Single cell PCR amplification of diatoms using fresh and preserved samples

Paul B. Hamilton, Keely E. Lefebvre and Roger D. Bull

Research and Collections, Canadian Museum of Nature, Ottawa, ON, Canada
Biology and Environmental Science, Center for Advanced Research in Environmental Genomics, University of Ottawa, Ottawa, ON, Canada

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 RNAlater. 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 and 96%. An rbcL consensus tree grouped morphologically similar specimens and was consistent across the two primary sample treatments: fresh and RNAlater. 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.

Keywords: single cells, phylogenetics, diatoms, RNAlater, fixation, barcoding

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 megakaryocyte-associated tyrosine kinase (MatK) serve as two of the preferred barcode genes for taxonomic identifications (CBOL Plant Work 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; Pouličková et al., 2010; Kaczmarska et al., 2014). The smaller (~500 bp) 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 bp) 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, Prymnesiophyceae, Raphidophyceae, Synurophyceae (e.g., Daugbjerg and Andersen, 1997; McCourt et al., 2000; Tomitani et al., 2006; Edvardsen et al., 2011; Bennett et al., 2014). However, compared to DNA research, and Bacillariophyceae, and evolution (Bertozzini and Richlen, 2005). In order to effectively pursue population studies, which has the potential for more detailed sequencing. Theriot et al., 2006) have shown that RNA amplification errors (Alverson et al., 2007; Godhe et al., 2002; Henrichs et al., 2007; Auinger et al., 2008) are prone to alterations in structural morphology. To date, there are only a few published examples using non-cultured single cell amplifications (e.g., Godhe et al., 2002; Auinger et al., 2008; 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., Bertozzini et al., 2005; Auinger et al., 2008). Chelex® resin is an effective DNA extraction tool with applications in molecular biology ranging from multicellular vertebrate tissues to single microbial cells (Hahn et al., 2000; Richlen and Barber, 2005). 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 RNAlater 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., Godhe et al., 2002; Henrichs et al., 2007; Lang and Kaczmarska, 2011; Ivanova et al., submitted). RNAlater 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). RNAlater 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 RNAlater 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 1, 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 RNAlater tissue storage buffer (http://www.protocol-online.org/prot/Protocols/RNALater-3999.html). The sample was hand shaken to mix, kept...
| Taxon                   | Accession number | GenBank ID for rbcL | GenBank ID for 18S | GenBank ID for psbA | Identifier | Sample medium | Locality and date of collection                                                                 | Collector |
|------------------------|------------------|---------------------|--------------------|--------------------|------------|---------------|------------------------------------------------------------------------------------------------|-----------|
| *Stauroneis cf. gracilis* | CANA 93655     | KM999045           | KM9990975          | 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           | KM9990976          | 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           | KM9990977          | 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           | KM9990978          | KM999013           | B5R10      | RINAlater     | Alymo, small pond (Quebec, Canada) 45.443007 N 75.810437 W, 30-Jan-2014                           | P.B. Hamilton |
| *Pinnularia sp.*        | CANA 93252     | KM999049           | KM9990979          | KM999014           | A2R10      | Water         | Alymo, small pond (Quebec, Canada) 45.443007 N 75.810437 W, 30-Jan-2014                           | P.B. Hamilton |
| *Pinnularia sp.*        | CANA 93251     | KM999050           | KM9990980          | KM999015           | A1R10      | Water         | Alymo, small pond (Ontario, Canada) 45.443097 N 75.810437 W, 30-Jan-2014                         | P.B. Hamilton |
| *Pinnularia sp.*        | CANA 93250     | KM999051           | KM9990981          | KM999016           | A3R10      | Water         | Alymo, small pond (Quebec, Canada) 54.443033 N 75.810378 W, 30-Jan-2014                         | P.B. Hamilton |
| *Pinnularia sp.*        | CANA 93249     | KM999052           | KM9990982          | KM999017           | A4R10      | Water         | Alymo, small pond (Quebec, Canada) 45.443022 N 75.810421 W, 30-Jan-2014                         | P.B. Hamilton |
| *Pinnularia sp.*        | CANA 93251     | KM999053           | KM9990983          | KM999018           | B1R10      | RINAlater     | Ditch, Hwy 56, Adirondack Park (New York, USA) 44.30692 N 74.71650 W, 26-Oct-2013              | P.B. Hamilton |
| *Pinnularia sp.*        | CANA 93641     | KM999054           | KM9990984          | KM999019           | B1R7       | RINAlater     | Alymo, small pond (Quebec, Canada) 45.443022 N 75.810421 W, 30-Jan-2014                         | P.B. Hamilton |
| *Pinnularia sp.*        | CANA 93250     | KM999055           | KM9990985          | KM999020           | A8R10      | RINAlater     | Alymo, small pond (Quebec, Canada) 54.443033 N 75.810378 W, 30-Jan-2014                         | P.B. Hamilton |
| *Pinnularia sp.*        | CANA 93660     | KM999056           | KM9990986          | KM999021           | E3R8       | Water         | Alymo, small pond (Quebec, Canada) 45.443033 N 75.810383 W, 15-Nov-2013                        | K.E. Lefebvre |
| *Navicula sp.*          | CANA 93656     | KM999057           | KM9990987          | 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           | KM9990988          | 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           | KM9990989          | 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           | KM9990990          | 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           | KM9990991          | 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           | KM9990992          | 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           | KM9990993          | 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           | KM9990994          | KM999029           | C3R6       | Water         | Small lake near Lac a Serpent (Quebec, Canada) 45.627633 N 75.601467 W, 28-Sep-2013             | F.D. Caron |

