfield. 2.11. ECD2237 Pelham. 2.12 ECD2247 Blue Hill Pond II. 2.13. ECD2234 Stratford. 2.14. ECD2243 Wrenham. 2.15. ECD2246 Blue Hill Pond I. Reproduction of these images under the copyright permission of the Institute of Paleontology, Museum für Naturkunde, Humboldt University of Berlin (BHUPM).



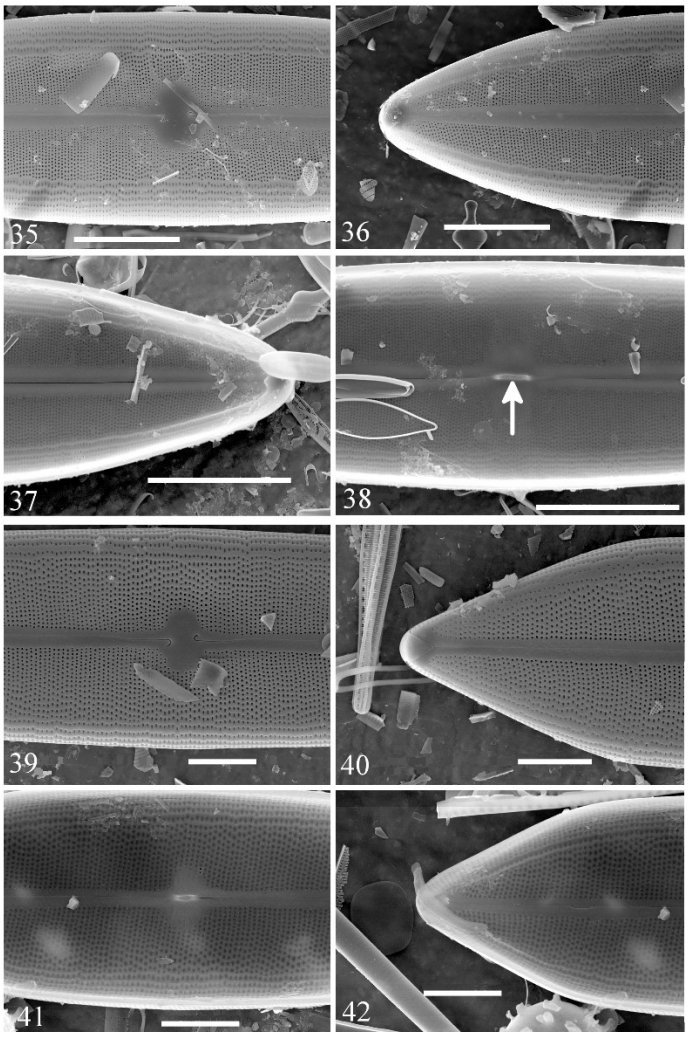
Figures 2.16–2.21. *Neidium* specimens (LM) original type materials. 2.16. *Neidium dilatatum* lectotype, Andover (Andower), Massachusetts (BHUPM! ECO–101). 2.17. *Neidium amphigomphus* lectotype, Bridgewater, Massachusetts (BHUPM! ECO–100). 2.18. *Neidium fossum* originally identified by Ehrenberg as *N. dilatatum*, Stratford, Connecticut. 2.19. *Neidium hitchcockii* lectotype, Bridgewater (BHUPM! ECO–102). 2.20. *Neidium fossum* originally identified by Ehrenberg as *N. dilatatum*. Pelham, Massachusetts. 2.21. *Neidium firma* lectotype (BM! 18414). Arrow indicates small pore in the center of the areolae. Scale bar = 20 μm.



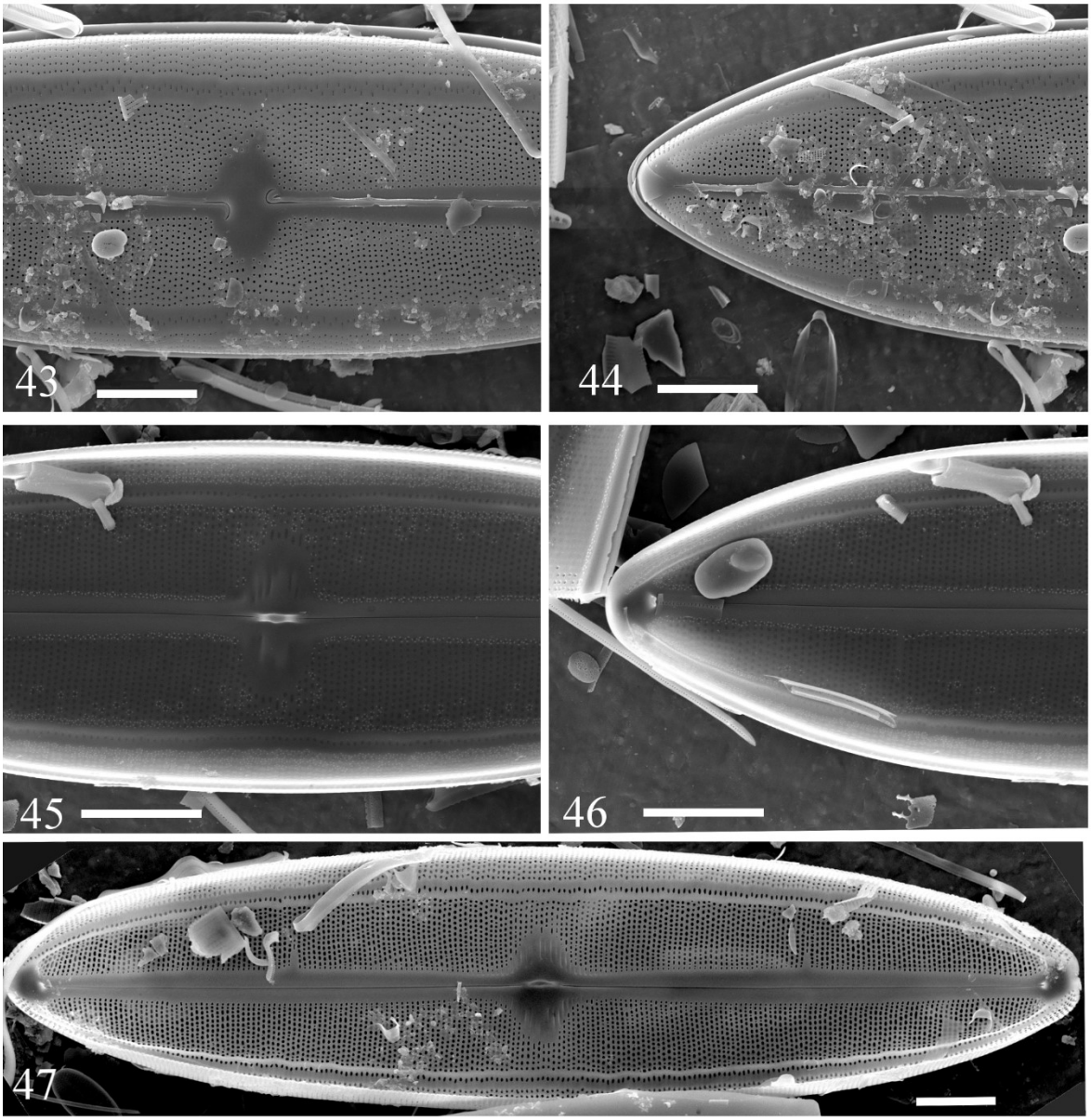
Figures 2.22–2.25. *Neidium* specimens (LM). 2.22. *Neidium tumescens* from Cherryfield, Maine. LM external view illustrating the whole valve. 2.23–2.24. *Neidium dilatatum*. LM external view illustrating the size range. 2.23. From CANA 93279. 2.24. From CANA 93281. 2.25. *Neidium hitchcockii* from CANA 100068. LM external view illustrating the whole valve. Scale bar = 20µm.



Figures 2.26–2.34. *Neidium amphigomphus* and *Neidium fossum* (LM). 2.26–2.30. *Neidium amphigomphus*. LM external view illustrating the size range of the valve. Note the straight pointed ends, small elliptical central area, and tight proximal raphe ends. 2.26. From CANA 93650. 2.27. From CANA 93279. 2.28. From CANA 93284. 2.29. From CANA 100066. 2.30. From CANA 100072. 2.31–2.34. *Neidium fossum* from CANA 100072. LM external view illustrating the size range of the valve. Note the large inner-most canal and subsequent small canals, arrow shaped ends, ovular central area and bent proximal raphe ends. 2.33. *Neidium fossum* holotype (CANA! 100072). Scale bar = 20 µm



Figures 2.35–2.42. *Neidium dilatatum* and *Neidium amphigomphus* (SEM). 2.35–2.36 *Neidium dilatatum* from CANA 93284. SEM external view of apex and central area, note tight circular central area, multiple small canals and small hooked proximal raphe ends. 2.37–2.38. *Neidium dilatatum* from CANA 100072. SEM internal view of apex and central area, note arrow-shaped ends with multiple small canals. Arrow indicates the fine silica ridge connecting the central helictoglossae. 2.39–2.40. *Neidium amphigomphus* from CANA 93284. SEM external view of apex and central area, note oblong central area and equal triangular ends. 2.41–2.42. *Neidium amphigomphus* from CANA 100072. SEM internal view of apex and central area, note the parallel sides and the equal triangular ends. Scale bars = 10 µm (39–42), 20 µm (35–38).



Figures 2.43–2.47. *Neidium fossum* (SEM). 2.43–2.44. *Neidium fossum* from CANA 100072. SEM external view of apex and central area, note ovular central area, large internal canal with smaller external canals, and long bent proximal raphe ends. 2.45–2.46. *Neidium fossum* from CANA 100072. SEM internal view of apex and central area. Note large internal canal. 2.47. *Neidium fossum* from CANA 100072. SEM internal view of whole valve. Note the large internal canal and oblong central area. Scale bar = 10 µm.

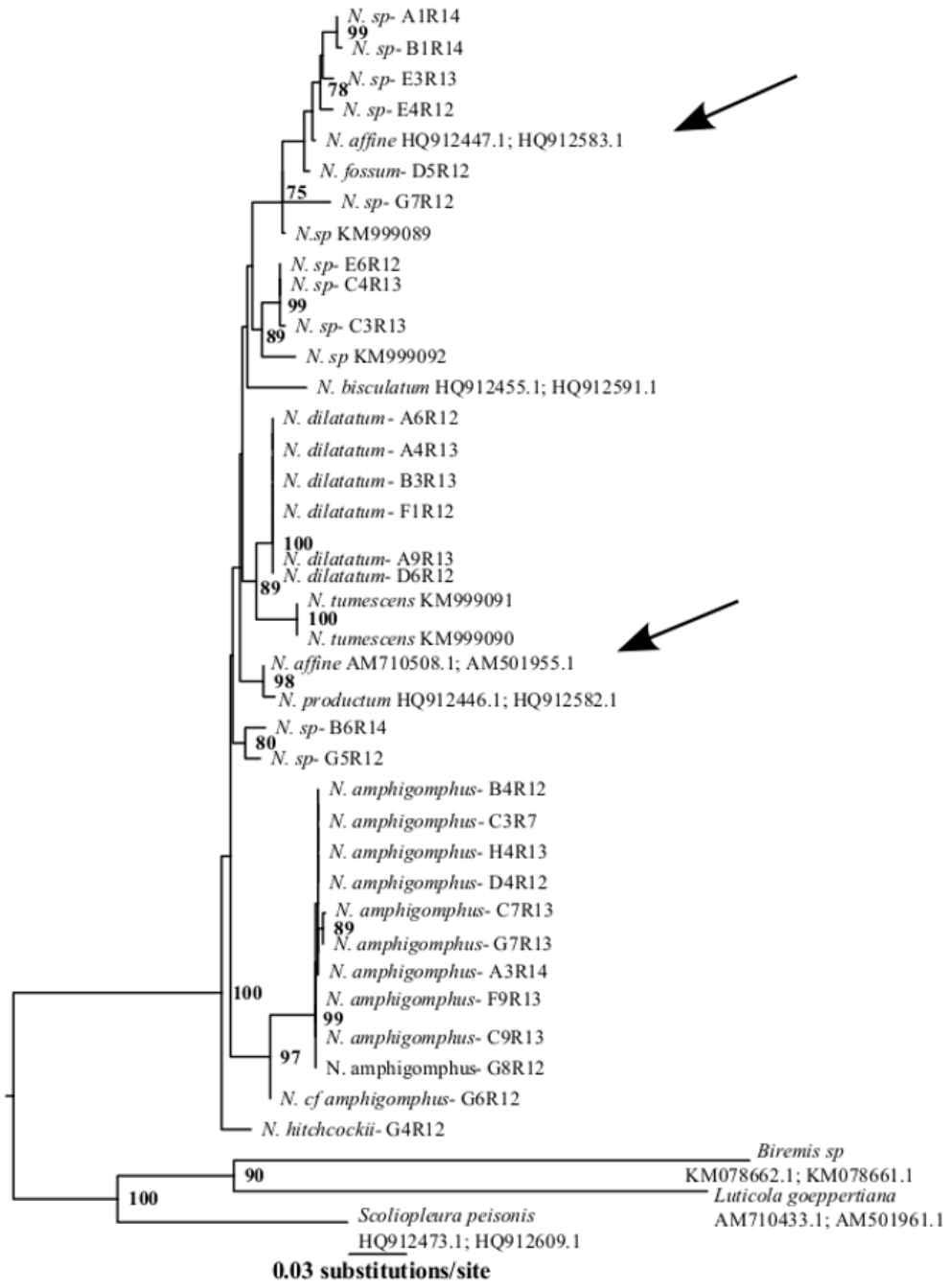


Figure 2.48. Phylogenetic relationships of species in the genus *Neidium* made using an *rcbL* + 18S alignment under the TIM3+I+G model. The tree was constructed using a bootstrapped Maximum Likelihood. The nodes indicate the level of support given as percentages of ML bootstrap values. Arrows indicate two branches showing *N. affine* sequences.

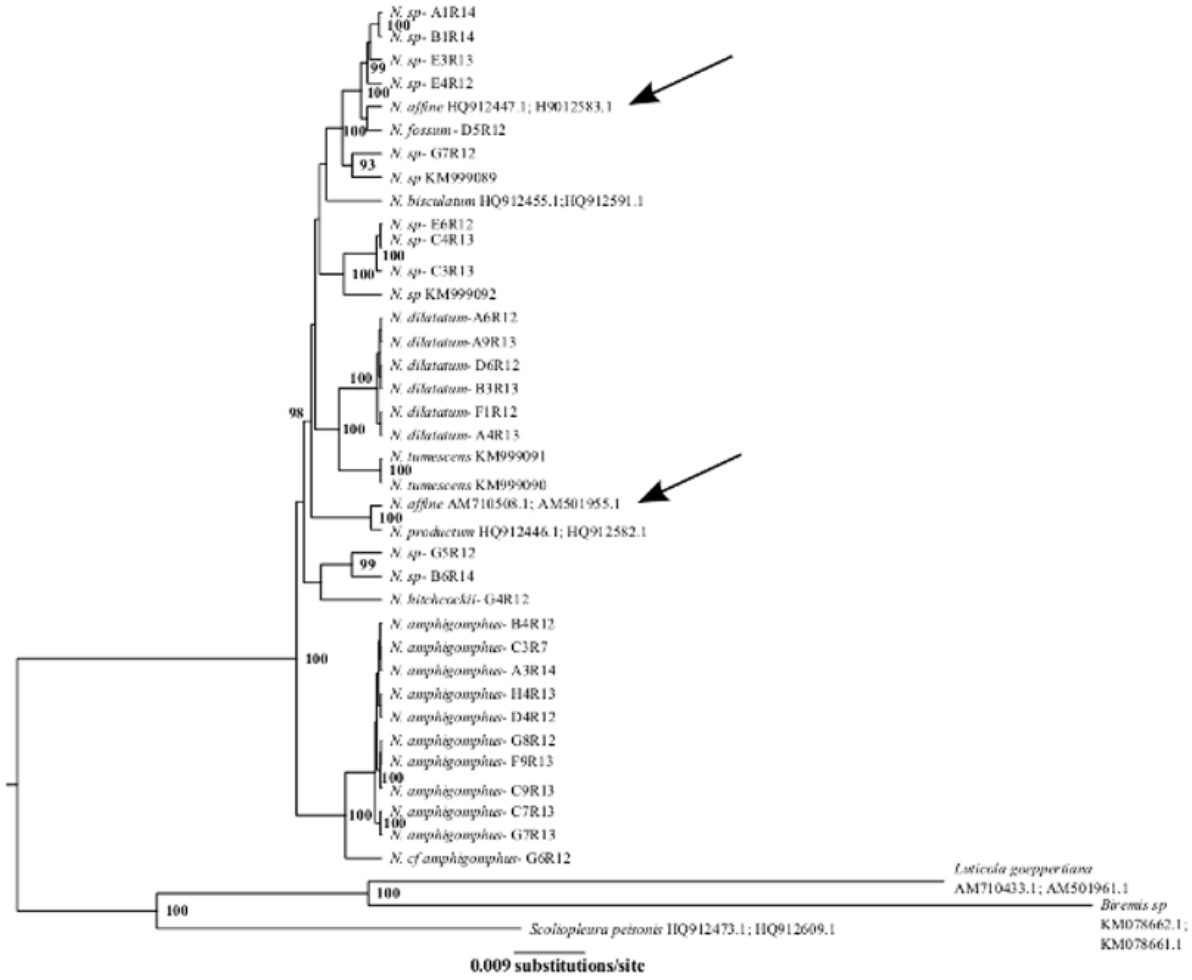


Figure 2.49. Phylogenetic relationships of species in the genus *Neidium* made using an *rcbL* + 18S alignment under the TIM3+I+G model. The tree was constructed using a Bayesian Analysis, with posterior probabilities listed at significant nodes to indicate the level of support. Arrows indicate two branches showing *N. affine* sequences.

Chapter 3: A comparison of molecular barcodes and morphology of small *Neidium* (Bacillariophyta) of North America

3.1 Abstract

Historically the morphospecies concept has been used in delimitation of diatom species. This has led to confusion between taxa within the benthic diatom genus *Neidium*. In this study samples from Eastern Ontario (Canada), Western Quebec (Canada), Nova Scotia (Canada), Avalon Peninsula, Newfoundland (Canada) and Adirondack Park, New York, (USA) were examined for *Neidium* taxa under LM and SEM. Isolated individuals from the samples were also amplified and sequenced for three chloroplast molecular markers (*rbcL, psbC,* and *psbA*) and one nuclear ribosomal molecular marker (18S). Phylogenetic reconstructions were completed with the concatenated chloroplast dataset and the 18S dataset using Maximum Likelihood and Bayesian analyses. The concatenated chloroplast dataset showed a species-level resolution phylogeny of *Neidium* taxa. In contrast, the 18S dataset had a much lower amount of sequence divergence and was unable to differentiate between *Neidium* taxa. We present emended species descriptions and sequence data of four previously described species: *N. sacoense*, *N. longiceps*, *N. fossum*, and *N. affine*. Additionally we describe three novel species, (*N. lowei, N. promontorium,* and *N. potapovii*) and also describe two additional species that have yet to be named. The distinguishing morphological features of *N. lowei* are its size, valve shape, and longitudinal canal. *Neidium promontorium* can be distinguished by its valve shape, longitudinal canal and apex along the raphe. *Neidium potapovii* can be distinguished by its size, valve shape and longitudinal canal. Future phylogenetic work which uses single cell multi-gene sequencing techniques may allow for taxonomic confusions within other diatom and protist genera to be resolved.

3.2 Introduction

Diatoms are well known for their distinct silica valve morphologies, which have been used to describe species for over 200 years (Mann, 2010). As these microorganisms play an important ecological role in carbon and biogeochemical cycling (Nelson *et al*., 1995) and are often used for biomonitoring and paleoreconstructions (Stoermer and Smol, 1999), a robust species-level taxonomy is critical. Most species are described using morphology alone. As technologies have improved in light microscopy (LM) and especially scanning electron microscopy (SEM), the differences in morphological traits used for species delimitation have become very subtle and difficult for non-experts to recognize (Alverson, 2008).

Not all species have been delimited strictly using morphological characteristics. There are some examples of species being delimited based on the biological species concept (Amato *et al*., 2007; Behnke *et al*., 2004; Mann *et al*., 2004; Quijano-Scheggia *et al*., 2009; Vanormelingen *et al*., 2008) whereby a species must be reproductively isolated in order to be differentiated as a species (Mayr, 1942). However, using reproductive isolation to delimit species can be problematic as mating experiments are difficult to conduct with diatoms, and species may be able to hybridize with other species in the lab that they were not in contact with in the natural environment (Alverson, 2008). Ecological species concepts are rarely used in diatom phylogenetics as there is a lack of knowledge of ecological niches of diatoms at the species level (Mann, 1999). This is due in part to a complicated classification system which is cumbersome and inaccessible to many ecologists.

More recently, DNA sequencing of molecular markers have been used within the phylogenetic species concepts to delimit species (Leliaert *et al*., 2014). DNA barcoding is defined as using short regions of DNA to differentiate between species (Hebert *et al*., 2003). In

the majority of diatom phylogenetic studies cultured material is used to obtain sequence data for species, but single diatom cells can also be sequenced for multiple molecular markers (Auinger *et al*., 2008; Godhe *et al*., 2002; Hamilton *et al*., 2015; Lang and Kaczmarska, 2011; Lefebvre and Hamilton, 2015). This allows analysis into inter- and intra-specific genetic distances of a species. As well, several different molecular markers are often used to construct more robust phylogenetic relationships between species. These markers include the chloroplast genes (*rbcL*, *psbA* and *psbC*), the mitochondrial gene *CO1*, and ribosomal complex genes and their spacers (18S, 28S, ITS1, ITS2) (Evans *et al*., 2008; Kaczmarska *et al*., 2014; Lefebvre and Hamilton, 2015; Moniz and Kaczmarska, 2010; Souffreau *et al*., 2011; Trobajo *et al*., 2010). At this point there is little consensus on which molecular markers should be used to delimit species of diatoms, and what the genetic distance between species should be (Leliaert *et al*., 2014; Mann *et al*., 2010; Moniz and Kaczmarska, 2009).

The genus *Neidium* Pfitzer (1871: 39) is a group of benthic freshwater diatoms containing over 300 taxa, some of which have been used as an important biomonitoring diatom in heavy metal pollution (Ivorra *et al*., 1999) and water quality (Herlory *et al*., 2013). Previous phylogenetic work focused on larger (>100 μm) *Neidium* taxa from North America (Lefebvre and Hamilton, 2015) which defined *N. amphigomphus, N. dilatatum, N. iridis, N. tumescens, N. fossum* and *N. hitchcockii*. There remains much confusion over the smaller *Neidium* taxa due to the similar valve shapes and overlapping size ranges between taxa populations. The result of this difficulty in distinguishing morphological characters has been that species like *N. affine* (Ehrenberg 1843: pl. 2/5, fig. 4) and *N. ampliatum* (Ehrenberg 1854: 16) have become broadly interpreted (Auct. Non.) categories for many small *Neidium* taxa.

The objective of this study was to validate and delimit several of the smaller (<100 μm) *Neidium* taxa from North America using both morphological and molecular methods. Within this study, our species criteria were that a species must: (1) show particular morphological characteristics which can be measurable and demonstrably different from other taxa, (2) have strong phylogenetic support from molecular analysis, and be monophyletic. Four molecular markers (*rbcL*, *psbA*, *psbC* and 18S), were used in the molecular methods, and their ability to delimit species was compared. Original species descriptions of *Neidium* taxa were compared to LM and SEM images of *Neidium* individuals which were sequenced.

3.3 Materials and Methods

Fifty live samples (Table 3.1) were collected from the benthos of rivers, ponds, bogs and streams in Newfoundland, Nova Scotia, Quebec and Ontario (Canada), as well as northern New York State (USA). The samples were kept in the lab at ambient light and temperature for approximately one week, during which specimens were taken for morphological and DNA analysis. The number of specimens selected varied from sample to sample depending on the availability of living individual cells and ability to isolate cells with photomicrographic documentation. Typically 3-8 cells were isolated from each sample. Subsamples were also fixed with RNAlater and freeze dried for deposition into the Canadian Museum of Nature Phycology Collection (CANA 93415-93416; 93279, 93375, 93650, 100021, 100066-100072, 108073, 108115-108131, 108578, 108593, 108601-108621, 108629, 108637).

3.3.1. Morphological Methods

All sample cleaning, light microscope slide preparations and SEM stub preparations followed the protocol from Chapter 2. Light microscopy was performed using a Leica DMR

microscope with a Pixelink digital camera. Specimens were examined at magnifications ranging from 640–1000x using NPlan or Plan Apo objectives (NA 1.25–1.35) with either brightfield (BF), phase contrast (PH), or differential interference contrast (DIC) optics. SEM stubs were observed with an FEI XL30 (FEI, Hillsboro, USA) tungsten filament ESEM with accelerated voltages between 10–20 kV and a working distance of 5–10 mm. The same morphological metrics were used as in Lefebvre and Hamilton (2015), and all terminology follows Krammer and Lange-Bertalot (1986), Siver *et al*. (2003) and von Stosch (1975).

3.3.2. Molecular Methods

The protocol used for single cell isolation, nested PCR and sequencing methods were followed from Hamilton *et al*., (2015). For nested PCR protocols and sequencing, the following primers were used (Table 3.2). In the nested PCR protocol Phusion ® High-Fidelity DNA Polymerase (New England Biolabs) was used in all amplifications, and the following cycling conditions were used: 98°C for 30s; followed by 34 cycles of 98°C for 10s, differing annealing temperatures for 30s, and 72°C for 30s; and a final elongation step at 72°C for 5 min. For the second amplification all concentrations and steps were the same as above except that 1 µl of the product from the previous amplification was used as a template. The optimized annealing temperatures for each marker used are listed for the first and second PCR amplifications respectively: *rbcL* 57.7°C and 60.3°C; 18S 60.3°C and 60.3°C; *psbC* 60.3°C and 57.7°C; *psbA* 63°C and 63°C. The *rbcL*, psbA, psbC, and 18S dataset's were trimmed to 1281, 576, 858, and 988 characters, respectively.

3.3.3. Phylogenetic Analyses

The outgroup used in the chloroplast concatenated and 18S analysis was *Scoliopleura peisonis*, and several *Neidium* cultures were used as sister groups (Table 3.1). *Neidium* sequences

previously published in Lefebvre and Hamilton (2015), and Hamilton *et al*. (2015) were also included in the phylogenetic analyses (Table 3.1). Sequences for all four genes were edited in Geneious v.8.0.5 and aligned using the MAAFT alignment algorithm. Pairwise percent identity between individuals were calculated in Geneious v.8.0.5 on both the 18S dataset and the chloroplast dataset to determine the percentage range of identical base pairs within a species. Likelihood values and AIC were used to determine the nucleotide substitution model for each individual gene and the concatenated analyses on JModel Test v.2.1.4 (Darriba *et al*., 2012). The nucleotide substitution model used in the 18S analysis, the individual and the concatenated chloroplast analysis was the General Time Reversible model (GTR+I+G) (Tavaré, 1986). For each of the chloroplast genes (*rbcL*, *psbA*, and *psbC*), approximate likelihood ratio tests with SH-like supports using NNI and SPR for tree topology search operations were run on PhyML v.3.1 (Guindon and Gascuel, 2003) to assess whether the data could be concatenated for analysis. As the individual chloroplast topologies showed no difference the data were concatenated, and using the GTR+I+G model, Bayesian analyses (BA) and Maximum likelihood analyses (ML) were completed on the concatenated chloroplast dataset and the 18S dataset. BA was performed with Beast v.3.1.2 (Drummond *et al*., 2012) where for each dataset a Monte Carlo Markov Chain (MCMC) was run for 10 million generations each with at least 3 independent replicates of the analysis completed. Runs were sampled every 1000^{th} generation, with 10% being discarded for burnin for each dataset. The convergence and stationarity of the analyses were assessed in Tracer v.1.6 (Rambaut *et al*., 2014) such that all ESS values were greater than 200. ML was performed with PhyML with 1000 bootstrapped replicates for each dataset. For the analysis, strong support indicates >90% BS and 1.00 PP at nodes, moderate support is indicated when there is 65 – 75% BS and 0.95 – 0.99 PP at nodes, weak support is indicated when there is <65% BS or <0.95 PP.

3.4 Results

3.4.1 Phylogenetic Analysis

For the phylogenetic analysis, 94 sequences were determined for *rbcL*, 124 sequences determined for both *psbC* and *psbA*, and 64 sequences determined for 18S. Sequence data from Chap. 2 were also used (Table 3.1). The datasets for *rbcL*, *psbA*, *psbC* and 18S contained 1281, 576, 858 and 988 characters respectively. LM of each of the diatoms sequenced in this manuscript can be seen in Appendix C, Fig. C.1. For each of the individual chloroplast genes all topologies of trees were the same, so only the concatenated analysis is shown. The ML phylogenetic trees for both the concatenated chloroplast analysis (-lnL=7028.26728) and the 18S analysis (-lnL=2209.06441) are shown with the BA posterior probabilities (Figs 3.1, 3.2). The BA trees for both the concatenated and 18S datasets can be seen in Figs C.2 and C.3 respectively. The chloroplast molecular markers (*rbcL*, *psbA*, and *psbC*) all had a lower amount of pairwise indentity (average % identity over the alignment) than the nuclear molecular marker 18S (Table 3.3). As well, the single cell diatom sequencing method was further validated by confirming that sequences obtained from single cells of UTEX *Neidium* cultures were identical to sequences obtained in Theriot *et al.* (2010) from whole culture extractions.

The concatenated chloroplast analysis showed greater support for a larger number of species within the *Neidium* genera than the 18S analysis. Within the concatenated analysis there was strong support (>90% BS, ≥0.99 PP) for the following species; *N. hitchcockii*, *N. amphigomphus*, *N. lowei* sp. nov., *N. tumescens*, *N. dilatatum, N. affine*, *N. sacoense*, *N. longiceps*, *N. fossum*, *N. potapovii* sp. nov., *N*. sp 1, and N. sp 6. There is a mix of strong (1.00 PP) and moderate support (88.5% BS) for *N. promontorium* sp. nov. However the 18S phylogenetic tree did not differentiate genetically between *N. affine*, *N. lowei*, *N. hitchcockii*, and *N. fossum*. As well, the

nodes separating other species such as *N. sacoense*, *N. longiceps*, *N. dilatatum*, *N*. sp 1 and *N*. sp 6 were weak in the Maximum Likelihood analysis (<65% BS), and moderate to strong in the Bayesain analysis (0.933 - 0.99 PP). The concatenated tree also showed some areas of particular genetic variants within a species group (*N. amphigomphus*), and others where there was a single genetic variant within a species (*N. dilatatum* and *N. tumescens*).

3.4.2 Species Descriptions

Using the phylogenetic and morphological analyses we describe three novel species below: *N. lowei*, *N. potapovii* and *N. promontorium*. We also emend descriptions of previously described species: *N. sacoense*, *N. longiceps* and *N. affine,* and identify two unnamed species (*N*. sp1, *N*.sp6).

