Can the diversity of Singapore Marine Planktonic Eukaryotes be uncovered using culture methods?

by
Rae Chua

Supervisor: Assistant Professor Adriana Lopes dos Santos
Asian School of Environment, Nanyang Technological University

Submitted to the School of Biological Sciences in partial fulfillment of the requirements for
Final Year Project

NANYANG TECHNOLOGICAL UNIVERSITY

Submission May, 2020

DECLARATION

I declare that,

in accordance with School requirements this thesis is under 6000 words in length;

all presented work was performed within the official project time frame as stated below;

all presented work was performed by me, unless otherwise specified;

all relevant work experience gained before the Final Year Project is stated below;

the input by my supervisor, or delegated supervisor, into this thesis was limited to reviewing of up to two hard copy drafts;

this thesis is my own work, unless otherwise referenced, as defined by the NTU policy on plagiarism and I have read the NTU Honour Code and Pledge;

the included Abstract, Introduction, Results, Discussion and Conclusions sections will be submitted to Turnitin no later than 24 hours after hardcopy submission.

Final Year Project start date: 13th Dec 2019
Final Year Project submission date: 4th May 2020
Total number of weeks: 20 weeks

Pre-Final Year Project experience
(List all <u>relevant</u> work experience gained before the Final Year Project, including paid employment, volunteer work and URECA experience. Include type of work and duration.)

BS9001 Research Experience- a 8-week attachment to SCELSE Kline Labs, bacterial culturing (growth assays) and creation of a complement mutant with electroporation and colony PCR.

Student's signature: Date: 3rd May 2020

TABLE OF CONTENTS

ACKNOLEDGEMENTS

ABBREVIATIONS

BSA   Bovine Albumin Solution

BLAST   Basic Local Alignment Search Tool

FCM   Flow Cytometry

NCBI   National Center for Biotechnology Information

OSD   Ocean Sampling Day

PCR   Polymerase Chain Reaction

PNS   Penicillin, neomycin and streptomycin

SMS   Singapore Marine Strains

TS   Time Series

Project Title: Can the diversity of Singapore Marine Planktonic Eukaryotes be uncovered using culture methods?

Student Name: Rae Chua (U1640646L)

ABSTRACT

Phytoplankton have key ecological significance and are suited to be bioindicators. There is a need to improve conventional morphology-based biomonitoring approaches and evaluate metabarcoding as a biomonitoring tool. This study aims to understand the culturable diversity of phytoplankton from Singapore waters. Several isolation strategies were used, including serial dilution, flow cytometry cell-sorting and single-cell pipette isolation, on marine samples collected from Singapore and Johor Straits. Isolated strains were genetically characterized by 18s rRNA sequencing and sequences were phylogenetically analyzed. Sequences obtained were compared against environmental metabarcoding data. A total of 22 strains were isolated, spanning 4 divisions- Chlorophyta, Cryptophyta, Dinoflagellata, Ochrophyta. Phylogenetic analysis revealed that several strains represented uncultured sequences in databases. Comparison with metabarcoding data demonstrated that culture methods uncover diversity not detected by metabarcoding. Culture methods can uncover the diversity of phytoplankton in Singapore waters and should be a continuous effort, to improve biomonitoring approaches and applicability.

INTRODUCTION

Morphology-based Biomonitoring

Marine planktonic communities can be divided by taxonomic groups and size fractions. They are composed of representatives from all domains of life- archaea, bacteria and eukaryota. Phytoplankton are a functional group of single-celled organisms, including groups of eukaryotes and prokaryotes like cyanobacteria, with photosynthetic capabilities. These cells span three orders of magnitude, ranging from mesoplankton (0.2-2mm) to picoplankton (0.2-2uM)(Not et al., 2012). In the marine environment, phytoplankton hold key ecological significance. Given their photosynthetic nature and abundance, they are the basis of marine food webs, contributing an estimated of 50% of global primary production (Field et al., 1998). Several works have shown that functional groups of phytoplankton are key players of major biogeochemical cycles, like carbon, phosphorous, nitrogen, silica and oxygen (Basu & Mackey, 2018; P. Falkowski, 2012; P. G. Falkowski, 1994; Litchman et al., 2015; Smith et al., 2014; Takao et al., 2020).

Since phytoplankton underpin many ecological processes and respond quickly and strongly to fluctuations in biotic and abiotic environmental conditions (Bode et al., 2015; Righetti et al., 2019), their use as reliable bioindicators, in both freshwater and marine ecosystems, has increased in recent years (Allende et al., 2019). Conventionally, morphology-based biomonitoring tools are used to assess environmental health. However, reviews and comparative studies have cited several drawbacks (Cordier et al., 2019). Traditional morphology-based biomonitoring involves use of light microscopy and taxonomy expertise to identify and enumerate species found in environmental samples (Weithoff & Beisner, 2019). There are three main factors hampering the effectiveness of morphology-based biomonitoring (McManus & Katz, 2009; Vuorio et al., 2020).

First, it is labor-intensive and time-consuming. Environmental samples are individually screened under a light microscope, often with epifluorescence, and cells are counted. Samples would also have to be processed for durability, either by preservation or culturing. The time-lag between sampling and results generation reduces the effectiveness of biomonitoring and rules out the possibility of implementing early-warning systems (Cristescu, 2014). Second, it relies on highly-trained individuals with expertise to identify species-specific structural details and accurately assign species. Given that the process of microscope identification is done by a single researcher, there is an element of subjectivity in documenting observed structures that may not be uniformed across studies (McManus & Katz, 2009). Lastly, morphology-based biomonitoring is limited by visual observation. Morphological characterization can only be

applied to easily-observable species with distinguishable structures making it impossible for some groups that lack clear morphological characteristics (Bazin et al., 2014; Rodríguez-Ramos et al., 2014). For example, it would be easy to document and identify, to morphospecies, a large diatom (>20µm) with hard parts and distinct surface structures but difficult to observe a picoplankton (<2µm) like *Micromonas pusilla* that consists of a single nucleus, mitochondrion, chloroplast and flagellum. Morphology-based combined with molecular phylogenetic studies have shown that cryptic diversity (species with undifferentiated morphologies) is often hidden by morphological plasticity. Due to these drawbacks, morphology-based monitoring approach has low automatable upscaling potential to detect significant changes in the phytoplankton community at speed and quality need by biomonitoring programs (Pawlowski et al., 2016; Porter & Hajibabaei, 2018).

Metabarcoding for Biomonitoring

Growing human pressure on coastal waters worldwide, including Singapore, has increased the demand for effective and fast monitoring programs with more cost-efficient and reliable tools to assess the quality of aquatic habitats. With advancement of molecular techniques, metabarcoding is a rapid and inexpensive approach for biomonitoring (Kim et al., 2017; Obiol et al., 2020; Penna & Galluzzi, 2008; Shokralla et al., 2012; Sunagawa et al., 2015). It involves high-throughput sequencing of selected gene markers from environmental samples. In brief, DNA from cells present in the environment is extracted, a target gene is amplified by polymerase chain reaction (PCR) and the sequences obtained compared against an accurate database (Figure 1). For eukaryotes, including phytoplankton, the 18s rRNA gene marker is readily adopted as it is variable enough to resolve genera and sometimes species yet is flanked by conserved regions, allowing the use of universal primers (Alemzadeh et al., 2014). There is growing database of studies and related literature supporting the use of metabarcoding by evaluating its efficacy against morphology-based approaches for uncovering phytoplankton species diversity (Cahill et al., 2018; Hajibabaei et al., 2011; Hamsher et al., 2013; Vuorio et al., 2020).

The advantages of metabarcoding are multi-fold. Fundamentally, it directly overcomes the limitations of morphology-based biomonitoring. PCR and sequencing techniques can be automated, lending to the objectiveness and ease of documentation/ data storage of this approach. Automation greatly decreases processing time and hence, removes the time-lag morphology-based methods face (Cordier et al., 2019). Furthermore, there are added benefits of metabarcoding. Amplification of target genes ensures that there is enough genomic material for sequencing, enabling detection of small populations and increases likelihood of detecting

rare and low-abundance species. Harnessing these advantages would allow more, spatially and temporally, extensive biomonitoring efforts (Porter & Hajibabaei, 2018).
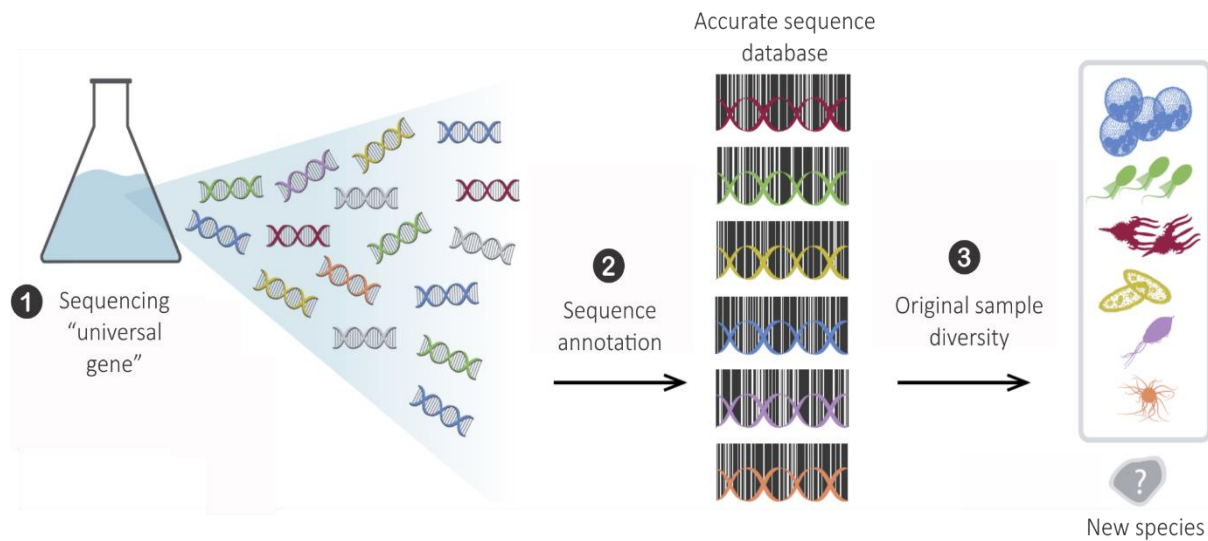

Figure 1. Schematic illustration of metabarcoding step modified from Ruiz-trillo 2018(Ruiz-trillo & Ferrer-bonet, 2018). DNA is extracted from samples obtained directly from the environment before amplification using universal gene primers. This is followed by sequencing "universal gene" and sequence annotation referencing an accurate database to assign strains to species level.

While promising, metabarcoding is highly dependent on taxon sampling and annotation accuracy of reference databases. To fully embrace metabarcoding to improve knowledge on phytoplankton diversity and ecology, there is an urgent need for high-quality reference sequences in the common-use databases (e.g. PR2 (Guillou et al., 2013)). To overcome some key challenges faced by metabarcoding, accuracy of databases must be reviewed and improved. This can be achieved by combining phylogenetic and morphology-based techniques. In fact, culture-based studies have been responsible for most of the new major lineages recently added to the eukaryotic tree. Good quality references sequences with accurate and detailed taxonomic description are needed to improve databases classifications; mostly obtained when organisms are cultivated.

Few studies have surveyed the marine phytoplankton community from Singapore waters. Tham Ah Know, 1973 described planktonic composition from Singapore waters based on microscopic observations (Tham, 1973). In 2016, also by microscope, Tan, et *al.* identified a large number of marine phytoplankton including pennate diatoms (Tan et al., 2016). These phytoplankton cells were missing from Chenard et *al.* which applied the metabarcoding approach (Chénard et al., 2019), and a great number of sequences were only assigned at family or genus level. These results prompted investigation of the culturable diversity of phytoplankton from Singapore waters.

MATERIALS AND METHODS

Sampling

From November 2019 to February 2020, marine water samples were collected from Pulau Kusu (1.2232°N, 103.8611°E), Pulau Hantu (1.2250°N, 103.7525°E) and Saint John's Island (1.2167°N, 103.8500°E) in the Singapore Strait. From the Johor Strait, sampling sites were Tanjong Chek Jawa (1.408731°N, 103.99103°E) and Pulau Ubin (1.4126°N, 103.9577°E).