(Continued)
| Taxon                     | Accession number | GenBank ID for rbcL | GenBank ID for 18S | GenBank ID for psbA | Identifier | Sample medium | Locality and date of collection                                                                 | Collector       |
|--------------------------|------------------|---------------------|--------------------|---------------------|------------|---------------|-------------------------------------------------------------------------------------------------|-----------------|
| Selphora sp.             | CANA 93647       | KM999065            | KM998995           | KM999030            | F4R6       | Water         | Milledgeville, south of airport (Georgia, USA) 33.1472 N 83.2505 W, 22-Oct-2013                  | P.B. Hamilton    |
| Craticula cf. cuspidata  | CANA 93655       | KM999066            | KM998996           | KM999031            | C4R6       | Water         | Small lake near Lac a Serpent (Quebec, Canada) 45.627633 N 75.601467 W, 28-Sep-2013            | F.D. Caron       |
| Craticula cf. cuspidata  | CANA 93655       | KM999067            | KM998997           | KM999032            | A3R6       | Water         | Small lake near Lac a Serpent (Quebec, Canada) 45.627633 N 75.601467 W, 07-Sep-2013            | F.D. Caron       |
| Craticula cf. cuspidata  | CANA 93655       | KM999068            | KM998998           | KM999033            | A2R6       | Water         | Small lake near Lac a Serpent (Quebec, Canada) 45.627633 N 75.601467 W, 07-Sep-2013            | F.D. Caron       |
| Craticula cf. cuspidata  | CANA 93655       | KM999069            | KM998999           | KM999034            | A1R6       | Water         | Small lake near Lac a Serpent (Quebec, Canada) 45.627633 N 75.601467 W, 07-Sep-2013            | F.D. Caron       |
| Craticula cuspidata      | CANA 93647       | KM999070            | KM999000           | KM999035            | O6R7       | RNALater      | Milledgeville, south of airport (Georgia, USA) 33.1472 N 83.2505 W, 22-Oct-2013                  | P.B. Hamilton    |
| Craticula cf. cuspidata  | CANA 93656       | KM999071            | KM999001           | KM999036            |            | Water         | Small lake near Lac a Serpent (Quebec, Canada) 45.627633 N 75.601467 W, 28-Sep-2013            | F.D. Caron       |
| Gyrosigma acuminaatum    | CANA 93251       | KM999072            | KM999002           | KM999037            | RNALater   | Water         | Alymer, small pond (Quebec, Canada) 45.442997 N 75.810437 W, 30-Jan-2014                     | P.B. Hamilton    |
| Gyrosigma acuminaatum    | CANA 93251       | KM999073            | KM999003           | KM999038            | Water      | Alymer, small pond (Quebec, Canada) 45.442997 N 75.810437 W, 30-Jan-2014                     | P.B. Hamilton    |
| Gyrosigma acuminaatum    | CANA 93251       | KM999074            | KM999004           | KM999039            | Water      | Alymer, small pond (Quebec, Canada) 45.442997 N 75.810437 W, 30-Jan-2014                     | P.B. Hamilton    |
| Gyrosigma acuminaatum    | CANA 93249       | KM999075            | KM999005           | KM999040            | RNALater   | Alymer, small pond (Quebec, Canada) 45.443022 N 75.810421 W, 30-Jan-2014                     | P.B. Hamilton    |
| Gyrosigma acuminaatum    | CANA 93663       | KM999076            | KM999006           | KM999041            | RNALater   | Alymer, small pond (Quebec, Canada) 45.443050 N 75.810367 W, 15-Nov-2013                    | K.E. Lefebvre    |
| Gyrosigma acuminaatum    | CANA 93663       | KM999077            | KM999007           | KM999042            | RNALater   | Alymer, small pond (Quebec, Canada) 45.443050 N 75.810367 W, 15-Nov-2013                    | K.E. Lefebvre    |
| Gyrosigma acuminaatum    | CANA 93663       | KM999078            | KM999008           | KM999043            | RNALater   | Alymer, small pond (Quebec, Canada) 45.443050 N 75.810367 W, 15-Nov-2013                    | K.E. Lefebvre    |
| Cymatopleura solea       | CANA 93251       | KM999079            | KM999009           | KM999044            | RNALater   | Alymer, small pond (Quebec, Canada) 45.442997 N 75.810437 W, 30-Jan-2014                     | P.B. Hamilton    |