Division Bacillariophyta

Subdivision Bacillariophytina

Class Bacillariophyceae

Order Naviculales

Sub-order Neidiineae

Family Neidiaceae

Neidium sacoense Reimer 1966: 402, (pl. 7:3) (Figs 3.3 – 3.7)

Emended species description: Number of individuals in morphological analysis = 45; Number of individuals sequenced = 8. Valves are broadly linear with ends apiculate to subrostrate. Valve length 55 – 118 µm; width 16 – 35 µm. Striae are parallel to weakly convergent towards the apices and evenly spaced with a density of 16 – 22 in 10 µm. Voigt faults are

present approximately 1/2 way between the edge of the central area and the apex on the secondary side of the valve. Areolae are evenly spaced as fine to larger puncta as they progress from the centre of the valve towards the edge with a density of 16 – 21 in 10 μm. Externally, in the SEM, the areolae are weakly recessed (Figs 3.3 – 3.4). Central area is elliptical in shape and extends over 1/2 of the valve face (Fig. 3.3). One large longitudinal canal is present at the margins of the valve. Areolae along the longitudinal canal are lineolate and are not evenly distributed. Externally the raphe is filiform linear with proximal raphe ends long, flat and oppositely bifurcating. The distal raphe ends terminate under an apical flap (Fig 3.4). Internally under SEM observation, the central helictoglossae endings of the raphes appear as one large connected helictoglossa (Fig. 3.5). The distal helictoglossae form small thickened nodules near the apex (Fig. 3.6). The areolae are slightly recessed internally (Fig. 3.5 – 3.6). The longitudinal canal terminates at the apices of the valve.

Type:—UNITED STATES OF AMERICA. Maine: York County, Saco Pond, Reimer, A-Shulze 367.

Observations: *Neidium sacoense* is found in slightly acidic (pH 4.92 – 5.53) ponds and wetlands around eastern North America. Confirmed identifications of this species are found in Patrick and Reimer 1966 (pl. 37: 3), Gaiser and Johansen 2000 (figs 58 – 59), Siver and Hamilton 2011 (pl. 190: 1-3; pl. 200: 1 – 6). The key features of *N. sacoense* specimens include straight valve sides similar to *N. amphigomphus*, a single large longitudinal canal similar to those seen in *N. fossum* and *N. lowei*, as well as the cuneate ends similar to *N. amphigomphus* and *N*. sp 6. Within the phylogenetic results, *N. sacoense* is a strongly supported monophyletic group, and is a sister taxa to *N. longiceps, N. fossum, N. promontorium* and *N*. sp6 (Fig. 3.1). The pairwise identity of *N.*

sacoense within the chloroplast phylogenetic analysis was between 99.8 – 100%, and was 100% within the 18S phylogenetic analysis.

Neidium affine (Ehrenberg) Pfitzer 1871: 39 Figs (3.8 – 3.13)

Basionym: *Navicula affinis* Ehrenberg, Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin 1841 pl. 2/5 fig. 4. 1843

Emended species description: Number of individuals used in morphological analysis = 38; Number of individuals sequenced = 17. Valves are linear to linear lanceolate with rostrate to slightly capitate ends. Valve length 31 – 84.3 µm; width 10 – 18 µm. Striae are evenly spaced and parallel, becoming convergent towards the apices with a density of 22 – 28 in 10 µm. Voigt faults are present 1/2 between the edge of the central area and the apex of the valve on the secondary valve. Areolae are fine rounded punctae to slightly elliptical and evenly spaced with a density of 20 – 28 in 10 µm. Central area is rounded and covers 1/3 to 1/2 of the valve width (Fig. 3.8, 3.10). There is one large longitudinal canal and two smaller canals along either side along the margins of the valve (Fig. 3.12 – 3.13). Raphe is linear in SEM with proximal raphe ends short and curved oppositely (Fig. 3.8). The distal raphe ends terminate in an apical flap (Fig. 3.9). Internally in the SEM, the areolae are rounded and larger around the apices. Renilimbia are present internally along the central raphe area and around areolae along the marginal longitudinal canal (Fig. 3.10 – 3.11). Central helictoglossae are distinct and connected by a fine silica ridge (Fig. 3.10). Distal helictoglossae are thickened nodes located at the apices (Fig. 3.11). The longitudinal canals terminate internally at the apices.

Type:—CANADA. Newfoundland (Lectotype – Geographical prep. 271715 BHUPM, Hamilton and Jahn 2005)

Observations: Within our study, *N. affine* was found in New York State and Newfoundland from acidic (pH 4.92 – 5.73) bogs, ponds and lakes. For this study we have used the same species concept of *N. affine* as was defined by Hamilton and Jahn (2005). *Neidium. affine* differs from *N. potapovii* sp. nov. as the valve sides are lanceolate-linear whereas *N. potapovii* features linear valve sides. As well, the ends of *N. affine* are rounded-rostrate to reducing-rostrate whereas the ends of *N. potapovii* are much more rostrate. In the phylogenetic analyses, the individuals of *N. affine* form a monophyletic group with *N. potapovii* as its sister taxa (Fig. 3.1). The pairwise identity of *N. affine* was between 99.3 – 100% within the concatenated chloroplast analysis, however *N. affine* was not a monophyletic group in the 18S analysis (Fig. 3.2).

Neidium longiceps (Gregory 1856) (Figs 3.19 – 3.29)

Basionym: *Navicula longiceps* Greg., Quart. Journal. Micr. Sci., 4:8, pl. 1, fig. 27. 1856

Synonym: *Neidium affine* var. *longiceps* (Greg.) Cl., K. Svenska Vet.-Akad. Handl. Ny Följd, 26(2): 68. 1894

Emended species description: Number of individuals used in morphological analysis = 28; Number of individuals sequenced = 4. Valves are broadly linear with ends from subcapitate to capitate. Valve length 28 – 47 µm; width 8 – 12 µm. Striae are parallel to slightly convergent toward the apices and evenly spaced with a density of 30 – 36 in 10 µm. One voigt fault is present 1/2 between the end of the central area and apex of valve. Areolae are evenly spaced fine rounded punctae from middle to edges of valve with a density of 30 – 36 in 10 µm. Central area is elliptical in shape and extends from1/2 to 2/3 of the valve face (Fig. 3.24). One large longitudinal canal is present along the valve margins (Fig. 3.3.24 –

3.26). Raphe is linear in SEM observations with proximal raphe ends flat and oppositely bifurcating (Fig. 3.24). Distal raphe ends terminate in an apical flap (Fig. 3.25). Internally, the areolae are rounded to elliptical. The proximal raphe ends terminate in central helictoglossae which are connected by a weakly connected silica ridge (Fig. 3.26). The distal helictoglossae form a thickened nodule at the apex of the valve (Fig. 3.27). Renilimbia are present around areolae located along the margins of the valve, and along the longitudinal central area surrounding the raphe (Figs 3.26 – 3.27, 3.29). The marginal longitudinal canal terminates at the apices (3.27).

Type:—UNITED KINGDOM, Scotland, Banffshire (Lectotype – BM.47646, Reimer 1959)

Observations: *Neidium longiceps* has a wide range, found in acidophilic lakes and rivers across the United States of America. Positive identifications of *N. longiceps* listed as *N. affine* var. *longiceps* can be found in Krammer and Langer-Bertalot (1986, pl. 103: 5), listed as *N*. spec. Metzeltin *et al*., (2005, pl.115:20-21) and listed as *N. longiceps* Werum and Lange-Bertalot (2004, pl.72: 8-13). The key morphological features of this species are subcapitate to capitate ends, linear sides featuring one large longitudinal canal. Several *N. longiceps* were found on the type slides from the British Museum. The individuals exhibited linear sides with capitate ends and a single large longitudinal canal (Figs 3.19 – 3.21) and had a length range of 35.1 – 38.4 µm, and width range of 9 – 10.1 µm. Within the phylogenetic analyses, *N. longiceps* individuals formed a monophyletic group with pairwise identity of 99.4 – 100% in the chloroplast sequences. Within the 18S phylogenetic analysis, *N. longiceps* did not form a monophyletic group.

Neidium potapovii K.Lefebvre and P.B.Hamilton, ***sp. nov.*** provisional (Figs 3.14 – 3.18, 3.30 – 3.35)

Number of individuals used in morphological analysis = 26; Number of individuals sequenced = 2. Valves are linear to linear-lanceolate with rostrate ends. Valve length 29.5 – 51.6 µm, valve width 7.5 – 12.6 µm. Striae are parallel and evenly spaced with a density of 24 – 30 in 10 µm. A Voigt fault is present 1/2 between the central area and the apex. Areolae are evenly spaced and rounded, with a density of 24 – 30 in 10 µm. Central area is elliptical in shape, covering 1/3 to 1/2 of the valve face (Fig. 3.30). One large longitudinal canal is present along the margins of the valve (Figs 3.30, 3.34 – 3.35). In the SEM, raphe is linear with proximal raphe ends bifurcating in opposite rounded hooks (Fig. 3.30). Distal raphe ends terminate in an apical flap (Figs 3.31, 3.34). Internally, proximal raphe ends terminate in distinct helictoglossae connected by a silica ridge (Fig. 3.32). The distal helictoglossae form thickened nodules at the apices of the valve (Fig. 3.33). The longitudinal canal terminates at the apices of the valve.

Type:—UNITED STATES OF AMERICA. New York: Adirondack Park, 43.90924 N 74.4397 W, *K. Lefebvre, 16 October 2014* (holotype: CANA! 108128, circled specimen, illustrated here as Fig. 3.18)

Etymology: This has been named in honour of Marina Potapova, curator of the most extensive diatom collection in North America (Academy of Natural Sciences of Drexel University).

Observations: This species was found in New York State in acidic (pH 4.92 – 5.33) wetlands and bogs. Examples of *N. potapovii* include Antoniades *et al*. 2008 (pl. 53: 4 – 7) (labelled as *N.* sp.), Metzeltin *et al*. 2005 (pl. 115: 16 – 18) (labelled as *N. longiceps*). The key features for this species are the rostrate ends and linear to linear lanceolate sides. It is similar to a species concept of *N. affine*, and can be differentiated by comparing the size ranges (*N. potapovii* has a smaller range) and very rostrate ends that it exhibits with the rounded rostrate sides of *N. affine.* The

phylogenetic analysis from the chloroplast genes show the individuals of *N. potapovii* in a strongly supported monophyletic group, and as the sister taxa to *N. affine* (Fig. 3.1). In the 18S phylogenetic reconstructions, *N. potapovii* is also in a strongly supported monophyletic group, but in this reconstruction, they are most similar to *N*. sp. individuals (Fig. 3.2). There chloroplast sequence data is identical for the two individuals of *N. potapovii*, and there is 99.9% pairwise identity in the 18S sequence data.

Neidium lowei K.Lefebvre and P.B.Hamilton, ***sp. nov.*** provisional (Figs 3.36 – 3.39, 3.44 – 3.49)

Number of individuals used in morphological analysis = 41; Number of individuals sequenced = 7. Valves are linear lanceolate with rounded rostrate ends. Valve length 78 – 135 µm; width 15 – 29 µm. Striae are parallel to slightly convergent towards the apices and evenly spaced with a density of 20 – 24 in 10 µm. Voigt faults are present 1/2 between the edge of the central area and apex on the secondary side of the valve. Areolae are fine rounded punctae, evenly spaced at a density of 17 – 24 in 10 µm. Central area is rounded and extends 1/2 to 1/3 across the width of the valve (Fig. 3.44). One large longitudinal canal is present, with associated areolae not following an orderly linear distribution (Figs 3.44, 3.48). Raphe is linear in SEM with proximal ends tightly rounded and oppositely bifurcating (3.44). Distal raphe ends terminate in an apical flap (Fig. 3.45). Internally, the central helictoglossae are distinct and connected by a fine silica ridge (3.46). The distal helictoglossae form rounded nodules at the apex (Fig. 3.47, 3.49). The longitudinal canal terminates at the apices. The areolae within the longitudinal canal are linear to elliptical; they are not evenly distributed along the canal, often alternating between one and two rows of areolae in the middle of the valve (Figs 3.46, 3.49).

Type:—CANADA. Quebec: Lac Jean Venne, 45.67893 N 76.0645 W, *P. Hamilton, 27 June 2015* (holotype: CANA! 100021, circled specimen, illustrated here as Fig. 3.38)

Etymology: This diatom has been named in honour of Rex L. Lowe, professor emeritus at Bowling Green State University, widely known for his extensive work freshwater algal ecology.

Observations: *N. lowei* is found from Ontario, Quebec, and Newfoundland within bogs and swamps of slightly acidic pH (6.02 – 6.63). It features a similar overall shape as *N. dilatatum* with a smaller size range. Central area size and shape is similar to *N. amphigomphus*. The internal raphe ends are similar to those of *N. dilatatum* and *N. amphigomphus*. Large longitudinal canal is similar to those seen in *N. fossum* and *N.* sp 6. It appears to be similar to *N. vanlandinghamii* Metzeltin and Lange-Bertalot (2007, pl. 182:1 – 3) with a smaller size range, but features the same cuneately rounded ends and similar lanceolate to elliptical-lanceolate sides. As well, it is similar in shape to *N. hamatum* Metzeltin and Lange-Bertalot (1998, pl. 121:1 – 2) from Brazil and *N. hamatum* var. *septentrionale* Metzeltin and Lange-Bertalot (2007, 181:1 – 3), although the ends are more rounded than the latter. In the chloroplast phylogenetic reconstruction, *N. lowei* individuals formed a strongly supported monophyletic group, but within the 18S reconstruction, *N. lowei* did not form a monophyletic group. Additionally, only two of the *N. lowei* individuals had successful sequence amplification and sequencing of the 18S region. The pairwise identity of the chloroplast sequences ranged from 99.8 – 100%, and for the 18S region the two sequences did not vary.

Neidium promontorium K.Lefebvre and P.B.Hamilton, ***sp. nov.*** provisional (Figs 3.40 – 3.43, 3.50 – 3.55)

Number of individuals used in morphological analysis = 40; Number of individuals sequenced = 6. Valves are linear to linear lanceolate with apiculate to subrostrate ends. Valve length 51 – 85 µm; width 14 – 22 µm. Striae are parallel, becoming curved radiate as they approach the apices with a density of 20 – 28 in 10 µm. Voigt faults are present 1/2 between the edge of the central area and the apex of the valve. Areolae are small and rounded with a density of 20 – 26 in 10 µm. Raphe is linear in SEM with proximal ends that are long, flat and oppositely bifurcating (3.50). The raphe is located on a thickened silica ridge which has a row of very fine rounded areolae present along the sides (Fig. 3.50 – 3.51, 3.54). There is one large longitudinal canal along the margins of the valve (Fig. 3.50). Distal raphe ends terminate in an apical flap (Fig. 3.51). Internally, the proximal raphe ends terminate in distinct helictoglossae which are connected with a thickened silica ridge (Fig. 3.52). The distal helictoglossae are thickened nodules at the apices of the valve (Fig. 3.53). Renilimbia are present internally along the sides of the central area and along the longitudinal raphe area (Figs 3.52 – 3.53, 3.55). The longitudinal canal contains internal areolae which are linear-elliptical in shape and terminates just before the apices (Figs 3.52 – 3.53).

Type:—CANADA. Quebec: Lac Jean Venne, 45.67893 N 76.0645 W, *P. Hamilton, 27 June 2015* (holotype: CANA! 100021, circled specimen, illustrated here as Fig. 3.42)

Etymology: The specific epithet, (*promontorium*), is Latin for apex, cape or peak, and it refers to the peak found longitudinally along the raphe of the valve face of *N. promontorium*.

Observations: *Neidium promontorium* was found in eastern Ontario and western Quebec in acidic (pH 6.59 – 6.63) wetlands and ponds. Examples of *N. promontorium* include Metzeltin and Lange-Bertalot 2007 (pl. 192: 4 – 5) (labeled as *N*. cf. *ampliatum*), Metzeltin et al. 2005 (pl.115: 1) (labeled as *N*. spec. cf. *ampliatum*). This species has similar linear sides as *N.*

amphigomphus, *N. lowei* and *N*. sp6. The single longitudinal canal is also similar to *N. lowei* and *N. fossum*. As well the areolae patterning on the large longitudinal canal is similar to that seen in *N. lowei*. *N. promontorium*'s key morphological features are the pronounced ridge along the external raphe, single large longitudinal canal and the distinctive apiculate to subrostrate ends. In the chloroplast phylogenetic reconstructions, the individuals of *N. promontorium* formed a strong monophyletic group (Fig. 3.1). Only one of the six individuals of *N. promontorium* was the 18S region successfully amplified and sequenced. The pairwise identity in the chloroplast sequences of the individuals of *N. promontorium* was between 99.9 – 100%.

Neidium sp 1 (Figs 3.56 – 3.60)

Number of individuals used in morphological analysis = 20; Number of individuals sequenced = 4. Valves are linear to linear-lanceolate with rostrate to sub-capitate ends. Valve length 64 – 98 µm; width 17 – 22 µm. Striae are parallel with a density of 24 – 28 in 10 µm. Voigt faults are present 1/3 between the apex of the valve and the central area. Central area is circular to slightly elliptical. Areolae are rounded with a density of 20 – 24 in 10 µm. One longitudinal canal is present. Raphe is linear in LM with proximal ends that bifurcate oppositely, and are rounded and hooked. Distal raphe ends terminate at the apical flap.

Type: not designated.

Observations: *Neidium* sp 1 was found in in eastern Ontario and northern New York State in acidic wetlands and ponds (pH 4.92 – 6.63). It is similar in overall shape to *N. affine*, with a slightly larger size range. Within this study we were unable to determine the specific morphological characteristics to define *N*. sp 1. In the phylogenetic reconstruction using the chloroplast dataset, the individuals of *N*. sp1 formed a strong monophyletic group, with a sister

group which included *N. lowei* (Fig. 3.1). This monophyletic grouping was also seen in the 18S phylogenetic reconstruction (Fig. 3.2). There was 100% pairwise identity between the individuals of *N*. sp 1 in both the chloroplast dataset and the 18S dataset.

Neidium sp 6 (Figs 3.61 – 3.64)

Number of individuals used in morphological analysis = 27; Number of individuals sequenced = 6. Valves are broadly linear with cuneate to obtuse ends. Valve length 72 – 118 µm; width 20 – 34 µm. Striae are parallel becoming slightly convergent as they reach the apex with a density of 17 – 25 in 10 µm. Voigt faults are present 1/2 between the edge of the central area and the apex of the valve. Areolae are small and rounded with a density of 17 – 22 in 10 µm. One large longitudinal canal is present. Raphe is linear in LM with proximal ends that are hooked and rounded and oppositely bifurcating. Distal raphe ends terminate in an apical flap.

Type: not designated.

Observations: *Neidium* sp 6 was found in eastern Ontario, and northern New York State in acidic wetlands (pH 4.92 – 6.2). It is similar overall shape to *N. amphigomphus*, and *N. fossum*. Within this study we were unable to definitively determine the morphological features of *N*. sp 6. With the chloroplast phylogenetic reconstruction, *N*. sp 6 individuals formed a strong monophyletic group (3.1), and this was also seen in the 18S phylogenetic reconstruction (3.2). The sister taxa to *N*. sp 6 was *N. promontorium* in the chloroplast analysis, and while there was only one promontorium individual used in the 18S analysis, it was located close to the group of *N*. sp 6 individuals. There was 100% pairwise identity in the chloroplast dataset, while the pairwise identity of the 18S dataset ranged between 99.9 – 100%.

3.5 Discussion

Within this study 124 *Neidium* individuals were analysed phylogenetically and morphologically. Two species, *N. sacoense* and *N. longiceps*, were matched to their original species descriptions. Three novel species were described (*N. lowei*, *N. potapovii* and *N. promontorium*) with two additional taxa identified but not named. For the purpose of this study we defined our species criteria to be that: (1) show particular morphological characteristics which are measurable and demonstrably different from other taxa (2) it must have strong (≥90% BS, 1.00 PP) phylogenetic support from our molecular analysis, and be monophyletic.

3.5.1 Cryptic Species

Cryptic species of diatoms are defined as taxa which have the same morphological characteristics but are shown to be genetically distinct taxa. Within this analysis there is an example of cryptic species between *N. affine* and the unnamed *N.* sp 1. These two taxonomic entities were found in the same locations, and were not distinguishable using morphological characteristics at the magnification level that we used in the single cell isolation imaging step. Despite this apparent cryptic species designation, it may be possible to discern these two taxa if a higher lens objective were used in single cell imaging as it may allow for greater insight into the morphologic characteristics that each of these species possess.

Within our phylogenetic analysis, *N. amphigomphus* showed three distinct genetic variants within the concatenated phylogenetic tree (Fig. 3.1). However, these three distinct groups were not distinguishable from each other in terms of geographic range or morphological characteristics, and the intraspecific variation which separated them was less than 1%. We suggest that these genetic variations are part of intraspecific variation and do not represent

variation present at the species level with our definition of a species. We propose to not describe the different variants as cryptic species, as they do not meet the morphological criteria for a species and are very closely related (phylogenetically) to each other.

3.5.2 Choice of Molecular Markers

The ML tree constructed from the concatenated chloroplast analysis (Fig. 3.1) has a species-level resolution of phylogenetic information. This is comparing it to the phylogenetic tree produced from only the 18S analysis, which was unable to distinguish between very well defined species. Likely part of this is due to the fact that the chloroplast analyses were using three genes instead of one. However from Table 3.3, 18S has a higher pairwise percent identity compared to the chloroplast genes. This conservative feature of 18S has been noted in other studies with other diatom taxa (Evans *et al*., 2007; Moniz and Kaczmarska, 2009). Within this study, we decided to analyse the chloroplast genes and 18S separately because they are from different genomes, and we did not want to assume a constant rate of mutation across the plastid and nuclear genomes.

While collecting the sequence data for this study, we attempted to also amplify several other genes from the nuclear ribosomal complex; 28S, ITS1, 5.8S, ITS2, and also *CO1* from the mitochondrial genome. We were unsuccessful at determining internal primers for the nuclear ribosomal complex genes. While we did create primers that worked on *Neidium* culture material for 28S, when we used the internal and external primers together in the nested PCR protocol, the resulting product was sequenced and found to be fungi. Later efforts will continue to see if there is a set of primers for the 28S region which can be optimized to work with the single cell

protocol, as some studies have suggested that 28S is a more variable region and possibly when combined with 18S, could be at a phylogenetic resolution to delimit species.

3.5.3 Limitations with single cell sequencing and morphology

The inability to find effective internal and external primers for this technique is one of the limitations of this approach. For molecular markers with increased variability it is often difficult to find conserved regions across a genus or group where internal primers can be used. However, when internal and external primers are found for diatoms, they appear to be good across families and orders (Appendix A). For future studies an increase in the number of diatom genomes available on GenBank and the lowering costs for whole genome sequencing will aid in the determination of new internal primer locations for diatom genera.

Another limitation that we experienced using this methodology is the quality of imaging live cells which are then destroyed in the DNA extraction process. The inability to image these cells at a higher resolution likely inhibits our ability to notice all minute distinct morphological characteristics of the individual. Future work could help mitigate these issues by using water immersion objectives which would allow for higher resolution of the live cells when imaging. This would be especially beneficial to the very small (<50 μm) diatoms, where it is very difficult to effectively describe their morphology using anything other than simple length and width measurements.

3.5.4 Conclusions

By combining molecular and morphological work this study was able to effectively distinguish species within the genus *Neidium*, and help to sort out some of the taxonomic confusion created by examining only the morphological features. Through a continued effort on

the part of diatom taxonomists, this type of approach can be used to clarify much of the taxonomic confusion within diatom taxa.

3.6 Acknowledgments

The author wishes to thank Mr. Paul Hamilton, Canadian Museum of Nature, Ottawa, Canada for his insights into *Neidium* taxonomy, and assistance with SEM and LM imaging of the samples, Dr. Frances Pick, Center for Advanced Research in Environmental Genomics, Ottawa, Canada for her thorough editing of the chapter, Dr. Julian Starr, Center for Advanced Research in Environmental Genomics, University of Ottawa, Ottawa, Canada for his helpful discussions on species concepts, Mr. Roger Bull, Canadian Museum of Nature, Ottawa, Canada for his technical assistance in the laboratory, and Miss Katie Baillie-David for her help with diatom isolations. Financial support was given to K.E.L. under a grant by the Canadian Museum of Nature obtained by P.B.H. Financial travel support to Newfoundland was given to K.E.L. under a grant given by the Faculty of Graduate and Postdoctoral Studies, University of Ottawa, Ottawa, Canada.

Table 3.1. List of *Neidium* specimens used in this study. The taxa are listed with GenBank Accession numbers, CANA Accession number, identifier, location and date collected. Sequences retrieved from GenBank as sister taxa and outgroups

GenBank Accession ID				CANA Accession Number	Identifier	Taxon	Locality and date of collection	Source
rbcL	***psbC***	***psbA***	**18S**					
HQ912473.1	HQ912302.1	KM009474.1	HQ912609.1	NA	FD13	*Scoliopleura peisonis*	UTEX FD 13	Theriot *et al.* (2010)
KP325181	KU674683	KU674465	KP325148	100068	G4R12	*Neidium hitchcockii*	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W 18/05/2014	Lefebvre & Hamilton (2015)
KU674586	KU674710	KU674492	—	93416	F1R14	*Neidium hitchcockii*	Petawawa Forestry Station (Ontario, Canada) 45.99275 N 77.39919 W 15/06/2014	
KU674592	KU674716	KU674498	KU674393	100021	C4R15	*Neidium hitchcockii*	Lac Jean Venne (Quebec, Canada) 45.67893 N 76.0645 W 27/06/2014	
KU674579	KU674698	KU674480	—	100066	G9R13	*Neidium amphigomphus*	Wilson Pond, Hwy 28, Adirondack Park (New York, USA) 43.84326 N 74.47681 W 10/05/2014	
KP325193	KU674697	KU674479	KP325160	100066	G7R13	*Neidium amphigomphus*	Wilson Pond, Hwy 28, Adirondack Park (New York, USA) 43.84326 N 74.47681 W 10/05/2014	Lefebvre & Hamilton (2015)
KU674611	KU674735	KU674517	KU674409	108124	A6R18	*Neidium amphigomphus*	Indian Pond Road, Hwy 56 (New York, USA) 44.44879 N 74.77396 W 16/10/2014	
KU674610	KU674734	KU674516	KU674408	108124	A2R18	*Neidium amphigomphus*	Indian Pond Road, Hwy 56 (New York, USA) 44.44879 N 74.77396 W 16/10/2014	
KU674599	KU674723	KU674505	KU674400	108124	K3R15	*Neidium amphigomphus*	Indian Pond Road, Hwy 56 (New York, USA) 44.44879 N 74.77396 W 16/10/2014	
KU674580	KU674699	KU674481	—	100066	H1R13	*Neidium amphigomphus*	Wilson Pond, Hwy 28, Adirondack Park (New York, USA) 43.84326 N 74.47681 W 10/05/2014	
KP325194	KU674693	KU674475	KP325161	93284	C7R13	*Neidium amphigomphus*	Big Moose Lake, Adirondack Park (New York, USA) 43.81723 N 74.88400 W 10/05/2014	Lefebvre & Hamilton (2015)
KP325188	KU674702	KU674484	KP325155	100068	A3R14	*Neidium amphigomphus*	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W 18/05/2014	Lefebvre & Hamilton (2015)
KU674638	KU674762	KU674544	KU674435	108073	E4R19	*Neidium amphigomphus*	Lake Charlie George (Ontario, Canada) 45.136325 N 78.184169 W 25/10/2014	
KP325186	KU674687	KU674469	KP325153	100068	G8R12	*Neidium amphigomphus*	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W 18/05/2014	Lefebvre & Hamilton (2015)
KU674619	KU674743	KU674525	KU674417	108131	C6R18	*Neidium amphigomphus*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KP325187	KU674696	KU674478	KP325154	100073	F9R13	*Neidium amphigomphus*	Boat Launch, Big Moose Lake (New York, USA) 43.8271 N 74.83831 W 10/05/2014	Lefebvre & Hamilton (2015)
KU674637	KU674761	KU674543	KU674434	108073	E3R19	*Neidium amphigomphus*	Lake Charlie George (Ontario, Canada) 45.136325 N 78.184169 W 25/10/2014	

Table 3.1. Continued

GenBank Accession ID				CANA Accession Number	Identifier	Taxon	Locality and date of collection	Source
rbcL	***psbC***	***psbA***	**18S**					
KP325184	KU674685	KU674467	KP325151	100068	G6R12	*Neidium amphigomphus*	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W 18/05/2014	Lefebvre & Hamilton (2015)
KP325185	KU674694	KU674476	KP325152	93284	C9R13	*Neidium amphigomphus*	Big Moose Lake, Adirondack Park (New York, USA) 43.81723 N 74.88400 W 10/05/2014	Lefebvre & Hamilton (2015)
KU674593	KU674717	KU674499	KU674394	108129	H4R15	*Neidium amphigomphus*	Big Brook, Hwy 30 (New York, USA) 44.00262 N 74.4809 W 26/10/2014	
KU674640	KU674764	KU674546	KU674437	108601	A3R20	*Neidium amphigomphus*	Bay Bulls (Newfoundland, Canada) 47.350125 N 52.808971 W 20/05/2015	
KU674661	KU674785	KU674567	—	108637	C7R25	*Neidium amphigomphus*	Maple Lake, Hwy 103 (Nova Scotia) 44.637249 N 64,081334 W 30/05/2015	
KU674641	KU674765	KU674547	KU674438	108601	A4R20	*Neidium amphigomphus*	Bay Bulls (Newfoundland, Canada) 47.350125 N 52.808971 W 20/05/2015	
KU674648	KU674772	KU674554	—	108603	C8R20	*Neidium amphigomphus*	Roadside Pond, Mobile (Newfoundland, Canada) 47.265771 N 52.835526 W 21/05/2015	
KP325189	KU674668	KU674450	KP325156	93650	C3R7	*Neidium amphigomphus*	Roadside Pond, Hwy 56 (New York, USA) 44.39273 N 74.43944 W 11/05/2014	Lefebvre & Hamilton (2015)
KU674643	KU674767	KU674549	KU674440	108602	A7R20	*Neidium amphigomphus*	Bay Bulls (Newfoundland, Canada) 47.350125 N 52.808971 W 20/05/2015	
KP325191	KU674700	KU674482	KP325158	100067	H4R13	*Neidium amphigomphus*	Owl Head Mountain Trail, Adirondack Park (New York, USA) 43.957394 N 74.464819 W 11/05/2014	Lefebvre & Hamilton (2015)
KU674574	KU674669	KU674451	KU674386	93375	A5R12	*Neidium amphigomphus*	Wetland, Hwy 30 south (New York, USA) 43.927 N 74.44318 W 11/05/2014	
KU674642	KU674766	KU674548	KU674439	108601	A5R20	*Neidium amphigomphus*	Bay Bulls (Newfoundland, Canada) 47.350125 N 52.808971 W 20/05/2015	
KU674573	KU674667	KU674449	KU674385	93650	C2R7	*Neidium amphigomphus*	Roadside Pond, Hwy 56 (New York, USA) 44.39273 N 74.43944 W 11/05/2014	
KU674575	KU674671	KU674453	KU674387	93375	B1R12	*Neidium amphigomphus*	Wetland, Hwy 30 south (New York, USA) 43.927 N 74.44318 W 11/05/2014	
KU674577	KU674674	KU674456	KU674389	100072	B8R12	*Neidium amphigomphus*	Wetland, Adirondack Park (New York, USA) 43.90870 N 74.43944 W 11/05/2014	
KU674576	KU674672	KU674454	KU674388	93375	B2R12	*Neidium amphigomphus*	Wetland, Hwy 30 south (New York, USA) 43.927 N 74.44318 W 11/05/2014	
KP325190	KU674677	KU674459	KP325157	100072	D4R12	*Neidium amphigomphus*	Wetland, Adirondack Park (New York, USA) 43.90870 N 74.43944 W 11/05/2014	Lefebvre & Hamilton (2015)
KU674581	KU674704	KU674486	KU674390	100068	B5R14	*Neidium amphigomphus*	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W 18/05/2014	