Water was collected from the surface, at 1m depth, with a Niskin bottle. Collected water was filtered through a 100µm mesh filter on site to eliminate larger fractions of plankton (e.g. fish larvae and marine invertebrates). Planktonic cells were also collected with a plankton net of 20µm to concentrate plankton larger than 20µm.

Collection bottles were acid-washed once and rinsed thrice with MilliQ water the day before sampling. On site, bottles were rinsed once with seawater. Water samples were kept away from light when transported back to the laboratory and processed on the same day.

Culture Maintenance- Growth Media

Media used for cultures was prepared from aged seawater with enrichment of nutrients, trace metals, and vitamins. Seawater was collected from the surface in clean, opaque 20L carboys and aged for minimally three months. Bottles for media preparation were autoclaved at 121°C for 20 minutes beforehand. Stock solutions of nitrates, phosphates, silicates, trace metals and vitamins were prepared and autoclaved. Chemicals required to make stock solutions were obtained from Sigma Chemical Company. Aged seawater was filtered through a Millipore filter tower with 0.2µm filter, before autoclaving at 121°C for 20 minutes. Autoclaving and filtration reduce contamination of bacteria or fungi.

L1 (Guillard & Hargraves, 1993) media is a basic defined media that encourages growth of fast-growing organisms. Addition of silicates supports growth of diatoms with thick silica walls. To enhance phytoplankton yield, soil extract (1%) was added as it provides additional nitrogen and phosphorous for growth.

Soil extract was prepared as per the following protocol (Vaulot, 2019). From an easily accessible and pesticide/ pollutant free region in the Bukit Timah Nature Reserve, 10g of dry soil was collected. Mixture of soil and 400mL of MilliQ water was boiled for 1 hour and filtered

through a 0.8μm filter before filtering through a 0.2μm filter. Filtrate was transferred to a clean glass bottle.

Pre-isolation Techniques

Filtration

To target different phytoplankton size fractions, 100μm pre-filtered samples were further filtered through 20μm and 3μm polycarbonate filters (47mm diameter) by gravity, in a Millipore filtration tower. Filtered fractions were used for serial dilution or enrichment.

Enrichment

Enrichment was performed by mixing 15mL of sample filtrates (20μm and 3μm) with 1mL of media. These enrichment cultures were incubated for 2 weeks at 22°C, with a 12:12 hour light:dark cycle.

Isolation Techniques

Serial Dilution

On the sampling day, 20μm and 3μm pre-filtered water samples were seeded into 96-well deep-well plates, in two dilution fractions per size fraction. 30mL of media was mixed with 100uL of filtrate to give a ~1cell/mL dilution and 30mL of media with 1mL of filtrate for a ~100 cell/mL dilution. Each well was filled with 1mL of each dilution. One 96-well deep-well plate was seeded for each water sample (Figure 2A).

Plates were incubated at 22°C, with a 12:12 hour light:dark cycle. Over 10 days, wells were periodically checked (~every 3 days) for coloration, which would indicate sufficient phytoplankton growth for further isolation.

When sufficient growth was observed, 100μL of culture from each well was transferred to a clear-bottom 96-well plate for observation with an inverted microscope. Contents of each well were described and recorded. Wells with distinct and novel strains were selected for further serial dilution. Epifluorescence illumination with blue light allowed for confirmation of photosynthetic phytoplankton strains with chlorophyll pigments that emit red light when excited by blue light.

To further improve isolation, serial dilution was carried out by seeding 100μL of culture in 900μL of media and pipetting 100μL down a column of six wells in a 48-well plate. Each time,

increasing the dilution factor by 10X (Figure 2B). For better growth of long chains and larger diatoms, 200μL of culture and 1800μL of media was used for 24-well plates, pipetting 200μL across a row of six wells.

Serial dilution plates were periodically checked (~every 3 days) for coloration, with the inverted microscope and epifluorescence illumination. Observation, record and selection of wells with strains of interest was repeated for new dilution plates until a well appeared to contain a single strain. These wells were seeded into 10mL of media in a 50mL culture flask and sub-cultured every 2 weeks to ensure maintenance and viability of a potentially clonal culture.
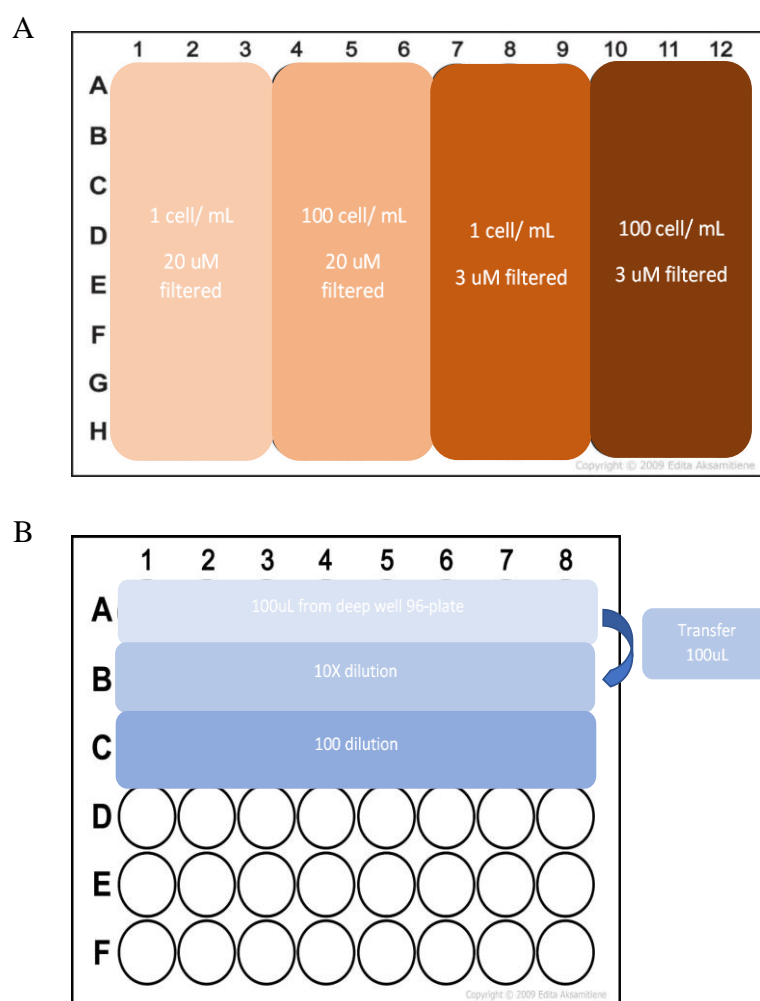

Figure 2: Isolation by serial dilution. A) Conformation of 96-well deep-well dilution plates. B) Seeding and serial dilution of a 48-well serial dilution plate.

Flow Cytometry Cell-sorting

Enrichment cultures were incubated for 2 weeks before being used for flow cytometry (FCM) cell-sorting into 48-well plates. FCM was also used for single-strain flasks, to ensure that

cultures were clonal. Protocol for FCM was optimized for the Roscoff Culture Collection (RCC) sorting of clonal strains (Marie et al., 2017).

A FACSAria cytometer (School of Biological Science, Flow Cytometer facility), equipped with a 488nm laser was used for sorting. Cultures were sorted into L1 media, consistent with media used for isolation, with 0.01% bovine albumin solution (BSA) that encourages recovery of viable cells and initiation of a viable culture. 4 wells were sorted with 100 cells to ensure that the sorting process was successful. Rest of the 44 wells were sorted with 1 cell each for establishment of a clonal culture from a single cell. Penicillin, neomycin and streptomycin (PNS) antibiotic mixture was added at a final concentration of 0.1% in each well, three days after sorting. This is to reduce bacterial contamination since sorting was done in non-sterile conditions of the flow cytometer.

Wells were monitored with an inverted microscope and/ or epifluorescence microscopy periodically (~every 3 days) for 2 weeks after sorting. When a clonal culture is initiated and established, 100μL of culture from the well was seeded into 10mL of media, and sub-cultured every 2 weeks to ensure maintenance and viability of a potentially clonal culture.

Single-cell Isolation
Plankton net samples were also used for culture and isolation. 15mL of plankton net samples were enriched with 1mL of media and incubated at 22°C, with 12:12 hour light:dark cycles for 2 weeks. Sample was then used for single-cell pipette isolation. A 2μL micropipette was fitted with a manually bent tip. Cells for isolation were observed and selected at 40X magnification. Tip was inserted into the well and observed with 5X magnification before uptake of 2μL of culture, including the targeted cells. Cells were transferred to a new well plate with 1mL of media in each well. 3 days after pipette isolation, PNS was added to a final concentration of 0.1%.

Media and incubation conditions were kept constant. Plates were monitored periodically (~every 3 days) after isolation to check for growth. Wells were then selected to be seeded into a serial dilution plate as in 'Serial Dilution'. Concurrently, pipette isolation was repeated for these wells. Observation, record and selection of wells for dilution or pipette isolation, was repeated with new plates until a well appeared to have a single strain. These wells were seeded into 10mL of media, in a 50mL culture flask, and sub-cultured every 2 weeks to ensure maintenance and viability of a potentially clonal culture

This method was also used as a complimentary isolation technique for serial dilution plates, especially for isolating chain-forming phytoplankton strains.

Microscope Observation

Microscope observation during and after isolation was performed with a CKX53 Olympus inverted light microscope. Brightfield, differential interference contrast and epifluorescence microscopy were used. Structural details captured allowed identification of some strains with distinct characteristics (e.g. presence of phycoerythrin in Cryptophytes which is bright orange under green light).

Molecular Analysis

Stable cultures were identified using partial 18S rRNA sequences and added to the Singapore Marine Strains (SMS) collection.

DNA was extracted directly from cultures by pipetting 0.2mL of fresh, health culture into PCR tubes, heated for 5 min at 95℃ and cooled to 4℃ in the PCR thermocycler. A DNA extraction with NucleoSpin Plant II kit (Macherey-Nagel) was performed for thick-walled or low-concentration strains. In brief, cells were harvested by centrifugation from 2mL of culture, with addition of 0.5µL of Pluronic F-127. Cultures were centrifuged at 11000 g for 1 minute. The supernatant was discarded in 10% bleach, and cells re-suspended in remaining media. DNA was then extracted from the pellet with the Nucleospin Plant II Kit (Machery-Nagel), following manufacturer given protocol. The concentration of extracted DNA was determined with an Invitrogen Qubit Fluorometer (Thermo Fisher Scientific).

For 18S rRNA amplification the primers 63F (5'ACGCTTGTCTCAAAGATTA3')(Lepere et al., 2011), 528F (5'-CCGCGGTAATTCCAGCTC3')(Zhu et al., 2005) and 1818R (5'ACGGAAACCTTGTTACGA3')(Lepere et al., 2011) were used. The combination of 528F and 1818R was used for PCR directly performed from cells. PCR amplification was performed in a 10µL mix containing 5µL of Phusion High-Fidelity PCR Master Mix® 2×, 0.3µM final concentration of primers, 1µL of DNA and H_2O. Thermal conditions were: 98℃ for 5min, followed by 35 cycles of 98℃ for 20s, 55℃ for 30s, 72℃ for 90s, and a final cycle of 72℃ for 5 min.

PCR products were purified using the QIAquick PCR purification kit (Qiagen). In brief, 50µL of PCR products were pipetted into 1.5mL Eppendorf tubes and purified following manufacturer's instructions, with one modification at the elution step; 12.5µL of elution buffer

was added instead of 25µL to bring total eluded volume to 25µL. Purified PCR product was then quantified with a NanoDrop spectrophotometer. Sanger sequencing was then carried out with Macrogen Singapore (https://dna.macrogen-singapore.com/eng/). All the sequences obtained have been deposited to GenBank under accession numbers MT489358: MT489378.

Sequence Analysis

Genotypic Characterization- Phylogenetic Analysis

Analysis of sequences and generation of phylogenetic trees were constructed using Geneious Prime (version 2019.2.3)(Meintjes et al., 2012). Consensus sequences were built from single-stranded sequences, using the *DeNovo* Assemble plugin. Partial 18S rRNA (~1Kb) consensus sequences generated were compared to those available in public databases with the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLASTn). Using the MAFFT Alignment plugin, consensus sequences and matches with the highest query match, percentage identities and lowest E value were used to build a reference alignment sequence list (~800bp - ~1.6Kb). Phylogenetic trees were constructed using the FastTree plugin in Geneious.