| Taxon                     | Accession number | GenBank ID   | Identifier | Sample medium | Locality and date of collection | Collector       |
|--------------------------|------------------|-------------|------------|---------------|---------------------------------|-----------------|
| Melosira varians         | CANA 100020      | F7R14       | I7R14      | Water         | Black Rapids, Rideau River (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | I14R13      | B7R13      | Water         | Black Rapids, Rideau River (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | A7R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | D5R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | D7R14       | G3R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | E8R15       | H7R15      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | B7R13      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | D5R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | H7R15       | G3R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | E8R15       | H7R15      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | B7R13      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | D5R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | H7R15       | G3R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | E8R15       | H7R15      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | B7R13      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | D5R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | H7R15       | G3R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | E8R15       | H7R15      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | B7R13      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | D5R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | H7R15       | G3R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | E8R15       | H7R15      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | B7R13      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | D5R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | H7R15       | G3R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | E8R15       | H7R15      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | B7R13      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | D5R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | H7R15       | G3R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | E8R15       | H7R15      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | B7R13      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | D5R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | H7R15       | G3R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | E8R15       | H7R15      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | B7R13      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | D5R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | H7R15       | G3R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | E8R15       | H7R15      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | B7R13      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | D5R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | H7R15       | G3R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | E8R15       | H7R15      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | B7R13      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | D5R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | H7R15       | G3R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Melosira varians         | CANA 100020      | E8R15       | H7R15      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | B7R13      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
| Synedra                 | CANA 93293       | Water       | D5R14      | Water         | Petawawa Forestry Station (Ontario, Canada) | P.B. Hamilton   |
at room temperature for 24 h, and then stored in the dark at 4°C between 5 and 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 I3, 20 g of KI, 200 mL ddH2O). 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 vs. 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 colorless).