Table 3.1. Continued

GenBank Accession ID				CANA Accession Number	Identifier	Taxon	Locality and date of collection	Source
rbcL	***psbC***	***psbA***	**18S**					
KP325192	KU674673	KU674455	KP325159	93375	B4R12	*Neidium amphigomphus*	Wetland, Hwy 30 south (New York, USA) 43.927 N 74.44318 W 11/05/2014	
KU674621	KU674745	KU674527	KU674419	108131	C8R18	*Neidium amphigomphus*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674623	KU674747	KU674529	KU674421	108131	D2R18	*Neidium* sp 1	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674631	KU674755	KU674537	KU674428	108131	F2R18	*Neidium* sp 1	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674617	KU674741	KU674523	KU674415	108131	C4R18	*Neidium* sp 1	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674627	KU674751	KU674533	KU674425	108131	D7R18	*Neidium* sp 1	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KP325183	KU674684	KU674466	KP325150	100068	G5R12	*Neidium* sp	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W 18/05/2014	Lefebvre & Hamilton (2015)
KP325182	KU674705	KU674487	KP325149	100068	B6R14	*Neidium* sp	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W 18/05/2014	Lefebvre & Hamilton (2015)
KU674635	KU674759	KU674541	KU674432	108115	G3R18	*Neidium* sp	Wetland, Hwy 17 (Ontario, Canada) 46.20602 N 77.97166 W 15/10/2014	
KU674660	KU674784	KU674566	—	108593	B4R23	*Neidium lowei*	Lac Jean Venne (Quebec, Canada) 45.6788855 N 76.0616434 W 16/05/2015	
KU674658	KU674782	KU674564	—	108593	B1R23	*Neidium lowei*	Lac Jean Venne (Quebec, Canada) 45.6788855 N 76.0616434 W 16/05/2015	
KU674649	KU674773	KU674555	—	108604	D3R20	*Neidium lowei*	Roadside Pond, Mobile (Newfoundland, Canada) 47.265771 N 52.835526 W 21/05/2015	
KU674588	KU674712	KU674494	—	100021	A5R15	*Neidium lowei*	Lac Jean Venne (Quebec, Canada) 45.67893 N 76.0645 W 27/06/2014	
KU674651	KU674775	KU674557	—	108606	D6R20	*Neidium lowei*	Roadside Pond, Mobile (Newfoundland, Canada) 47.265771 N 52.835526 W 21/05/2015	
KU674650	KU674774	KU674556	KU674445	108606	D4R20	*Neidium lowei*	Roadside Pond, Mobile (Newfoundland, Canada) 47.265771 N 52.835526 W 21/05/2015	
KU674589	KU674713	KU674495	KU674392	100021	A6R15	*Neidium lowei*	Lac Jean Venne (Quebec, Canada) 45.67893 N 76.0645 W 27/06/2014	
KU674603	KU674727	KU674509	—	NA	A8R16	*Neidium productum*	UTEX LB FD116	
HQ912446.1	HQ912275.1	KM009447.1	HQ912582.1	NA		*Neidium productum*	UTEX LB FD116	Theriot *et al.* (2010)
KM999091	KU674670	KU674452	KP325162	100072	A7R12	*Neidium*	Wetland, Adirondack Park (New York, USA)	Hamilton *et al.*

Table 3.1. Continued

GenBank Accession ID				CANA Accession Number	Identifier	Taxon	Locality and date of collection	Source
rbcL	***psbC***	***psbA***	**18S**					
						tumescens	43.90870 N 74.43944 W 11/05/2014	(2015)
KM999090	KU674675	KU674457	KP325163	100072	C7R12	*Neidium tumescens*	Wetland, Adirondack Park (New York, USA) 43.90870 N 74.43944 W 11/05/2014	Hamilton *et al.* (2015)
KU674596	KU674720	KU674502	KU674397	108128	J3R15	*Neidium tumescens*	Creek, Hwy 30 (New York, USA) 43.90924 N 74.4397 W 16/10/2014	
KU674664	KU674788	KU674570	—	108629	D4R25	*Neidium dilatatum*	Little Indian Lake, Hwy 103 (Nova Scotia, Canada) 44.710208 N 63.904635 W 30/05/2015	
KU674628	KU674752	KU674534	KU674426	108131	D8R18	*Neidium dilatatum*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674666	KU674790	KU674572	KU674448	108629	D6R25	*Neidium dilatatum*	Little Indian Lake, Hwy 103 (Nova Scotia, Canada) 44.710208 N 63.904635 W 30/05/2015	
KU674659	KU674783	KU674565	—	108593	B2R23	*Neidium dilatatum*	Lac Jean Venne (Quebec, Canada) 45.6788855 N 76.0616434 W 16/05/2015	
KP325199	KU674690	KU674472	KP325168	93279	A9R13	*Neidium dilatatum*	Wetland, Adirondack Park (New York, USA) 44.44806 N 74.77439 W 10/05/2014	Lefebvre & Hamilton (2015)
KP325200	KU674681	KU674463	KP325169	93372	F1R12	*Neidium dilatatum*	Pond, Hwy 56 (New York, USA) 44.39536 N 74.75690 W 11/05/2014	Lefebvre & Hamilton (2015)
KP325196	KU674679	KU674461	KP325165	100072	D6R12	*Neidium dilatatum*	Wetland, Adirondack Park (New York, USA) 43.90870 N 74.43944 W 11/05/2014	Lefebvre & Hamilton (2015)
KU674665	KU674789	KU674571	—	108629	D5R25	*Neidium dilatatum*	Little Indian Lake, Hwy 103 (Nova Scotia, Canada) 44.710208 N 63.904635 W 30/05/2015	
KU674663	KU674787	KU674569	—	108629	D3R25	*Neidium dilatatum*	Little Indian Lake, Hwy 103 (Nova Scotia, Canada) 44.710208 N 63.904635 W 30/05/2015	
KU674653	KU674777	KU674559	—	108605	E5R20	*Neidium dilatatum*	Roadside Pond, Mobile (Newfoundland, Canada) 47.265771 N 52.835526 W 21/05/2015	
KU674662	KU674786	KU674568	KU674447	108629	C8R25	*Neidium dilatatum*	Little Indian Lake, Hwy 103 (Nova Scotia, Canada) 44.710208 N 63.904635 W 30/05/2015	
KU674654	KU674778	KU674560	KU674446	108605	E7R20	*Neidium dilatatum*	Roadside Pond, Mobile (Newfoundland, Canada) 47.265771 N 52.835526 W 21/05/2015	
KU674647	KU674771	KU674553	KU674444	108619	C6R20	*Neidium dilatatum*	Bog, Hwy 90 (Newfoundland, Canada) 47.087781 N 53.474651 W 22/05/2015	
KU674620	KU674744	KU674526	KU674418	108131	C7R18	*Neidium dilatatum*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674618	KU674742	KU674524	KU674416	108131	C5R18	*Neidium dilatatum*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674615	KU674739	KU674521	KU674413	108131	C2R18	*Neidium dilatatum*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	

Table 3.1. Continued

GenBank Accession ID				CANA Accession Number	Identifier	Taxon	Locality and date of collection	Source
rbcL	***psbC***	***psbA***	**18S**					
KU674608	KU674732	KU674514	KU674406	108122	C2R17	*Neidium dilatatum*	Indian Pond Road, Hwy 56 (New York, USA) 44.44879 N 74.77396 W 16/10/2014	
KU674600	KU674724	KU674506	KU674401	108124	K4R15	*Neidium dilatatum*	Indian Pond Road, Hwy 56 (New York, USA) 44.44879 N 74.77396 W 16/10/2014	
KU674595	KU674719	KU674501	KU674396	108128	I6R15	*Neidium dilatatum*	Creek, Hwy 30 (New York, USA) 43.90924 N 74.4397 W 16/10/2014	
KP325195	KU674688	KU674470	KP325164	93279	A4R13	*Neidium dilatatum*	Wetland, Adirondack Park (New York, USA) 44.44806 N 74.77439 W 10/05/2014	Lefebvre & Hamilton (2015)
KP325198	KU674691	KU674473	KP325167	93281	B3R13	*Neidium dilatatum*	Creek, Adirondack Park (New York, USA) 44.00715 N 74.49155 W 11/05/2014	Lefebvre & Hamilton (2015)
KM999092	KU674676	KU674458	KP325170	93372	D1R12	*Neidium* sp	Pond, Hwy 56 (New York, USA) 44.39536 N 74.75690 W 11/05/2014	Hamilton et al. (2015)
KU674594	KU674718	KU674500	KU674395	108128	I5R15	*Neidium* sp	Creek, Hwy 30 (New York, USA) 43.90924 N 74.4397 W 16/10/2014	
KU674598	KU674722	KU674504	KU674399	108128	J6R15	*Neidium* sp	Creek, Hwy 30 (New York, USA) 43.90924 N 74.4397 W 16/10/2014	
KU674597	KU674721	KU674503	KU674398	108128	J4R15	*Neidium potapovii*	Creek, Hwy 30 (New York, USA) 43.90924 N 74.4397 W 16/10/2014	
KU674612	KU674736	KU674518	KU674410	108131	B6R18	*Neidium potapovii*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674645	KU674769	KU674551	KU674442	108616	B5R20	*Neidium affine*	Bog, Hwy 90 (Newfoundland, Canada) 47.087781 N 53.474651 W 22/05/2015	
KP325202	KU674690	KU674462	KP325172	100072	E6R12	*Neidium affine*	Wetland, Adirondack Park (New York, USA) 43.90870 N 74.43944 W 11/05/2014	Lefebvre & Hamilton (2015)
KU674644	KU674768	KU674550	KU674441	108621	B1R20	*Neidium affine*	Cape St Mary's Ecological Reserve (Newfoundland, Canada) 46.844263 N 54.178485 W 22/05/2015	
KU674652	KU674776	KU674558	—	108613	E1R20	*Neidium affine*	Bog near Ferryland, Hwy 10 (Newfoundland, Canada) 46.958412 N 52.9566 W 21/05/2015	
KU674646	KU674770	KU674552	KU674443	108617	B7R20	*Neidium affine*	Bog, Hwy 90 (Newfoundland, Canada) 47.087781 N 53.474651 W 22/05/2015	
KP325201	KU674692	KU674474	KP325171	93284	C4R13	*Neidium affine*	Big Moose Lake, Adirondack Park (New York, USA) 43.81723 N 74.88400 W 10/05/2014	Lefebvre & Hamilton (2015)
KU674622	KU674746	KU674528	KU674420	108131	D1R18	*Neidium affine*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674616	KU674740	KU674522	KU674414	108131	C3R18	*Neidium affine*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	

Table 3.1. Continued

GenBank Accession ID				CANA Accession Number	Identifier	Taxon	Locality and date of collection	Source
rbcL	***psbC***	***psbA***	**18S**					
HQ912455.1	HQ912284.1	KM009456.1	HQ912591.1	NA		*Neidium bisulcatum*	UTEX LB FD417	Theriot *et al.* (2010)
KU674602	KU674726	KU674508	—	NA	A5R16	*Neidium bisulcatum*	UTEX LB FD417	
KU674601	KU674725	KU674507	—	NA	A4R16	*Neidium bisulcatum*	UTEX LB FD417	
KU674587	KU674711	KU674493	—	100021	A2R15	*Neidium sacoense*	Lac Jean Venne (Quebec, Canada) 45.67893 N 76.0645 W 27/06/2014	
KU674624	KU674748	KU674530	KU674422	108131	D4R18	*Neidium sacoense*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674632	KU674756	KU674538	KU674429	108131	F4R18	*Neidium sacoense*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674607	KU674731	KU674513	KU674405	108122	C1R17	*Neidium sacoense*	Indian Pond Road, Hwy 56 (New York, USA) 44.44879 N 74.77396 W 16/10/2014	
KU674604	KU674728	KU674510	KU674402	108122	A5R17	*Neidium sacoense*	Indian Pond Road, Hwy 56 (New York, USA) 44.44879 N 74.77396 W 16/10/2014	
KU674605	KU674729	KU674511	KU674403	108122	A7R17	*Neidium sacoense*	Indian Pond Road, Hwy 56 (New York, USA) 44.44879 N 74.77396 W 16/10/2014	
KU674578	KU674689	KU674471	—	93279	A8R13	*Neidium sacoense*	Wetland, Adirondack Park (New York, USA) 44.44806 N 74.77439 W 10/05/2014	
KU674606	KU674730	KU674512	KU674404	108122	A8R17	*Neidium sacoense*	Indian Pond Road, Hwy 56 (New York, USA) 44.44879 N 74.77396 W 16/10/2014	
KU674609	KU674733	KU674515	KU674407	108122	C5R17	*Neidium sacoense*	Indian Pond Road, Hwy 56 (New York, USA) 44.44879 N 74.77396 W 16/10/2014	
HQ912447.1	HQ912276.1	KM009448.1	HQ912583.1	NA		*Neidium affine var. longiceps*	UTEX LB FD127- Minnesota, USA	Theriot *et al.* (2010)
KU674629	KU674753	KU674535	KU674427	108131	E7R18	*Neidium longiceps*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674639	KU674763	KU674545	KU674436	108073	E8R19	*Neidium longiceps*	Lake Charlie George (Ontario, Canada) 45.136325 N 78.184169 W 25/10/2014	
KU674630	KU674754	KU674536	—	108131	E8R18	*Neidium longiceps*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674625	KU674749	KU674531	KU674423	108131	D5R18	*Neidium longiceps*	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KP325206	KU674695	KU674477	KP325177	93375	E3R13	*Neidium* sp	Wetland, Hwy 30 south (New York, USA) 43.927 N 74.44318 W 11/05/2014	Lefebvre & Hamilton (2015)
KP325208	KU674701	KU674483	KP325179	100068	A1R14	*Neidium* sp	Petawawa Forestry Station (Ontario, Canada)	Lefebvre &

Table 3.1. Continued

GenBank Accession ID				CANA Accession Number	Identifier	Taxon	Locality and date of collection	Source
rbcL	***psbC***	***psbA***	**18S**					
							45.992966 N 77.399194 W 18/05/2014	Hamilton (2015)
KP325209	KU674703	KU674485	KP325180	100068	B1R14	*Neidium* sp	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W 18/05/2014	Lefebvre & Hamilton (2015)
KP325205	KU674678	KU674460	KP325176	100072	D5R12	*Neidium fossum*	Wetland, Adirondack Park (New York, USA) 43.90870 N 74.43944 W 11/05/2014	Lefebvre & Hamilton (2015)
KU674657	KU674781	KU674563	—	108598	A3R23	*Neidium fossum*	Lac a L'eau Claire (Quebec, Canada) 45.663005 N 75.622338 W 16/05/2015	
KU674655	KU674779	KU674561	—	108578	A2R21	*Neidium fossum*	near Actinolyte, Hwy 7 (Ontario, Canada) 44.62236 N 77.17145 W 02/05/2015	
KU674565	KU674780	KU674562	—	108578	A4R21	*Neidium fossum*	near Actinolyte, Hwy 7 (Ontario, Canada) 44.62236 N 77.17145 W 02/05/2015	
KM999089	KU674682	KU674464	KP325174	100068	G3R12	*Neidium fossum*	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W 18/05/2014	Hamilton *et al.* (2015)
KP325205	KU674686	KU674468	KP325175	100068	G7R12	*Neidium fossum*	Petawawa Forestry Station (Ontario, Canada) 45.992966 N 77.399194 W 18/05/2014	Lefebvre & Hamilton (2015)
KU674636	KU674760	KU674542	KU674433	108115	G8R18	*Neidium* sp 6	Wetland, Hwy 17 (Ontario, Canada) 46.20602 N 77.97166 W 15/10/2014	
KU674633	KU674757	KU674539	KU674430	108115	G1R18	*Neidium* sp 6	Wetland, Hwy 17 (Ontario, Canada) 46.20602 N 77.97166 W 15/10/2014	
KU674626	KU674750	KU674532	KU674424	108131	D6R18	*Neidium* sp 6	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674614	KU674738	KU674520	KU674412	108131	C1R18	*Neidium* sp 6	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674613	KU674737	KU674519	KU674411	108131	B8R18	*Neidium* sp 6	Wetland, Adirondack Park (New York, USA) 44.02056 N 74.50678 W 16/10/2014	
KU674634	KU674758	KU674540	KU674431	108115	G2R18	*Neidium* sp 6	Wetland, Hwy 17 (Ontario, Canada) 46.20602 N 77.97166 W 15/10/2014	
KU674590	KU674714	KU674496	—	100021	B5R15	*Neidium promontorium*	Lac Jean Venne (Quebec, Canada) 45.67893 N 76.0645 W 27/06/2014	
KU674591	KU674715	KU674497	—	100021	B7R15	*Neidium promontorium*	Lac Jean Venne (Quebec, Canada) 45.67893 N 76.0645 W 27/06/2014	
KU674583	KU674707	KU674489	KU674391	93415	E1R14	*Neidium promontorium*	Pond, Hwy 17 Picnic Area (Ontario, Canada) 46.246422 N 78.180147 W 15/06/2014	
KU674582	KU674706	KU674488	—	93415	D8R14	*Neidium promontorium*	Pond, Hwy 17 Picnic Area (Ontario, Canada) 46.246422 N 78.180147 W 15/06/2014	
KU674584	KU674708	KU674490	—	93416	E5R14	*Neidium promontorium*	Petawawa Forestry Station (Ontario, Canada) 45.99275 N 77.39919 W 15/06/2014	

Table 3.1. Continued

GenBank Accession ID				CANA Accession Number	Identifier	Taxon	Locality and date of collection	Source
rbcL	***psbC***	***psbA***	**18S**					
KU674585	KU674709	KU674491	—	93416	E8R14	*Neidium promontorium*	Petawawa Forestry Station (Ontario, Canada) 45.99275 N 77.39919 W 15/06/2014	
—	—	—	KP325178	100072	E4R12	*Neidium* sp.	Wetland, Adirondack Park (New York, USA) 43.90870 N 74.43944 W 11/05/2014	Lefebvre & Hamilton (2015)

Table 3.2 List of primer sequences used in the nested amplification of *rbcL*, *psbA*, 18S and *psbC*.

Primer Name	Primer Sequence (5'-3')	Reference
RBCL66+ [a]	TTTAAGGAGAAATAAATGTCTCAATCTG	Alverson *et al.* 2007
RBCLdp7- [a]	AAASHDCCTTGTGTWAGTVTC	Daugbjerg and Andersen, 1997
RBCL40+ [b]	GGACTCGAATVAAAAGTGAACG	Ruck and Theriot, 2011
RBCL1444- [b]	GCGAAATCAGCTGTATCTGTWG	Ruck and Theriot, 2011
psbA-F [a]	ATGACTGCTACT TTAGAAAGACG	Yoon *et al.* 2002
psbA-R [a]	GCTAAATCTARWGGGAAGTTGTG	Yoon *et al.* 2002
psba_45F_dg [b]	WTTYATCGCWGCTCCWCCAG	Appendix A
psba_648R_dg [b]	RTGWGCTGCAACGATRTTRT	Appendix A
18S_P2F [a]	CTGGTTGATTCTGCCAGT	Hannen van *et al*. 1999
18S_P4R [a]	TGATCCTTCYGCAGGTTCAC	Guillou *et al.* 1999
18S200F [b]	YGGSRWGAYRTGGTGADTCA	Appendix A
18S1444 [b]	GVRTRCATCAGTGTAGCGCG	Appendix A
psbC+ [a]	ACAGGMTTYGCTTGGTGGAGTGG	Alverson *et al.* 2007
psbC- [a]	CACGACCWGAATGCCACCAATG	Alverson *et al.* 2007
psbC22+ [b]	CGTGGTGATACATAGTTA	Ruck and Theriot, 2011
psbC1154- [b]	GCDCAYGCTGGYTTAATGG	Ruck and Theriot, 2011

[a] denotes external primers; [b] denotes internal primers

Table 3.3 Summary of pairwise identity character across the four molecular markers. Each molecular marker is shown along with the percentage of pairwise identity (number of identical bp).

		Pairwise Identity	
Molecular Marker	**Total number of sites**	in *Neidium* taxa (Identical sites)	in Alignment (Identical sites)
rbcL	1281	96.8% (1098)	96.8% (1069)
psbA	576	98.4% (505)	98.3% (500)
psbC	858	96.1% (582)	96.1% (563)
18S	988	98.6% (929)	98.5% (883)

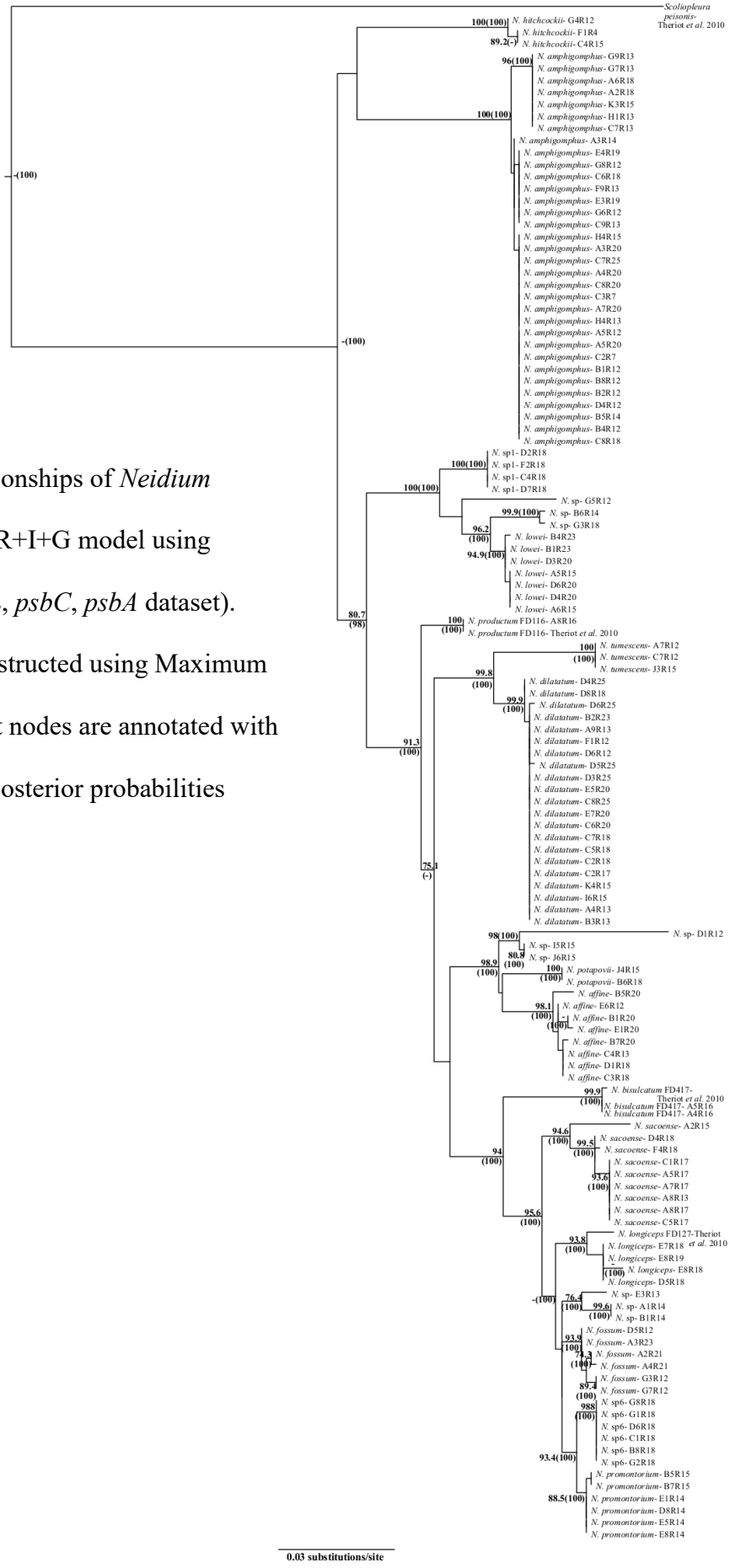


Figure 3.1. Phylogenetic relationships of *Neidium* taxa constructed under the GTR+I+G model using concatenated chloroplast (*rbcL*, *psbC*, *psbA* dataset). The phylogenetic tree was constructed using Maximum likelihood analysis. Significant nodes are annotated with ML bootstrap values and BA posterior probabilities (MLBS/BAPP).

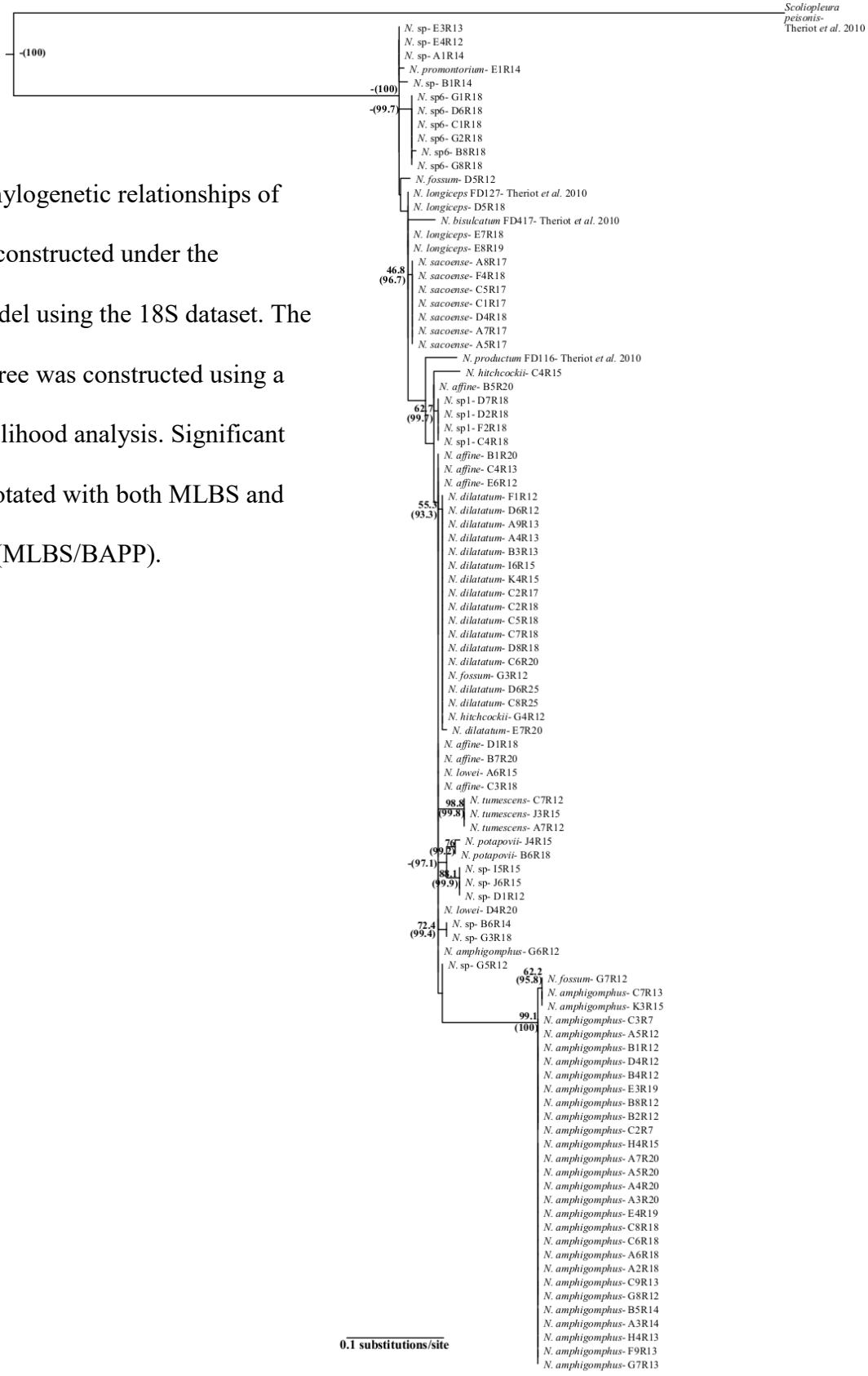
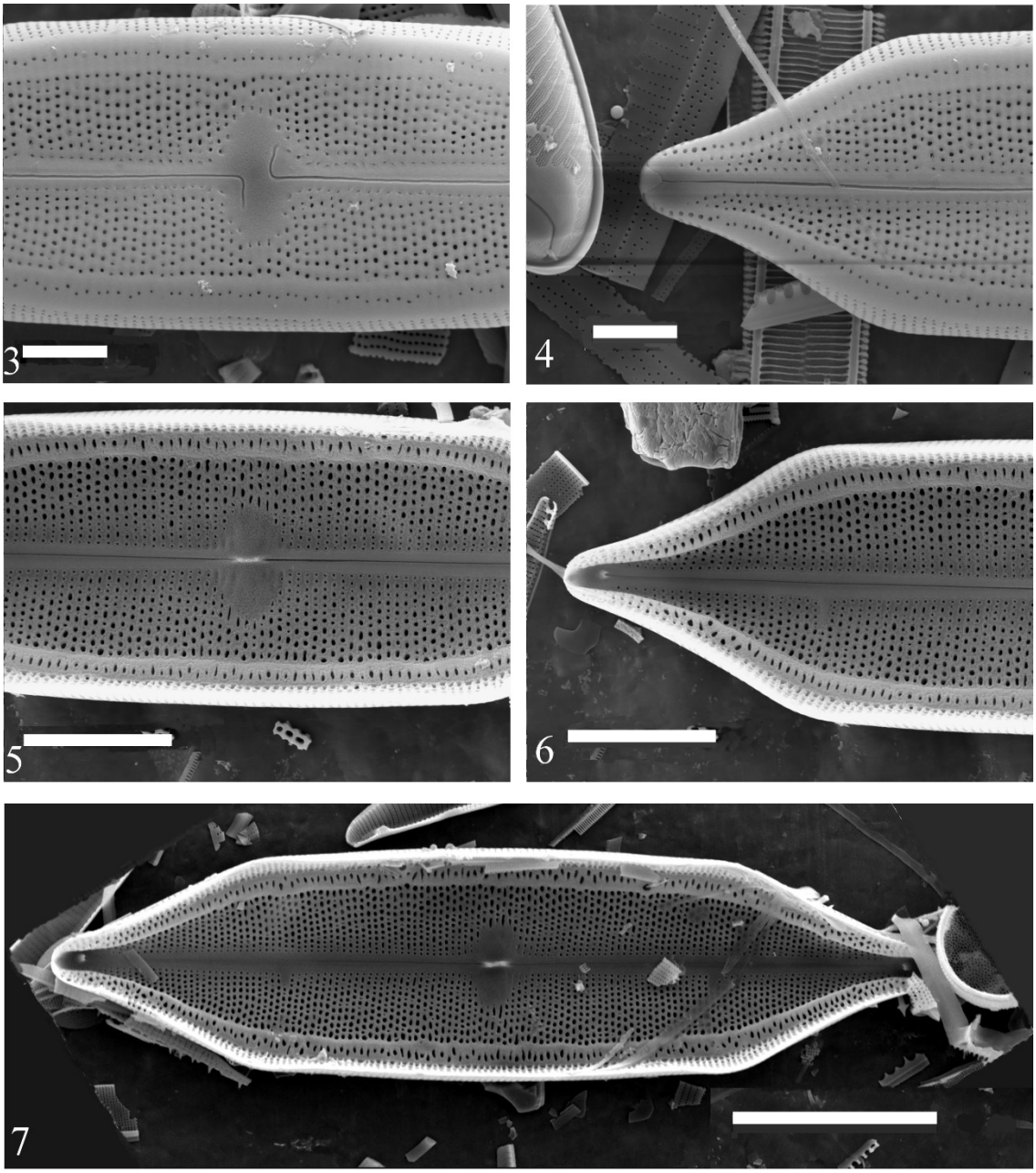
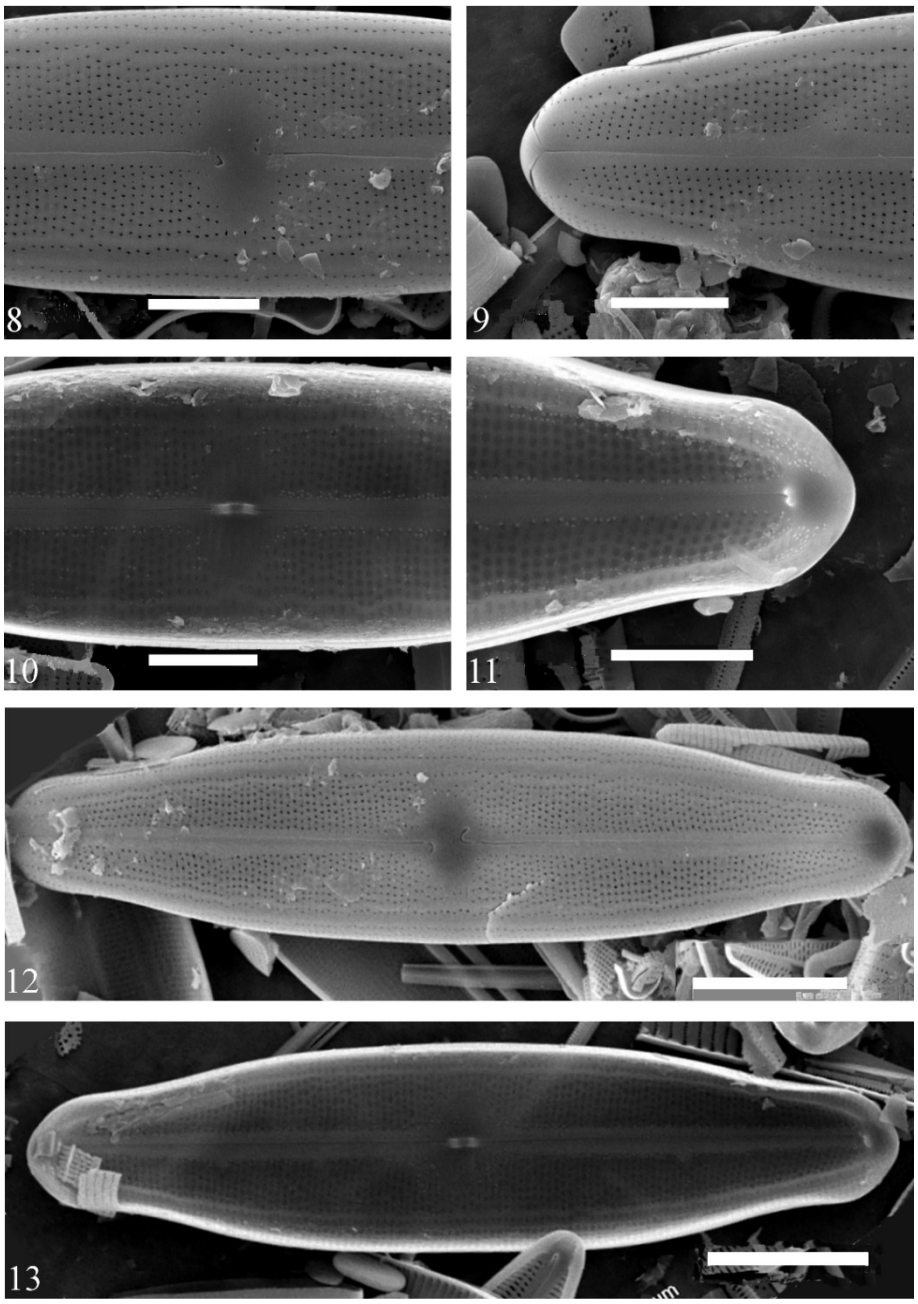