Comparing with Metabarcoding Data

Diversity of strains isolated in the project was compared to publicly available metabarcoding data obtained in frame of Ocean Sampling Day (OSD) project (2014 and 2015)(Tragin et al., 2018a) and time series (TS) metabarcoding study carried out from February 2017 to July 2018 (Chénard et al., 2019). Fasta files with representative sequences of the operational taxonomic units published from those datasets created by sorting the datasets into Cryptophyta, Dinoflagellata, Bacilariophyta centric and Bacillariophyta pennate groups. These were chosen based on the sequencing results obtained.

Focusing on Bacillariophyta pennate, phylogenetic trees were constructed using both reference alignment and environmental sequence lists to compare strains identified. Alignments and corresponding phylogenetic trees were generated separately for OSD and TS datasets as different regions of the 18s rRNA gene were sequenced. The OSD sequenced the V4 region while the TS sequenced the V6 region of the 18S rRNA gene. When combined with reference alignment sequence lists, gave alignments of ~400bp and ~500bp respectively.

RESULTS & DISCUSSION

Isolated Strains- Overall Diversity

In this study, by combining three pre-isolation and three isolation techniques, 22 strains assigned to four taxa were isolated from Singapore waters. Strains were incorporated into the Singapore Marine Strains (SMS) culture collection and assigned a reference number (Table 1). **SMS54** and **SMS56** did not match sequences from existing cultured strains in GenBank. **SMS57** and **SMS75** matched environmental sequences from uncultured eukaryotes in GenBank. Strains belonged to four divisions- Chlorophyta (1), Cryptophyta (2), Dinoflagellata (1), Ochrophyta (18). Ochrophyta was further analyzed at family level, dividing strains into Centric and Pennate diatoms; belonging to the class Bacillariophyta. The most abundant group represented was diatoms with genera *Minutocellus* (4), *Skeletonema* (2) and *Thalassiosira* (5) for Centric diatoms. Pennate diatom genera included *Cymbella* (1), *Navicula* (2), *Nitzchia* (1) and *Cylindrotheca* (3). The Chlorophyta genus *Picochlorum*, Cryptophyta genera *Proteomonas* and *Hemiselmis*, and Dinoflagellata genus *Biecheleiopsis* were also uncovered. The relative abundance of these four phytoplankton divisions mirror the abundance of main phytoplankton players in natural marine environments (Effendi et al., 2016; Engelen et al., 2015; Pierella Karlusich et al., 2020), with diatoms being the most abundant (Malviya et al., 2016).

All the strains in this work were isolated by traditional methods such as serial dilution or combination of single cell with further serial dilution. FCM was only efficient to obtain clonal strains. The application of FCM for phytoplankton isolation and cultivation is believe to be an intermediate solution to overcome inefficiencies of traditional isolation techniques (serial dilution and single-cell pipetting). However, the success is compromised by potential cell-damage caused by physical stresses like fluidic pressure, exposure to laser beam, deflection on metal plates, interaction with high-voltage fields and collision with well surfaces. Marie et. *al.* demonstrated that the addition of BSA enabled better cell recovery and there were genus-specific optimum concentrations. Another consideration cited was media choice; environmental taxa have unknown nutritional requirements and groups can be outcompeted by fast-growing species. Hence, there should be a variety of media used to uncover greater diversity (Marie et al., 2017). Given that this study did not compare different BSA concentration or media, future experiments should aim to improve FCM protocol for tropical waters.

Table 1: List of isolated strains, with isolation details, best BLASTn match and assigned SMS reference number.

Taxa	Strains	ID	Isolation Location	Isolation Fraction	BLAST best match	Accession Number
Centric	SMS39	*Thalassiosira sp.*	Pulau Ubin	20 um/ 1 cell/mL	*Thalassiosira allenii*	HM991688
	SMS55	*Minutocellus sp.*	Pulau Kusu	3 um, 1 cell/ mL	*Minutocellus* sp. P1 12-8-13 FL LS-H1	MF001989
	SMS57	*Arcocellus sp.*	Pulau Hantu	20 um, 100 cell/ mL	*Arcocellulus mammifer* CCMP132	HQ912569
	SMS58	*Thalassiosira sp.*	Pulau Hantu	20 um, 1 cell/ mL	*Thalassiosira oceanica* CCMP1001	DQ514878
	SMS59	*Thalassiosira sp.*	Pulau Hantu	20 um, 100 cell/ mL	*Thalassiosira profunda* RCC4663	MN528651
	SMS60	*Minutocellus sp.*	Pulau Hantu	20 um, 100 cell/ mL	*Minutocellus* sp. P1 12-8-13 FL LS-H1	MF001989
	SMS61	*Thalassisora sp.*	Pulau Hantu	20 um, 100 cell/ mL	*Thalassiosira minima* CCMP990	DQ514876
	SMS62	*Minutocellus sp.*	Pulau Hantu	20 um, 1 cell/ mL	*Minutocellus sp*. P1 12-8-13 FL LS-H1	MF001989
	SMS63	*Thalassiosira sp.*	Pulau Hantu	20 um, 100 cell/ mL	*Thalassiosira oceanica* CCMP1001	DQ514878
	SMS74	*Skeletonema sp.*	Chek Jawa	20 um/ 1 cell/mL	*Skeletonema tropicum*	KU363218
	SMS75	*Skeletonema sp.*	Chek Jawa	20 um/ 100 cell/mL	uncultured *Skeletonema* clone st1-ske-5	KY817215
Pennate	SMS41	*Cylindrotheca sp.*	Pulau Ubin	20 um/ 1 cell/mL*	*Cylindrotheca closterium* MGB0501	DQ019446
	SMS45	*Nitzchia sp.*	Pulau Ubin	20 um/ I cell/mL*	*Cylindrotheca closterium*	AF289049
	SMS53	*Cylindrotheca sp.*	Pulau Hantu	3 um/ 1 cell/mL	*Cylindrotheca gracilis* strain TA46	KY320374
	SMS54	*Navicula sp.*	Pulau Kusu	20um, 100 cell/ mL	*Navicula* sp. FLMan1 capNavA7	MN977831
	SMS56	*Navicula sp.*	Pulau Kusu	20um, 100 cell/ mL	*Navicula sp*. FLMan1 capNavA7	MN977831

	SMS64	*Cylindrotheca sp.*	Pulau Hantu	20 um, 1 cell/ mL	*Cylindrotheca closterium* UPMC-A0076	MH166733
	SMS65	*Nitzchia sp.*	Pulau Hantu	20 um, 1 cell/ mL	*Nitzschia paleaeformis* TA394	KY320383
Chlorophyta	SMS40	*Picochlorum sp.*	Pulau Ubin	20 um/ 1 cell/mL	*Picochlorum maculatum* isolate DHmm1W1	KU561115
Cryptophyta	SMS50	*Hemiselmis sp*	Pulau Ubin	20 um/ 100 cell/mL*	*Hemiselmis* sp. RCC5942	MK295687
	SMS51	*Proteomonas sp*	Pulau Kusu	3 um/ 1 cell/ mL	*Proteomonas* sp. CCMP2715	MK828429
Dinoflagellata	SMS52	*Biecheleriopsis sp.*	Pulau Kusu	3 um/ 1 cell/mL	*Biecheleriopsis adriatica* strain trd278-kt	LC068843

Accession numbers and sequence names were obtained from running the generated sequence through a BLASTn search with GenBank, the top hit was chosen. Names of strains were incorporated into the SMS was assigned after phylogenetic analysis was done.

* Isolated and stable cultures were further made clonal by FCM cell-sorting.

Isolated Strains- Phylogenetic Analysis of Culture Diversity

Centric Diatoms

SMS39, SMS55, **SMS57**, **SMS58**, **SMS59**, **SMS60**, **SMS61**, **SMS62**, **SMS63**, **SMS74** and **SMS75** were identified to belong to the Polar-centric-Mediophyceae and Radial-centric-basal-Coscinodiscophyceae families; class Bacillariophyceae. This was the most abundant group of isolates.

The *Minutocellus sp.* genotype, represented by **SMS55**, was cultured from Pulau Kusu. Sequence obtained grouped with *Minutocellus sp.* (GenBank MF001989) isolated from coastal South Africa with 100% percentage identity (Dąbek et al., 2019)(Figure 3).

SMS60 and **SMS62** represent the *Minutocellus sp.* genotype, cultured from Pulau Hantu. Both are likely to be the same species given that reference sequences obtained from BLASTn were similar, with top hits for query cover, E value and percentage identity being the same. They grouped with *M. polymorphus* strains and *Arcocellulus cornucervis* with 100% percentage identity (GenBank KU561128). The two *M. polymorphus* (GenBank KY054966, KY054967) and *A. cornucervis* strains were obtained from Zhanjiang coast, China. Additionally, all other reference sequences belong to the *Minutocellus* genus, hence lending more confidence to assigning **SMS60** and **SMS62** as *Minutocellus sp.* instead. This could indicate that

18

A.cornucervis was wrongly assigned (Figure 3). Hence, exemplifying that culture-based methods supported with phylogenetic analysis are needed to identify these inaccuracies and rectify them. Re-assignments can be done with sufficient morphological evidence (Chen et al., 2007; Wang et al., 2016; Zingone et al., 2005).

SMS57 was cultured from Pulau Hantu. The 18s rRNA sequence matched, with 100% percentage identity, an uncultured Stramenopile clone obtained from a subtropical coastal ecosystem in Taiwan (GenBank KU743647). Close geographical proximity and climate similarity to Singapore suggests that it could be the same strain uncovered by metabarcoding, which is now available in culture (Figure 3). **SMS57** could henceforth be characterized, described and incorporated into the database, increasing the assignment ability for future matches (Bazin et al., 2014).

Skeletonema tropicum **SMS74** was cultured from Chek Jawa. It shared 100% percentage identity with *S. tropicum* strains B205 and B210 (GenBank DQ396515, DQ396516) from Urguguay (Sarno et al., 2007) as well as *S. tropicum* from Kaoshiung Habor, Taiwan (GenBank KU363218). From Chek Jawa the *Skeletonema sp.* genotype is represented by **SMS75** which grouped with *S. grevilleii* from Fukuoka, Japan (GenBank AB948143) and an uncultured Skeletonema sequence from Keelung, Taiwan. The geographical proximity and climate similarity between and with the latter two suggest that **SMS75** is closely related to *S. grevilleii* but could be an undescribed strain (Figure 3).

SMS58 and **SMS63** represent the *Thalassiosira sp.* genotype cultured from Pulau Hantu. Both are likely to be the same species given that reference sequences obtained from BLASTn were similar, with top hits for query cover, E value and percentage identity being the same. With only 99.90% percentage identity, they grouped with from *T. oceanica* from Seoul, South Korea (GenBank HM991696). Sequences match, with 100% percentage identity, with *T. oceanica* (DQ514878), isolated from the North Atlantic Ocean (Alverson et al., 2007) (Figure 3).

The *Thalassiosira sp.* genotype **SMS61** also obtained from Pulau Hantu clustered with several *Thalassiosira* strains, but it seemed to be close related to *T. minima* strain CCMP990 (GenBank DQ514876) for which the isolation location is unknown. **SMS59**, another *Thalassiosira sp.* genotype obtained from Pulau Hantu form a clade with three *T. profunda* strains. *T. profunda* strain X9III12 (GenBank KC284713), *T. profunda* strain RCC5883 and strain RCC4663 (MN52865, MN528654) isolated from the Western English Chanel form a sub-clade; suggesting that **SMS59** might be an undescribed species of the *Thalassiosira* genus (Figure 3).

SMS39, the *Thalassiosira sp*. genotype from Pulau Ubin shared 100% percentage identity with *T. allenii* obtained from Seoul, South Korea (GenBank HM991688) (Figure 3).

Pennate Diatoms

SMS41, SMS45, SMS53, SMS54, SMS56, SMS64 and **SMS65** were identified to belong to the Raphid-Pennate and Araphid-Pennate families; class Bacillariophyceae. The phylogenetic tree is presented in Figure 4.

