UNIVERSITY
OF OSLO

Master thesis

Biodiversity and distribution of Arctic phytoplankton and ice algae

Khrystyna Gryn

Marine biology and limnology

60 credits

Department of Biosciences

The Faculty of Mathematics and Natural Sciences

Biodiversity and distribution of Arctic phytoplankton and ice algae

Khrystyna Gryn

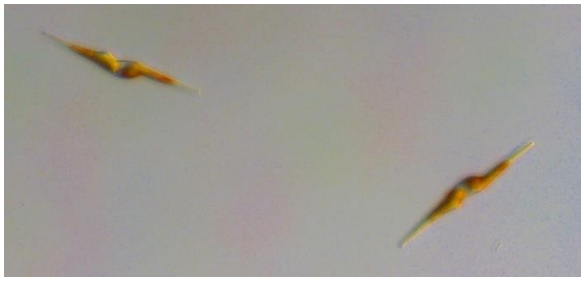
Thesis submitted for the degree of Master in Bioscience
Marine biology and limnology
60 credits

Section for aquatic biology and toxicology
Department of Biosciences
Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO

May 2023

Biodiversity and distribution of Arctic phytoplankton and ice algae



Khrystyna Gryn

grin.kristina@hotmail.com

+47 92882647

Supervisors

Bente Edvardsen

Department of Biosciences, UiO

bente.edvardsen@ibv.uio.no

Luka Supraha

Department of Biosciences, UiO/Norwegian Institute of Water Research

luka.spraha@ibv.uio.no

Wenche Eikrem

Department of Biosciences, UiO/Norwegian Institute of Water Research

wenche.eikrem@ibv.uio.no

Karoline Saubrekka

Department of Biosciences, UiO

karoline.saubrekka@ibv.uio.no

Thesis submitted for the degree of Master in Bioscience
Marine biology and limnology
60 credits

Section for aquatic biology and toxicology
Department of biosciences
Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO

May 2023

2023

The biodiversity and distribution of Arctic phytoplankton and ice algae

http://www.duo.uio.no/

Printing: Reprosentralen, Universitetet i Oslo

VI

Abstract

The project aims to enhance our understanding of Arctic marine ecosystems by researching the diversity, distribution, and functioning of phytoplankton and ice algae in response to rapid environmental changes. The diversity and biogeography of phytoplankton and ice algae from the Barents Sea were studied by combining algal cultivation and 18S and 28S rRNA gene sequencing. During the Nansen legacy project cruise in August 2019 (AeN706), a total of 46 algal strains were isolated from stations along a S-N transect in the Barents Sea into the Arctic Ocean. The identifications obtained from DNA sequencing, phylogenetic analysis and morphological examination indicated the presence of common Arctic algal species, such as *Attheya septentrionalis*, *Thalassiosira gravida*, and *Chaetoceros neogracilis*, among the identified genotypes. Additionally, this study has also characterised potentially novel species or species that have not been adequately described before. Taxonomic identification of the strains was performed using molecular and morphological approaches, and the biogeographic distribution was mapped using a global compilation of published metabarcoding datasets. It was demonstrated by the biogeographic analyses that four general biogeographic distribution types exist for arctic phytoplankton and ice algae: polar, arctic, arctic-temperate, and cosmopolitan. Genotypes with arctic-temperate distribution was most common among the characterised strains, and genotypes endemic to the arctic was also found. Only two genotypes with cosmopolitan distribution and one genotype with polar distribution were discovered. The results highlight that the phytoplankton and ice algal communities in the Barents Sea and Arctic Ocean consist of genotypes endemic to the Arctic and genotypes with broader biogeographic distributions.

Acknowledgements

This study was supported by the Nansen Legacy (Arven etter Nansen) research project aimed at gathering knowledge on the Barents Sea and Arctic Basin amidst ongoing environmental changes.

I would like to express my gratitude to my supervisors, Bente Edvardsen, Wenche Eikrem, Karoline Saubrekka, and Luka Supraha, for their valuable assistance and expertise spanning from theoretical to practical aspects. With their guidance and support, I acquired valuable practical skills in areas such as algae cultivation, light microscopy, laboratory work, data structuring, and constructing and interpreting phylogenetic trees.