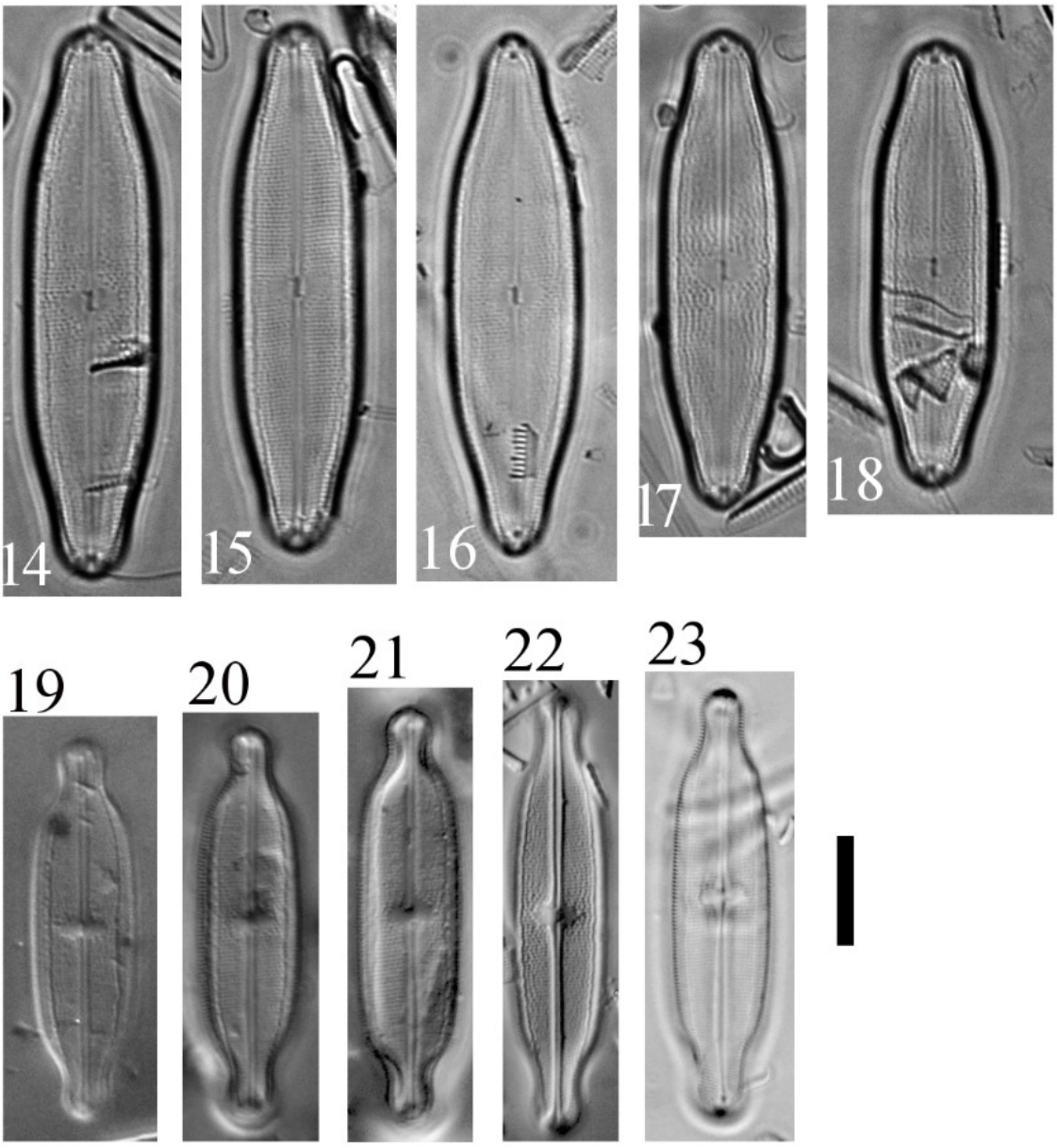
Figure 3.2. Phylogenetic relationships of *Neidium* taxa constructed under the GTR+I+G model using the 18S dataset. The phylogenetic tree was constructed using a maximum likelihood analysis. Significant nodes are annotated with both MLBS and BAPP values (MLBS/BAPP).



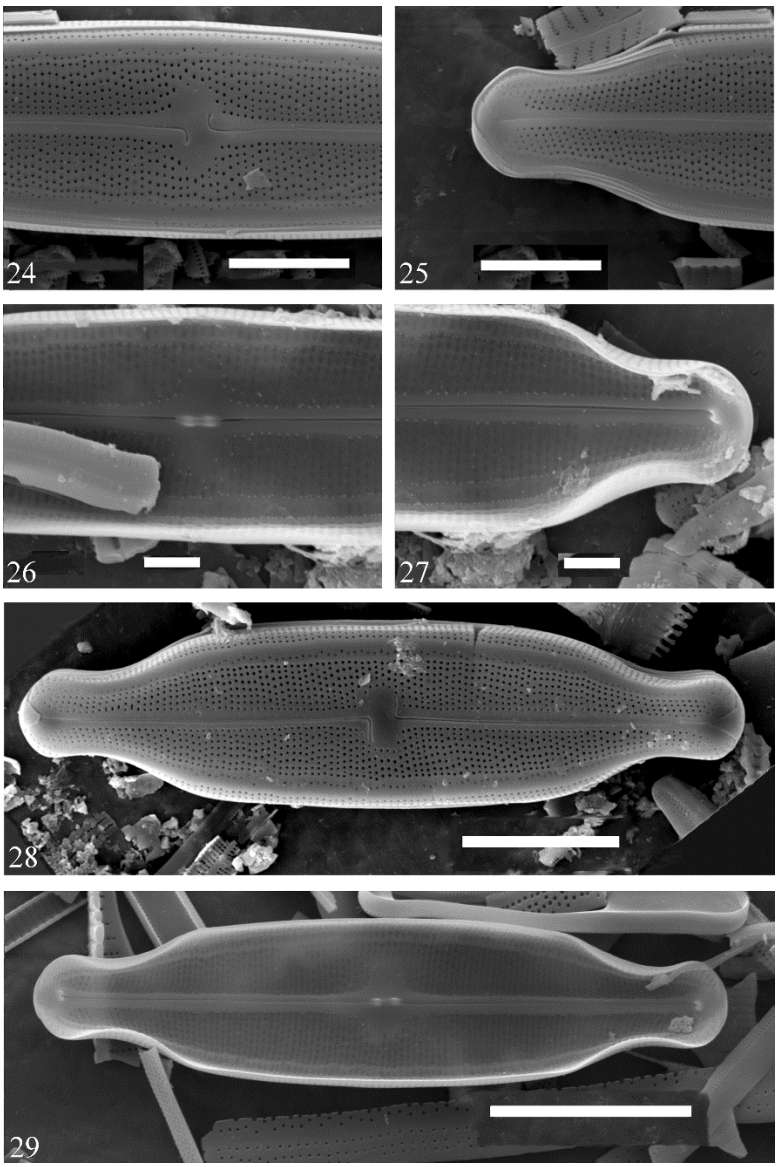
Figures 3.3 – 3.7. *Neidium sacoense* from CANA 108131. 3.3 – 3.4. SEM external view of midsection and apex. 3.5 – 3.6. SEM internal view of midsection and apex. 3.7. SEM internal view of entire valve. Note single large longitudinal canal and cuneate to apiculate ends. Scale bar = 5 µm (3.3 – 3.4), 10 µm (3.5 – 3.6), 20 µm (3.7).



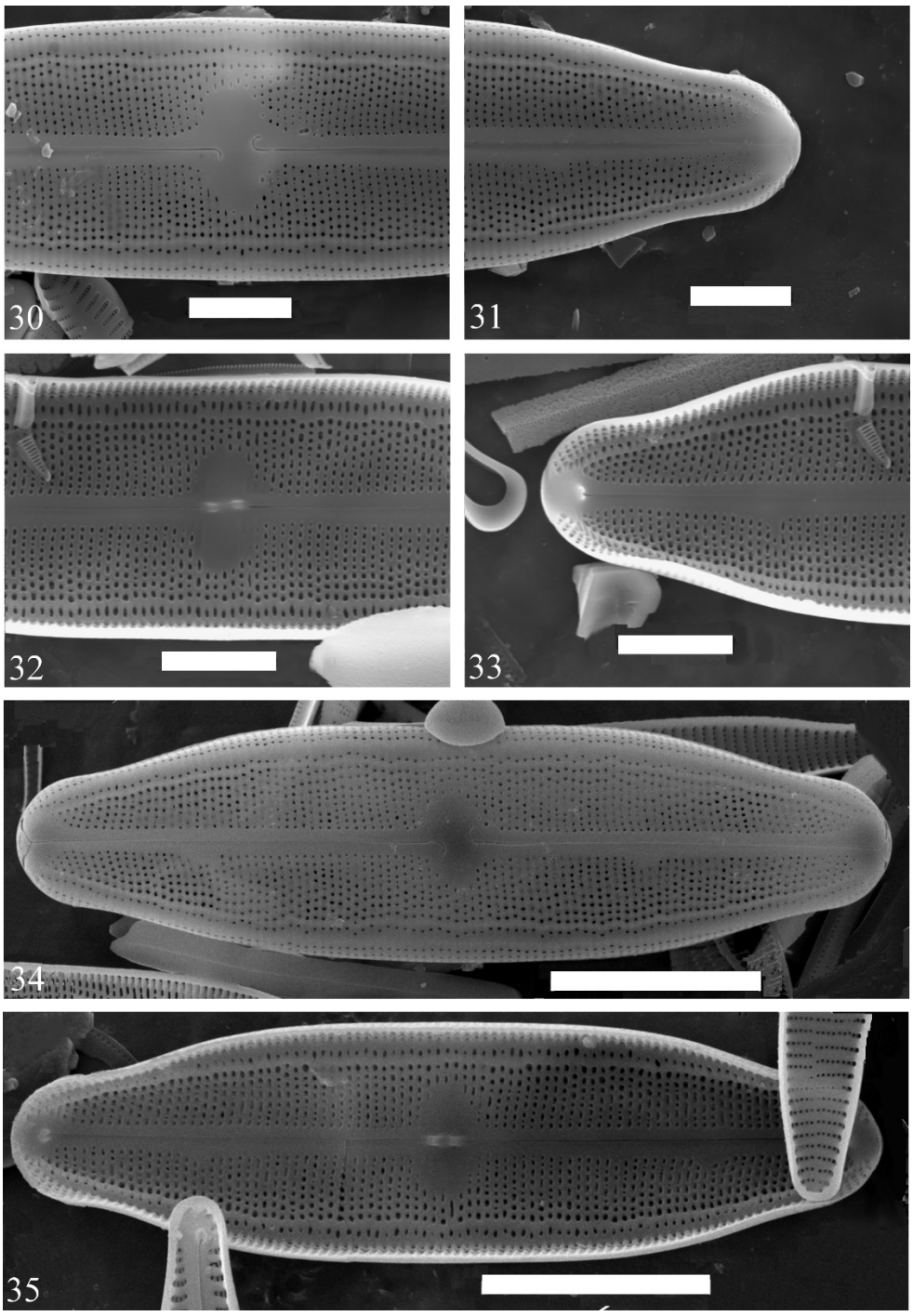
Figures 3.8 – 3.13. *Neidium affine* from CANA 108621. 3.8 – 3.9. SEM external view of midsection and apex. 3.10 – 3.11. SEM internal view of midsection and apex. 3.12. SEM external view of entire valve. 3.13. SEM internal view of entire valve. Note rostrate to slightly capitate ends, large longitudinal canal with two smaller longitudinal canals and distinct helictoglossae. Scale bar = 5 µm (3.8 – 3.11), 10 µm (3.12 – 3.13).



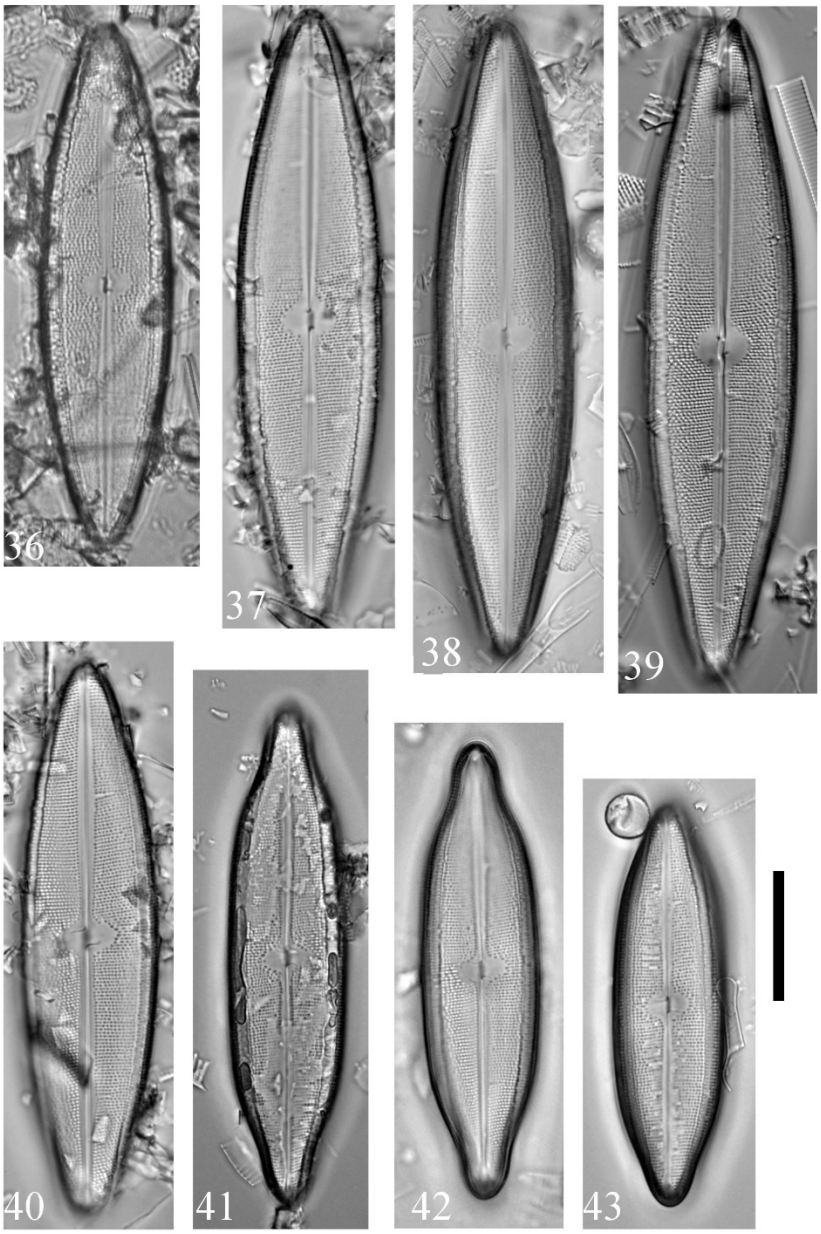
Figures 3.14 – 3.23. *Neidium potapovii* and *Neidium longiceps* LM external view illustrating the size range. 3.14 – 3.15, 3.17 – 3.18 *Neidium potapovii* from CANA 108128. 3.16 *Neidium potapovii* from CANA 108131. 3.18. *Neidium potapovii* holotype (CANA! 108128). 3.19 – 3.21 *Neidium longiceps* from BM47646. 3.22 – 3.23 *Neidium potapovii* from CANA 108131. Scale bar = 10 µm.



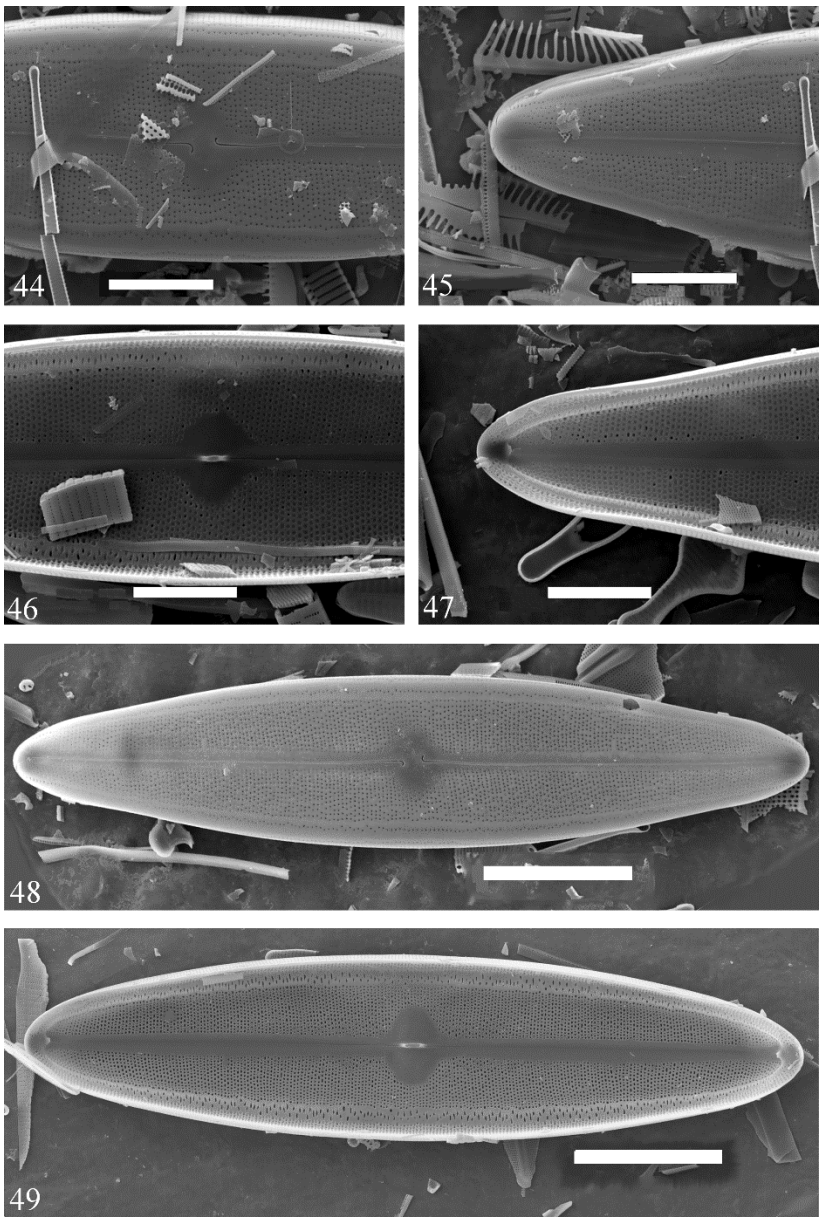
Figures 3.24 – 3.29. SEM images of *Neidium longiceps*. 3.24 – 3.25, 3.29. *Neidium longiceps* from CANA 108129. SEM external view of midsection and apex. 3.26 – 3.28. *Neidium longiceps* from CANA 93446. SEM internal view of midsection and apex. 3.28. SEM external view of entire valve. 3.29. SEM internal view of entire valve. Note linear sides, single longitudinal canal and capitate to subcapitate ends. Scale bar = 10 µm (3.28 – 3.29), 5 µm (3.24 – 3.25), 2 µm (3.26 – 3.27).



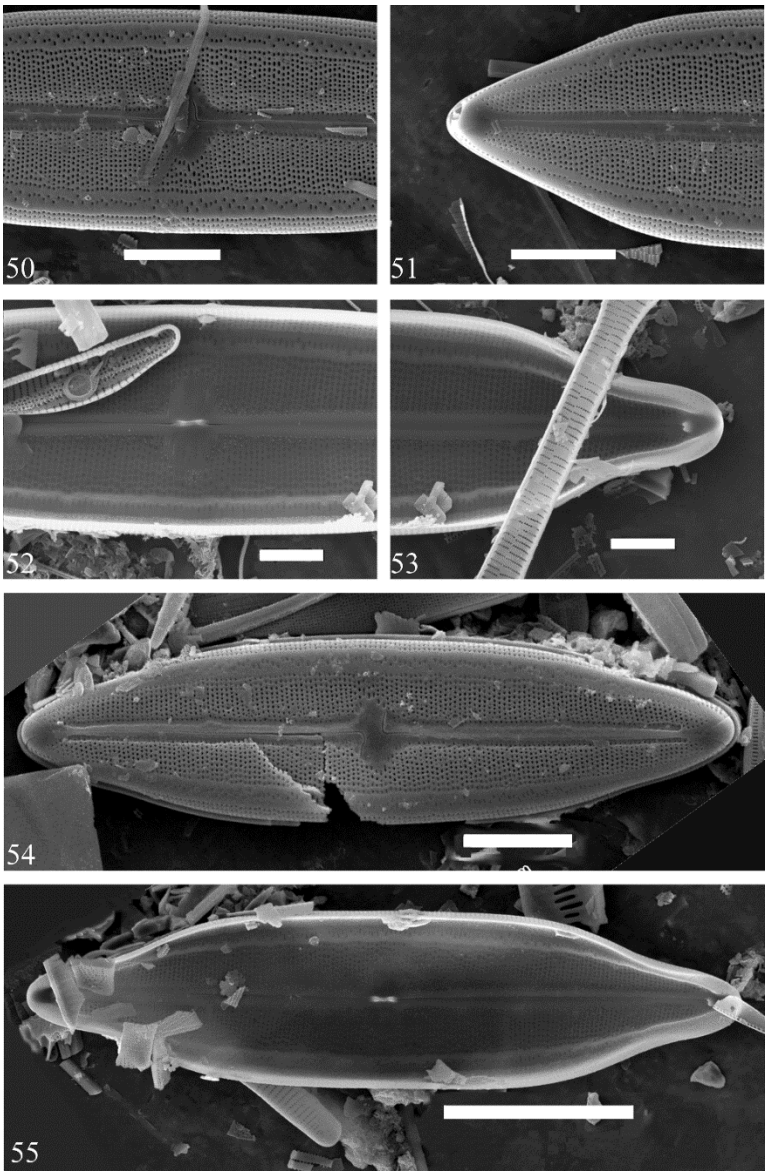
Figures 3.30 – 3.35. SEM images of *Neidium potapovii* from CANA 108128. 3.30 – 3.31. SEM external view of midsection and apex. 3.32 – 3.33. SEM internal view of midsection and apex. 3.34. SEM external view of entire valve. 3.35. SEM internal view of entire valve. Note single longitudinal canal and rostrate ends. Scale bar = 5 µm (3.30 – 3.33), 10 µm (3.34 – 3.35).



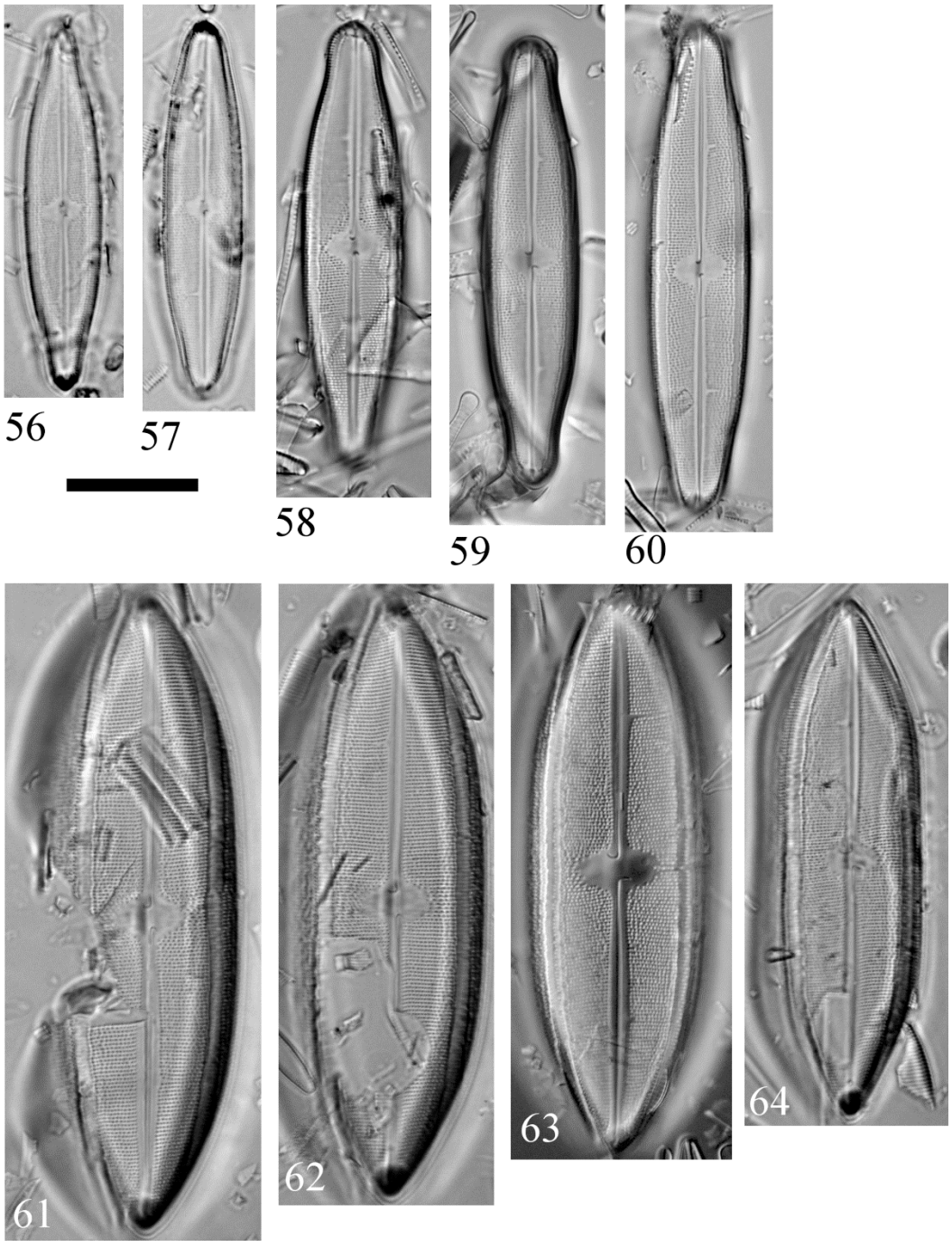
Figures 3.36 – 3.43. *Neidium lowei* and *Neidium promontorium* LM external view illustrating size range. 3.36 *Neidium lowei* from CANA 108606. 3.37. *Neidium lowei* from CANA 93416. 3.38. *Neidium lowei* from CANA 100021. 3.39. *Neidium lowei* from CANA 108593. 3.38. *Neidium lowei* holotype (CANA! 100021). 3.40. *Neidium promontorium* from CANA 93416. 3.41 – 3.43 *Neidium promontorium* from CANA 100021. 3.42. *Neidium promontorium* holotype (CANA! 100021). Scale bar = 20 µm.