The *Cymbella sp*. genotype, represented by **SMS45** obtained from Pulau Ubin clustered with several *Cymbella sp.* strains and the uncultured stramenopile sequence (GenBank AY179994) from the Great Sippewisset salt marsh (Stoeck & Epstein, 2003). Within *Cymbella* clade, two sub-clades were formed, one by *C. cimbebasiae* (GenBank HQ680722, HQ680523) strains while the second by **SMS45** and two *C. cistuliformis* strains (JF90983, JF90980) isolated from the Gulf of Mexico (Rocke et al., 2013)(Figure 4).
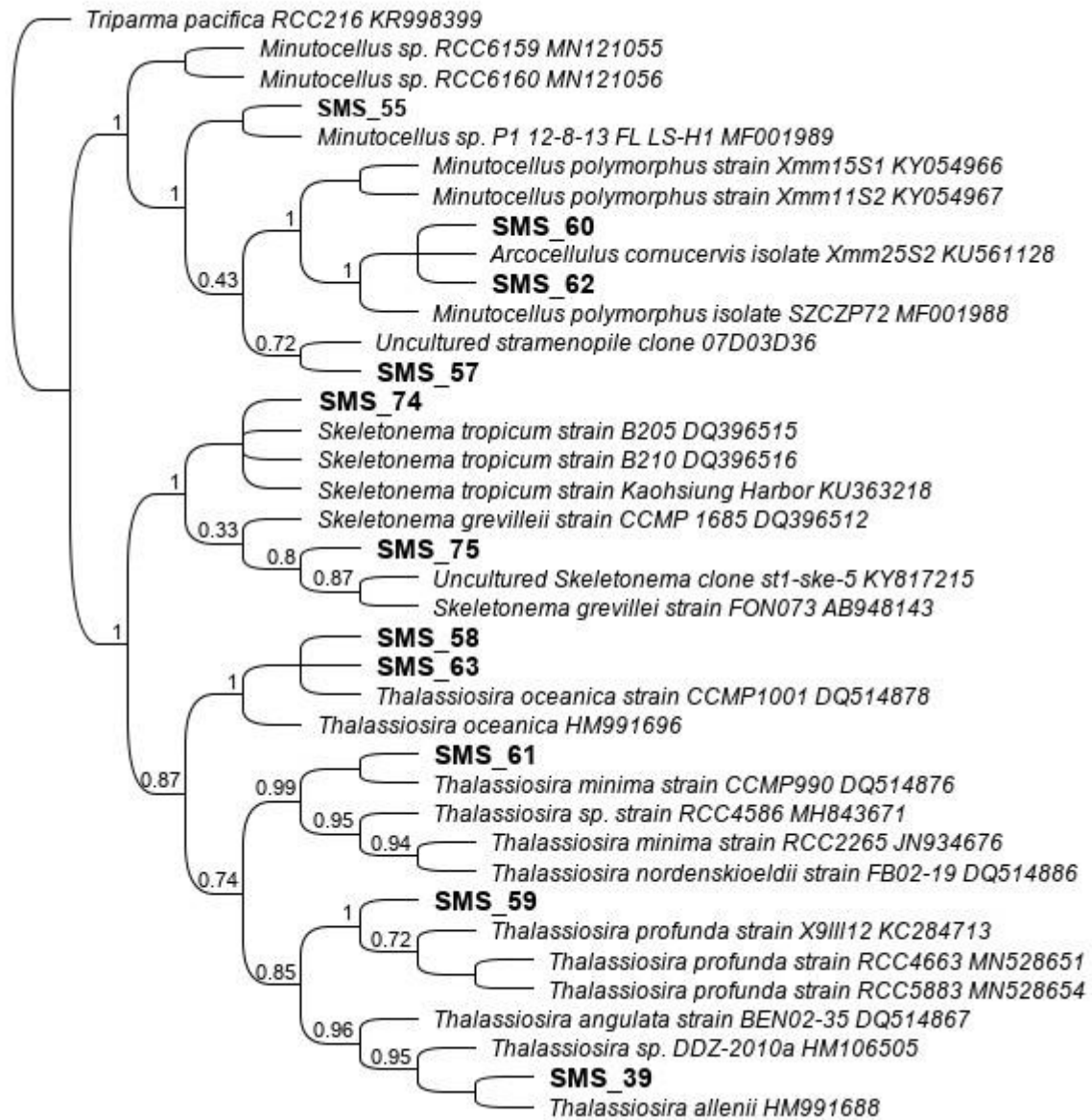
Figure 3: Phylogenetic Tree of Centric Diatoms. *Triparma pacifica* RCC216 KR998399 was used as an outgroup (Kuwata et al., 2018).

SMS54 and **SMS56** clustered with several *Navicula sp.* strains, including- *N. radiosa*, *N. lanceolata*, *N. virudula*. Sequences for *N. radiosa* and N. *lanceolata* were obtained from the PR2 reference database (Guillou et al., 2013). Within *Navicula* clade, **SMS54** and **SMS56** form a sub-clade with an uncultured eukaryote clone (GenBank GU824512) from the Cariaco Basin, Caribbean (Edgcomb et al., 2011), suggesting that these strains are undescribed, with no representatives in culture (Figure 4).

SMS65 was isolated from Pulau Hantu and represents the *Nitzchia sp.* genotype. This strain forms a cluster with an uncultured marine eukaryote sequence from the Bering Sea (GenBank KC771171) and *N. paleaeformis* and *Nitzchia cf. paleacea* (KY320380, KY320383) species.

However, **SMS65** was not closely related to any existing sequenced strain, indicating that **SMS65** might represent an undescribed strain (Figure 4).

Cylindrotheca sp. representatives, **SMS41**, **SMS53**, **SMS64** and from a cluster. **SMS53** formed an independent lineage suggesting that it might be a novel strain. **SMS41** and **SMS64** formed a clade with strong support with *C. closterium* strains including UPMC-A0076 (GenBank MH166733) isolated from Port Dickson, Malaysia (Khaw et al., 2020), MGB0501 (DQ019446) from Zhanqiao, Qingdao (Li et al., 2007) and KMMCC:B-119 from Busan, South Korea (GQ468536)(Figure 4).
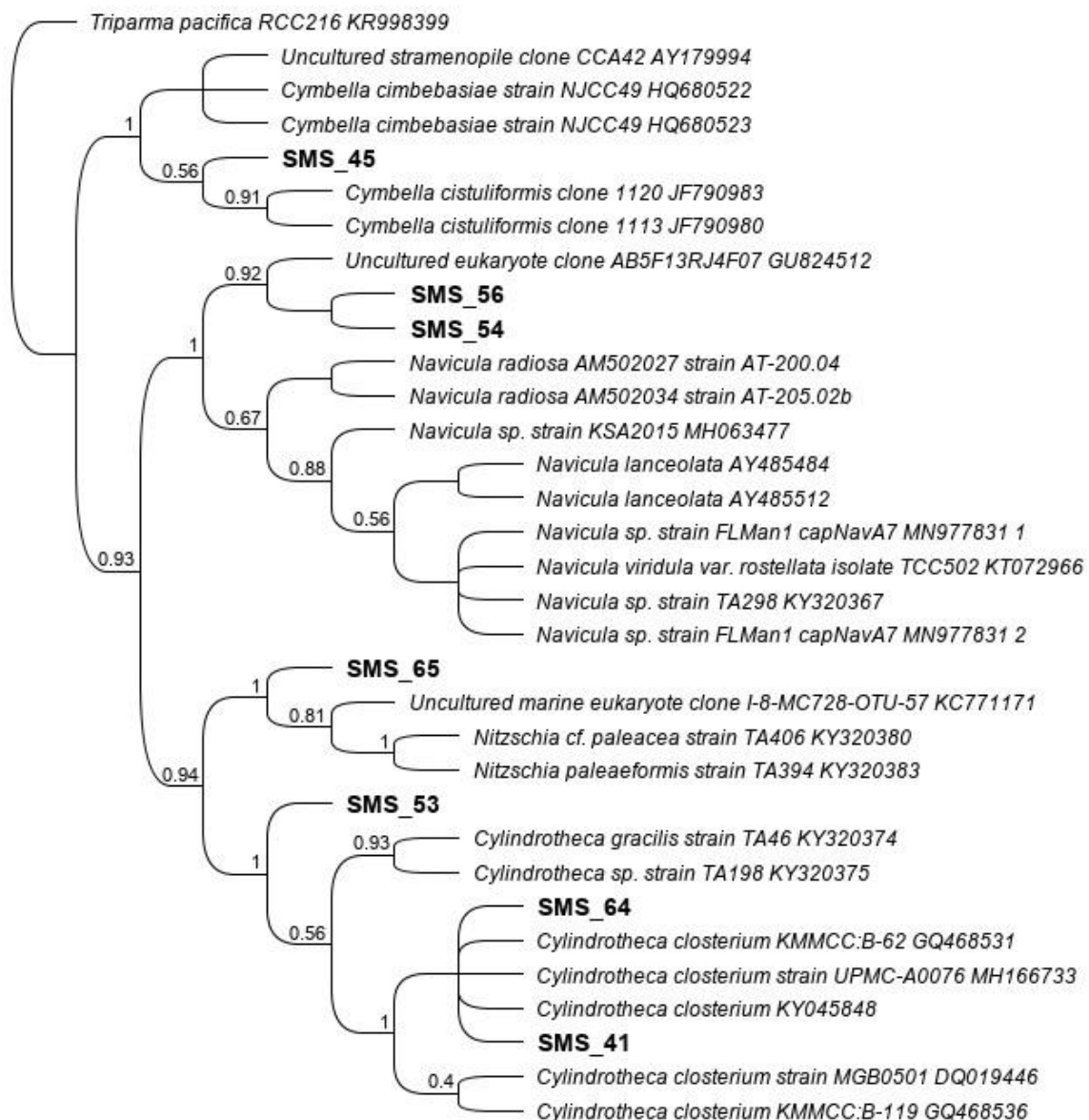

Figure 4. Phylogenetic Tree of Pennate Diatoms. *Triparma pacifica* RCC216 KR998399 was used as an outgroup (Kuwata et al., 2018).

Chlorophyta

Only **SMS40** was identified to belong to the Chlorophyta division. Given that this group is undergoing taxonomic and phylogenetic revision, reference sequences in the GenBank BLASTn database are less reliable. Reference sequence assignments were obtained from three separate papers, where mis-assignments were corrected (Foflonker et al., 2015; Henley et al., 2004; Krasovec et al., 2018)(Figure 5).

The reference sequences showed a clear distinction between freshwater and marine strains (Figure 5). Isolated from marine coastal waters off Pulau Ubin, **SMS40** form a cluster with *Picochlorum* marine and brackish water strains. However, the sequence from **SMS40** did not match the three *Picochlorum* species described today, *P. oklahomensis*, *P. maculatum* and *P. costavermella* (Figure 5).

Studies highlight that smaller fractions like picoplankton are often only detected by metabarcoding, corroborating results that only one *Picochlorum sp.* strain was obtained from this short culturing effort. Vuorio et. *al.* found that high-throughput sequencing (HTS) outperformed microscopic identification in uncovering genera represented in databases but lack morphological characteristics (Vuorio et al., 2020).