Thanks to my friends Katrine Brynildsrud and Simon Hasselø Kline, whose engaging conversations and shared enthusiasm for both scientific and everyday matters have been a constant source of comfort and inspiration.

A special thanks to my family, Mariya and Daryna, for their support and encouragement throughout my master's degree journey. Their presence and assistance were invaluable as I navigated through various challenges, including the global pandemic, personal life changes, and the difficult situation in my home country. Their unwavering support helped me stay focused and motivated during this demanding time. Thank you for being there for me every step of the way.

Table of contents

Biodiversity and distribution of Arctic phytoplankton and ice algae V
Abstract VII
Acknowledgements VIII
Table of contents IX
Introduction 1
1.1 Research subject 1
1.2 Research area. Arctic and impact from the climate change 1
1.3 Aims and questions 4
The main questions that this thesis aims to answer are: 4
1.4 The hypotheses that will be tested in the thesis are: 4
2 Materials and Methods 5
2.1 Sampling and strain isolation 5
2.2 Cultivation 7
2.3 Molecular analysis 8
2.3.1 DNA Extraction 8
2.3.2 Measurement of DNA concentration 8
2.3.3 Polymerase chain reaction (PCR) 9
2.3.4 Agarose Gel Electrophoresis 10
2.3.5 Purification of PCR-product and preparation for sequencing 10
2.3.6 Processing of the sequences, BLAST 11
2.4 Phylogenetic analysis 11
2.5 Morphological analysis 12
2.6 Biogeographic analysis 14
3 Results 15
3.1 Identification by DNA sequencing 15
3.2 Phylogenetic analysis 22
3.3 Morphological analysis 35
3.4 Biogeographic analysis 41
4 Discussion 46

4.1 BLAST Analysis 46
4.2 Phylogenetic Analysis 47
4.3 Light Microscopy of Arctic Phytoplankton and Ice Algae 48
4.4 Taxonomy and phylogenetic placement 48
4.5 Station, habitat, and water depth of the isolated genotypes 49
4.6 Species novel to science 49
4.7 Biogeographical distribution based on metaPR2 50
4.8 Unique to the Arctic species (endemic species) 51
4.9 Alternative data on the global distribution 51
4.10 The Rejection of Hypotheses 52
4.11 Conclusion 53
References 54
Appendix 1. 61
Appendix 2. 62
Appendix 3. 64
Appendix 4. 65

Introduction

1.1 Research subject

Marine phytoplankton is an essential component of the ocean's ecosystem, responsible for nearly half of all photosynthesis and oxygen production on Earth (Field at al., 1998). Additionally, it serves as the main nutritional source for a wide range of marine organisms, thereby playing a critical role in maintaining the biodiversity of aquatic life worldwide (Simon et al., 2009; Thyrring et al., 2017). Since phytoplankton are the main primary producers in the ocean, changes in their composition can have a significant impact on all other trophic levels in the marine ecosystem (Behrenfeld et al., 1997). Composition of phytoplankton is important for understanding primary production and the subsequent production at higher trophic levels in the ocean (Krumhardt et al., 2020). One of the key questions in this field is how climate change will impact the composition of phytoplankton in the ocean. Additionally, understanding how different environmental factors affect the occurrence, diversity, and production of marine phytoplankton (microalgae) is crucial (Basu et al., 2018).

In order to address these issues, we need comprehensive data that covers the biodiversity of microalgae in oceanic communities. It is also crucial to collect data on how changing environmental factors can potentially impact the dynamics of microalgal community biodiversity, including their variations in time and space. However, the biodiversity of planktonic microalgae and sea ice algae is a relatively understudied, particularly in the Arctic, where new species are continually being discovered (Poulin et al., 2011). Some of these species are unique to the Arctic, making them especially vulnerable to the dramatic climate changes currently taking place in the region.

1.2 Research area. Arctic and impact from the climate change

The Arctic is mainly covered by ice and is characterised by extreme cold temperatures and a unique ecosystem. A key feature of the Arctic is the permanent icecap (Polyak et al., 2010). Currently, following general global climate change patterns, the Arctic is undergoing a considerable increase in temperature. As a result, the sea ice cover is constantly decreasing (Figure 1 and 2.). It is predicted that by the second half of the 21st century the Arctic may see a see-ice-free summer (Notz and SIMIP, 2020).