Figures 3.44 – 3.49. *Neidium lowei* (SEM). 3.44 – 3.45. *Neidium lowei* from CANA 100021. SEM external view of midsection and apex. 3.46 – 3.47. *Neidium lowei* from CANA 108129. SEM internal view of midsection and apex. 3.48 – 3.49. *Neidium lowei* from CANA 108131. 3.48. SEM external view of entire valve. 3.49. SEM internal view of entire valve. Note size range, single large longitudinal canal and rounded rostrate ends. Scale bar = 10 µm (3.44 – 3.47), 20 µm (3.48 – 3.49).



Figures 3.50 – 3.55. *Neidium promontorium* (SEM). 3.50 – 3.51. *Neidium promontorium* from CANA 108131. SEM external view of midsection and apex. 3.52 – 3.53. *Neidium promontorium* from CANA 100021. SEM internal view of midsection and apex. 3.54. *Neidium promontorium* from CANA 93415. SEM external view of entire valve. 3.55. *Neidium promontorium* from CANA 93416. SEM internal view of entire valve. Note raised ridge along external raphe, single large longitudinal canal and apiculate to subrostrate ends. Scale bar = 10 µm (3.50 – 3.51, 3.54), 5 µm (3.52 – 3.53), 20 µm (3.55).



Figures 3.56 – 3.64. *Neidium* sp1 and *Neidium* sp6 LM external view illustrating size range. 3.56 – 3.57 *Neidium* sp1 from CANA 108131. 3.58 – 3.60 *Neidium* sp1 from CANA 93417. 3.61 – 3.62, 3.64 *Neidium* sp6 from CANA 108131. 3.63 *Neidium* sp6 from CANA 108115. Scale bar = 20 µm.

Chapter 4: Conclusions

4.1 General Conclusions

The overall objective of this thesis was to use a multidisciplinary approach of molecular and morphological methods to delimit species in the diatom genus *Neidium*. In pre-thesis work the single cell methodology was verified using a large range of diatom species (Appendix A). Large *Neidium* taxa (>100 µm) were then studied in fine detail using two molecular markers, *rbcL* and 18S (Chapter 2). Smaller *Neidium* taxa (<100 µm) as well as *Neidium* taxa from a wider geographic range, were also examined using the multidisciplinary approach with an increased number of molecular markers, *rbcL*, *psbA*, *psbC*, and 18S (Chapter 3).

Of the large *Neidium* taxa examined one new taxon was described, *N. fossum*, and other taxa were combined into one likely species, *N. dilatatum* and *N. iridis* into *N. dilatatum*. By re-examining the descriptions of *Neidium* species with molecular methods we were able to reduce confusion surrounding some of the larger species commonly found in North America. As well, with our treatment of the smaller *Neidium* taxa we were also able to determine and delimit species which lessened the phylogenetic uncertainty surrounding *N. affine* and *N. longiceps*. By emending unclear descriptions and attaching sequence data to these species we have made it easier for ecologists and taxonomists to effectively identify these species when they are present. This may allow for species-level ecological observations to be made which would not have been possible before.

This type of phylogenetic analyses can also give insights into the evolutionary history of diatoms. Within our phylogenetic reconstructions in Chapter 3, it was seen that the single, large, longitudinal canal is not a synapomorphy for a particular subset of taxa, but possibly an ancestral

trait as it appears in all clades of the *Neidium* species analysed. Two of the large species, *N. dilatatum* and *N. tumescens,* did form a monophyletic group, but the other large species, N. amphigomphus, was removed from this clade. Phylogenetic reconstructions can be used with single-celled DNA amplifications to determine other ancestral and derived traits within other diatom genera.

Within this work, several molecular markers were used (*rbcL*, *psbA*, *psbC* and 18S). During the experiments, it was shown that the chloroplast markers exhibited a higher level of divergence which was useful in interpreting species complexes of *Neidium* taxa. While these markers were effective at delimiting the *Neidium* genus, careful study must be made for other diatom genera before deciding to delimit species based on a particular set of markers. As well, multi-locus analysis should always be used when delimiting species to avoid assumptions based on a single molecular marker. Further phylogenetic analyses using portioned models on the alignments could also provide insight into relationships between taxa. Using the single cell sequencing method to obtain sequence data from individual diatoms has some limitations. This methodology is quite laborious for very small cells, and due larger working distance needed between the microscope's objective and stage, imaging resolution available for cells being isolated is low. This makes discerning the morphological characteristics of small cells difficult. Despite the limitations, single cell sequencing method can be useful when applied carefully to assist in species-level delimitation of diatoms.

While the accessibility of single cell sequencing allows for a massive increase in the amount of genetic data available about a species, having a clear definition of the species concept is critical for this information to be put to use. De Clerk *et al*. (2013) found that within diatom sequences posted on GenBank, the percentage of properly named algae fell from ~ 90% in 1993

to less than 20% in 2010. Future work in diatom taxonomy should strive to fill these knowledge gaps in the molecular sequence banks through a similar species level molecular taxonomical approach as was presented here.

This work serves as an example of what methods future species delimitations in diatoms should consider using. The molecular technique of single cell sequencing has an important role to play in the future of diatom taxonomy and it may be able to assist in answering larger questions of diatom evolution, however it has had limited applications to date. By using effective molecular markers for a particular genus, this methodology can help to group individuals together into species complexes. As well, this technique allows for a population level molecular approach to be applied, instead of relying solely on laboratory cultures. This methodology can allow for the diatom species-level taxonomy to be delimited and better understood which could allow for greater insight into the water quality of current and past aquatic ecosystems.

References

Alverson, A. J. (2008). Molecular systematics and the diatom species. *Protist*, *159*(3), 339–53. http://doi.org/10.1016/j.protis.2008.04.001

Amato, A., Kooistra, W., Ghiron, J. H. L., Mann, D. G., Proschold, T., & Montresor, M. (2007). Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist*, *158*(2), 193–207. http://doi.org/10.1016/j.protis.2006.10.001

Antoniades, D., Hamilton, P. B., Douglas, M. S. V., & Smol, J. P. (2008). Diatoms of North America. *Iconographia Diatomologica*, *17*, 1–649.

Auinger, B. M., Pfandl, K., & Boenigk, J. (2008). Improved methodology for identification of protists and microalgae from plankton samples preserved in Lugol's iodine solution: combining microscopic analysis with single-cell PCR. *Applied and Environmental Microbiology*, *74*(8), 2505–2510. http://doi.org/10.1128/AEM.01803-07

Behnke, A., Friedl, T., Chepurnov, V. A., & Mann, D. G. (2004). Reproductive compatibility and rDNA sequence analyses in the *Sellephora pupula* species complex (Bacillariophyta). *Journal of Phycology*, *40*(1), 193–208. http://doi.org/10.1046/j.1529-8817.2004.03037.x

Bory de Saint-Vincent, J. B. G. M. (1822). Bacillarées. *Dictionnaire Classique d'Histoire Naturelle*, *2*, 127–129.

Bruder, K., & Medlin, L. K. (2007). Molecular assessment of phylogenetic relationships in selected species/genera in the naviculoid diatoms (Bacillariophyta). I. The genus *Placoneis*. *Nova Hedwigia*, *85*(3), 331–352. http://doi.org/10.1127/0029-5035/2007/0085-0331

Camburn, K. E., & Charles, D. F. (2000). *Diatoms of Low-Alkalinity Lakes in the Northeastern United States*. Philadelphia: Scientific Publications.

Cantonati, M., Lange-Bertalot, H., & Angeli, N. (2010). *Neidiomorpha* gen. nov. (Bacillariophyta): A new freshwater diatom genus separated from *Neidium* Pfitzer. *Botanical Studies*, *51*, 195–202.

Chipman, M. L., Clarke, G. H., Clegg, B. F., Gregory-Eaves, I., & Hu, F. S. (2008). A 2000 year record of climatic change at Ongoke Lake, southwest Alaska. *Journal of Paleolimnology*, *41*, 57–75. http://doi.org/10.1007/s10933-008-9257-8

Cleve, P. T. (1894). Synopsis of the Naviculoid diatoms. *Kongliga Svenska Vetenskaps-Akademiens Handlingar*, *26*(2), 1–194. http://doi.org/10.5962/bhl.title.54740

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772.

Daugbjerg, N., & Andersen, R. A. (1997). A Molecular phylogeny of the heterokont algae based on analysis of chloroplast-encoded rbcL sequence data. *Journal of Phycology*, *1041*(August), 1031–1041.

De Clerck, O., Guiry, M. D., Leliaert, F., Samyn, Y., & Verbruggen, H. (2013). Algal taxonomy: a road to nowhere? *Journal of Phycology*, *49*(2), 215–225. http://doi.org/10.1111/jpy.12020

DeSalle, R., Egan, M. G., & Siddall, M. (2005). The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1462), 1905–16. http://doi.org/10.1098/rstb.2005.1722

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*, 1969–1973.

Ehrenberg, C. G. (1843). Verbreitung und Einfluss des mikroskopischen Lebens in Sud- und Nord-Amerika. *Abhandlungen Der Königlichen Akademie Der Wissenschaften Zu Berlin 1841*, 291–446.

Ehrenberg, C. G. (1854). Mikrogeologie. In *Einundvierzig tafeln mit über viertausend grossentheils colorirten figuren Gezeicnet vom Verfasser* (p. 40 pls.). Leipzig: Leopold Voss.

Estes, A., & Dute, R. R. (1994). Valve abnormalities in Diatom clones maintained in long-term culture. *Diatom Research*, *9*(2), 249–258. http://doi.org/10.1080/0269249X.1994.9705305

Evans, K. M., Chepurnov, V. A., Sluiman, H. J., Thomas, S. J., Spears, B. M., & Mann, D. G. (2009). Highly differentiated populations of the freshwater diatom *Sellaphora capitata* suggest limited dispersal and opportunities for allopatric speciation. *Protist*, *160*(3), 386–96. http://doi.org/10.1016/j.protis.2009.02.001

Evans, K. M., Wortley, A. H., & Mann, D. G. (2007). An assessment of potential Diatom "barcode" genes (*cox1*, *rbcL*, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). *Protist*, *158*(3), 349–364. http://doi.org/10.1016/j.protis.2007.04.001

Evans, K. M., Wortley, A. H., Simpson, G. E., Chepurnov, V. A., & Mann, D. G. (2008). A molecular systematic approach to explore diversity within the *Sellaphora pupula* species complex (Bacillariophyta). *Journal of Phycology*, *44*(1), 215–231. http://doi.org/10.1111/j.1529-8817.2007.00454.x

Felsenstein, J. (1985). Confidence Limits on Phylogenies : An Approach Using the Bootstrap. *Evolution*, *39*(4), 783–791. Retrieved from http://www.jstor.org/stable/2408678 .

Gaiser, E. E., & Johansen, J. (2000). Freshwater diatoms from Carolina Bays and other isolated wetlands on the Atlantic coastal plain of South Carolina, U.S.A., with descriptions of seven taxa new to science. *Diatom Research*, *15*(1), 75 – 130. http://doi.org/10.1080/0269249X.2000.9705487

Godhe, A., Anderson, D. M., & Rehnstam-Holm, A.-S. (2002). PCR amplification of microalgal DNA for sequencing and species identification: studies on fixatives and algal growth stages. *Harmful Algae*, *1*(4), 375–382. http://doi.org/10.1016/S1568-9883(02)00049-5

Gregory-Eaves, I., Smol, J. P., Douglas, M. S. V, & Finney, B. P. (2003). Diatoms and sockeye salmon (*Oncorhynchus nerka*) population dynamics : Reconstructions of salmon-derived nutrients over the past 2200 years in two lakes from Kodiak Island, Alaska. *Journal of Paleolimnology*, *30*, 35–53.

Guillou, L., Medlin, L. K., Oce, O., & Vaulot, D. (1999). *Bolidomonas*: A new genus with two

species belonging to a new algal class, the Bolidophyceae (Heterokonta). *Journal of Phycology*, *35*, 368–381.

Guindon, S., & Gascuel, O. (2003). A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology*, *52*, 696–704.

Güttinger, W. (1989). *Collection of SEM micrographs of diatoms series 1*. Soriscio, Pura: CH-6984.

Hamilton, P. B., De Haan, M., Kopalova, K., Zidarova, R., & Van de Vijver, B. (2014). An evaluation of selected *Neidium* species from the Antarctic region. *Diatom Research*, *29*(1), 27–40.

Hamilton, P. B., & Jahn, R. (2005). Typification of *Navicula affinis* Ehrenberg: Type for the name of the genus *Neidium* Pfitzer. *Diatom Research*, *20*(2), 281–294.

Hamilton, P. B., Lefebvre, K. E., & Bull, R. D. (2015). Single cell PCR amplification of diatoms using fresh and preserved samples. *Frontiers in Microbiology*, *6*(October), 1084. http://doi.org/10.3389/fmicb.2015.01084

Hamilton, P. B., Poulin, M., & Walker, D. (1995). *Neidium hitchcockii* (Ehrenberg) Cleve a morphologically complex taxon within the genus *Neidium* (Naviculales Bacillariophyta). In J. P. Kociolek & M. J. Sullivan (Eds.), *A Century of Diatom Research in North America: a Tribute to the Distinguished Careers of Charles W. Reimer and Ruth Patrick* (pp. 66–77 pp.). Koeltz Scientific Books Campaign.

Hannen, E. J. Van, Mooij, W., Agterveld, M. P. Van, Gons, H. J., & Laanbroek, H. J. (1999). Detritus-dependent development of the microbial community in an experimental system: qualitative analysis by denaturing gradient gel electrophoresis. *Applied and Environmental Microbiology*, *65*(6), 2478.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings. Biological Sciences / The Royal Society*, *270*(1512), 313–21. http://doi.org/10.1098/rspb.2002.2218

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(41), 14812–7. http://doi.org/10.1073/pnas.0406166101

Henrichs, D. W., Renshaw, M. A., Santamaria, C. A., Richardson, B., Gold, J. R., & Campbell, L. (2008). PCR amplification of microsatellites from single cells of *Karenia brevis* preserved in Lugol's iodine solution. *Marine Biotechnology*, *10*(2), 122–7. http://doi.org/10.1007/s10126-007-9044-y

Herlory, O., Bonzom, J.-M., Gilbin, R., Frelon, S., Fayolle, S., Delmas, F., & Coste, M. (2013). Use of diatom assemblages as biomonitor of the impact of treated uranium mining effluent discharge on a stream: case study of the Ritord watershed (Center-West France). *Ecotoxicology (London, England)*, *22*(8), 1186–99. http://doi.org/10.1007/s10646-013-1106-5

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic

trees. *Bioinformatics (Oxford, England)*, *17*(8), 754–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11524383

Ivorra, N., Hettelaar, J., Tubbing, G. M. J., Kraak, M. H. S., Sabater, S., & Admiraal, W. (1999). Translocation of microbenthic algal assemblages used for in situ analysis of metal pollution in rivers. *Archives Environmental Contamination and Toxicology*, *37*, 19–28.

Jasnitsky, V. (1936). Neue und interessante Arten der Diatomeen aus dem Baikalsee. *Botanischeskii Zhurnal*, *21*(6), 689–703.

Kaczmarska, I., Mather, L., Luddington, I. A., Muise, F., & Ehrman, J. M. (2014). Cryptic diversity in a cosmopolitan diatom known as *Asterionellopsis glacialis* (Fragilariaceae): Implications for ecology, biogeography, and taxonomy. *American Journal of Botany*, *101*(2), 267–86. http://doi.org/10.3732/ajb.1300306

Kociolek, J. P., Laslandes, B., Benett, D., Thomas, E., Brady, M., & Greff, C. (2014). Diatoms of the United States 1. *Bibliotheca Diatomologica*, *1*, 1–31.

Krammer, K., & Lange-Bertalot, H. (1985). Naviculaceae neue und wenig bekannte taxa, neue kombinationen und synonym sowie bemerkungen zu einigene gattungen. *Bibliotheca Diatomologica*, *9*, 1–230.

Krammer, K., & Lange-Bertalot, H. (1986). Bacillariophyceae. 1. Teil: Naviculaceae. In H. Ettl, J. Gerloff, H. Heynig, & D. Mollenhauer (Eds.), (pp. 1–876). Stuttgart: Gustav Fischer.

Kützing, F. T. (1844). Die Kieselschaligen. In *Bacillarien oder Diatomeen* (p. 152 pp.). Nordhausen: Köhne, W.

Lagerstedt, N. G. W. (1873). Sötvattens-Diatomaceer från Spetsbergen och Beeren Eiland. *Bihang till Kongliga Svenska Vetenskaps-Akademiens Handlingar*, *1*(14), 1–52.

Lang, I., & Kaczmarska, I. (2011). A protocol for a single-cell PCR of diatoms from fixed samples: method validation using *Ditylum brightwellii* (T. West) Grunow. *Diatom Research*, *26*(1), 43–49. http://doi.org/10.1080/0269249X.2011.573703

Lange-Bertalot, H., & Genkal, S. I. (1999). Diatoms from Siberia I – Islands in the Arctic Ocean (Yugorsky-Shar Strait). *Iconographia Diatomologica*, *6*, 1–303.

Lazarus, D., & Jahn, R. (1998). Using the Ehrenberg collection. *Diatom Research*, *13*, 273–291. http://doi.org/10.1080/0269249X.1998.9705451

Lefebvre, K. E., & Hamilton, P. B. (2015). Morphology and molecular studies on large *Neidium* species (Bacillariophyta) of North America, including an examination of Ehrenberg's types. *Phytotaxa*, *220*(3), 201–223. http://doi.org/http://dx.doi.org/10.1646/phytotaxa.220.3.1

Leliaert, F., Verbruggen, H., Vanormelingen, P., Steen, F., López-Bautista, J. M., Zuccarello, G. C., & De Clerck, O. (2014). DNA-based species delimitation in algae. *European Journal of Phycology*, *49*(2), 179–196. http://doi.org/10.1080/09670262.2014.904524

Leliaert, F., Verbruggen, H., Wysor, B., & De Clerck, O. (2009). DNA taxonomy in morphologically plastic taxa: algorithmic species delimitation in the *Boodlea* complex (Chlorophyta: Cladophorales). *Molecular Phylogenetics and Evolution*, *53*(1), 122–33.

http://doi.org/10.1016/j.ympev.2009.06.004

MacGillivary, M. L., & Kaczmarska, I. (2011). Survey of the efficacy of a short fragment of the rbcL gene as a supplemental DNA barcode for diatoms. *The Journal of Eukaryotic Microbiology*, *58*(6), 529–36. http://doi.org/10.1111/j.1550-7408.2011.00585.x

Mann, D. G. (1984). Auxospore formation and development in *Neidium* (Bacillariophyta). *British Phycological Journal*, *19*(4), 319–331. http://doi.org/10.1080/00071618400650371

Mann, D. G. (1999). The species concept in diatoms. *Phycologia*, *38*, 437–495.

Mann, D. G. (2010). Discovering diatom species: is a long history of disagreements about species-level taxonomy now at an end? *Plant Ecology and Evolution*, *143*(3), 251–264. http://doi.org/10.5091/plecevo.2010.405

Mann, D. G., & Droop, S. J. M. (1996). Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia*, *336*, 19–32.

Mann, D. G., Mcdonald, S. M., Bayer, M. M., & Droop, S. J. M. (2004). The *Sellaphora pupula* species complex (Bacillariophyceae): morphometric analysis, ultrastructure and mating data provide evidence for five new species. *Phycologia*, *43*(4), 459–482.

Mann, D. G., Sato, A., Trobajo, R., Vanormelingen, P., & Souffreau, C. (2010). DNA barcoding for species identification and discovery in diatoms. *Cryptogamie, Algologie*, *31*(4), 557–577.

Mayr, E. (1942). *Systematics and the Origins of Species, from the Viewpoint of a Zoologist*. Cambridge: Harvard University Press.

McManus, G. B., & Katz, L. A. (2009). Molecular and morphological methods for identifying plankton: what makes a successful marriage? *Journal of Plankton Research*, *31*(10), 1119–1129. http://doi.org/10.1093/plankt/fbp061

Medlin, L. K., & Kaczmarska, I. (2004). Evolution of the diatoms: V. Morphological and cytological support for the major clades and a taxonomic revision. *Phycologia*, *43*(3), pp. 245–270.

Meister, F. (1912). *Die Kieselalgen der Schweiz. Beitrage zur Kryptogamenflora der Schweiz. Vol. IV, fasc. 1.* Bern: Wyss, K.J.

Metzeltin, D., & Lange-Bertalot, H. (1998). Tropical Diatoms of South America I. *Iconographia Diatomologica*, *5*, 1–695.

Metzeltin, D., & Lange-Bertalot, H. (2002). Diatoms from the “island continent” Madagascar. *Iconographia Diatomologica*, *11*, 1–286.

Metzeltin, D., & Lange-Bertalot, H. (2007). Tropical diatoms of South America II: Special remarks on the biogeographic disjunction. *Iconographia Diatomologica*, *18*, 1–877.

Metzeltin, D., & Lange-Bertalot, H. (2009). Diatoms in Mongolia: *Sichuania lacustris* spec. & gen. nov. *Iconographia Diatomologica*, *11*, 1–70.

Metzeltin, D., Lange-Bertalot, H., & Garcia-Rodriguez, F. (2005). Diatoms of Uruguay.

Iconographia Diatomologica, *15*, 1–736.

Meyer, C. P., & Paulay, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, *3*, 2229–2238. http://doi.org/10.1371/journal.pbio.0030422

Mölder, K. (1937). Einige neue Diatomeen aus Finnland und Estland. *Annales Botanici Societatis Zoologicae Botanicae Fennicae*, *8*(7), 30–32.

Moniz, M. B. J., & Kaczmarska, I. (2009). Barcoding diatoms: is there a good marker? *Molecular Ecology Resources*, *9*, 65–74. http://doi.org/10.1111/j.1755-0998.2009.02633.x

Moniz, M. B. J., & Kaczmarska, I. (2010). Barcoding of Diatoms: nuclear encoded ITS revisited. *Protist*, *161*(1), 7–34. http://doi.org/10.1016/j.protis.2009.07.001

Moritz, C., & Cicero, C. (2004). DNA barcoding: promise and pitfalls. *PLoS Biology*, *2*(10), e354. http://doi.org/10.1371/journal.pbio.0020354

Nelson, D. M., Tréguer, P., Brzezinski, M. A., Leynaert, A., & Quéguiner, B. (1995). Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycles*, *9*(3), 359. http://doi.org/10.1029/95GB01070

Patrick, R. M., & Reimer, C. W. (1966). *The diatoms of the United States exclusive of Alaska and Hawaii. Volume 1. Fragilariaceae, Eunotoniaceae, Achnanthaceae, Naviculaceae. Monographs of the Academy of Natural Sciences of Philadelphia., Vol. 13*.

Pfitzer, E. (1871). Untersuchungen über Bau und Entwicklung der Bacillariaceen (Diatomaceen). In J. Hanstein (Ed.), *Botanische Abhandlungen aus dem Gebiet der Morphologie und Physiologie* (p. 189). Bonn.

Potapova, M. G., & Charles, D. F. (2002). Benthic diatoms in USA rivers: Distributions along spatial and environmental gradients. *Journal of Biogeography*, *29*, 167–187. http://doi.org/10.1046/j.1365-2699.2002.00668.x

Quijano-Scheggia, S. I., Garcés, E., Lundholm, N., Moestrup, Ø., Andree, K., & Camp, J. (2009). Morphology, physiology, molecular phylogeny and sexual compatibility of the cryptic *Pseudo-nitzschia delicatissima* complex (Bacillariophyta), including the description of *P. arenysensis* sp. nov. *Phycologia*, *48*(6), 492–509. http://doi.org/10.2216/08-21.1

Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). Tracer v1.6, available from http://beast.bio.ed.ac.uk/Tracer.

Richlen, M. L., & Barber, P. H. (2005). A technique for the rapid extraction of microalgal DNA from single live and preserved cells. *Molecular Ecology Notes*, *5*(3), 688–691. http://doi.org/10.1111/j.1471-8286.2005.01032.x

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572–1574. http://doi.org/10.1093/bioinformatics/btg180

Round, F. E., Crawford, R. M., & Mann, D. G. (1990). *The Diatoms- Biology & Morphology of the Genera*. Cambridge: Cambridge University Press.

Ruck, E. C., Nakov, T., Jansen, R. K., Theriot, E. C., & Alverson, A. J. (2014). Serial gene losses and foreign DNA underlie size and sequence variation in the plastid genomes of diatoms. *Genome Biology and Evolution*, *6*(3), 644–54. http://doi.org/10.1093/gbe/evu039

Ruck, E. C., & Theriot, E. C. (2011). Origin and evolution of the canal raphe system in diatoms. *Protist*, *162*(5), 723–37. http://doi.org/10.1016/j.protis.2011.02.003

Saunders, G. W. (2005). Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1462), 1879–88. http://doi.org/10.1098/rstb.2005.1719

Savolainen, V., Cowan, R. S., Vogler, A. P., Roderick, G. K., & Lane, R. (2005). Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1462), 1805–11. http://doi.org/10.1098/rstb.2005.1730

Schmidt, A. (1877). *Atlas der Diatomaceen-kunde*. Aschersleben: Verlag Von Ludwig Siever's Buchandlung. http://doi.org/10.5962/bhl.title.64396

Schroeder, H. (1939). Die algenflora der Mulde. *Pflanzenforschung*, *21*, 1–88.

Siver, P. A., & Hamilton, P. B. (2011). *Iconographia Diatomologica. Vol. 22. Diatoms of North America*. (H. Lange-Bertalot, Ed.). Ruggell: A.R.G. Gantner Verlag K.G.

Siver, P. A., Hamilton, P. B., Stachura-Suchoples, K., & Kociolek, J. P. (2003). Morphological Observations of *Neidium* species with sagittate apices, including the description of *N. capecodii* sp. nov. *Diatom Research*, *18*(1), 131–148.

Siver, P. A., Hamilton, P. B., Stachura-Suchoples, K., & Kociolek, J. P. (2005). Diatoms of North America: The freshwater flora of Cape Cod, Massachusetts, U.S.A. *Iconographia Diatomologica*, *14*, 1–463.

Smith, W. (1853). *Synopsis of the British Diatomaceae, with remarks on their structure, function and distribution, and instructions for collecting and preserving specimens. Vol. 1*. London: John van Voorst.

Souffreau, C., Vanormelingen, P., Van de Vijver, B., Isheva, T., Verleyen, E., Sabbe, K., & Vyverman, W. (2013). Molecular evidence for distinct Antarctic lineages in the cosmopolitan terrestrial Diatoms *Pinnularia borealis* and *Hantzschia amphioxys*. *Protist*, *164*(1), 101–115. http://doi.org/10.1016/j.protis.2012.04.001

Souffreau, C., Verbruggen, H., Wolfe, A. P., Vanormelingen, P., Siver, P. A., Cox, E. J., Mann, D.G., Van de Vijiver, B., Sabbe, K., Vyverman, W. (2011). A time-calibrated multi-gene phylogeny of the diatom genus *Pinnularia*. *Molecular Phylogenetics and Evolution*, *61*(3), 866–79. http://doi.org/10.1016/j.ympev.2011.08.031

Stepanek, J. G., & Kociolek, J. P. (2014). Molecular phylogeny of *Amphora* sensu lato (Bacillariophyta): an investigation into the monophyly and classification of the amphoroid diatoms. *Protist*, *165*(2), 177–95. http://doi.org/10.1016/j.protis.2014.02.002

Stoermer, E. F., & Smol, J. P. (1999). *The Diatoms. Applications for the Environmental and*

Earth Sciences. Cambridge: Cambridge University Press.

Stosch, H. A. von. (1975). An amended terminology of the diatom girdle. *Nova Hedwigia*, *53*, 1–35.

Summerbell, R. C., Lévesque, C. A., Seifert, K. A., Bovers, M., Fell, J. W., Diaz, M. R., Boekhout, T., de Hoog, G.S., Stalpers, J., Crous, P. W. (2005). Microcoding: the second step in DNA barcoding. *Philosophical Transactions of the Royal Society of Londoncal Biological Sciences*, *360*, 1897–1903. http://doi.org/10.1098/rstb.2005.1721

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. In *Some mathematical questions in biology- DNA sequence analysis* (pp. 57–86). Providence, RI: American Mathematical Society.

Theriot, E. C., Ashworth, M., Ruck, E., Nakov, T., & Jansen, R. K. (2010). A preliminary multigene phylogeny of the diatoms (Bacillariophyta): challenges for future research. *Plant Ecology and Evolution*, *143*(3), 278–296. http://doi.org/10.5091/plecevo.2010.418

Trobajo, R., Mann, D. G., Clavero, E., Evans, K. M., & Mcgregor, R. C. (2010). The use of partial cox1 , rbcL and LSU rDNA sequences for phylogenetics and species identification within the *Nitzschia palea* species complex (Bacillariophyceae) The use of partial cox1, rbcL, and LSU rDNA sequences for phylogenetics and species identification. *European Journal of Phycology*, *45*(4), 413–425. http://doi.org/10.1080/09670262.2010.498586

Urbánková, P., & Veselá, J. (2013). DNA-barcoding: a case study in the diatom genus *Frustulia* (Bacillariophyceae). *Nova Hedwigia, Beiheft*, *142*(September), 147–162.

Vanormelingen, P., Chepurnov, V. A., Mann, D. G., Sabbe, K., & Vyverman, W. (2008). Genetic divergence and reproductive barriers among morphologically heterogeneous sympatric clones of *Eunotia bilunaris* sensu lato (Bacillariophyta). *Protist*, *159*(1), 73–90. http://doi.org/10.1016/j.protis.2007.08.004

Werum, M., & Lange-Bertalot, H. (2004). Diatoms in Springs. *Iconographia Diatomologica*, *13*, 1–480.

Will, K. W., & Rubinoff, D. (2004). Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics*, *20*(1), 47–55. http://doi.org/10.1111/j.1096-0031.2003.00008.x

Witkowski, A., Barka, F., Mann, D. G., Li, C., Weisenborn, J. L. F., Ashworth, M. P., Kurzydlowski, J. K., Zglobikca, I., Dobosz, S. (2014). A description of *Biremis panamae* sp. nov., a new Diatom species from the marine littoral, with an account of the phylogenetic position of *Biremis* D. G. Mann. *PLoS Biology*, *9*(12), 1–24. http://doi.org/10.1371/journal.pone.0114508

Witt, J. D. S., Threloff, D. L., & Hebert, P. D. N. (2006). DNA-barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. *Molecular Ecology*, *15*, 3073–3082. http://doi.org/10.111/j.1365-294X.006.02999.x

Wu, J. (1999). A generic index of diatom assemblages as bioindicator of pollution in the Keelung River of Taiwan. *Hydrobiologia*, *397*, 79–87.