Photosynthetic marine micro-eukaryotes have known economic value, for example biotechnological applications in aquaculture. *Picocholorum* strains, like **SMS40** have high lipid and protein content. These characteristics led to the use of several *Picochlorum* strains in biotechnological applications in aquaculture (Dinesh Kumar et al., 2017) and biofuel production (Islam et al., 2013).
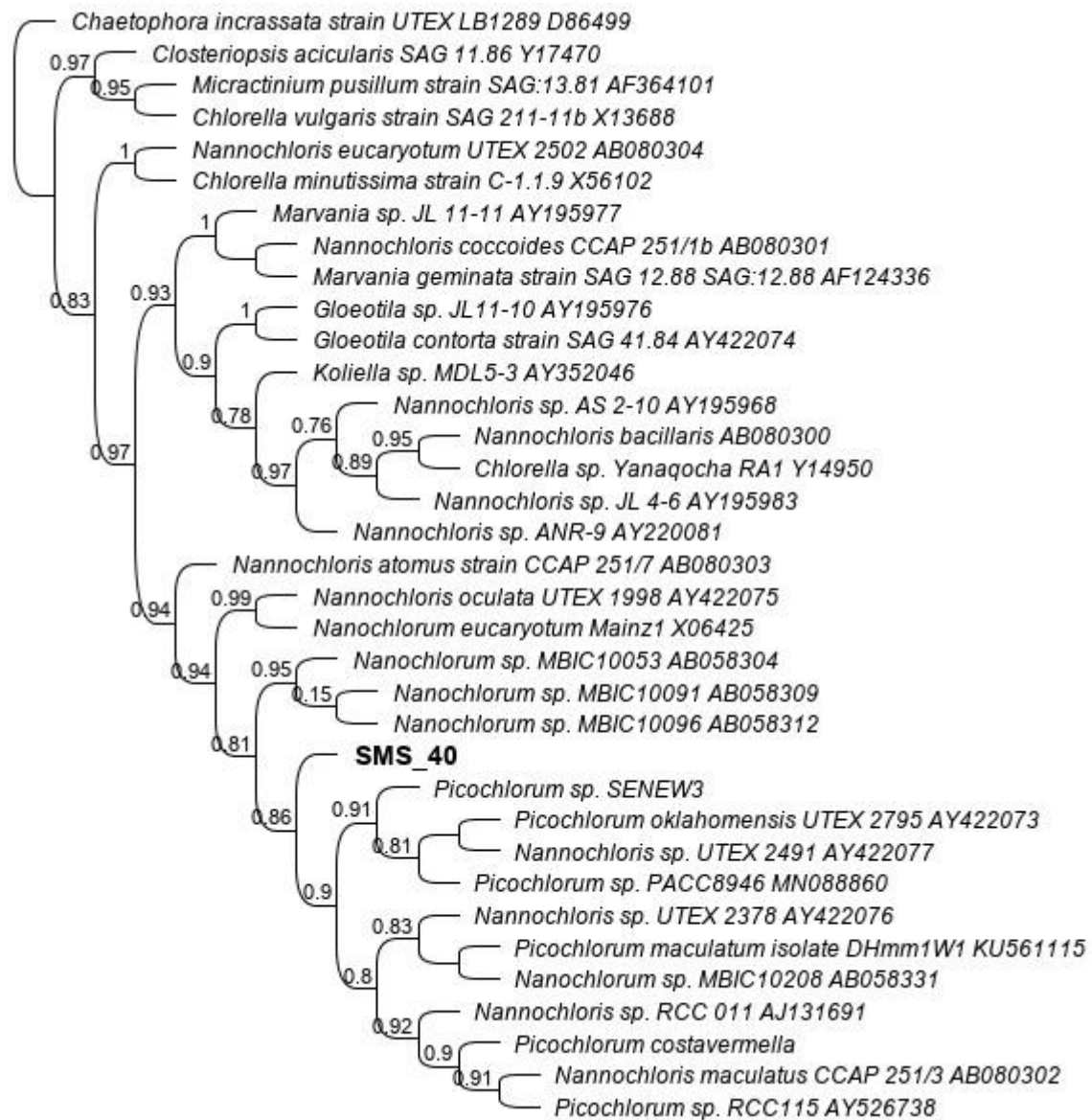
Figure 5. Phylogenetic Tree of Chlorophyta Strains. *Chaetophora incrassate* UTEX LB1289 (GenBank D86499) was used as outgroup.

Cryptophyta

SMS50 and **SMS51** were identified to belong to the Cryptophyta division. Similar to Chlorophyta, reference sequence assignments were obtained from two papers where mis-assignments were corrected (Balzano et al., 2012; Ribeiro et al., 2020)(Figure 6).

SMS50 was cultured from waters off Pulau Ubin. The sequence from **SMS50** clustered with *Hemiselmis sp.* and *Hemiselmidaceae sp.* undescribed strains isolated from temperate waters (GenBank MK295687, MK295692) (Figure 6). **SMS51** was cultured from Pulau Kusu and the sequence formed a cluster is nearly 100% identical to the sequence from the strain

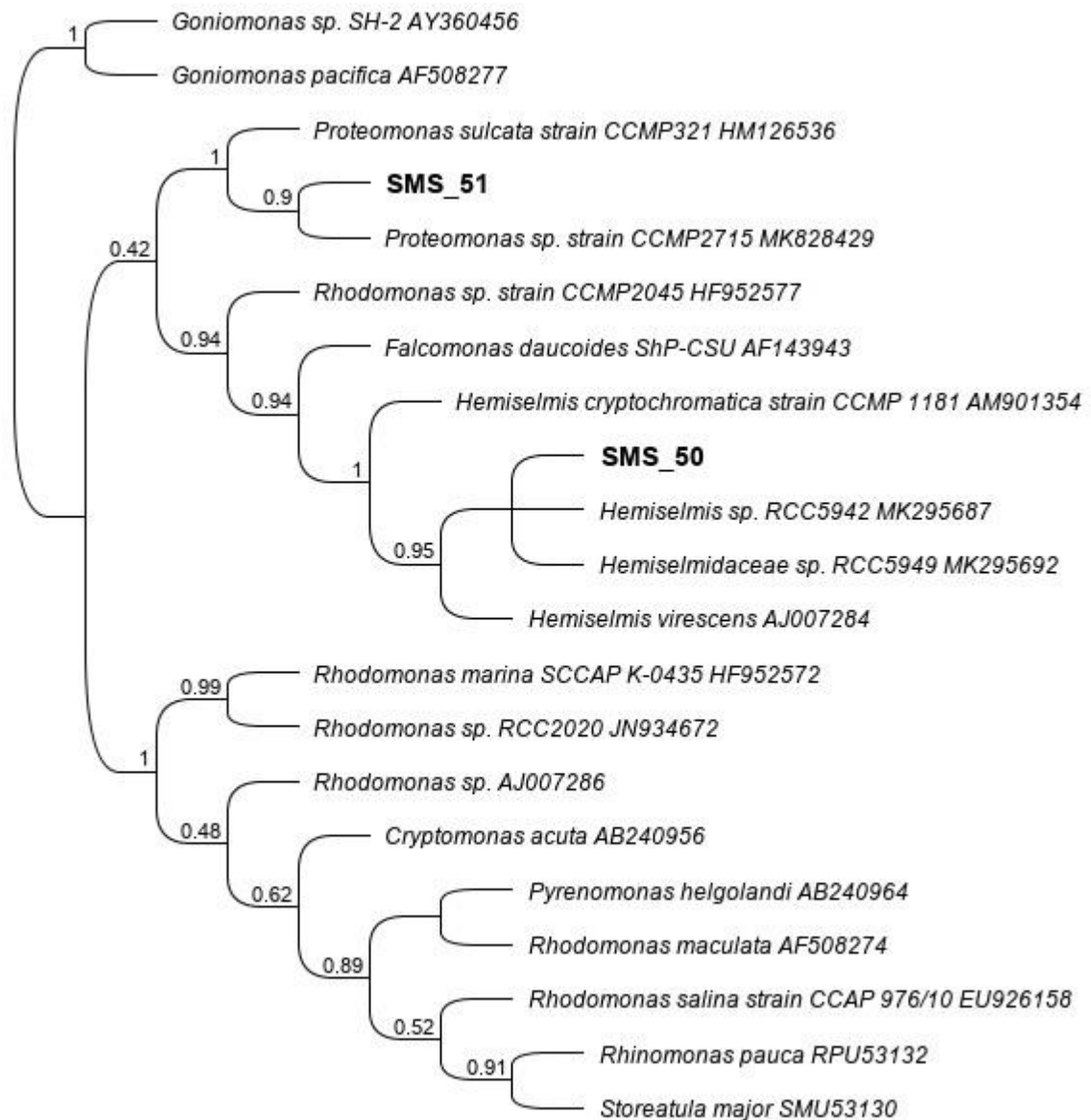*Proteomonas sp.* strain CCMP2715, isolated from Imugya Marine Garden, Miyako-Jima, Okinawa (Figure 6).

Figure 6. Phylogenetic Tree of Cryptophyta Strains. *Goniomonas* sp. SH-2 AY360456 and *Goniomonas pacifica* AF508277 were used as outgroups.

Dinoflagellata

SMS52, isolated from Pulau Kusu, was identified to belong to the Dinophyceae division. Reference sequences were obtained from two papers where mis-assignments were corrected (Balzano et al., 2012; Ribeiro et al., 2020). **SMS52** formed a strong clade with sequences belonging to the dinoflagellate species *Biecheleriopsis adriatica* (Figure 7) isolated from South Korea coastal waters (GenBank LM992904) and Japan (LC068843).

B. adriatica has been isolated and characterized from South Korea (Jang et al., 2015), China and several South East Asian regions, including Singapore (Benico et al., 2019). Interestingly, there was no conclusive molecular analysis done for the Singapore strain- which is presented here. Both studies absolve *B. adriatica* from toxic or harmful bloom involvements, indicating that it is likely to be non-harmful to Singapore marine ecosystems. However, there is still a need to investigate the spatial distribution of *B. adriatica* vegetative cells, given that many dinoflagellate species form red tides or harmful algal blooms and can be transported (as vegetative cells) by ship ballasts (Kang et al., 2019).
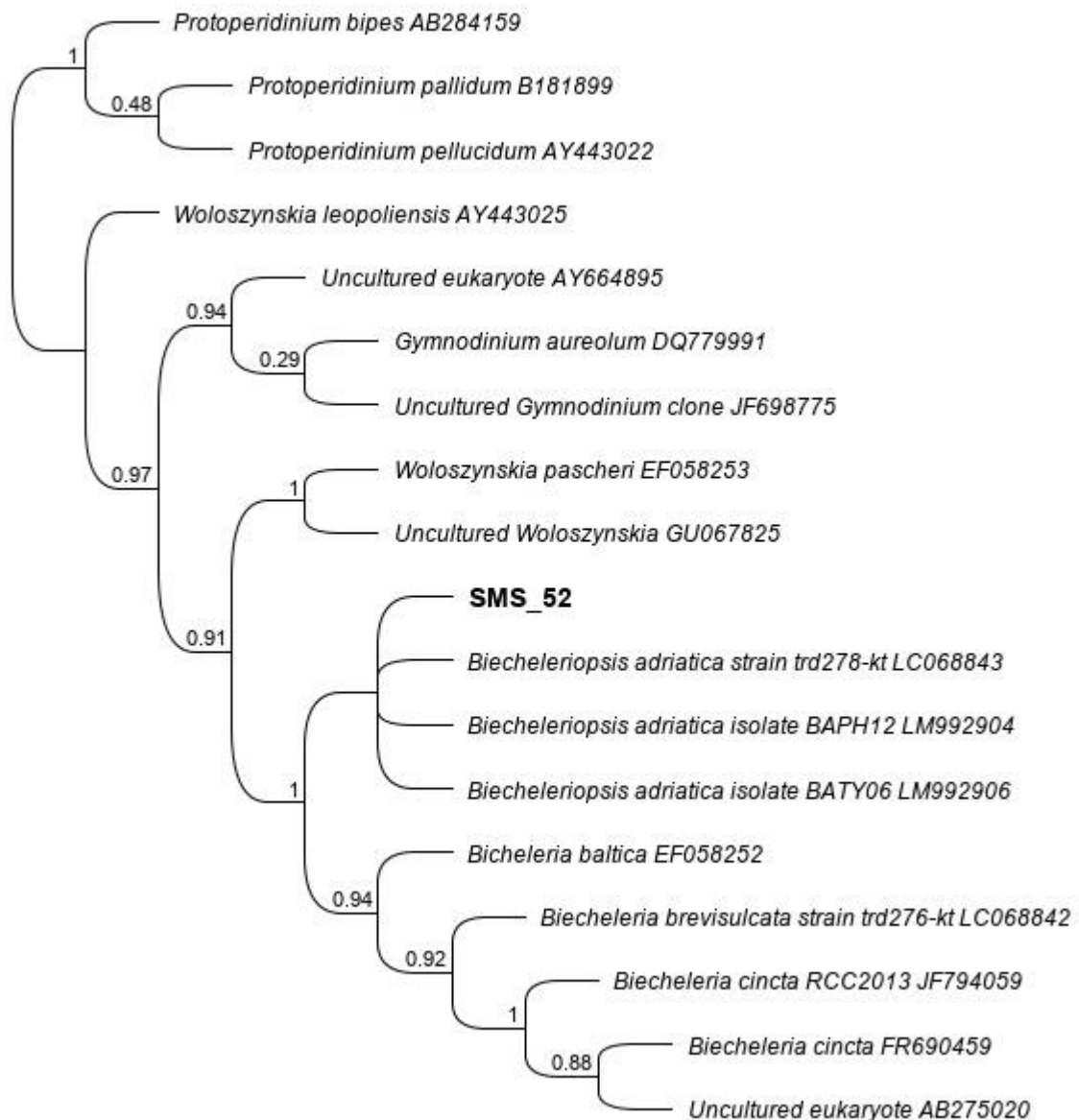

Figure 7. Phylogenetic Tree of Dinoflagellate Strains. *Protoperidinium bipes* AB284159, *Protoperidinium pallidum* B181899 and *Protoperidinium pellucidum* AY443022 were used as outgroups.