**Single Diatom Cell Isolation**
One to two milliliters of each sample (living, RNAlater 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 1, 2; Supplement 1 in Appendix).

**DNA Extraction, PCR and Sequencing Protocol**
For DNA extraction, Chelex-stored samples were incubated for 20 min 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 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 MgCl2, 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 penate 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 ModelTest v.2.1.4 (Guindon and Gascuel, 2003; Darriba et al., 2012) 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+1+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+1+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
TABLE 3 | Oligonucleotide primer sequences used in the nested amplifications.

| Nested | Forward primers: name, (sequence) | Reverse primers: name, (sequence) | Author |
|--------|-----------------------------------|-----------------------------------|--------|
| a      | rbcL66+ (5′-TTAAGAGGAGTAAATTGCTCAATCTG-3′) | rbcLdp7- (5′-AAASHDGCTGTGTGWATC-3′) | Alverson et al., 2007; Daugbjerg and Andersen, 1997 |
| a      | rbcL40+ (5′-GGAGTCTATTGAAGGTAACG-3′) | rbcL1444-(5′-GGGAACTACGTTATCTGTWG-3′) | Zwickl, 2006 |
| b      | rbcL181F (5′-ACCGTGCTAAGCTT-3′) | rbcL1174R (5′-ACCRATTGTAACACC-3′) | Ruck and Theriot, 2011 |
| a      | 18SP2F (5′-CTGGTTGATTCTGCCAGT-3′) | 18SP4R (5′-GTAGCTTCGCGGGTACTC-3′) | This study |
| b      | 18S200F (5′-YGGSRWGVAYRGTGTAADTC-3′) | 18S1444F (5′-GVRTRCATCACTGTAACGCC-3′) | van Haninen et al., 1999; Guilhou et al., 1999 |
| a      | psbAF (5′-ATGACTGCTACTTTAGAAGACG-3′) | psbAR (5′-OCTAATCTARWGGAAATGTTG-3′) | Yoon et al., 2002 |
| b      | psbA45Fd (5′-WTYTATCGGTCCWCCAG-3′) | psbA648R (5′-RTGWCGTAACGTRTRTTR-3′) | This study |

**Table Notes:** a refers to the first primer amplifications; b, the primer used in the second amplification.

Results

Partial sequences for rbcL (1202–1305 bp), 18S (811–1144 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, Eugenemia Ehrenberg, Navicula Bory, Neidium Pfitzer, Placoneis C.Mereschkowsky, Frustulia C.Ag., Gyrosigma Hassall, Stauroesiis Ehrenberg, Craticula Grunow, Sellaphora C.Mereschkowsky, Pinnularia Ehrenberg, Cymatopleura W.Sm., Encyonema Kütz., Gomphonema Ehrenberg, Nitzschia Hassall, Hantzschia Grunow, and Surirella Turpin.

Of the diatom sequences, 60 were from fresh living samples, 12 were from RNA-later fixed samples, and one from a Lugol's fixed sample (Tables 1, 2, Supplement 1 in Appendix). NCBI Blast searches using the new sequences resulted in matches consistent with the genus-level morphological identifications of our specimens. Method validation of the new sequences found on GenBank 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 1). Our dataset showed significant separation at the family level for the following: Pinnulariaceae (PP = 100, BS = 100), Selphoraceae (100, 74), Stauroeisaceae (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 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 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 intragenic 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 intragenic variation (Figure 1). Individual cells were principally collected...
TABLE 5 | Amplification success rates for live, RNAlater, and Lugol’s solution samples.

| Gene | Live 5–7 days storage | Live 8–14 days storage | Live 15–21 days storage | Live 22–30 days storage | RNAlater 5–7 days storage | RNAlater 8–14 days storage | RNAlater 15–21 days storage | RNAlater 22–30 days storage | Lugol’s 5–7 days storage | Lugol’s 8–14 days storage | Lugol’s 15–21 days storage | Lugol’s 22–30 days storage |
|------|------------------------|------------------------|------------------------|------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| rbcl | 348 (448)              | 9 (15)                 | 4 (13)                 | 4 (7)                  | 4 (53)                   | 4 (53)                   | 4 (53)                   | 4 (53)                   | 4 (53)                   | 4 (53)                   | 4 (53)                   | 4 (53)                   |
| 18S  | 183 (201)              | 7 (8)                  | 3 (4)                  | 1 (3)                  | 13* (19)                 | 13* (19)                 | 13* (19)                 | 13* (19)                 | 13* (19)                 | 13* (19)                 | 13* (19)                 | 13* (19)                 |
| psbA | 65 (86)                | 7 (7)                  | 4 (4)                  | NA                     | 1 (1)                    | 1 (1)                    | 1 (1)                    | 1 (1)                    | 1 (1)                    | 1 (1)                    | 1 (1)                    | 1 (1)                    |

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

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 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 *Stauroeis* (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, RNAlater 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 RNAlater preserved isolates were on the same terminal branch (Figure 2, black arrows).
FIGURE 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.

(Continued)
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).

Figures 3A–E). Additional specimens fixed in RNAlater and frozen (−17°C) for 2 days maintained frustule integrity with cytoplasm and chloroplast structure (Figure 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., Oedogonium 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; Pouličková et al., 2010). Nested DNA amplifications have the potential to open 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 RNAlater 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, Cricraticula to Stauromesis, and Pinnularia to Sellaphora (Figures 1, 2) were in agreement with other studies (Evans et al., 2007; Theriot et al., 2010). 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 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, 10 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 (Figure A1). 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 3, Figure A1). 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., Pouličková et al., 2010).