Zimmermann, J., Jahn, R., & Gemeinholzer, B. (2011). Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity & Evolution*, *11*(3), 173–192. http://doi.org/10.1007/s13127-011-0050-6

Appendix A: Single cell PCR amplification of diatoms using fresh and preserved samples

Paul B. Hamilton[1*], **Keely Lefebvre[2]**, Roger D. Bull[1]

[1]Research and Collections, Canadian Museum of Nature

[2]Ottawa-Carleton Institute of Biology, University of Ottawa

This appendix was published as:

Hamilton, P.B., Lefebvre, K.E., Bull, R.D. 2015. Single cell PCR amplification of diatoms using fresh and preserved samples. *Front. Microbiol*. 6:1084.

DOI: 10.3389/fmicb.2015.01084

Abstract

Single cell Chelex® DNA extraction and nested PCR amplification were used to examine partial gene sequences from natural diatom populations for taxonomic and phylogenetic studies at and above the level of species. DNA was extracted from cells that were either fresh collected or stored in RNA*later*. Extractions from Lugol's fixation were also attempted with limited success. Three partial gene sequences (*rbcL*, 18S and *psbA*) were recovered using existing and new primers with a nested or double nested PCR approach with amplification and success rates between 70–96%. An *rbcL* consensus tree grouped morphologically similar specimens and was consistent across the two primary sample treatments: fresh and RNA*later*. This tool will greatly enhance the number of microscopic diatom taxa (and potentially other microbes) available for barcoding and phylogenetic studies. The near-term increase in sequence data for diatoms generated via routine single cell extractions and PCR will act as a multiproxy validation of longer-term next generation genomics.

Introduction

DNA barcoding has become common practice in animal and plant taxonomy (Hebert *et al*., 2003) with cytochrome c oxidase 1 (*CO1*), a mitochondrial gene, serving as the main animal barcoding gene (Hebert *et al.,* 2004). In plants the chloroplast genes ribulose 1,5-biphosphate (*rbcL*) and maturase K (*MatK*) serve as two of the preferred barcode genes for taxonomic identifications (CBOL working group 2009). This is in contrast with the situation faced in diatom barcoding where several regions are presently identified as prominent taxonomic markers (e.g., Yoon *et al.,* 2002; Evans *et al*., 2007). In some studies the conservative *rbcL*, *CO1*, and ribosomal complex gene 18S are considered to be good taxonomic characters for species determinations (Mann *et al*., 2010; Hamsher *et al*., 2011; Zimmermann *et al*., 2011). Ribosomal complex genes ITS, 18S and 28S were also used both individually and in multi-gene studies to evaluate cryptic taxonomic variations at the genus and species levels (e.g. Amato *et al*., 2007; Vanormelingen *et al*., 2008; Poulickova *et al*., 2010; Kaczmarska *et al*., 2014). The smaller (~500 base pair) chloroplast *psbA* gene was used in evolutionary studies but is not variable enough to be informative for species-level taxonomic studies (Souffreau *et al*., 2011, pers. obs.). In contrast, chloroplast gene *psbC* (>1000 base pair) is more widely used across all the major orders (Theriot *et al*., 2010). With little consensus as to which marker best delimits diatom groups, the ability to amplify several genes including new genes from a single cell is essential for diatom taxonomy using DNA barcodes.

In microbial studies, genetic sequencing has been successful across all the primary algal families Bacillariophyceae, Cyanophyceae, Chlorophyceae, Chrysophyceae, Cryptophyceae, Desmidiaceae, Euglenophyceae, Haptophyceae, Pyrrhophyceae, Raphidophyceae, Synurophyceae (e.g. Daugbjerg and Anderson, 1997; McCourt *et al*., 2000; Tomitani *et al*.,

2006; Edvardsen *et al*., 2011; Bennett *et al*., 2014). However, compared to DNA research in Plantae and Animalia, there are much fewer sequences available for diatom taxa, leaving large taxonomic holes in DNA databases. As well, microbial genetic studies in algae are limited by the ability to collect enough material, and population genetic studies are all but absent. Culturing algae and other microbes to accumulate sufficient DNA has been the time-limiting step in microbial genetics. Cultures supply extra material for morphological identification and validation; however cultures are prone to alterations in structural morphology (Trainor *et al*., 1971; Estes and Dute 1994).

Single cell extraction and PCR protocols have been advanced in microbe research on live and fixed materials, although no single approach meets all aspects of routine enhanced multi-gene taxonomic research (e.g. Sherbakova *et al*., 2000; Ruiz Sebastián and O'Ryan 2001; Lang and Kaczmarska, 2011). To date single cell extractions have been successfully completed with the Chrysophyceae (Auinger *et al*., 2008) Pyrrhophyceae (Richlen and Barber, 2005; Henrichs *et al*., 2007) and Bacillariophyceae (Lang and Kaczmarska, 2011). There are only a few published examples using non-cultured single-cell amplifications (e.g. Auinger *et al*., 2008; Godhe *et al*., 2002; Lang and Kaczmarska, 2011). These amplifications are successful if there is sufficient initial DNA within the cell and the primer set is effective at maximizing amplification efficiency. In most cases, amplification of DNA for sequencing requires a nested amplification protocol. This nested approach is effective but has the potential to generate amplification errors (Ruck *et al*., 2014). In order to effectively utilize this nested approach, clear protocols for error checking and sequence validation must be established. Single cell genetic determinations are presently a novel and potentially efficient way to generate a large reference library of taxonomically informative data from single algal cells in complex environmental systems. This reference

library will also contain an extensive database for population genetics and genetic biogeography studies (e.g. Alverson *et al*., 2007).

There are a number of reagents which can be used in single cell DNA amplifications with algae (e.g. Auinger *et al*., 2008; Bertozzini *et al*., 2005). Chelex® resin is an effective DNA extraction tool with applications in molecular biology ranging from multicellular vertebrate tissues to single microbial cells (Richlen and Barber, 2005; Hahn *et al*., 2000). This extraction method was used in genetic investigations in forensic science (Legrand *et al*., 2002), population studies (Richlen and Barber 2005), and evolution (Theriot *et al*., 2010). In mammalian research Chelex extractions have been used on fresh, frozen, alcohol preserved, and to a limited degree on formalin-fixed tissues (Legrand *et al*., 2002). The simplicity of this technique coupled with relatively cheap cost allows for quick PCR assays (e.g. Bowers *et al*., 2000; Reyes-Escogido *et al*., 2010) which has the potential for more detailed taxonomic barcoding initiatives and finer population genetic studies.

The recovery of gene sequences from structurally fixed and preserved material has had some success across all the biological groups, but has not developed into a routine protocol used for sequencing (Connell, 2002; Godhe *et al*., 2002; Henrichs *et al*., 2007). Institutions with collections of fixed biological materials, like museums, are extremely interested in the recovery of genes and genomes from their historic collections. The temporal record hidden in fixed biological collections—with regard to speciation and population genetics—is waiting to be mined. In microbial genetics, gene sequences have been recovered from ethanol, Lugol's solution, buffered formalin and RNA*later* fixations (Ambion, 1999; Connell, 2002; Godhe *et al*., 2002; Auinger *et al*., 2008; this study). The general recipe for success is removal or dilution of the traditional fixation solution followed by standard sample processing, DNA extraction, and

gene sequencing. Buffered formalin preserved materials can be treated with cold methanol to minimize the impact of the fixation, while sodium thiosulfate is effective in capturing iodine from solutions and biological materials (Godhe *et al*., 2002; Auinger *et al*., 2008). In the Protista, ethanol preparation and fixation prior to sequencing is less common, although has been successful with some limitations (e.g. Lang and Kaczmarska, 2011; Godhe *et al*., 2002; Henrichs *et al*., 2007; Ivanova *et al*., 2014). RNA*later* is a more recent stabilizer suitable for the storage of material prior to RNA and DNA sequencing with a high rate of recovery (Ambion, 1999). RNA*later* has the advantage of being a good fixative and further prevents the degradation of RNA and DNA. At 4°C viable fixation can be maintained for a month and at -20°C samples can be maintained indefinitely (Ambion, 1999).

The objective of this study was to evaluate a nested amplification protocol for multiple genes in diatoms from single cells under live and RNA*later* preserved conditions. This will establish a standard multiproxy routine for reproducible barcoding with morphological analysis of natural populations. In this study new primer pairs for nested amplification of *rbcL*, 18S and *psbA* were designed and different DNA polymerases and cycling protocols were compared. Finally, examples are presented for how this protocol can be used to establish a more comprehensive reference library of taxonomic and physiological genetic data.

Materials and methods

Collection of Samples

All samples were benthic or planktonic and collected from freshwaters with a wide range in pH (5.1 – 7.8) and eutrophication states (oligotrophic–mesotrophic) (Tables A.1, A.2). The samples

were kept cool in transport and at room temperature in the lab in natural light prior to single cell isolations.

RNAlater Preservation

A 0.2 mL volume of fresh benthic sample was aliquoted into a 1.5 mL tube containing 1 mL of RNA*later* tissue storage buffer (http://www.protocol-online.org/prot/Protocols/RNAlater-3999.html). The sample was hand shaken to mix, kept at room temperature for 24 hours, and then stored in the dark at 4°C between 5 to 21 days before single cell isolation.

Lugol's fixation

Following the protocol from Henrichs *et al*. (2007), 0.2 mL of living benthic sample was aliquoted into a 1.5 mL tube and then fixed with Lugol's iodine solution (10 g I_2, 20 g of KI, 200 mL ddH_2O). In this study non-acidified Lugol's fixation (no glacial acetic acid) was used while Henrichs *et al*. (2007) used acidified Lugol's fixation. The merits of using non-acidified versus acidified Lugol's fixation, was discussed in Throndsen (1978). Samples fixed with non-acidified Lugol's iodine solution varied in storage duration from 12 days to 20 years. Just prior to isolation, 20 µL of 1 M sodium thiosulfate was added to the samples and hand mixed until the Lugol's iodine solution was dissipated (became colourless).

Single diatom cell isolation

1–2 mL of each sample (living, RNA*later* preserved or Lugol's preserved) was placed on a large microscope slide. The sample was then diluted with ~1 mL of sterile, nuclease free water (Bioshop Canada Inc.) and examined under an inverted microscope (Leica) or a compound microscope (Nikon, with long working distance objectives) at 10x magnification. Individual cells were isolated through suction using 20–40 µl drawn-out disposable pipets, either with a

Narishige micromanipulator or simple manual suction. This isolation procedure was modified from Throndsen (1978) by removing the use of Formvar film. The isolated cell with associated contaminants was transferred to a new water droplet of DNA nuclease free water. This isolation and transfer was repeated 2–5 times to remove any contaminants and/or preservative residue. Individual cells were then isolated for the final time and transferred to a 0.2 mL PCR tube containing 200 µL of 10% (w/v) Chelex® 100 solution (Richlen and Barber, 2005). The samples were stored from 1 to 51 days at 4°C in the dark until DNA extraction (Tables A.1, A.2; Fig. A.Supplement).

DNA extraction, PCR and sequencing protocol

For DNA extraction, Chelex-stored samples were incubated for 20 minutes at 95°C. They were then vortexed for 15 s and centrifuged for 15 s at 14,000 rcf. We also tested extraction without the incubation step on some samples, and had similar success with both protocols. For PCR study, the following primers were used (Table A.3).

The first amplifications were performed in a 25 µL volume with a final concentration of 1x PCR buffer (Bioshop Canada Inc.), 2 mmol L^{-1} $MgCl_2$, 0.3 mmol L^{-1} dNTP, 0.4 µmol L^{-1} of each primer, 1 unit of *Taq* DNA polymerase (BioShop), and 5.0 µl of Chelex DNA extract supernatant. The following cycling conditions were used: 94°C for 210 s; followed by 36 cycles of 94°C for 50 s, 52°C for 50 s, and 72°C for 80 s; and then a final elongation step at 72°C for 15 min. For the second amplification (and the third amplification step when performed), all steps and concentrations were the same as above except that 1 µl of the product from the previous amplification was used as template. The success of the PCR was assessed by visualizing the products on a 1.5% agarose gel. Successful PCR products were purified using the enzymes Exonuclease I and shrimp alkaline phosphatase (USB Corporation). Big Dye version 3.1 (Life

Technologies Corporation) was used for sequencing reactions using 0.6 µL of Big Dye in a 10 µL reaction. Sequencing reaction products were purified via ethanol-EDTA-sodium acetate precipitation. Nucleotide sequences were generated using automated cycle-sequencing on an Applied Biosystems 3130xl automated sequencer. To validate the use of Bioshop Taq DNA polymerase, seven of the samples were re-amplified for all three genes using Phusion® High-Fidelity DNA Polymerase (New England Biolabs). The optimized annealing temperatures used in the first and second Phusion amplifications were as follows: *rbcL* 57.7°C and 60.3°C; 18S 52.2°C and 60.3°C; *psbA* 52.2°C and 63°C. Further, to ensure that the error rate was not affecting our sequences, seven samples were re-amplified for *rbcL* using Phusion with 20, 25, and 30 cycles in the second amplification. The final PCR products were then sequenced and compared for base pair differences with the sequences obtained with Bioshop Taq using the standard 34 cycle second amplification step.

Seventy-two *rbcL* sequences were studied from a wide range of diatom taxa. Within the pennate raphe bearing diatoms, 35 diatom cells were sequenced for all three genes to ensure that the method would allow for multi-gene sequencing. Sequences were assembled and edited in Geneious version 6.1.5 and consensus sequences were aligned using the MAFFT alignment tool. Consensus sequences were compared to the GenBank database using the Basic Local Alignment Search Tool (BLAST) to verify and ensure that no contaminants were sequenced. Initial Maximum Parsimony (MP) tree topologies of each gene were assessed in PAUP v.4.0 (Swofford, 2003), and phylogenetic model testing (using likelihood scores and AIC calculations) of each region was analyzed in jModel Test v.2.1.4 (Darriba *et al*., 2012; Guindon and Gascuel, 2003) to ensure that the data could be concatenated for analysis. Datasets had the preferred General Time Reversible model (GTR+I+G) (Tavaré, 1986) except the 18S dataset which had

the Transitional Model (TIM3+I+G) (Posada, 2003). However, the initial topology tree of the 18S matched both the *rbcL* and *psbA* initial topology trees, and, using a chi-squared distribution, the delta values from TIM3+I+G and GTR+I+G (delta = 0.0000, K=78; delta = 3.9436, K=80, respectively) were not shown to be statistically different (P<0.15). Therefore the model GTR+I+G was used for Bayesian analysis (BI) and Maximum Likelihood (ML) for both the concatenated data set (*rbcL*, 18S, *psbA*) and the single gene dataset (*rbcL*). The BI was carried out with MrBayes v.3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003), with a Monte Carlo Markov Chain (MCMC) run for 1 million generations for the concatenated gene data set and 5 million generations for the *rbcL* dataset with the default settings. Runs were sampled every 1000th generation. The first 250,000 and 1,250,000 were discarded as burn-in for the concatenated gene dataset and *rbcL* dataset, respectively. The convergence and stationarity of the BI results were analyzed in Tracer v1.6 (Rambaut *et al*., 2013) and topology convergences were analyzed in AWTY (Wilgenbusch *et al*., 2004). ML Bootstrap analysis (Felsenstein, 1985) was done in Garli v.2.01 (Zwickl, 2006), using the GTR+I+G model, with 1,000 bootstrap replicates for both the concatenated dataset and *rbcL* data set. *Fragilaria bidens* (GenBank Acquisition AB430716.1) was used as the outgroup for our concatenated dataset as it was a close sister species to the taxa used in the concatenated analysis. *Bolidomonas pacifica* (GenBank Acquisition HQ912421.1) and *Cyclotella meneghiniana* (GenBank Acquisition KF959651.1) were used as outgroup and sister taxa, respectively, for the *rbcL* dataset as they were the closest relatives to the dataset taxa available on GenBank.

Results

Partial sequences for *rbcL* (1,202-1,305 bp), 18S (811-1,144 bp), and *psbA* (537-578 bp) were determined for 35 single cell freshwater diatom isolates. An additional 37 partial *rbcL* sequences were determined for a variety of diatom genera including *Melosira* C.Ag., *Aulacoseira* Thwaites*, Synedra* (*Ulnaria)* Ehrenberg, *Eunotia* Ehrenberg, *Navicula* Bory*, Neidium* Pfitzer, *Placoneis* C.Mereschkowsky, *Frustulia* C.Ag., *Gyrosigma* Hassall, *Stauroneis* Ehrenberg, *Craticula* Grunow, *Sellaphora* C.Mereschkowsky*, Pinnularia* Ehrenberg, *Cymatopleura* W.Sm., *Encyonema* Kütz., *Gomphonema* Ehrenberg, *Nitzschia* Hassall, *Hantzschia* Grunow *and Surirella* Turpin.

Of the diatoms sequenced, 60 were from fresh living samples, 12 were from RNA*later* fixed samples, and one from a Lugol's fixed sample (Tables A.1 and A.2, A.Supplement). NCBI Blast searches using the new sequences resulted in matches consistent with the genus-level morphological identifications of our specimens. Method validation of the number of cycles (20, 25, 30, and 34) and type of DNA polymerase using 5 different taxa showed only one instance of base pair substitutions, though there we no differences in the overall sequence alignments. The recovered sequence lengths for *rbcL* and 18S were both within the average range for the diatom sequences found on GenBank (Table A.4). The length of *psbA* sequence recovered was slightly below the gene length found for diatom sequences on Genbank (Table A.4). The amplification success was 70%, 90%, and 96% for *rbcL*, 18S, and *psbA*, respectively (Table A.5). The recovery success of 18S and *psbA* was higher than *rbcL* because only samples that amplified successfully for *rbcL* were processed for these two regions. We had very low amplification success with the Lugol's fixed samples (Table A.5). In addition, within the Lugol's fixed

samples 13 contained fungi 18S nuclear DNA. RNA*later* amplification success was consistent across sample storage periods ranging from 5 to 21 days (Table A.5).

Multiple gene analysis

Individual topologies of the three genes (*rbcL*, 18S, *psbA*) showed no differences, neither did the ML (-LnL=10058.4412) nor BI analyses, thus only the BI tree was shown with both the BI posterior probabilities (PP) and the ML bootstrap values (BS) (Figure A.1). Our dataset showed significant separation at the family level for the following: Pinnulariaceae (PP = 100, BS = 100), Sellaphoraceae (100, 74), Stauroneidaceae (100, 85), Pleurosigmataceae (100, 100), Naviculaceae (100, 100). RNA*later* preserved and fresh samples of the same taxa were found within the correct clades. Examples of this can be seen in the genera *Craticula*, *Gyrosigma* and *Pinnularia* (Figure A.1). In the genus *Gyrosigma,* a fresh sample and an RNA*later* sample were significantly similar (100, 87), and came out on the same terminal branch (Figure A.1, stars). The small branching of individual taxa within this genera were due to ≤5 bp differences. Although difficult to determine, low number of base pair differences could be either base pair substitution error or intragenetic variation (0.001%) between the concatenated sequences (>3000 bp). Slight branching between individuals was also present in the genera of *Craticula* (≤ 3 bp differences, 0.001%) and *Pinnularia* (≤ 14 bp differences, 0.004%), showing low levels of intragenetic variation (Figure A.1). Individual cells were principally collected from benthic sediments, leading to the larger representation of Naviculaceae taxa.

Single gene analysis

For the single gene dataset using only *rbcL*, neither the ML (-lnL=9690.8210) nor the BI trees showed any differences, thus only the BI tree was shown with both the BI PP and the ML BS values (Figure A.2). All isolates from the same genus showed strongly supported

monophyletic taxa. In particular, the genera *Craticula, Pinnularia, Neidium, Frustulia, Cymatopleura, Surirella, Gyrosigma, Melosira* and *Aulocoseira* all had very high support values (PP=100; BS=100), while the genera *Stauroneis* (100, 65) and *Navicula* (100, 85) had supported monophyletic taxa groups. Cells of the same taxon, collected from the same location (+/- 10 m area) were also more closely aligned in the tree compared to similar cell isolates of the same taxon from other locales (Figure 2, red arrows). Specimens from the same genus which were isolated from either fresh, RNA*later* preserved or in one case Lugol's preserved samples were always in the same monophyletic group. The *Gyrosigma* specimen which was isolated from iodine fixation was placed with all other *Gyrosigma* isolates (PP=100, BS=100). As well, for both *Gyrosigma* and *Pinnularia*, fresh and RNA*later* preserved isolates were on the same terminal branch (Figure 2, black arrows).

Specimen and sample fixation

Diatom specimens or populations for morphological study in association with DNA were cleaned and mounted for light microscopy (LM) and scanning electron microscopy (SEM) validation. A population of *Gyrosigma acuminatum* (Kütz.) Rabenh. for example (n=30) illustrated a natural size diminution series (Length: 106.5 – 163.5; width: 18.5 – 27.5; stria density: 17 – 19 μm; areola density: 17 – 18 in 10 μm, Fig. A.3,a-e). Additional specimens fixed in RNA*later* and frozen (-17 °C) for 2 days maintained frustule integrity with cytoplasm and chloroplast structure (Fig. 3f). Chloroplasts were intact however alterations in the structure were observed; in some specimens there was slight shrinkage apically and transapically, while others had poorly defined chloroplast walls. Cyanophyceae (e.g. *Phormidium* sp., *Oscillatoria* cf. *princeps* Vaucher ex Gomont) and Chlorophyceae (e.g. *Oedigonium* sp., *Pediastrum boryanum*

(Turp.) Menegh.) cells also maintained cell structure. The Chlorophyceae had intact but sometimes slightly altered chloroplasts (pers. obs.)

Discussion

The development of novel DNA extraction protocols has accelerated the exploitation of microbial genetic studies in health (e.g. Richlen and Barber, 2005), environment (e.g. Neilan, 1995; Kermarrec *et al*., 2013), and even diatom taxonomic research (e.g. Evans *et al*., 2007; Poulickova *et al*., 2010). Nested DNA amplifications have the potential to open the genetic vault of taxonomic information from single microbial cells. The simple methodology of single cell Chelex® DNA extraction followed by nested PCR has great implications for expanding the genetic reference library of information in algal research. This study uses diatoms as test organisms; preliminary PCR success using dinoflagellates (pers. obs.) is also recorded.
The 70 – 96% amplification success rates using live and (31-100%) using RNA*later* fixed samples for single cell PCRs is similar to recovery rates using cultures (Lang and Kaczmarska, 2011). DNA polymerases of different quality and price points (BioShop Taq® and Phusion®) produced (93%) the same sequence results. Comparisons of sequences using NCBI BLAST also supported the morphological taxa identifications of the specimens. In this study, the systematic associations of *Gyrosigma* to *Navicula*, *Craticula* to *Stauroneis*, and *Pinnularia* to *Sellaphora* (Figures A.1 and A.2) were in agreement with other studies (Theriot *et al*., 2010; Evans *et al*., 2007). This success rate increases the utility of conserved genes known to be good for species level taxonomic discriminations (Hamsher *et al*., 2011).

In this study, one existing nested primer set for *rbcL* was used while two new primer pairs were developed for the nested amplification of 18S and *psbA* in diatoms. The recovered sequence lengths for *rbcL* and 18S were within the average range of the lengths of diatom

sequences found on GenBank while *psbA* sequence lengths were slightly below the average gene length found for diatom sequences (Table A.4). One single cell DNA extraction provided enough amplification template for multiple DNA amplifications making the approach compatible for robust multi-gene analyses. In this study, ten amplifications from a single cell extract were successfully performed. Based upon the replicated amplifications done, a conservative estimate suggests there could be enough template for 30 – 35 PCR amplifications per extraction. There is even potential that nested amplifications can be used to generate large-scale genetic datasets using next generation sequencing protocols optimized for low concentration DNA templates (e.g., initial amplification using the Multiple Displacement Amplification reaction, Lasken, 2007; but see Ning *et al*., 2014 for a review of current challenges). Some research suggests that the Chelex extraction method, which also conserves proteins, may not be the most suitable for Multiple Displacement Amplification reactions. However, one study found that keeping proteins was not an issue for whole gene amplification (Lepere *et al*., 2011).

One concern related to using this technique is the potential for contamination during isolation of a single cell. Often diatom cells, like taxa in the genera *Asterionella* Hassall, *Tabellaria* Ehrenberg and *Surirella*, have epiphytes which may contaminate amplifications. In addition, single cells often have organic extracellular polymeric substances (EPS) which capture bacteria, fungi and loose organics with remnant DNA (Das *et al*., 2013). Problems related to the amplification of DNA from non-target sources were observed in this study. In the Lugol's fixed samples, 18S nuclear DNA recovery from fungi was amplified in 13 samples. In this instance, the contamination problem was identified by BLASTing the sequencing product. Contamination from non-algae sources can be easily identified using this protocol and removed from study. Since contamination was only observed in selected Lugol's samples, we can conclude that with

good isolations, fungi can be effectively removed from sample concerns using fresh and freshly fixed material. Contamination from more genetically associated algal sources maybe more problematic given the potential for cross-contamination during amplifications (Zhang *et al*., 1992; Ruck *et al*., 2014) and a limited genetic reference library available for comparison. However, in this study there was no evidence of contamination.

Time consuming recovery of DNA from micro-organism cultures has limited the success of barcoding in the Protista. Ruck and Theriot (2011) developed a single cell diatom field isolation and *rbcL* DNA extraction procedure using Chelex. This approach is effective, although limited by time requirements for isolating cells in the field and no ability to reinvestigate the samples collected. Collecting live and fixed samples in the field for subsequent isolation back in the lab gives greater success in DNA recovery. Using this single cell extraction protocol alone, or with a limited number of replicated daughter cells (short-term cultures), will greatly increase the database of DNA sequences for diatoms and microbes.

In cell sequencing, there are inherent problems with the destruction of the voucher specimen (A.Supplement). Although reported that diatom valves can be recovered after DNA extraction in ethanol (Lang and Kaczmarska, 2011), the glass beads in the Chelex solution destroy diatom valves after centrifugation (pers. obs.). To limit the possibility of erroneous specimen identifications, many photomicrographs were taken of each single cell prior to DNA extraction. These were linked to morphological vouchers (cleaned diatom valves) collected from the same respective population. For example, diatom specimens of *G. acuminatum* were matched using DNA results and recoverable specimens from the natural population (Figure A.3, A.Supplement). To further reduce identification errors, replication of DNA sequencing results with comparison to more morphological specimens from a population can improve the validation

of species identifications in association with both genetic and morphological variability. In the case of *G*. *acuminatum*, these gene sequences can be directly related to a morphological study of the species type (Sterrenberg, 1995). However, this approach can be highly problematic when taxa within a community have overlapping morphological characteristics. In these cases, additional detailed multi-gene studies of populations could enhance the resolution and identification of cryptic species (e.g. Poulickova *et al*. 2010).

Traditional fixation of microbial samples for morphological, ecological and physiological study has a long history (Simmons, 2014; Throndsen, 1978). Recent studies have demonstrated that gene sequences can be recovered from a variety of traditionally fixed samples (Connell, 2002; Godhe *et al.,* 2002; Henrichs *et al*., 2007; Lang and Kaczmarska, 2011). Ethanol fixed samples are not commonly found in microbial museum collections because they are subject to evaporation and have negative extraction effects on cell pigments. In the current study, recovery of gene product from Lugol's fixed samples was poor 7.5% and 68% for *rbcL* and 18S genes respectively. With these product recoveries sequence success was reduced further with many isolates have low quality sequence results which were rejected. Only one *psbC* amplification and sequence determination was attempted and successful. No success was observed from formalin-fixed samples, although we did not immediately transfer formalin-fixed material into methanol storage as described by Godhe *et al*. (2002). Although Bertozzini *et al*. (2005), and Auinger *et al*. (2008) have successfully recovered DNA from dinoflagellates and chrysophyte fixed with Lugol's solution, we need more detailed methodological studies to improve percent success in the routine recover of DNA from diatoms in Lugol's and formalin fixation. However, Godhe *et al*. (2002) suggest that Lugol's solution and varying ethanol fixations have other shortcomings. In this study we also noted the presence of fungi in museum collections fixed with Lugol's, a

potential problem with historically fixed collections. At present RNA*later* represents an excellent genetic fixation protocol for sample collection, short term storage and long-term archiving of microbial collections. Cell wall structure in Chlorophyceae (*Oedigonium* sp., *Pediastrum boryanum*), Cyanophyceae (*Phormidium* sp., *Oscillatoria* cf. *princeps*) and diatoms were maintained in RNA*later* fixed samples under cold and frozen conditions. Chloroplast integrity was even maintained for *G. acuminatum* during freezing at -17°C (Figure A.3, f). The specifications for RNA*later* indicate that treated tissues can be stored at 25°C for one week, at 4°C for one month, or at -20°C indefinitely (Ambion 1999). In this study, both fresh and RNAlater fixed samples had predictable extraction success (Table A.5). This supports the adoption of RNAlater as a long-term diatom storage media.

The recovery of DNA from archived museum and research collections is currently poor but quickly advancing, especially with vertebrate collections (e.g. Payne and Sorenson, 2002). However, museums and large collections should prioritize the implementation of storage and fixation techniques that maintain the molecular integrity of the samples. RNA*later* preserved algae, including diatoms, subjected to freeze-thaw cycles showed some internal cell cytoplasmic alterations; however the chloroplasts and associated pyrenoids remained intact. RNA*later* represents a good alternative for specimen, tissue and single cell preservation. DNA barcoding can help with species delimitation and refining the concept of cryptic species. For example in this rudimentary study with a small population of *G. acuminatum* (Figures A.1, A.3), gene sequences for *rbcL* showed 1–5 base pairs differences between the 4 specimen clades, collected from three different sites within our primary pond (NHC-1). *C.* cf. *cuspidata* showed no variability (no base pair differences) in specimens from another sample site. In contrast, up to 117 base pair differences were observed within the *Navicula* clade from 3 different locations

within a lake and up to 88 base pair differences were noted in the *Pinnularia* clade from 4 different lakes and pond locations. These results suggest that by expanding the use of barcodes to many individuals within a diatom population, inter- and intraspecific questions can be routinely addressed.

Historical problems in extracting, amplifying, and sequencing DNA from single-cells have limited the development of genetics as a tool in the study of global microbial diversity, biogeography, and physiology. In diatoms, DNA sequencing from single cells is a logical step forward in population, taxonomy and environmental genetic studies. More conventional morphometric studies routinely use sample populations to determine size diminution series and variability of morphological expression (e.g. Lange-Bertalot *et al*., 2011; Levkov *et al*., 2013). With detailed genetic studies of single cells, links to match morphological populations will be informative in understanding variations in genotypic and phenotypic expression. At the species level, single cell multi-gene sequences along with associated morphometrics can act as multi-proxy validation datasets for species identifications. Future developments with single cell sequencing may even advance next generation genomic research.

Acknowledgements

This study was financially supported by a Canadian Museum of Nature RAC grant (2012-15) to P.B. Hamilton. K.E. Lefebvre is funded by a graduate student scholarship from the Canadian Museum of Nature. We wish to thank Frédérick Caron for assistance with cell isolations, amplifications and sequencing of specimens. Finally we thank two reviewers in an earlier version of this manuscript and three reviewers from this submission for their valued comments, which greatly improved this publication.