Comparing with Environmental Metabarcoding Data

Metabarcoding datasets of 18S rRNA gene variable regions 4 and 6 sequences from Singapore marine waters were obtained in the frame of Ocean Sampling Day (OSD)(Tragin et al., 2018b) project and by Chenard et *al.* 2019 (Chénard et al., 2019). The sequences from these datasets were compared with the Pennate diatom phylogenetic tree generated above. Despite being one of the most abundant groups isolated in this study (7 strains), the diversity of pennate diatoms was largely overlooked by metabarcoding with only two sequences identified in the TS dataset. In the OSD dataset, there were 21 pennate diatoms sequences identified; out of which 8 could not be assigned to genus level resolution.

Consistent with comparative studies, this highlights that metabarcoding reveals great diversity, but not all groups are successfully detected/ represented. Cahill et. *al.* revealed the discrepancies of metabarcoding with the mitochondrial cytochrome oxidase gene compared to traditional morphology to uncover marine phytoplankton diversity. They highlight that despite correspondence in overall patterns identified in both approaches, metabarcoding was not able to detect all groups equally. This is largely attributed to the limitations and inconsistencies of available reference databases, hindering accurate species-level assignments. As well as, limited sensitivity of universal primers to amplify certain groups (Cahill et al., 2018). Hamsher et. *al.* found that morphology-based survey outperforms metabarcoding in uncovering rare taxa which could be exemplified by the strains isolated in this study. The paper also cautions the possible ecological significance of rare taxa and consequences of overlooking them (Hamsher et al., 2013).

None of the two pennate sequences obtained from the TS dataset (**Seq_asv_1654** and **Seq_asv_1817**) clustered with the pennate strains we isolated (**SMS45, SMS54**, **SMS56**, and **SMS65**)(Figure 8). This result shows that by cultivation we were able complement and uncover diversity that environmental sequencing method failed to obtain. It is noteworthy that **SMS54** and **SMS56** form a single branch, independent of any reference sequence, TS or OSD sequences (Figure 4, 8 and 9), indicating that they may be novel strains that have not been described, assigned phylogeny or included in any database.

Among the 21 pennate sequences identified in OSD database, **OSD_Seq_226** and **OSD_Seq_468** formed a cluster with **SMS65** (Figure 9). These two reads were assigned at family level in this database, confirming that **SMS65** might represent an undescribed species of pennate diatoms. It also exemplifies that without proper reference sequence database, the

assignation of metabarcoding datasets cannot be performed with sufficient resolution. Further phylogenetic and morphological studies will be required to identify the taxonomy of **SMS65**. **OSD_Seq_316** and **OSD_Seq_139** sequences clustered with the sequences from **SMS54** and **SMS56** an uncultured eukaryote clone (GenBank GU824512), confirming that these are undescribed strains that have been isolated for the first time (Figure 9).

This indicates potential for culture methods to uncover entirely novel strains for further characterization, contributing to the robustness of databases. Especially in tropical regions like Singapore, with little seasonal variation in temperature and other abiotic factors, species resolution is important to represent or uncover any seasonality in our waters. Through an extensive analysis of DNA reference libraries- Barcode of Life Data Systems (BOLD) and NCBI GenBank- Weigand et. *al.* identified gaps hindering biomonitoring with metabarcoding for aquatic biota. This paper attributes the limited application of metabarcoding for biomonitoring to be the lack of a robust reference library; it acknowledges the need for a well-curated, stringent database. Lack of described sequences and culturability of strains are cited as underlying challenges (Weigand et al., 2019).

Future Direction

Combining Morphology-based and Molecular Methods

Studies comparing morphology-based and molecular methods indicate the potential of metabarcoding for biomonitoring. Groendahl, Kahlert and Fink, propose the use of both metabarcoding and morphological identification to assess microalgae after demonstrating how both methods were not able to detect all species in mesocosm biofilms (Groendahl et al., 2017). McManus and Ktaz, review literature documenting comparisons of the methods and recommend that molecular approaches should be advanced, while using morphology-based identification to validate species identification (McManus & Katz, 2009). The need for a coordinated advancement of molecular and morphology-based approaches is addressed by Cristescu; this cooperation should guide future research to fully capitalize on well-developed taxonomic knowledge and extensive sequences databases (Cristescu, 2014). Jang et. *al.* demonstrated how morphological, swimming speed and pigment analyses improved the species resolution of *Biecheleiopsis adriatica,* distinguishing strains isolated in Korea from the Adriatic Sea (Jang et al., 2015).

These studies also highlight the need for cooperation, in order to achieve several goals that would advance phytoplankton research and application to biomonitoring efforts. Firstly, to bridge the gap between taxonomic and genetic information to uncover the true phytoplankton

diversity. Thus, reducing the oversight of cryptic and rare taxa that could be economically and ecologically significant. Secondly, to improve molecular and metabarcoding approaches, high-quality databases need to be curated. Culturing and morphological characterization with phylogenetic analyses would increase the accuracy and resolution of genus/ species assignments. Thirdly, to focus on uncovering cryptic diversity and classify morphologically indistinct strains; largely enabled by molecular sequencing efforts.

Sarno et. *al.* exemplifies how both methods would practically tie in with each other. Cryptic diversity was unveiled within the *Skeletonema costatum* species assignment that was initially described by morphological methods. New distinct species were phylogenetically classified by their rDNA sequences and culture methods enabled detailed taxonomy to be described for several new species (Sarno et al., 2007).

Machine Learning

The inability to index sequences to environmental indicator values heavily limits the use of metabarcoding, without linking sequences to taxonomy, for biomonitoring. (Vuorio et al., 2020). Cordier et. *al.* proposes the application of machine learning to accelerate the application of phytoplankton species composition to biomonitoring efforts. However, this paper recognizes the need for enhancing the robustness and applicability of current databases to bridge the gap between traditional and novel approaches (Cordier et al., 2018, 2019).

More pioneering studies should be done with machine learning approaches to improve association between sequences and their environmental indicator values. To do so, more sample sets of metabarcoding data must be accumulated. Despite advances in metabarcoding and machine learning, lacking reference databases remain an influential limitation (McGee et al., 2019). Hence necessitating the complementary use of molecular and morphology-based approaches to uncover phytoplankton diversity.
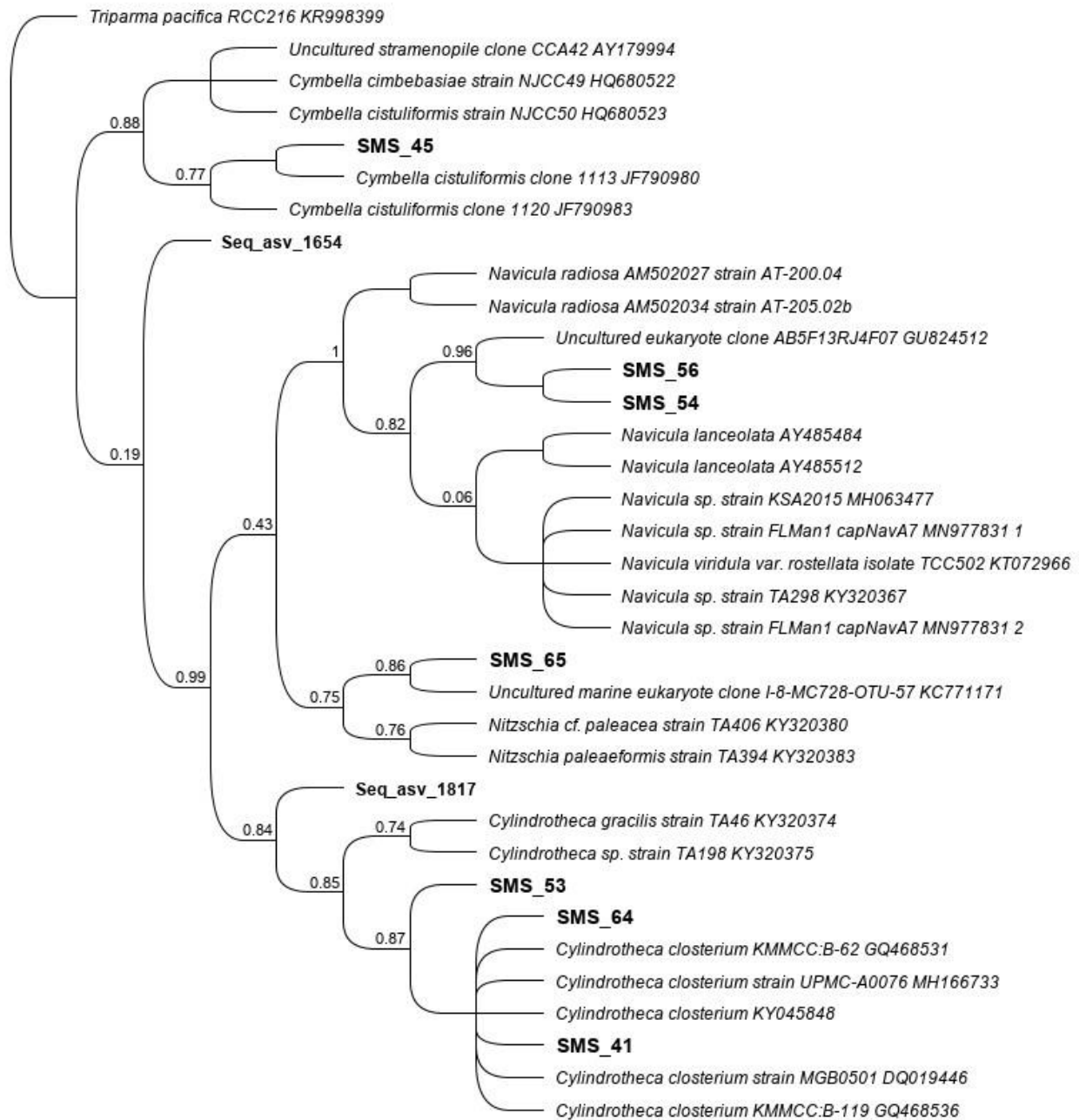
Figure 8. Phylogenetic Tree of Pennate Diatoms, with Time Series Environmental Sequences.
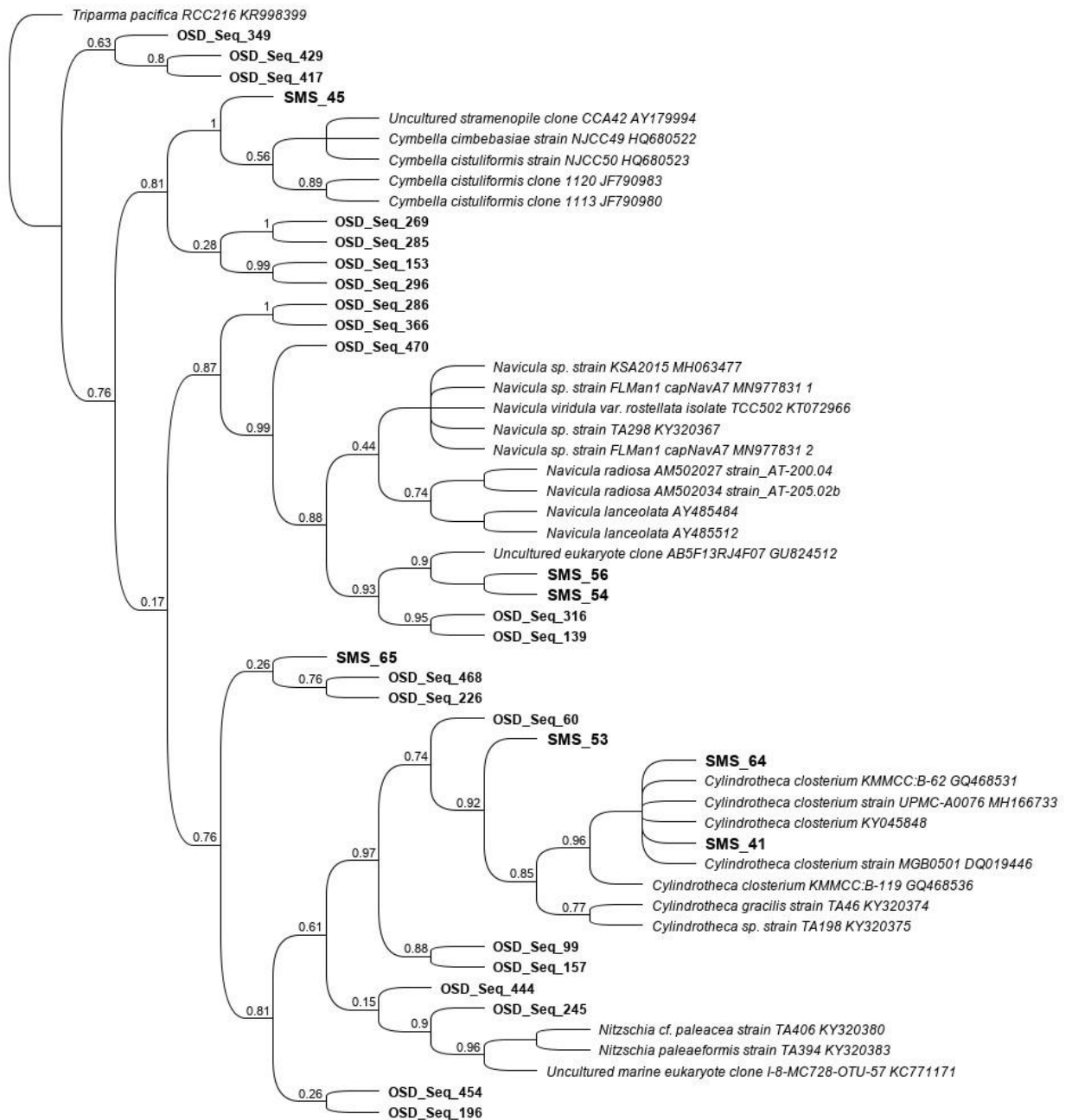
Figure 9. Phylogenetic Tree of Pennate Diatoms, with OSD Environmental Sequences.