Traditional fixation of microbial samples for morphological, ecological and physiological study has a long history (Thronsden, 1978; Simmons, 2014). 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 RNAlater represents an excellent genetic fixation protocol for sample collection, short term storage and long-term archiving of microbial collections. Cell wall structure in Chlorophyceae (Oedogonium sp., Pediastrum boryanum), Cyanophyceae (Phormidium cf. princeps) and diatoms were maintained in RNAlater fixed samples under cold and frozen conditions. Chloroplast integrity was even maintained for G. acuminatum during freezing at −17°C (Figure 3F). The specifications for RNAlater indicate that treated tissues can be stored at 25°C for 1 week, at 4°C for 1 month, or at −20°C indefinitely (Ambion, 1999). In this study, both fresh and RNAlater fixed samples had predictable extraction success (Table 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 Sorensen, 2002). However, museums and large collections should prioritize the implementation of storage and fixation techniques that maintain the molecular integrity of the samples. RNAlater preserved algae, including diatoms, subjected to freeze-thaw cycles showed some internal cell cytoplasmic alterations; however the chloroplasts and associated pyrenoids remained intact. RNAlater 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 1, 3), gene sequences for rbcL showed 1–5 base pairs differences between the four 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 bp differences were observed within the Navicula clade from three different locations within a lake and up to 88 bp differences were noted in the Pinnularia clade from four different lakes and pond locations. These results suggest that by expanding the use of barcodes to many individuals within a diatom population, inter- and intra-specific 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.

Author Contributions

PH funded the project, developed in collaboration with RB the experimental design, identified the development with fixed samples, initiated the expansion to multiple gene and extractions, completed cell isolations and wrote the drafts of the manuscript. KL conducted cell isolations and the majority of the sequencing, edited and corrected the manuscript. This author developed new ideas on recovery of viable DNA, produced the figures and tables, wrote most of the methods and results. RB conducted the initial work in developing the single cell isolation protocol for the Canadian Museum of Nature Laboratory of Molecular Biodiversity, developed new primers for the nested amplification, experimental ideas for the development of the manuscript and contributed extensively to the writing and final editing of the manuscript.

Acknowledgments

This study was financially supported by a Canadian Museum of Nature RAC grant (2012-15) to PH. KL 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.
References

Alverson, A. J., Jansen, R. K., and Theriot, E. C. (2007). Bridging the rubicon: phylogenetic analysis reveals repeated colonization of marine and fresh waters by thalassiosiroid diatoms. Mol. Phylogenet. Evol. 45, 193–210. doi: 10.1016/j.ympev.2007.03.024

Amato, A., Kooistra, W. H. C. F., Ghiron, J. H. L., Mann, D. G., Pröschold, T., and Messersor, M. (2007). Reproductive isolation among sympatric cryptic species in marine diatoms. Protist 158, 193–207. doi: 10.1016/j.protis.2006.10.001

Ambion. (1999). Preserve RNA and tissue cell samples with RNAlater®. Ambion TechNotes NewsL. 5, 7–8.

Auinger, B. M., Pfandl, K., and 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. App. Environ. Microbiol. 74, 2505–2510. doi: 10.1128/AEM.01803-07

Bennett, M. S., Wiegert, K. E., and Trierrie, R. E. (2014). Characterization of Eugleniformis gen. nov. and the chloroplast genome of Eugleniformis [Euglena] proxima (Euglenophyta). Physiolgia 53, 66–73. doi: 10.2216/13-198.1

Bertozzini, E., Penna, A., Pierboni, E., Bruce, I., and Magnani, M. (2005). Development of new procedures for the isolation of phytoplankton DNA from fixed samples. J. Appl. Phycol. 17, 223–229. doi: 10.1007/s10811-005-2130-5

Bowers, H. A., Tengs, T., Glsaglow, H. B. Jr., Burkholder, J. M., Rublee, P. A., and Godhe, A., Anderson, D. M., and Rehnstam-Holm, A.-S. (2002). PCR sequence data. Phycologia 41, 15–21. doi: 10.2216/i0031-8884-41-1-15.1