Statement of Contributions

P.B. Hamilton was responsible for the original project design, sampling efforts and much of the written manuscript. R.D. Bull was responsible for some of the molecular work and editing the manuscript. I was responsible for the bulk of the molecular lab work, all data analysis, writing up the methodology and results section and designing the figures.

Table A.1. List of 35 taxa isolated and sequenced for genes *rbcL*, *psbA* and 18S, with Canadian Museum of Nature Accession numbers, GenBank ID numbers, sample media, source locality with date of collection and collector.

Taxon	CANA Accession Number	GenBank ID			Identifier	Sample Medium	Locality and date of collection	Collector
		rbcL	18S	*psbA*				
Stauroneis cf. *gracilis*	CANA 93655	KM999045	KM998975	KM999010	A6R6	Water	small lake near Lac a Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 07-Sep-2013	F.D. Caron
Stauroneis cf. *gracilis*	CANA 93655	KM999046	KM998976	KM999011	A5R6	Water	small lake near Lac a Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 07-Sep-2013	F.D. Caron
Stauroneis cf. *anceps*	CANA 93655	KM999047	KM998977	KM999012	B3R6	Water	small lake near Lac a Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 07-Sep-2013	F.D. Caron
Pinnularia sp.	CANA 93252	KM999048	KM998978	KM999013	B5R10	RNALater	Alymer, small pond (Quebec,Canada) 45.443007 N 75.810437 W, 30-Jan-2014	P.B. Hamilton
Pinnularia sp.	CANA 93252	KM999049	KM998979	KM999014	A2R10	Water	Alymer, small pond (Quebec,Canada) 45.443007 N 75.810437 W, 30-Jan-2014	P.B. Hamilton
Pinnularia sp.	CANA 93251	KM999050	KM998980	KM999015	A1R10	Water	Alymer, small pond (Ontario,Canada) 45.442997 N 75.810437 W, 30-Jan-2014	P.B. Hamilton
Pinnularia sp.	CANA 93250	KM999051	KM998981	KM999016	A3R10	Water	Alymer, small pond (Quebec,Canada) 54.443033 N 75.810378 W, 30-Jan-2014	P.B. Hamilton
Pinnularia sp.	CANA 93249	KM999052	KM998982	KM999017	A4R10	Water	Alymer, small pond (Quebec,Canada) 45.443022 N 75.810421 W, 30-Jan-2014	P.B. Hamilton
Pinnularia sp.	CANA 93251	KM999053	KM998983	KM999018	B1R10	RNALater	Alymer, small pond (Quebec,Canada) 45.442997 N 75.810437 W, 30-Jan-2014	P.B. Hamilton
Pinnularia sp.	CANA 93641	KM999054	KM998984	KM999019	B1R7	RNALater	Ditch, Hwy 56, Adirondack Park (New York,USA) 44.30692 N 74.71650 W, 26-Oct-2013	P.B. Hamilton
Pinnularia sp.	CANA 93250	KM999055	KM998985	KM999020	A8R10	RNALater	Alymer, small pond (Quebec,Canada) 54.443033 N 75.810378 W, 30-Jan-2014	P.B. Hamilton
Pinnularia sp.	CANA 93660	KM999056	KM998986	KM999021	E3R8	Water	Alymer, small pond (Quebec,Canada) 45.443033 N 75.810383 W, 15-Nov-2013	K.E. Lefebvre
Navicula sp.	CANA 93656	KM999057	KM998987	KM999022	B8R6	Water	small lake near Lac a Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 28-Sep-2013	F.D. Caron
Navicula sp.	CANA 93655	KM999058	KM998988	KM999023	B7R6	Water	small lake near Lac a Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 07-Sep-2013	F.D. Caron
Navicula sp.	CANA 93655	KM999059	KM998989	KM999024	B6R6	Water	small lake near Lac a Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 07-Sep-2013	F.D. Caron
Navicula sp.	CANA 93655	KM999060	KM998990	KM999025	B5R6	Water	small lake near Lac a Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 07-Sep-2013	F.D. Caron
Navicula sp.	CANA 93655	KM999061	KM998991	KM999026	B4R6	Water	small lake near Lac a Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 07-Sep-2013	F.D. Caron
Navicula sp.	CANA 93655	KM999062	KM998992	KM999027	B1R6	Water	small lake near Lac a Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 07-Sep-2013	F.D. Caron
Navicula sp.	CANA 93656	KM999063	KM998993	KM999028	C6R6	Water	small lake near Lac a Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 28-Sep-2013	F.D. Caron
Navicula sp.	CANA 93656	KM999064	KM998994	KM999029	C3R6	Water	small lake near Lac a Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 28-Sep-2013	F.D. Caron
Sellephora sp.	CANA 93647	KM999065	KM998995	KM999030	F4R6	Water	Milledgeville, south of airport (Georgia,USA)	P.B.

Table A.1. Continued

Taxon	CANA Accession Number	GenBank ID			Identifier	Sample Medium	Locality and date of collection	Collector
		rbcL	18S	*psbA*				
Craticula cf. *cuspidata*	CANA 93656	KM999066	KM998996	KM999031	C4R6	Water	33.1472 N 83.2505 W, 22-Oct-2013 small lake near Lac a Serpent (Quebec,Canada)	Hamilton
Craticula cf. *cuspidata*	CANA 93655	KM999067	KM998997	KM999032	A3R6	Water	45.627633 N 75.601467 W, 28-Sep-2013 small lake near Lac a Serpent (Quebec,Canada)	F.D. Caron
Craticula cf. *cuspidata*	CANA 93655	KM999068	KM998998	KM999033	A2R6	Water	45.627633 N 75.601467 W, 07-Sep-2013 small lake near Lac a Serpent (Quebec,Canada)	F.D. Caron
Craticula cf. *cuspidata*	CANA 93655	KM999069	KM998999	KM999034	A1R6	Water	45.627633 N 75.601467 W, 07-Sep-2013 small lake near Lac a Serpent (Quebec,Canada)	F.D. Caron
Craticula cuspidata	CANA 93647	KM999070	KM999000	KM999035	C6R7	RNALater	45.627633 N 75.601467 W, 07-Sep-2013 Milledgeville, south of airport (Georgia,USA)	P.B. Hamilton
Craticula cf. *cuspidata*	CANA 93656	KM999071	KM999001	KM999036		Water	33.1472 N 83.2505 W, 22-Oct-2013 small lake near Lac a Serpent (Quebec,Canada)	Hamilton
Gyrosigma accuminatum	CANA 93251	KM999072	KM999002	KM999037		RNALater	45.627633 N 75.601467 W, 28-Sep-2013 Alymer, small pond (Quebec,Canada)	F.D. Caron P.B.
Gyrosigma accuminatum	CANA 93251	KM999073	KM999003	KM999038		Water	45.442997 N 75.810437 W, 30-Jan-2014 Alymer, small pond (Quebec,Canada)	Hamilton P.B.
Gyrosigma accuminatum	CANA 93251	KM999074	KM999004	KM999039		Water	45.442997 N 75.810437 W, 30-Jan-2014 Alymer, small pond (Quebec,Canada)	Hamilton P.B.
Gyrosigma accuminatum	CANA 93249	KM999075	KM999005	KM999040		RNALater	45.442997 N 75.810437 W, 30-Jan-2014 Alymer, small pond (Quebec,Canada)	Hamilton P.B.
Gyrosigma accuminatum	CANA 93663	KM999076	KM999006	KM999041		RNALater	45.443022 N 75.810421 W, 30-Jan-2014 Alymer, small pond (Quebec,Canada)	Hamilton K.E.
Gyrosigma accuminatum	CANA 93663	KM999077	KM999007	KM999042		RNALater	45.443050 N 75.810367 W, 15-Nov-2013 Alymer, small pond (Quebec,Canada)	Lefebvre K.E.
Gyrosigma accuminatum	CANA 93663	KM999078	KM999008	KM999043		RNALater	45.443050 N 75.810367 W, 15-Nov-2013 Alymer, small pond (Quebec,Canada)	Lefebvre K.E.
Cymatopleura solea	CANA 93251	KM999079	KM999009	KM999044		RNALater	45.443050 N 75.810367 W, 15-Nov-2013 Alymer, small pond (Quebec,Canada)	Lefebvre P.B.
							45.442997 N 75.810437 W, 30-Jan-2014	Hamilton

Table A.2. List of 38 taxa isolated and sequenced for the gene *rbcL*, with Canadian Museum of Nature Accession numbers, GenBank ID numbers, sample media, source locality with date of collection and collector.

Taxon	Accession Number	GenBank ID	Identifier	Sample Medium	Locality and date of collection	Collector
Melosira varians	CANA 100020	KM999080	I7R14	Water	Black Rapids, Rideau River (Ontario,Canada) 45.321990 N 75.699427 W, 19-Jun-2014	P.B. Hamilton
Melosira varians	CANA 100020	KM999081	I4R14	Water	Black Rapids, Rideau River (Ontario,Canada) 45.321990 N 75.699427 W, 19-Jun-2014	P.B. Hamilton
Synedra sp.	CANA 93622	KM999082	C7R5	Water	Alymer, small pond (Quebec,Canada) 45.442838 N 75.810296 W, 04-Oct-2013	P.B. Hamilton
Surirella	CANA 100068	KM999083	A7R14	Water	Petawawa Forestry Station (Ontario,Canada) 45.992966 N 77.399194 W, 18-May-2014	P.B. Hamilton
Surirella	CANA 93281	KM999084	B6R13	Water	Creek, Adirondack Park (New York,USA) 44 00.715 N 74 49.155 W, 11-May-2014	P.B. Hamilton
Cymatopleura solea	CANA 93636	KM999085	G3R5	Water	Alymer, small pond (Quebec,Canada) 45.442970 N 75.810277 W, 10-Oct-2013	P.B. Hamilton
Cymatopleura solea	CANA 93622	KM999086	C2R5	Water	Alymer, small pond (Quebec,Canada) 45.442838 N 75.810296 W, 04-Oct-2013	P.B. Hamilton
Stauroneis cf. *gracilis*	CANA 93639	KM999087	B8R5	Water	Ditch, Hwy 36 (Ontario,Canada) 44.803911 W 76.509686 W, 28-Sep-2013	P.B. Hamilton
Sellephora cf. laevissima	CANA 100068	KM999088	B2R14	Water	Petawawa Forestry Station (Ontario,Canada) 45.992966 N 77.399194 W, 18-May-2014	P.B. Hamilton
Neidium sp.	CANA 100068	KM999089	G3R12	Water	Petawawa Forestry Station (Ontario,Canada) 45.992966 N 77.399194 W, 18-May-2014	P.B. Hamilton
Neidium tumescens	CANA 100072	KM999090	C7R12	Water	Wetland, Adirondack Park (New York,USA) 43.90870 N 74.43944 W, 11-May-2014	P.B. Hamilton
Neidium tumescens	CANA 100072	KM999091	A7R12	Water	Wetland, Adirondack Park (New York,USA) 43.90870 N 74.43944 W, 11-May-2014	P.B. Hamilton
Neidium sp.	CANA 93372	KM999092	D1R12	Water	Pond, Hwy 56 (New York,USA) 44.39536 N 74.75690 W, 11-May-2014	P.B. Hamilton
Cymbella sp.	CANA 100068	KM999093	H1R12	Water	Petawawa Forestry Station (Ontario,Canada) 45.992966 N 77.399194 W, 18-May-2014	P.B. Hamilton
Cymbopleura subcuspidata	CANA 100068	KM999094	H4R12	Water	Petawawa Forestry Station (Ontario,Canada) 45.992966 N 77.399194 W, 18-May-2014	P.B. Hamilton
Cymbopleura subcuspidata	CANA 93374	KM999095	D7R13	Water	Wetland, Adirondack Park (New York,USA) 44.31925 N 74.72392 W, 10-May-2014	P.B. Hamilton
Cymbopleura subcuspidata	CANA 93374	KM999096	D6R13	Water	Wetland, Adirondack Park (New York,USA) 44.31925 N 74.72392 W, 10-May-2014	P.B. Hamilton
Gomphonema cf. parvulum	CANA 93654	KM999097	A6R8	Water	Alymer, small pond (Quebec,Canada) 45.44288 N 75.81045 W, 05-Nov-2013	K.E. Lefebvre
Encyonema sp.	CANA 93281	KM999098	B8R13	Water	Creek, Adirondack Park (New York,USA) 44.00715 N 74.49155 W, 11-May-2014	P.B. Hamilton
Frustulia bahlsii	CANA 93284	KM999099	C6R13	Water	Big Moose Lake, Adirondack Park (New	P.B. Hamilton

Table A.2. Continued

Taxon	Accession Number	GenBank ID	Identifier	Sample Medium	Locality and date of collection	Collector
					York,USA) 43.81723 N 74.88400 W, 10-May-2014	
Frustulia bahlsii	CANA 93279	KM999100	A5R13	Water	Wetland, Adirondack Park (New York,USA) 44.44806 N 74 .77439 W, 10-May-2014	P.B. Hamilton
Frustulia bahlsii	CANA 93284	KM999101	C5R13	Water	Big Moose Lake, Adirondack Park (New York,USA) 43.81723 N 74.88400 W, 10-May-2014	P.B. Hamilton
Frustulia cf. *saxonica*	CANA 93281	KM999102	B5R13	Water	Creek, Adirondack Park (New York,USA) 44.00715 W 74.49155 W, 11-May-2014	P.B. Hamilton
Gyrosigma accuminatum	CANA 93253	KM999103	E5R10	Iodine	Alymer, small pond (Quebec,Canada) 45.443031 N 75.810467 W, 21-Mar-2014	P.B. Hamilton
Gyrosigma accuminatum	CANA 93663	KM999104	G3R9	RNALater	Alymer, small pond (Quebec,Canada) 45.26583 N 75.48622 W, 15-Nov-2013	K.E. Lefebvre
Navicula sp.	CANA 93656	KM999105	C5R6	Water	small lake near Lac A Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 28-Sep-2013	F.D. Caron
Navicula sp.	CANA 93656	KM999106	D1R6	Water	small lake near Lac A Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 28-Sep-2013	F.D. Caron
Navicula sp.	CANA 93656	KM999107	C1R6	Water	small lake near Lac A Serpent (Quebec,Canada) 45.627633 N 75.601467 W, 28-Sep-2013	F.D. Caron
Navicula sp.	CANA 93645	KM999108	E5R6	Water	Alymer, small pond (Quebec,Canada) 45.44305 N 75.81043 W, 08-Oct-2013	P.B. Hamilton
Navicula sp.	CANA 93645	KM999109	E3R6	Water	Alymer, small pond (Quebec,Canada) 45.44305 N 75.81043 W, 08-Oct-2013	P.B. Hamilton
Navicula sp.	CANA 93644	KM999110	D3R6	Water	Alymer, small pond (Quebec,Canada) 45.44305 N 75.81043 W, 08-Oct-2013	P.B. Hamilton
Navicula sp.	CANA 93644	KM999111	D5R6	Water	Alymer, small pond (Quebec,Canada) 45.44305 N 75.81043 W, 08-Oct-2013	P.B. Hamilton
Navicula cf. *cryptocephala*	CANA 93655	KM999112	A8R6	Water	small lake near Lac A Serpent (Quebec,Canada) 45 37.658 N 75 36.088 W, 07-Sep-2013	P.B. Hamilton
Nitzschia cf. *sigmoidea*	CANA 100068	KM999113	G2R12	Water	Petawawa Forestry Station (Ontario,Canada) 45.992966 N 77.399194 W, 18-May-2014	P.B. Hamilton
Nitzschia linearis	CANA 93374	KM999114	D8R13	Water	Adirondack Park (New York,USA) 44.31925 N 74.72392 W, 10-May-2014	P.B. Hamilton
Eunotia sp.	CANA 93373	KM999115	D2R13	Water	Big Moose Lake, Adirondack Park (New York,USA) 43.81782 N 74.88404 W, 10-May-2014	P.B. Hamilton
Aulacoseria granulata	CANA 100020	KM999116	I6R14	Water	Black Rapids, Rideau River (Ontario,Canada) 45.321990 N 75.699427 W, 19-Jun-2014	P.B. Hamilton
Aulacoseria granulata	CANA 100020	KM999117	I3R14	Water	Black Rapids, Rideau River (Ontario,Canada) 45.321990 N 75.699427 W, 19-Jun-2014	P.B. Hamilton

Table A.3. Oligonucleotide primer sequences used in the nested amplifications. Nested, a: refers to the first primer amplifications; b: the primer used in the second amplification.

Nested	Forward primers: name, (sequence)	Reverse primers: name, (sequence)	Author
a	*rbcL*66+ (5'-TTTAAGGAGAAATAAATGTCTCAATCTG-3')	*rbcL*dp7- (5'-AAASHDCCTTGTGTWAGTVTC-3')	Alverson *et al.* 2007 / Daugbjerg and Andersen 1997
a	*rbcL*40+ (5'-GGACTCGAATVAAAAGTGAACG-3')	*rbcL*1444-(5'-GCGAAATCAGCTGTATCTGTWG-3')	Ruck and Theriot 2011
b	*rbcL*181F (5'-ACCGTGCTAAAGCTT-3')	*rbcL*1174R (5'-ACCRATTGTACCACC-3')	This study
a	18SP2F (5'-CTGGTTGATTCTGCCAGT-3')	18SP4R (5'-TGATCCTTCYGCAGGTTCAC-3'	Hannen *et al.* 1999 / Guillou *et al.* 1999
b	18S200F (5'-YGGSRWGAYRTGGTGADTCA-3')	18S1444R (5'-GVRTRCATCAGTGTAGCGCG-3')	This study
a	*psbA*F (5'-ATGACTGCTACTTTAGAAAGACG-3')	*psbA*R (5'-GCTAAATCTARWGGGAAGTTGTG-3')	Yoon *et al.* 2002
b	*psbA*45Fdg (5'-WTTYATCGWGCTCCWCCAG-3)	*psbA*648R (5'-RTGWGCTGCAACGATRTTRT-3')	This study

Table A.4. Range of sequence base pair lengths for the 3 genes studied from our samples and those reported from GenBank. Mean ± SD represents base pair lengths from GenBank sequences.

	Our samples	**GenBank samples**	**Mean**
rbcL	1202-1298	364-1475	1181 (+/- 386)
18S	996-1145	3-1915	818 (+/- 556)
psbA	537-578	660-1082	790 (+/- 110)

Table A.5. Amplification success rates for live, RNAlater and Lugol's solution samples.

	Live	RNAlater				Lugol's
		5-7 days storage	8-14 days storage	15-21 days storage	22-30 days storage	
Gene	**Successful (total attempts)**	**Successful (total attempts)**	**Successful (total attempts)**	**Successful (total attempts)**	**Successful (total attempts)**	**Successful (total attempts)**
rbcL	348 (446)	9 (15)	1 (3)	4 (13)	4 (7)	4 (53)
18S	183 (201)	7 (8)	1 (1)	3 (4)	1 (3)	13* (19)
psbA	65 (66)	7 (7)	1 (1)	4 (4)	NA	1 (1)

* All of the successful samples were found to contain residual fungi DNA.

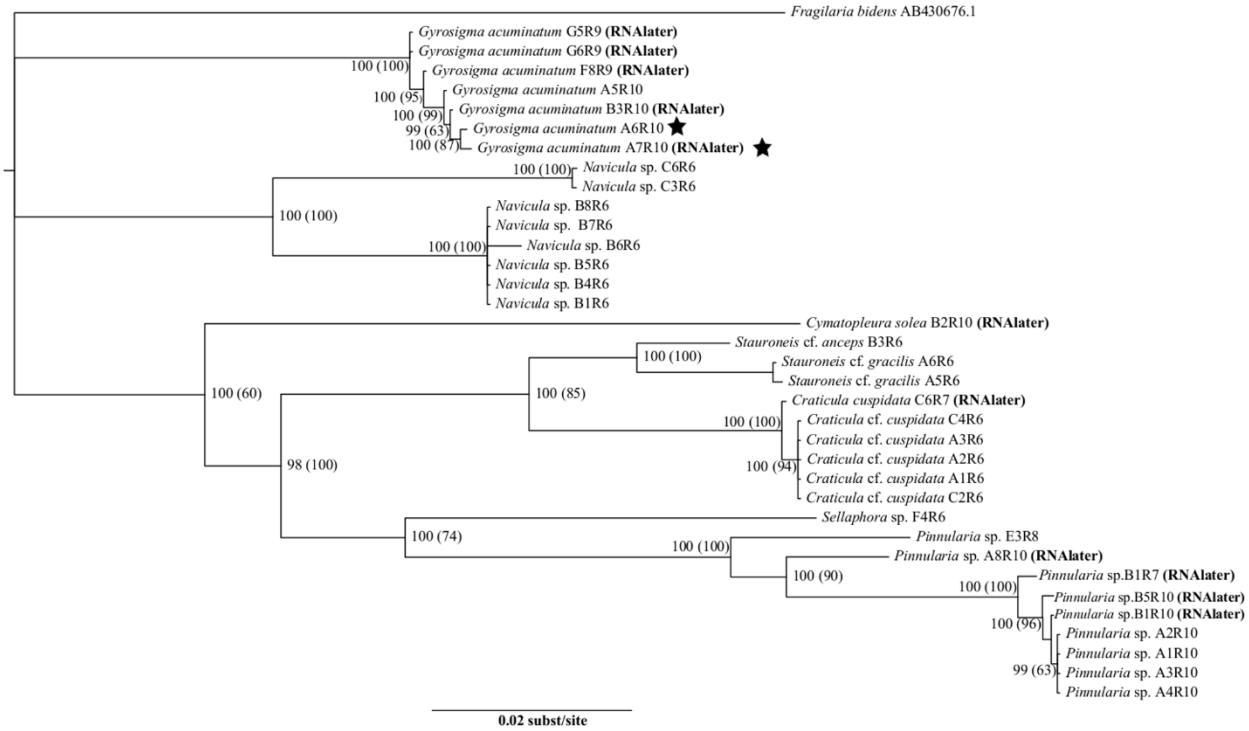


Fig. A.1. Tree showing relationships of *rbcL*, 18S and *psbA* using the best fit model, GTR+I+G. Statistical support is shown with numbers at nodes: Bayesian posterior probabilities (Maximum Likelihood bootstrap values). The stars indicates two *Gyrosigma* cells, one from a fresh sample and the other from an RNAlater preserved samples which had significantly similar sequences. Taxa are indicated to be from either RNA*later* or fresh samples.

Fig. A.2. Phylogenetic relationships of Bacillariophyta rooted with *Bolidomonas pacifica* based on a *rbcL* dataset using best fit model GTR+I+G with BI and MLBS analyses. Numbers at the nodes indicate statistical support if both methods resulted in >50%: Bayesian Inference posterior probabilities (Maximum Likelihood bootstrap values). Taxa are indicated to be from either fresh, RNALater or iodine samples. Colour blocks indicate the different orders of Bacillariophyta (A=Melosirales; B= Aulacoseirales;C= Fragilariales; D= Thalassiosirales; E= Eunotiales; F= Bacillariales; G= Naviculales; H= Cymbellales; I= Surirellales). The black arrows show two instances in which sequences from preserved cells had very similar sequences from fresh cell samples. The red arrows show three instances where taxa collected from the same location had identical sequences. *Bolidomonas pacifica* (GenBank Acquisition HQ912421.1) and *Cyclotella meneghiniana* (GenBank Acquisition KF959651.1) were used as an outgroup and sister group respectively, and their *rbcL* sequences were obtained from GenBank.

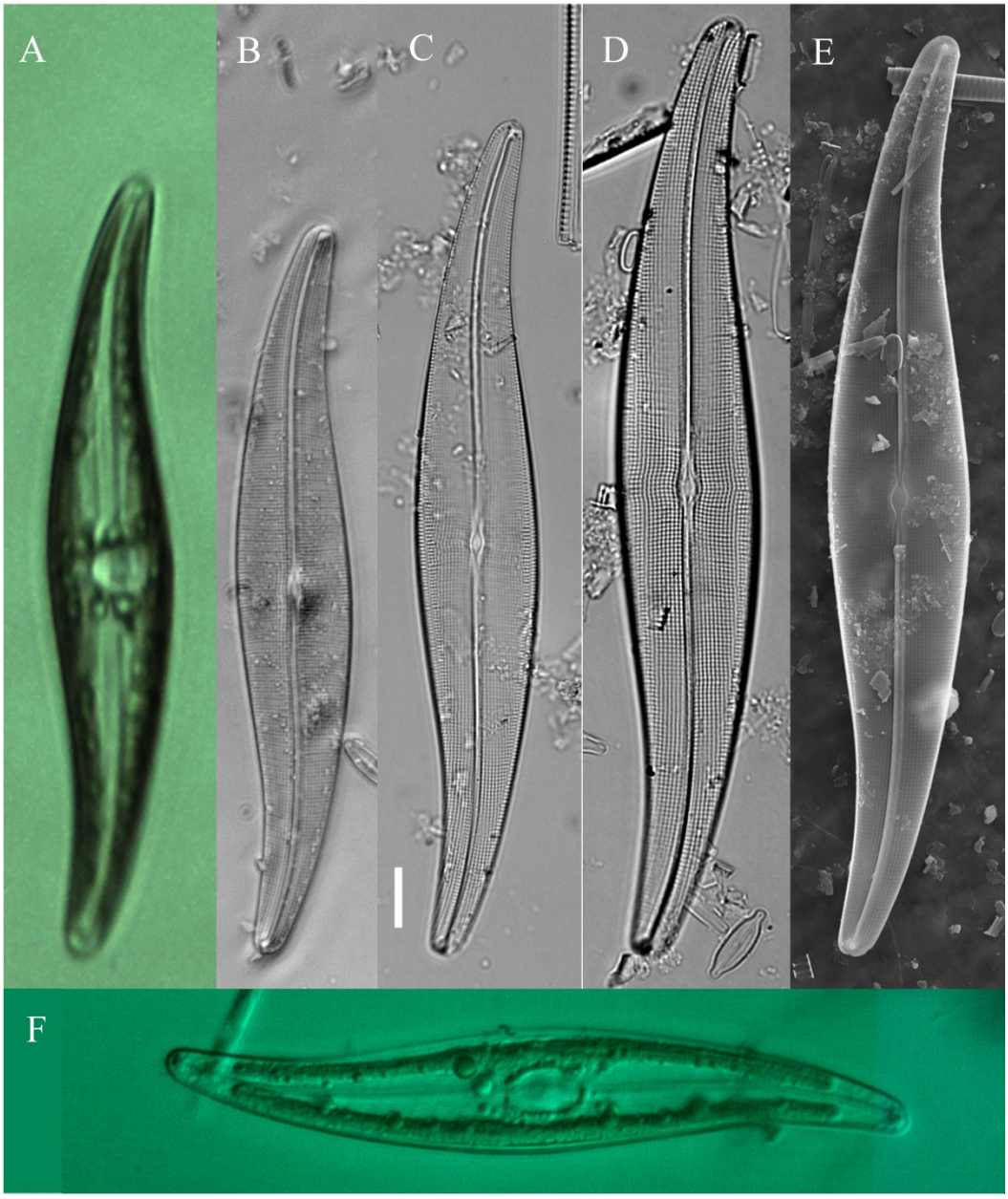


Fig. A.3. *Gyrosigma acuminatum* shown as live, fixed and cleaned specimens with light microscopy (LM) and scanning Electron microscopy (SEM). **(A)** live specimen from NHC pond (sequenced for *rbcL*), low magnification showing valve outline and cytoplasmic structure. **(B-D)** LM micrographs showing the size reduction series from one sample location. **(E)** SEM illustration the internal view of the valve. **(F)** LM image of an RNA*later* fixed specimen after freezing at -20°C and thawing. Scale bar = 10 µm.