CONCLUSION

Although most cultures obtained by this study showed sequence similarity with strains recovered elsewhere, either formerly described (e.g. *Biecheleriopsis adriatica* **SMS52**) or undescribed (e.g. *Hemiselmis* sp. **SMS50**), 3 strains (**SMS54**, **SMS56** and **SMS65**) represented completely new isolates. This highlights that culture efforts must continue, especially in under-sampled biodiversity hotspots like South East Asian marine waters.

In this study, high-throughput culture isolation by flow cytometry sorting only worked with dense mixed cultures. No strain was obtained when flow cytometry was applied to natural samples, indicating that more tests are required to establish a protocol for Singapore waters.

Climate change and population growth will impact plankton diversity, dynamics and community structures in coastal waters worldwide. As diversity within culture collections improves to reflect the complexity of the environment, an increased amount of validated reference sequences will help scientists better assess eukaryotic plankton distribution patterns across these affected ecosystems. The combination of metabarcoding and supervised machine learning has proven effective in monitoring the impact on benthic communities within and around salmon farming activities in Norway (Cordier et al., 2018).

REFERENCES

Alemzadeh, E., Haddad, R., & Ahmadi, A. R. (2014). Phytoplanktons and DNA barcoding: Characterization and molecular analysis of phytoplanktons on the Persian Gulf. *Iranian Journal of Microbiology*, *6*(4), 296–302.

Allende, L., Fontanarrosa, M. S., Murno, A., & Sinistro, R. (2019). Phytoplankton functional group classifications as a tool for biomonitoring shallow lakes: a case study. *Knowledge and Management of Aquatic Ecosystems*, *2019-Janua*(420). https://doi.org/10.1051/kmae/2018044

Alverson, A. J., Jansen, R. K., & Theriot, E. C. (2007). Bridging the Rubicon: Phylogenetic analysis reveals repeated colonizations of marine and fresh waters by thalassiosiroid diatoms. *Molecular Phylogenetics and Evolution*, *45*(1), 193–210. https://doi.org/10.1016/j.ympev.2007.03.024

Balzano, S., Gourvil, P., Siano, R., Chanoine, M., Marie, D., Lessard, S., Sarno, D., & Vaulot, D. (2012). Diversity of cultured photosynthetic flagellates in the northeast Pacific and Arctic Oceans in summer. *Biogeosciences*, *9*(11), 4553–4571. https://doi.org/10.5194/bg-9-4553-2012

Basu, S., & Mackey, K. R. M. (2018). Phytoplankton as key mediators of the biological carbon pump: Their responses to a changing climate. *Sustainability (Switzerland)*, *10*(3). https://doi.org/10.3390/su10030869

Bazin, P., Jouenne, F., Friedl, T., Deton-Cabanillas, A. F., Le Roy, B., & Véron, B. (2014). Phytoplankton diversity and community composition along the estuarine gradient of a temperate macrotidal ecosystem: Combined morphological and molecular approaches. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0094110

Benico, G. A., Takahashi, K., Lum, W. M., Yñiguez, A. T., Azanza, R. V., Leong, S. C. Y., Lim, P. T., & Iwataki, M. (2019). First report of Biecheleriopsis adriatica in Bolinao, Northwestern Philippines and its wide distribution in Southeast Asia and adjacent waters. *Philippine Journal of Natural Sciences*, *24*(March), 34–41.

Bode, A., Estévez, M. G., Varela, M., & Vilar, J. A. (2015). Annual trend patterns of phytoplankton species abundance belie homogeneous taxonomical group responses to climate in the NE Atlantic upwelling. *Marine Environmental Research*, *110*, 81–91. https://doi.org/10.1016/j.marenvres.2015.07.017

Cahill, A. E., Pearman, J. K., Borja, A., Carugati, L., Carvalho, S., Danovaro, R., Dashfield, S., David, R., Féral, J. P., Olenin, S., Šiaulys, A., Somerfield, P. J., Trayanova, A., Uyarra, M. C., & Chenuil, A. (2018). A comparative analysis of metabarcoding and morphology-based identification of benthic communities across different regional seas. *Ecology and Evolution*, *8*(17), 8908–8920. https://doi.org/10.1002/ece3.4283

Chen, G. F., Wang, G. C., Zhang, B. Y., & Fan, X. L. (2007). Morphological and phylogenetic analysis of Skeletonema costatum-like diatoms (Bacillariophyta) from the China Sea. *European Journal of Phycology*, *42*(2), 163–175. https://doi.org/10.1080/09670260601149784

Chénard, C., Wijaya, W., Vaulot, D., Lopes dos Santos, A., Martin, P., Kaur, A., & Lauro, F. M. (2019). Temporal and spatial dynamics of Bacteria, Archaea and protists in equatorial

coastal waters. *Scientific Reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-52648-x

Cordier, T., Forster, D., Dufresne, Y., Martins, C. I. M., Stoeck, T., & Pawlowski, J. (2018). Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding applied to biomonitoring. *Molecular Ecology Resources*, *18*(6), 1381–1391. https://doi.org/10.1111/1755-0998.12926

Cordier, T., Lanzén, A., Apothéloz-Perret-Gentil, L., Stoeck, T., & Pawlowski, J. (2019). Embracing Environmental Genomics and Machine Learning for Routine Biomonitoring. *Trends in Microbiology*, *27*(5), 387–397. https://doi.org/10.1016/j.tim.2018.10.012

Cristescu, M. E. (2014). From barcoding single individuals to metabarcoding biological communities: Towards an integrative approach to the study of global biodiversity. *Trends in Ecology and Evolution*, *29*(10), 566–571. https://doi.org/10.1016/j.tree.2014.08.001

Dąbek, P., Ashworth, M. P., Górecka, E., Krzywda, M., Bornman, T. G., Sato, S., & Witkowski, A. (2019). Toward a multigene phylogeny of the Cymatosiraceae (Bacillariophyta, Mediophyceae) II: morphological and molecular insights into the taxonomy of the forgotten species Campylosira africana and of Extubocellulus, with a description of two new taxa. *Journal of Phycology*, *55*(2), 425–441. https://doi.org/10.1111/jpy.12831

Dinesh Kumar, S., Santhanam, P., Ananth, S., Kaviyarasan, M., Nithya, P., Dhanalakshmi, B., Park, M. S., & Kim, M. K. (2017). Evaluation of suitability of wastewater-grown microalgae (*Picochlorum maculatum*) and copepod (*Oithona rigida*) as live feed for white leg shrimp Litopenaeus vannamei post-larvae. *Aquaculture International*. https://doi.org/10.1007/s10499-016-0037-6

Edgcomb, V., Orsi, W., Bunge, J., Jeon, S., Christen, R., Leslin, C., Holder, M., Taylor, G. T., Suarez, P., Varela, R., & Epstein, S. (2011). Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME Journal*, *5*(8), 1344–1356. https://doi.org/10.1038/ismej.2011.6

Effendi, H., Kawaroe, M., Lestari, D. F., Mursalin, & Permadi, T. (2016). Distribution of Phytoplankton Diversity and Abundance in Mahakam Delta, East Kalimantan. *Procedia Environmental Sciences*, *33*, 496–504. https://doi.org/10.1016/j.proenv.2016.03.102

Engelen, S., Hingamp, P., Sieracki, M., Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Bescot, N., Probert, I., Carmichael, M., Poulain, J., & Romac, S. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*(6237), 1261605-1/11. https://doi.org/10.1007/s13398-014-0173-7.2

Falkowski, P. (2012). Ocean Science: The power of plankton. *Nature*, *483*(7387), S17–S20. https://doi.org/10.1038/483S17a

Falkowski, P. G. (1994). The role of phytoplankton photosynthesis in global biogeochemical cycles. *Photosynthesis Research*, *39*(3), 235–258. https://doi.org/10.1007/BF00014586

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*, *281*(5374), 237–240. https://doi.org/10.1126/science.281.5374.237

Foflonker, F., Price, D. C., Qiu, H., Palenik, B., Wang, S., & Bhattacharya, D. (2015). Genome of the halotolerant green alga Picochlorum sp. Reveals strategies for thriving under fluctuating environmental conditions. *Environmental Microbiology*, *17*(2), 412–426. https://doi.org/10.1111/1462-2920.12541

Groendahl, S., Kahlert, M., & Fink, P. (2017). The best of both worlds: A combined approach for analyzing microalgal diversity via metabarcoding and morphology-based methods. *PLoS ONE*, *12*(2), 1–15. https://doi.org/10.1371/journal.pone.0172808

Guillard, R. R. L., & Hargraves, P. E. (1993). Stichochrysis immobilis is a diatom, not a chrysophyte. *Phycologia*, *32*(3), 234–236. https://doi.org/10.2216/i0031-8884-32-3-234.1

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J., Dunthorn, M., Bente, E., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Lebescot, N., Logares, R., … Christen, R. (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small SubUnit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, *41*, D597–D604.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C., & Baird, D. J. (2011). Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE*, *6*(4). https://doi.org/10.1371/journal.pone.0017497

Hamsher, S. E., LeGresley, M. M., Martin, J. L., & Saunders, G. W. (2013). A comparison of morphological and molecular-based surveys to estimate the species richness of Chaetoceros and Thalassiosira (bacillariophyta), in the Bay of Fundy. *PloS One*, *8*(10), e73521–e73521. https://doi.org/10.1371/journal.pone.0073521

Henley, W. J., Hironaka, J. L., Guillou, L., Buchheim, M. A., Buchheim, J. A., Fawley, M. W., & Fawley, K. P. (2004). Phylogenetic analysis of the "Nannochloris-like" algae and diagnoses of Picochlorum oklahomensis gen. et sp. nov. (Trebouxiophyceae, Chlorophyta). *Phycologia*, *43*(6), 641–652. https://doi.org/10.2216/i0031-8884-43-6-641.1

Islam, M. A., Magnusson, M., Brown, R. J., Ayoko, G. A., Nabi, M. N., & Heimann, K. (2013). Microalgal species selection for biodiesel production based on fuel properties derived from fatty acid profiles. *Energies*, *6*(11), 5676–5702. https://doi.org/10.3390/en6115676