Diatom Res. 9, 149–258. doi: 10.1080/0269249X.2011.573703

Lang, I., and Kaczmarska, I. (2011). A protocol for a single-cell PCR of diatoms from fixed samples: method validation using Ditylum brightwellii (T. West) Grunow. Diatom Res. 26, 43–49. doi: 10.1006/2062949.X.2011.573703

Lang-Bertalot, H., Bak, M., and Witkowski, A. (2011). “Diatoms of the European inland waters and comparable habitats: EUnotia and some related taxa” in Diatoms of Europe, Vol. 6, ed H. Lange-Bertalot (Koenigstein: Koeltz Scientific Books), 1–747.

Lasken, R. S. (2007). Single-cell genomic sequencing using multiple displacement amplification.Curr. Opin. Microbiol. 10, 510–516. doi: 10.1016/j.mib.2007.08.005

Legrand, B., Mazancourt, P. d., Durigon, M., Khalifat, V., and Crainic, K. (2002). DNA genotyping of unbuffered formalin fixed paraffin embedded tissues. Forensic Sci. Int. 125, 205–211. doi: 10.1016/S0338-0291(02)00211-7

Lepere, C., Demura, M., Kawachi, M., Romac, S., Probert, I., and Vaulot, D. (2011). Whole-genome amplification (WGA) of marine photosynthetic eukaryotic populations. MicrobIol. Ecol. 67, 513–523. doi: 10.1111/j.1574-6941.2011.0172x.2

Levkov, Z., Metzelin, D., and Pavlov, A. (2013). “Diatoms of the European inland waters and comparable habitats: Luticola and Luticola,” in Diatoms of Europe, Vol. 7, ed H. Lange-Bertalot (Koenigstein: Koeltz Scientific Books), 1–697.

Mann, D. G., Sato, S., Tobajo, R., Vanormelingen, P., and Souffreau, C. (2010). DNA barcoding for species identification and discovery in diatoms. Cryptogam. Algol. 31, 557–577.

McCourt, R. M., Karol, K. G., Bell, J., Helm-Bychowski, K. M., Grajewska, A., Wojciechowski, M. F., et al. (2000). Phylogeny of the conjugating green algae (Zygnemataceae) based on rbcl sequences. J. Phycol. 36, 747–758. doi: 10.1111/j.0022-3646.2000.01715.x

McCourt, R. M., Karol, K. G., Bell, J., Helm-Bychowski, K. M., Grajewska, A., Wojciechowski, M. F., et al. (2000). Phylogeny of the conjugating green algae (Zygnemataceae) based on rbcl sequences. J. Phycol. 36, 747–758. doi: 10.1111/j.0022-3646.2000.01715.x

Meng, J., Wang, J., Zhao, Y., Liu, J., and Luo, Y. (2010). A simple DNA barcoding protocol for species identification in microalgae. Appl. Environ. Microbiol. 76, 513–523. doi: 10.1128/AEM.00461-14

Oldach, D. W. (2000). Development of real-time PCR assays for rapid detection of Pfiesteria piscicida. Appl. Environ. Microbiol. 66, 4641–4648. doi: 10.1128/AEM.66.11.4641-4648.2000

Posada, D. (2003). “Using Modeltest and PAUP* to select a model of nucleotide substitution,” in Current Protocols in Bioinformatics, eds A. D. Baxevanis, D. B. Steinberg, M. E. Hughes, R. D. M. Page, G. A. Petsko, L. D. Stern, and G. D. Stormo (New York, NY: John Wiley and Sons), 6.5.1–6.5.14.
Pouličková, A., Veselá, J., Neustupa, J., and Škaloud, P. (2010). Pseudocryptic diversity versus cosmopolitanism in diatoms: a case study on *Navicula cryptocephala* Kütz. (*Bacillariophyceae*) and morphologically similar taxa. *Protist* 161, 353–369. doi: 10.1016/j.protis.2009.12.003

Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2013). *Tracer* v1.5. Available online at: http://beast.bio.ed.ac.uk/Tracer

Reyes-Escogido, L., Balam-Chi, M., Rodríguez-Buenfil, I., Valdés, J., Kameyama, L., and Martínez-Pérez, F. (2010). Purification of bacterial genomic DNA in less than 20 min using Chelex-100 microwave: examples from strains of lactic acid bacteria isolated from soil samples. *Antonie van Leeuwenhoek* 98, 465–474. doi: 10.1007/s10482-010-9462-0