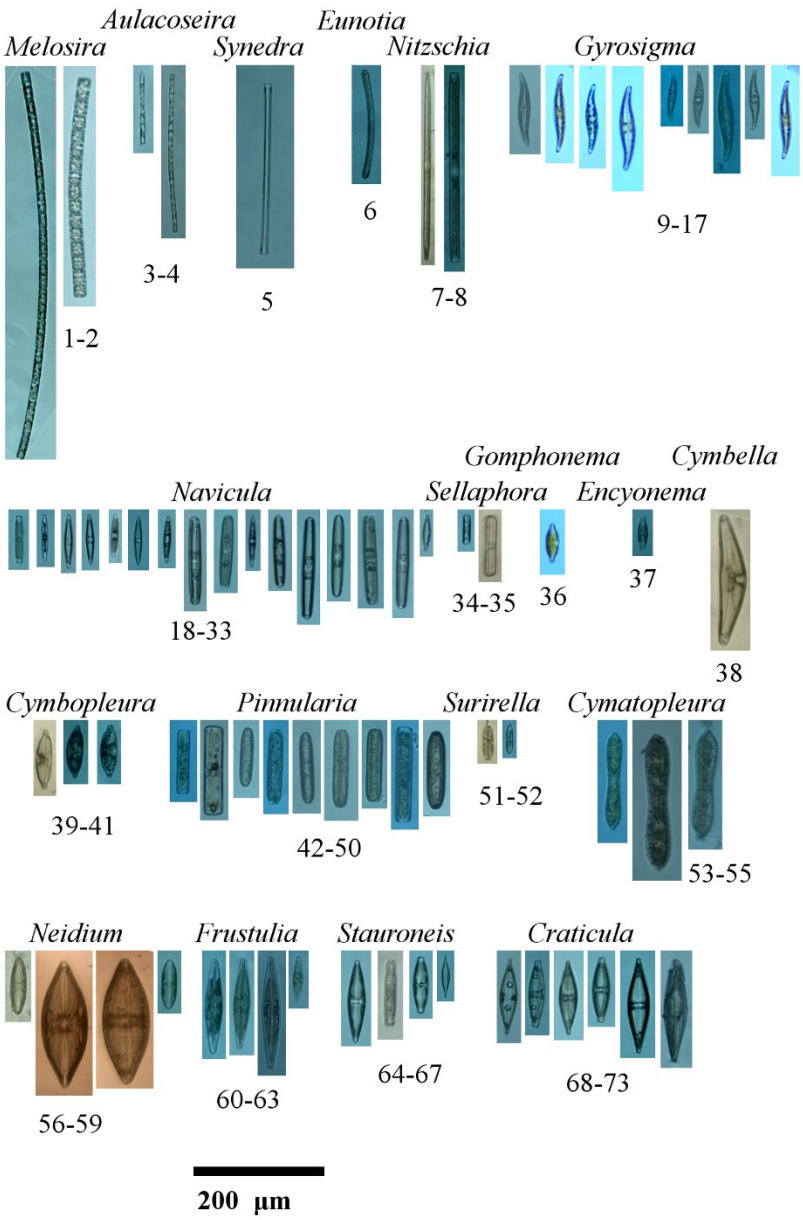


Fig. A.Supplement. Isolation photos of sequenced taxa aligned to their placement in Fig. 2. *Bolidomonas pacifica* (GenBank Acquisition HQ912421.1) and *Cyclotella meneghiniana* (GenBank Acquisition KF959651.1) are not pictured as their sequences were used as an outgroup and sister group from GenBank respectively. The isolated taxa are aligned as in Fig. 2 and shown with names and identifier. 1) *Melosira varians* I7R14; 2) *Melosira varians* I4R14; 3) *Aulacoseira granulata* I6R14; 4) *Aulacoseira granulata* I3R14; 5) *Synedra* sp. C7R5; 6) *Eunotia*

sp. D2R13; 7) *Nitzschia* cf. *sigmoidea* G2R12; 8) *Nitzschia linearis* var. *tenuis* D8R13; 9) *Gyrosigma acuminatum* E5R10; 10) *Gyrosigma acuminatum* G3R9 ; 11) *Gyrosigma acuminatum* G6R9; 12) *Gyrosigma acuminatum* G5R9; 13) *Gyrosigma acuminatum* F8R9; 14) *Gyrosigma acuminatum* A5R10; 15) *Gyrosigma acuminatum* B3R10; 16) *Gyrosigma acuminatum* A6R10; 17) *Gyrosigma acuminatum* A7R10; 18) *Navicula* sp. C5R6; 19) *Navicula* sp. B8R6; 20) *Navicula* sp. B7R6; 21) *Navicula* sp. B5R6; 22) *Navicula* sp. B4R6; 23) *Navicula* sp. B1R6 ; 24) *Navicula* sp. B6R6; 25) *Navicula* cf. *cryptocephala* A8R6; 26) *Navicula* sp. C6R6; 27) *Navicula* sp. D1R6; 28) *Navicula* sp. C1R6; 29) *Navicula* sp. E5R6; 30) *Navicula* sp. E3R6; 31) *Navicula* sp. D3R6; 32) *Navicula* sp. C3R6; 33) *Navicula* sp. D5R6; 34) *Sellaphora* sp. F6R4; 35) *Sellaphora* cf. *laevissima* B2R14; 36) *Gomphonema* cf. *parvulum* A6R8; 37) *Encyonema* sp. B8R13; 38) *Cymbella* sp. H1R12; 39) *Cymbopleura subcuspidata* H4R12; 40) *Cymbopleura subcuspidata* D7R13; 41) *Cymbopleura subcuspidata* D6R13; 42) *Pinnularia* sp. E3R8; 43) *Pinnularia* sp. A8R10; 44) *Pinnularia* sp. A2R10; 45) *Pinnularia* sp. B1R10; 46) *Pinnularia* sp. A1R10; 47) *Pinnularia* sp. A3R10; 48) *Pinnularia* sp. A4R10; 49) *Pinnularia* sp. B5R10; 50) *Pinnularia* sp. B1R7; 51) *Surirella* sp. A7R14; 52) *Surirella angusta* B6R13; 53) *Cymatopleura solea* B2R10; 54) *Cymatopleura solea* G3R5; 55) *Cymatopleura solea* C2R5; 56) *Neidium* sp. G3R12; 57) *Neidium tumescens* C7R12; 58) *Neidium tumescens* A7R12; 59) *Neidium* sp. D1R12; 60) *Frustulia bahlsii* C6R13; 61) *Frustulia bahlsii* A5R13; 62) *Frustulia bahlsii* C5R13; 63) *Frustulia saxonica* B5R13; 64) *Stauroneis* cf. *gracilis* A6R6; 65) *Stauroneis* cf. *gracilis* B8R5; 66) *Stauroneis* cf. *gracilis* A5R6; 67) *Stauroneis* cf. *anceps* B3R6; 68) *Craticula* cf. *cuspidata* C4R6; 69) *Craticula* cf. *cuspidata* C2R6; 70) *Craticula* cf. *cuspidata* A3R6; 71) *Craticula* cf. *cuspidata* A2R6; 72) *Craticula* cf. *cuspidata* A1R6; 73) *Craticula cuspidata* C6R7.

Appendix B: Supplemental Material from Chapter 2

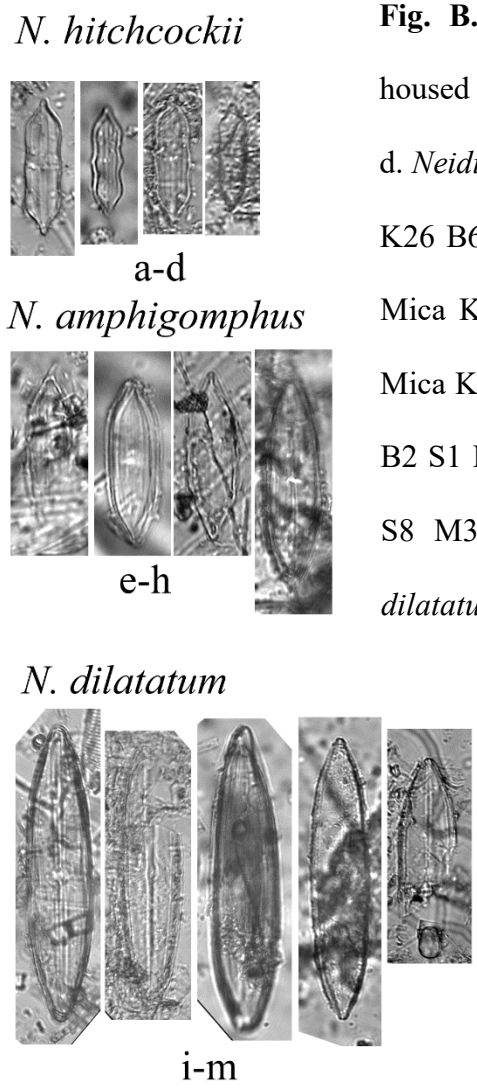


Fig. B.1. LM images from Ehrenberg's original mica mounts housed in the Museum fur NatureKunde (*BHUPM*) Collection. a–d. *Neidium hitchcockii*. a. Mica K26 B6 S3 M3 white ring. b. Mica K26 B6 S2 M3 white ring. c. Mica K26 B6 S3 M4 blue ring. d. Mica K26 B6 S8 M4 red ring. e–h. *Neidium amphigomphus*. e. Mica K26 B1 S8 M3, specimen outside missing ring. f. Mica K26 B2 S1 M4 in circle. g. Mica K26 B8 S8 M1 in circle. h. K26 B3 S8 M3, blue circle [could also be *N. fossum*]. i–l. *Neidium dilatatum*. i. Mica K26 B3 S3 M3, blue circle. j. Mica K26 B2 S4 M3, white circle. k. Mica K26 B5 S3 M2, red circle. l. Mica K26 B8 S3 M5 in circle. m–p. *Neidium fossum* (*Neidium dilatatum* sensu Ehrenberg). m. Mica K26 B5 S7 M5, blue circle. n. Mica K26 B5 S1 M,3 blue circle. o. Mica K26 B2 S1 M4 in circle. p. Mica K26 B1 S1 M1, white circle. q. Mica K26 B3 S6 M5, red circle (Ehrenberg has it listed in the blue circle).

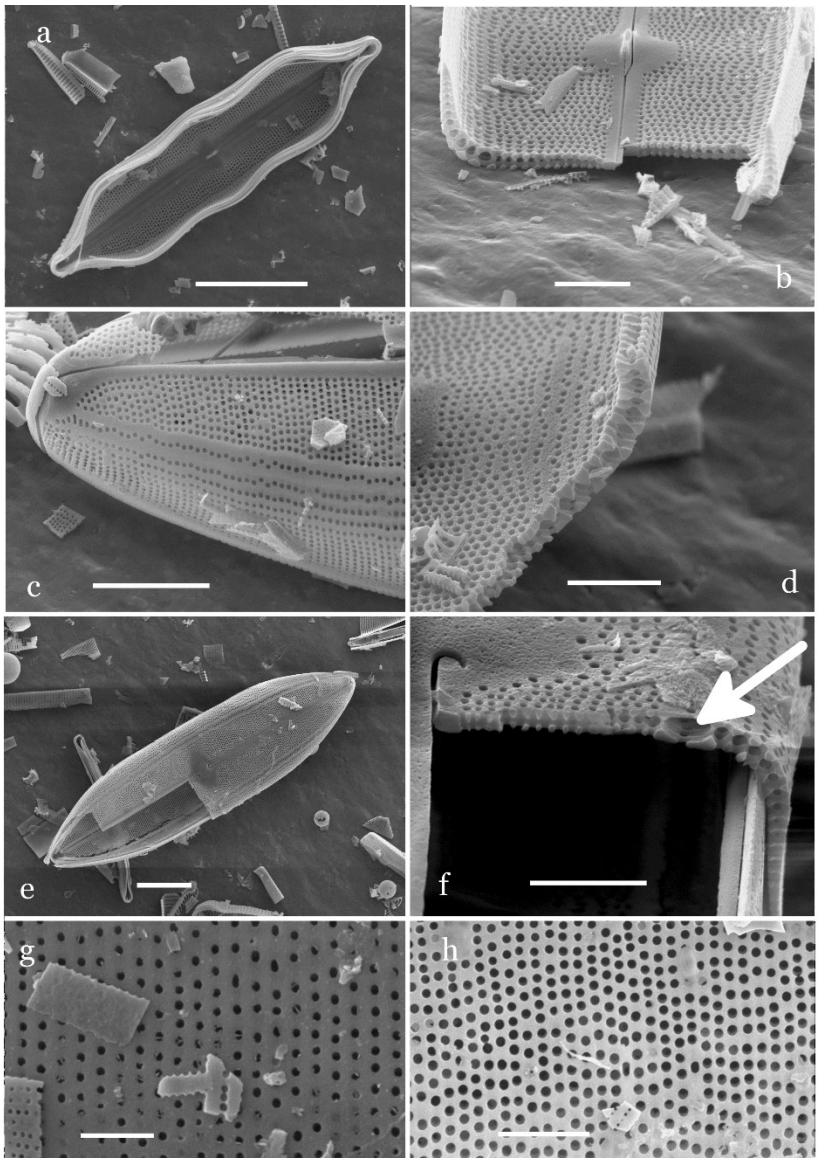


Fig. B.2. SEM images of *Neidium* specimens from the Ehrenberg samples housed in the Museum fur NatureKunde (BHUPM) Collection. a. Internal view of *N. hitchcockii* including 3 copulae. One large longitudinal band along each margin. b. External view central valve face of *N. dilatatum*. c. External view of the apex-mantle for *N. dilatatum*. Five longitudinal canal are merging down to one at the apex. d. Internal view, fractures wall of *N. dilatatum* showing chambered structure of the wall for areolae and the lonitudinal canals. e. Whole frustule, partly broken for *N. fossum*, *sp. nov*. f. Extenal view, fractured wall of *N. fossum* showing the interconnected formation of the chambered areolae and fomation of the longitudinal canals. Arrow indicates the large longitudinal canal. g. External view of areloae for *N. fossum* sp. nov. h. External view of areolae for *N. dilatatum*. a, b, g. From Bridgewater, Massachusetts, peat, Sample 1769. c, d, e, f, h. Samples from Andover (Andower), Massachusetts, fossil, sample 1771. Scale bars = 20 µm (a, e), 10 µm (c), 5 µm (b, d, f), 2 µm (g, h).

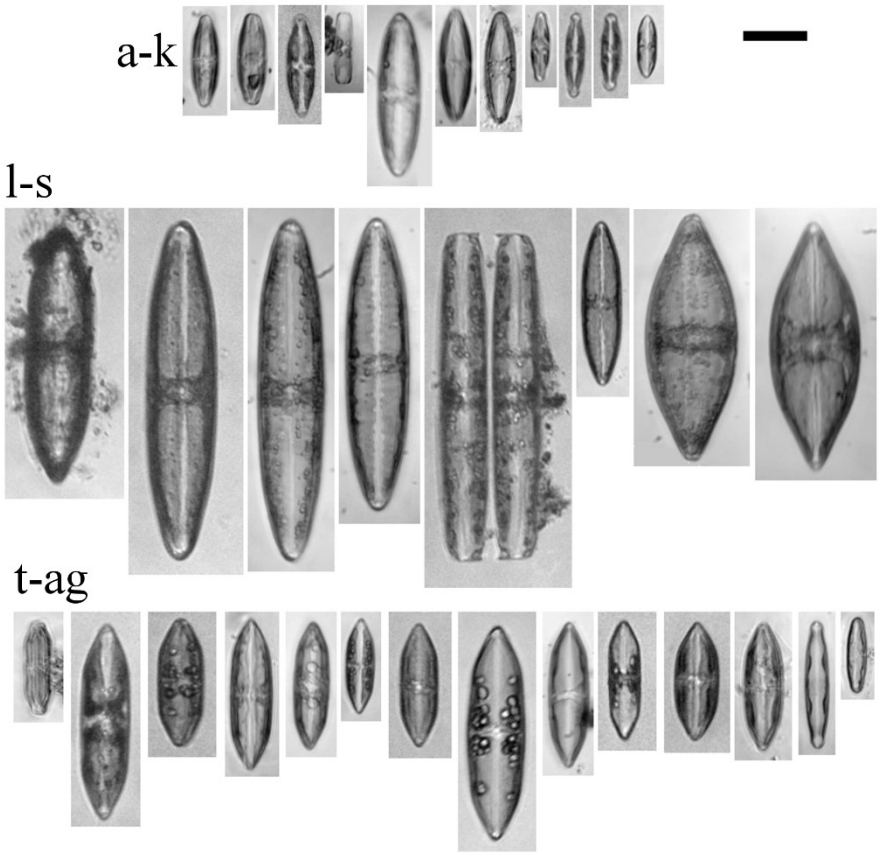


Fig. B.3. Isolation photos of sequenced *Neidium* specimens. a. *Neidium* sp. A1R14. b. *Neidium* sp. B1R14. c. *Neidium* sp. E3R13. d. *Neidium* sp. E4R12. e. *Neidium fossum* D5R12. f. *Neidium* sp. G7R12. g. *Neidium* sp. KM999089. h. *Neidium* sp. E6R12. i. *Neidium* sp. C4R13. j. *Neidium* sp. C3R13. k. *Neidium* sp. KM999092. l. *Neidium dilatatum* A4R13. m. *Neidium dilatatum* B3R13. n. *Neidium dilatatum* F1R12. o. *Neidium dilatatum* D6R12. p. *Neidium dilatatum* A9R13. q. *Neidium dilatatum* A6R12. r. *Neidium tumescens* KM999091. s. *Neidium tumescens* KM999090. t. *Neidium hitchcockii* G4R12. u. *Neidium amphigomphus* C7R13. v. *Neidium amphigomphus* G7R13. w. *Neidium amphigomphus* D4R12. x. *Neidium amphigomphus* A3R14. y. *Neidium amphigomphus* B4R12. z. *Neidium amphigomphus* H4R13. aa. *Neidium amphigomphus* C3R7. ab. *Neidium amphigomphus* G8R12. ac. *Neidium amphigomphus* F9R13. ad. *Neidium amphigomphus* C9R13. ae. *Neidium cf. amphigomphus* G6R12. af. *Neidium* sp B6R14. ag. *Neidium* sp G5R12. Scale bar = 50 μm.

Appendix C: Supplemental Material from Chapter 3

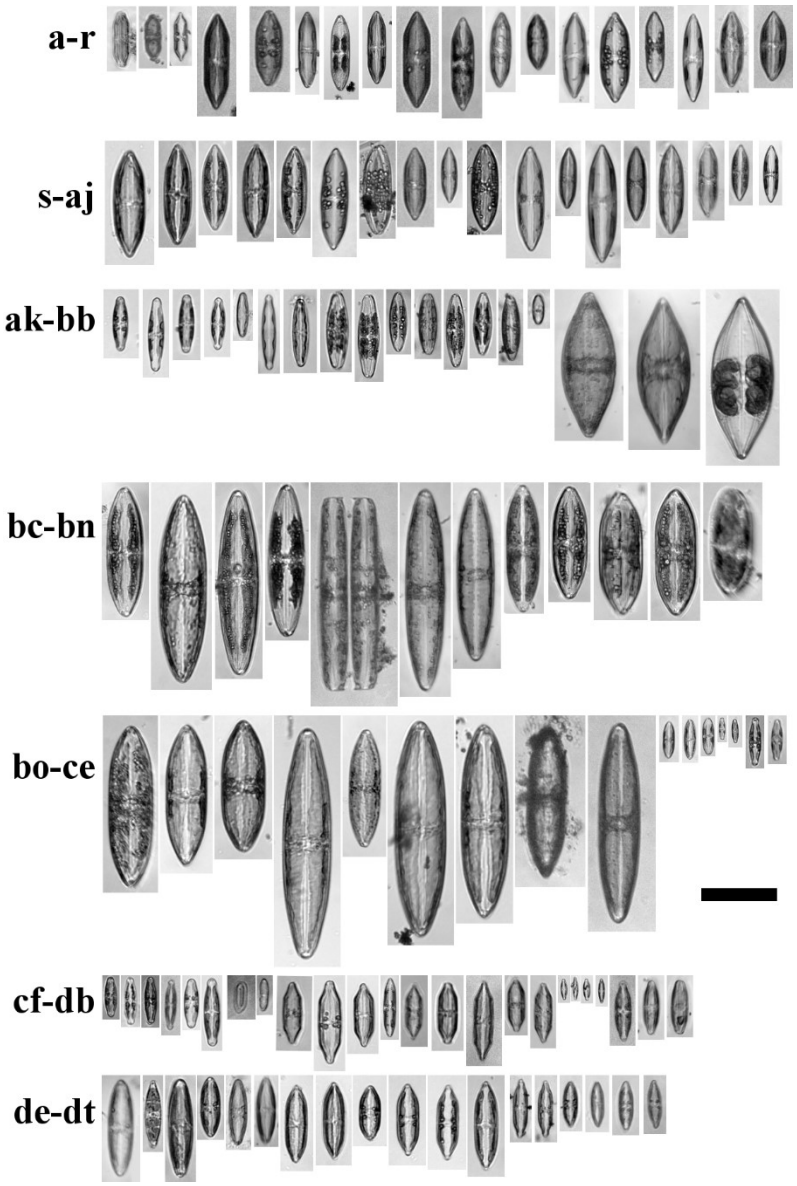


Fig. C.1 – Isolation photos of sequenced *Neidium* specimens. a. *Neidium hitchcockii* G4R14. b. *Neidium hitchcockii* F1R14. c. *Neidium hitchcockii* C4R15. d. *Neidium amphigomphus* G9R13. e. *Neidium amphigomphus* G7R13. f. *Neidium amphigomphus* A6R18. g. *Neidium amphigomphus* A2R18. h. *Neidium amphigomphus* K3R15. i. *Neidium amphigomphus* H1R13. j. *Neidium amphigomphus* C7R13. k. *Neidium amphigomphus* A3R14. l. *Neidium amphigomphus*

E4R19. m. *Neidium amphigomphus* G8R12. n. *Neidium amphigomphus* C6R18. o. *Neidium amphigomphus* F9R13. p. *Neidium amphigomphus* E3R19. q. *Neidium amphigomphus* G6R12. r. *Neidium amphigomphus* C9R13. s. *Neidium amphigomphus* H4R15. t. *Neidium amphigomphus* A3R20. u. *Neidium amphigomphus* C7R25. v. *Neidium amphigomphus* A4R20. w. *Neidium amphigomphus* C8R20. x. *Neidium amphigomphus* C3R7. y. *Neidium amphigomphus* A7R20. z. *Neidium amphigomphus* H4R13. aa. *Neidium amphigomphus* A5R12. ab. *Neidium amphigomphus* A5R20. ac. *Neidium amphigomphus* C2R7. ad. *Neidium amphigomphus* B1R12. ae. *Neidium amphigomphus* B8R12. af. *Neidium amphigomphus* B2R12. ag. *Neidium amphigomphus* D4R12. ah. *Neidium amphigomphus* B5R14. ai. *Neidium amphigomphus* B4R12. aj. *Neidium amphigomphus* C8R18. ak. *Neidium* sp1 D2R18. al. *Neidium* sp1 F2R18. am. *Neidium* sp1 C4R18. an. *Neidium* sp1 D7R18. ao. *Neidium* sp G5R12. ap. *Neidium* sp B6R14. aq. *Neidium* sp G3R18. ar. *Neidium lowei* B4R23. as. *Neidium lowei* B1R23. at. *Neidium lowei* D3R20. au. *Neidium lowei* A5R15. av. *Neidium lowei* D6R20. aw. *Neidium lowei* D4R20. ax. *Neidium lowei* A6R15. ay. *Neidium productum* A8R16. az. *Neidium tumescens* A7R12. ba. *Neidium tumescens* C7R12. bb. *Neidium tumescens* J3R15. bc. *Neidium dilatatum* D4R25. bd. *Neidium dilatatum* D8R18. be. *Neidium dilatatum* D6R25. bf. *Neidium dilatatum* B2R23. bg. *Neidium dilatatum* A9R13. bh. *Neidium dilatatum* F1R12. bi. *Neidium dilatatum* D6R12. bj. *Neidium dilatatum* D5R25. bk. *Neidium dilatatum* D3R25. bl. *Neidium dilatatum* E5R20. bm. *Neidium dilatatum* C8R25. bn. *Neidium dilatatum* E7R20. bo. *Neidium dilatatum* C6R20. bp. *Neidium dilatatum* C7R18. bq. *Neidium dilatatum* C5R18. br. *Neidium dilatatum* C2R18. bs. *Neidium dilatatum* C2R17. bt. *Neidium dilatatum* K4R15. bu. *Neidium dilatatum* I6R15. bv. *Neidium dilatatum* A4R13. bw. *Neidium dilatatum* B3R13. bx. *Neidium* sp D1R12. by. *Neidium* sp I5R15. bz. *Neidium* sp J6R15. ca. *Neidium potapovii* J4R15. cb. *Neidium potapovii* B6R18.

cc. *Neidium affine* B5R20. cd. *Neidium affine* E6R12. ce. *Neidium affine* B1R20. cf. *Neidium affine* E1R20. cg. *Neidium affine* B7R20. ch. *Neidium affine* C4R13. ci. *Neidium affine* D1R18. cj. *Neidium affine* C3R18. ck. *Neidium bisulcatum* A5R16. cl. *Neidium bisulcatum* A4R16. cm. *Neidium sacoense* A5R15. cn. *Neidium sacoense* D4R18. co. *Neidium sacoense* F4R18. cp. *Neidium sacoense* C1R17. cq. *Neidium sacoense* A5R17. cr. *Neidium sacoense* A7R17. cs. *Neidium sacoense* A8R13. ct. *Neidium sacoense* A8R17. cu. *Neidium sacoense* C5R17. cv. *Neidium longiceps* E7R18. cw. *Neidium longiceps* E8R19. cx. *Neidium longiceps* E8R18. cy. *Neidium longiceps* D5R18. cz. *Neidium* sp E3R13. da. *Neidium* sp A1R14. db. *Neidium* sp B1R14. dc. *Neidium fossum* D5R12. dd. *Neidium fossum* A3R23. de. *Neidium fossum* A2R21. df. *Neidium fossum* A4R21. dg. *Neidium fossum* G3R12. dh. *Neidium fossum* G7R12. di. *Neidium* sp6 G8R18. dj. *Neidium* sp6 G1R18. dk. *Neidium* sp6 D6R18. dl. *Neidium* sp6 C1R18. dm. *Neidium* sp6 B8R18. dn. *Neidium* sp6 G2R18. do. *Neidium promontorium* B5R15. dp. *Neidium promontorium* B7R15. dq. *Neidium promontorium* E1R14. dr. *Neidium promontorium* D8R14. ds. *Neidium promontorium* E5R14. dt. *Neidium promontorium* E8R14.

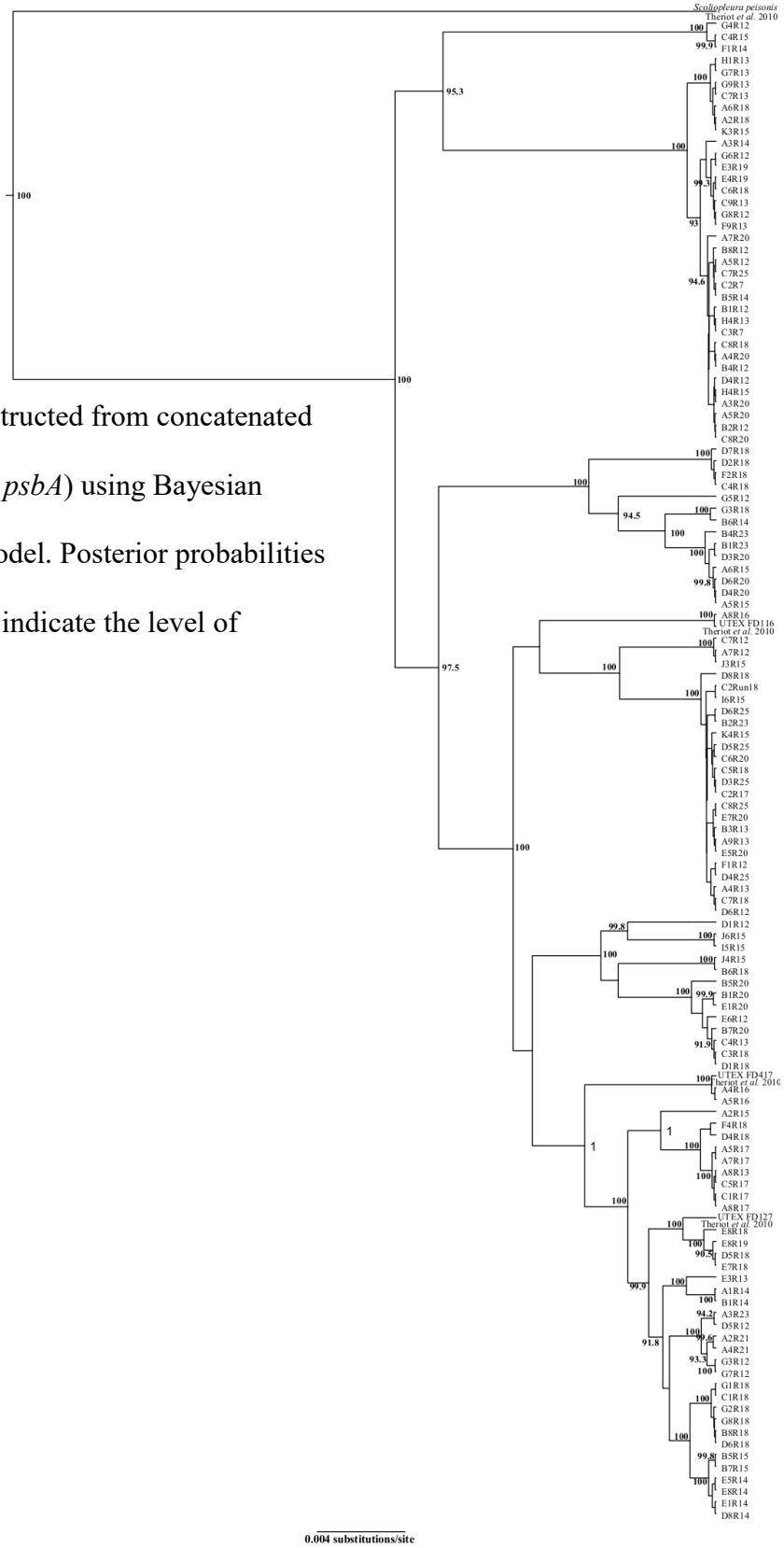


Fig. C.2. Phylogenetic tree constructed from concatenated chloroplast dataset (*rbcL*, *psbC*, *psbA*) using Bayesian analysis under the GTR+I+G model. Posterior probabilities are listed at significant nodes to indicate the level of support.

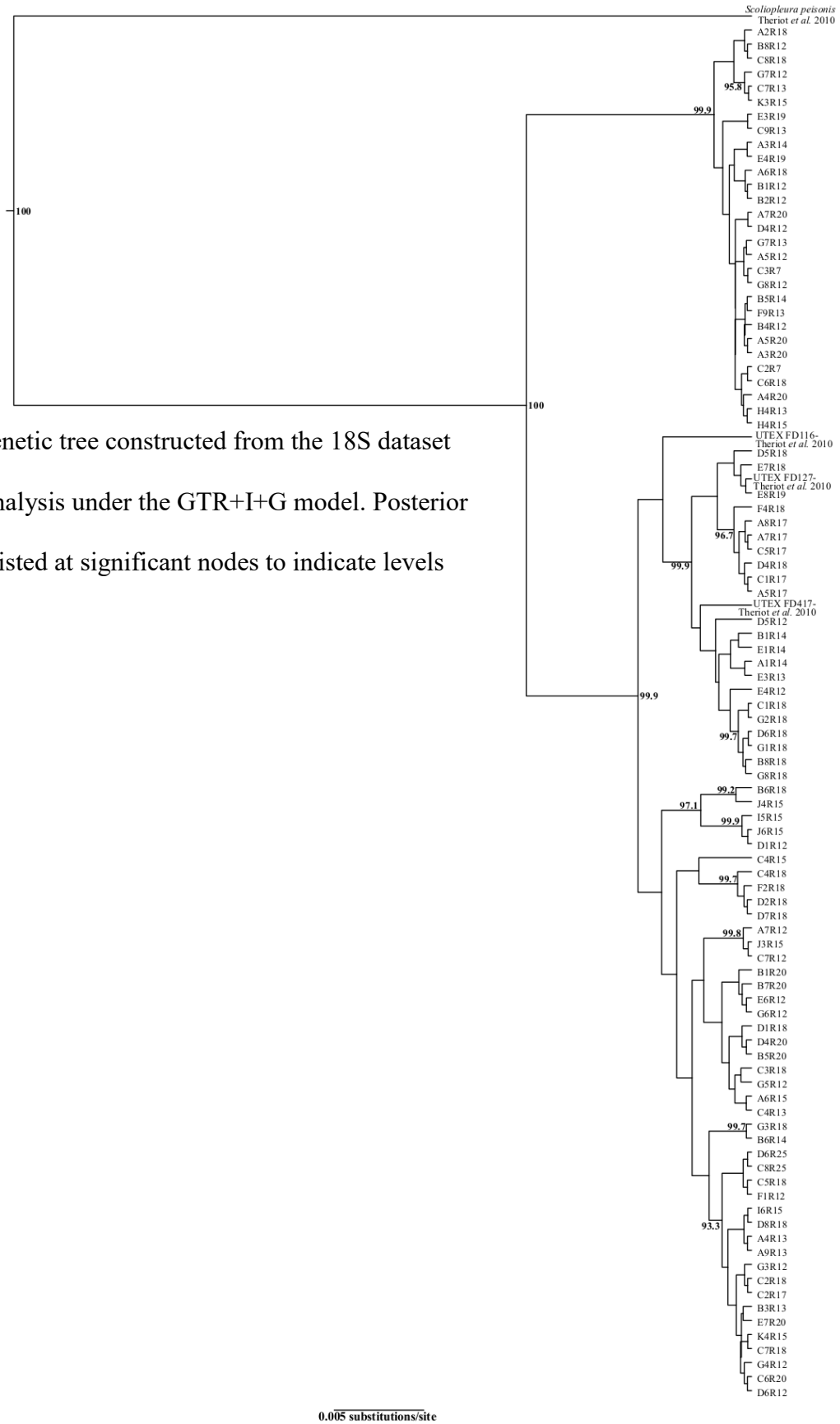


Fig. C.3. Phylogenetic tree constructed from the 18S dataset using Bayesian analysis under the GTR+I+G model. Posterior probabilities are listed at significant nodes to indicate levels of support.