Jang, S. H., Jeong, H. J., Moestrup, Ø., Kang, N. S., Lee, S. Y., Lee, K. H., Lee, M. J., & Noh, J. H. (2015). Morphological, molecular and ecophysiological characterization of the phototrophic dinoflagellate Biecheleriopsis adriatica from Korean coastal waters. *European Journal of Phycology*, *50*(3), 301–317. https://doi.org/10.1080/09670262.2015.1054892

Johnson, Z. I., & Martiny, A. C. (n.d.). *Techniques for Quantifying Phytoplankton Biodiversity*. https://doi.org/10.1146/annurev-marine-010814-015902

Kang, H. C., Jeong, H. J., Ok, J. H., You, J. H., Jang, S. H., Lee, S. Y., Lee, K. H., Park, J. Y., & Rho, J. R. (2019). Spatial and seasonal distributions of the phototrophic dinoflagellate Biecheleriopsis adriatica (Suessiaceae) in Korea: Quantification using qPCR. *Algae*, *34*(2), 111–126. https://doi.org/10.4490/algae.2019.34.5.25

Khaw, Y. S., Khong, N. M. H., Shaharuddin, N. A., & Yusoff, F. M. (2020). A simple 18S rDNA approach for the identification of cultured eukaryotic microalgae with an emphasis on primers. *Journal of Microbiological Methods*, *172*(August 2019). https://doi.org/10.1016/j.mimet.2020.105890

Kim, H., Kim, H., Hwang, H. S., & Kim, W. (2017). Metagenomic analysis of the marine coastal invertebrates of South Korea as assessed by Ilumina MiSeq. *Animal Cells and Systems*, *21*(1), 37–44. https://doi.org/10.1080/19768354.2016.1271012

Krasovec, M., Vancaester, E., Rombauts, S., Bucchini, F., Yau, S., Hemon, C., Lebredonchel, H., Grimsley, N., Moreau, H., Sanchez-Brosseau, S., Vandepoele, K., & Piganeau, G. (2018). Genome analyses of the microalga picochlorum provide insights into the evolution of thermotolerance in the green lineage. *Genome Biology and Evolution*, *10*(9), 2347–2365. https://doi.org/10.1093/gbe/evy167

Kuwata, A., Yamada, K., Ichinomiya, M., Yoshikawa, S., Tragin, M., Vaulot, D., & Santos, A. L. dos. (2018). Bolidophyceae, a sister picoplanktonic group of diatoms - A review. *Frontiers in Marine Science*, *5*(OCT), 1–17. https://doi.org/10.3389/fmars.2018.00370

Lepere, C., Demura, M., Kawachi, M., Romac, S., Probert, I., & Vaulot, D. (2011). Whole-genome amplification (WGA) of marine photosynthetic eukaryote populations. *FEMS Microbiology Ecology*. https://doi.org/10.1111/j.1574-6941.2011.01072.x

Li, H., Yang, G., Sun, Y., Wu, S., & Zhang, X. (2007). Cylindrotheca closterium is a species complex as was evidenced by the variations of rbcL gene and SSU rDNA. *Journal of Ocean University of China*, *6*(2), 167–174. https://doi.org/10.1007/s11802-007-0167-6

Litchman, E., de Tezanos Pinto, P., Edwards, K. F., Klausmeier, C. A., Kremer, C. T., & Thomas, M. K. (2015). Global biogeochemical impacts of phytoplankton: A trait-based perspective. *Journal of Ecology*, *103*(6), 1384–1396. https://doi.org/10.1111/1365-2745.12438

Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., De Vargas, C., Bittner, L., Zingone, A., & Bowler, C. (2016). Insights into global diatom distribution and diversity in the world's ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(11), E1516–E1525. https://doi.org/10.1073/pnas.1509523113

Marie, D., Le Gall, F., Edern, R., Gourvil, P., & Vaulot, D. (2017). Improvement of phytoplankton culture isolation using single cell sorting by flow cytometry. *Journal of Phycology*, *53*, 271–282. https://doi.org/10.1111/jpy.12495-16-076

McGee, K. M., Robinson, C. V., & Hajibabaei, M. (2019). Gaps in DNA-Based Biomonitoring Across the Globe. *Frontiers in Ecology and Evolution*, *7*(September), 1–7. https://doi.org/10.3389/fevo.2019.00337

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. https://doi.org/10.1093/plankt/fbp061

Meintjes, P., Duran, C., Kearse, M., Moir, R., Wilson, A., Stones-Havas, S.-V., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Thierer, T., Ashton, B., & Heled, J. (2012). Geneious Basic: An integrated and extendable desktop software platform for the

organization and analysis of sequence data. *Bioinformatics*.

Not, F., Siano, R., Kooistra, W. H. C. F., Simon, N., Vaulot, D., & Probert, I. (2012). Diversity and Ecology of Eukaryotic Marine Phytoplankton. In *Advances in Botanical Research* (Vol. 64). Elsevier. https://doi.org/10.1016/B978-0-12-391499-6.00001-3

Obiol, A., Giner, C. R., Sánchez, P., Duarte, C. M., Acinas, S. G., & Massana, R. (2020). A metagenomic assessment of microbial eukaryotic diversity in the global ocean. *Molecular Ecology Resources*, *January*, 1–14. https://doi.org/10.1111/1755-0998.13147

Pawlowski, J., Lejzerowicz, F., Apotheloz-Perret-Gentil, L., Visco, J., & Esling, P. (2016). Protist metabarcoding and environmental biomonitoring: Time for change. *European Journal of Protistology*, *55*, 12–25. https://doi.org/10.1016/j.ejop.2016.02.003

Penna, A., & Galluzzi, L. (2008). PCR techniques as diagnostic tools for the identification and enumeration of toxic marine phytoplankton species. *NATO Science for Peace and Security Series A: Chemistry and Biology*, 261–283.

Pierella Karlusich, J. J., Ibarbalz, F. M., & Bowler, C. (2020). Phytoplankton in the Tara Ocean . *Annual Review of Marine Science*, *12*(1), 233–265. https://doi.org/10.1146/annurev-marine-010419-010706

Porter, T. M., & Hajibabaei, M. (2018). Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. *Molecular Ecology*, *27*(2), 313–338. https://doi.org/10.1111/mec.14478

Ribeiro, C. G., Dos Santos, A. L., Gourvil, P., Le Gall, F., Marie, D., Tragin, M., Probert, I., & Vaulot, D. (2020). Culturable diversity of Arctic phytoplankton during pack ice melting. *Elem Sci Anth*, *8*(1), 6. https://doi.org/10.1525/elementa.401

Righetti, D., Vogt, M., Gruber, N., Psomas, A., & Zimmermann, N. E. (2019). Global pattern of phytoplankton diversity driven by temperature and environmental variability. *Science Advances*, *5*(5), 1–11. https://doi.org/10.1126/sciadv.aau6253

Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., Dmitrieff, E., Malmstrom, R., Stepanauskas, R., & Woyke, T. (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nature Protocols*, *9*(5), 1038–1048. https://doi.org/10.1038/nprot.2014.067

Rocke, E., Jing, H., & Liu, H. (2013). Phylogenetic composition and distribution of picoeukaryotes in the hypoxic northwestern coast of the Gulf of Mexico. *MicrobiologyOpen*, *2*(1), 130–143. https://doi.org/10.1002/mbo3.57

Rodríguez-Ramos, T., Dornelas, M., Marañeón, E., & Cermeñeo, P. (2014). Conventional sampling methods severely underestimate phytoplankton species richness. *Journal of Plankton Research*, *36*(2), 334–343. https://doi.org/10.1093/plankt/fbt115

Ruiz-trillo, I., & Ferrer-bonet, M. (2018). ¿Con quién compartimos el planeta? *Investigacion y Ciencia*, 56–60.

Sarno, D., Kooistra, W. H. C. F., Balzano, S., Hargraves, P. E., & Zingone, A. (2007). Diversity in the genus Skeletonema (bacillariophyceae): III. Phylogenetic position and morphological variability of Skeletonema costatum and Skeletonema grevillei, with the

description of Skeletonema ardens sp. nov. *Journal of Phycology*, *43*(1), 156–170. https://doi.org/10.1111/j.1529-8817.2006.00305.x

Shokralla, S., Spall, J. L., Gibson, J. F., & Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, *21*(8), 1794–1805. https://doi.org/10.1111/j.1365-294X.2012.05538.x

Sieracki, M. E., Poulton, N. J., Jaillon, O., Wincker, P., de Vargas, C., Rubinat-Ripoll, L., Stepanauskas, R., Logares, R., & Massana, R. (2019). Single cell genomics yields a wide diversity of small planktonic protists across major ocean ecosystems. *Scientific Reports*, *9*(1), 1–11. https://doi.org/10.1038/s41598-019-42487-1

Smith, J. M., Chavez, F. P., & Francis, C. A. (2014). Ammonium uptake by phytoplankton regulates nitrification in the sunlit ocean. *PLoS ONE*, *9*(9). https://doi.org/10.1371/journal.pone.0108173

Stoeck, T., & Epstein, S. (2003). Novel eukaryotic lineages inferred from small-subunit rRNA analyses of oxygen-depleted marine environments. *Applied and Environmental Microbiology*, *69*(5), 2657–2663. https://doi.org/10.1128/aem.69.5.2657-2663.2003

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., D'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., … Bork, P. (2015). Structure and function of the global ocean microbiome. *Science*, *348*(6237), 1–10. https://doi.org/10.1126/science.1261359

Takao, S., Nakaoka, S.-I., Hashihama, F., Shimada, K., Yoshikawa-Inoue, H., Hirawake, T., Kanda, J., Hashida, G., & Suzuki, K. (2020). Effects of phytoplankton community composition and productivity on sea surface pCO2 variations in the Southern Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, *September 2019*, 103263. https://doi.org/10.1016/j.dsr.2020.103263

Tan, T. H., Leaw, C. P., Leong, S. C. Y., Lim, L. P., Chew, S. M., Teng, S. T., & Lim, P. T. (2016). Marine micro-phytoplankton of Singapore, with a review of harmful microalgae in the region. *Raffles Bulletin of Zoology*, *2016*(Part I), 78–96.

Tham, A. K. (1973). Seasonal distribution of the plankton in Singapore Straits. *Special Publicagtion Dedicated to Dr N K Panikkar*, 15.

Tragin, M., Zingone, A., & Vaulot, D. (2018a). Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. *Environmental Microbiology*, *20*(2), 506–520. https://doi.org/10.1111/1462-2920.13952

Tragin, M., Zingone, A., & Vaulot, D. (2018b). Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. *Environmental Microbiology*, *20*(2), 506–520. https://doi.org/10.1111/1462-2920.13952

Vaulot, D. (2019). *Soil extract for algal media*. 17504.

Vuorio, K., Mäki, A., Salmi, P., Aalto, S. L., & Tiirola, M. (2020). Consistency of Targeted Metatranscriptomics and Morphological Characterization of Phytoplankton

Communities. *Frontiers in Microbiology*, *11*(February). https://doi.org/10.3389/fmicb.2020.00096

Wang, L., Lin, X., Goes, J. I., & Lin, S. (2016). Phylogenetic Analyses of Three Genes of Pedinomonas noctilucae, the Green Endosymbiont of the Marine Dinoflagellate Noctiluca scintillans, Reveal its Affiliation to the Order Marsupiomonadales (Chlorophyta, Pedinophyceae) under the Reinstated Name Protoeuglena noctilucae. *Protist*, *167*(2), 205–216. https://doi.org/10.1016/j.protis.2016.02.005

Weithoff, G., & Beisner, B. E. (2019). Measures and approaches in trait-based phytoplankton community ecology - From freshwater to marine ecosystems. *Frontiers in Marine Science*, *6*(FEB), 1–11. https://doi.org/10.3389/fmars.2019.00040

Zhu, F., Massana, R., Not, F., Marie, D., & Vaulot, D. (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology*. https://doi.org/10.1016/j.femsec.2004.10.006

Zingone, A., Percopo, I., Sims, P. A., & Sarno, D. (2005). Diversity in the genus Skeletonema (Bacillariophyceae). I. A reexamination of the type material of S. costatum with the description of S. grevillei sp. nov. *Journal of Phycology*, *41*(1), 140–150. https://doi.org/10.1111/j.1529-8817.2005.04066.x