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

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

Ruck, E. C., Nakov, T., Jansen, R. K., Theriot, E. C., and Alversen, A. J. (2014). Serial genome of diatoms. *Genome Biol. Evol.* 6, 644–654. doi: 10.1093/gbe/evu039

Ruck, E. C., and Theriot, E. C. (2011). Origin and evolution of the canal raphe system in diatoms. *Protist* 162, 723–737. doi: 10.1016/j.protis.2011.02.003

Ruiz Sebastián, C., and O’Ryan, C. (2001). Single-cell sequencing of dinoflagellate (*Dinophyceae*) nuclear ribosomal genes. *Mol. Ecol. Notes* 1, 329–331. doi: 10.1046/j.1471-8278.2001.00084.x

Sherbakova, T. A., Rubtsov, N. B., Likhoshway, Y. V., and Grachev, M. A. (2000). Combined SEM ultrastructure studies and PCR with individual diatom cells. *Diatom Res.* 15, 349–354. doi: 10.1080/0269249X.2000.9705501

Simmons, J. E. (2014). *Fluid Preservation a Comprehensive Reference*. Plymouth: Rowman and Littlefield.

Souffreau, C., Verbruggen, H., Wolfe, A. P., Vanormalingen, P., Siver, P. A., Cox, E. J., et al. (2011). A time-calibrated multi-gene phylogeny of the diatom genus *Pinnularia*. *Mol. Phylogenet. Evol.* 61, 866–879. doi: 10.1016/j.ympev.2011.08.031

Sterrenberg, F. A. S. (1995). Studies on the genus *Gyrosigma* and *Pleurosigma* (*Bacillariophyceae*): *Gyrosigma acuminatum* (Kützing) Rabenhorst, *G. spenceri* (Quckett) Griffith, and *G. rautenbuchiae* Cholnoky. *Proc. Acad. Nat. Sci. Phila.* 146, 467–480.

Swofford, D. L. (2003). *PAUP*:* Phylogenetic Analysis using Parsimony* (*and other Methods*). Version 4. Sunderland, MA: Sinauer Associates.

Tavare, S. (1986). "Some probabilistic and statistical problems in the analyses of DNA sequences," in *Some Mathematical Questions in Biology—DNA Sequence Analysis*, ed R. M. Miura (Providence, RI: American Mathematical Society), 57–86.

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

Zwickl, D. J. (2006). Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion. Ph.D. dissertation, The University of Texas at Austin, USA, 1–115.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Hamilton, Lefebvre and Bull. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.
FIGURE A1 | Isolation photos of sequenced taxa aligned to their placement in Figure 2. Bolidomonas pacifica (GenBank Acquisition HQ912421.1) and Cyclotella meneghiniana (GenBank Acquisition KF959651.1) are not pictured as their sequences were used as an out group and sister group from GenBank.

(Continued)
respectively. The isolated taxa are aligned as in Figure 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, Nitschia cf. sigmoidea G2R12; 8, Nitschia 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. C2R6; 33, Navicula sp. D5R6; 34, Sellaphora sp. F6R4; 35, Sellaphora cf. laevisima B2R14; 36, Gomphonema cf. parvulum A6R6; 37, Encyonema sp. B8R13; 38, Cymbellina sp. H1R12; 39, Cymbopleura subcuspida H4R12; 40, Cymbopleura subcuspida D7R13; 41, Cymbopleura subcuspida 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, Sunrella sp. A7R14; 52, Sunrella 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, Cricula cf. cuspidata C4R6; 69, Cricula cf. cuspidata C2R6; 70, Cricula cf. cuspidata A3R6; 71, Cricula cf. cuspidata A2R6; 72, Cricula cf. cuspidata A1R6; 73, Cricula cuspidata C6R7.

**Data Accessibility**

DNA sequence: Genbank accessions KM999080-KM999117.
Final DNA sequence assembly: online support information.

Supplement 1: Light microscope images of the single cells used in the DNA sequencing, in Figures 1, 2.

Supplement 2: All original samples (CANA_100043-100099) are housed in the National Algae Collection of Canada (CANA), http://www.nature-cana.ca/databases/index.php.