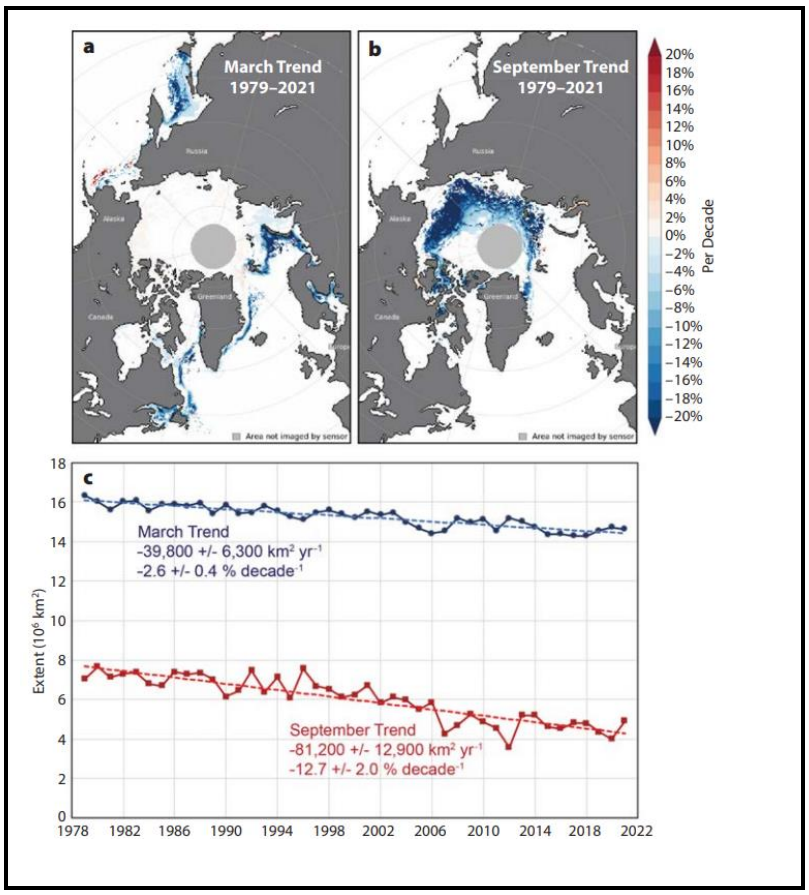


Figure 1. The decrease in the Arctic Sea ice has been documented from 1978 to 2021. The figure shows the trends in ice extent during March and September over a 30-year period. Figure is from an article "An Updated Assessment of the Changing Arctic Sea Ice Cove" by Meier et al., 2022.

However, the lack of sufficient data makes it challenging to evaluate the impact of climate change on Arctic marine ecosystems, including species distribution, long-term monitoring, oceanographic parameters, remote sensing, and paleoecological records, which are crucial for understanding and assessing the consequences of climate change (Mousing at al., 2017). It is expected that marine ecosystems will be subject to changes due to increasing ocean temperatures resulting in the increased melting of the sea ice. The loss of ice in the Arctic will lead to a significant decrease in biodiversity due to factors such as habitat loss, as ice serves as a critical habitat and platform for various species, impacting their ability to survive and thrive in the changing ecosystem (Duarte et al., 2012).

Global warming will also have an impact on the biodiversity of phytoplankton in the Arctic Ocean. A warmer Arctic may lead to changes in ocean currents and weather patterns, which can affect the growth and survival of phytoplankton. Changes in water temperature and ocean currents can alter the distribution and abundance of phytoplankton species in the Arctic Ocean, resulting in a decline in overall species diversity. Warmer water temperatures can also favour the growth of certain phytoplankton species over others, leading to a shift in the dominant species and a decrease in overall biodiversity (Verde at al., 2016). Melting sea ice could alter the amount of light available for photosynthesis and access to nutrients. This could

impact phytoplankton diversity and effect the water column's pH and salinity (Langbehn & Varpe, 2017).

Changes in ocean acidification due to increased carbon dioxide levels can also impact phytoplankton species diversity, as some species may be more adapted to varying acidity levels than others (Riebesell & Gattuso, 2015).

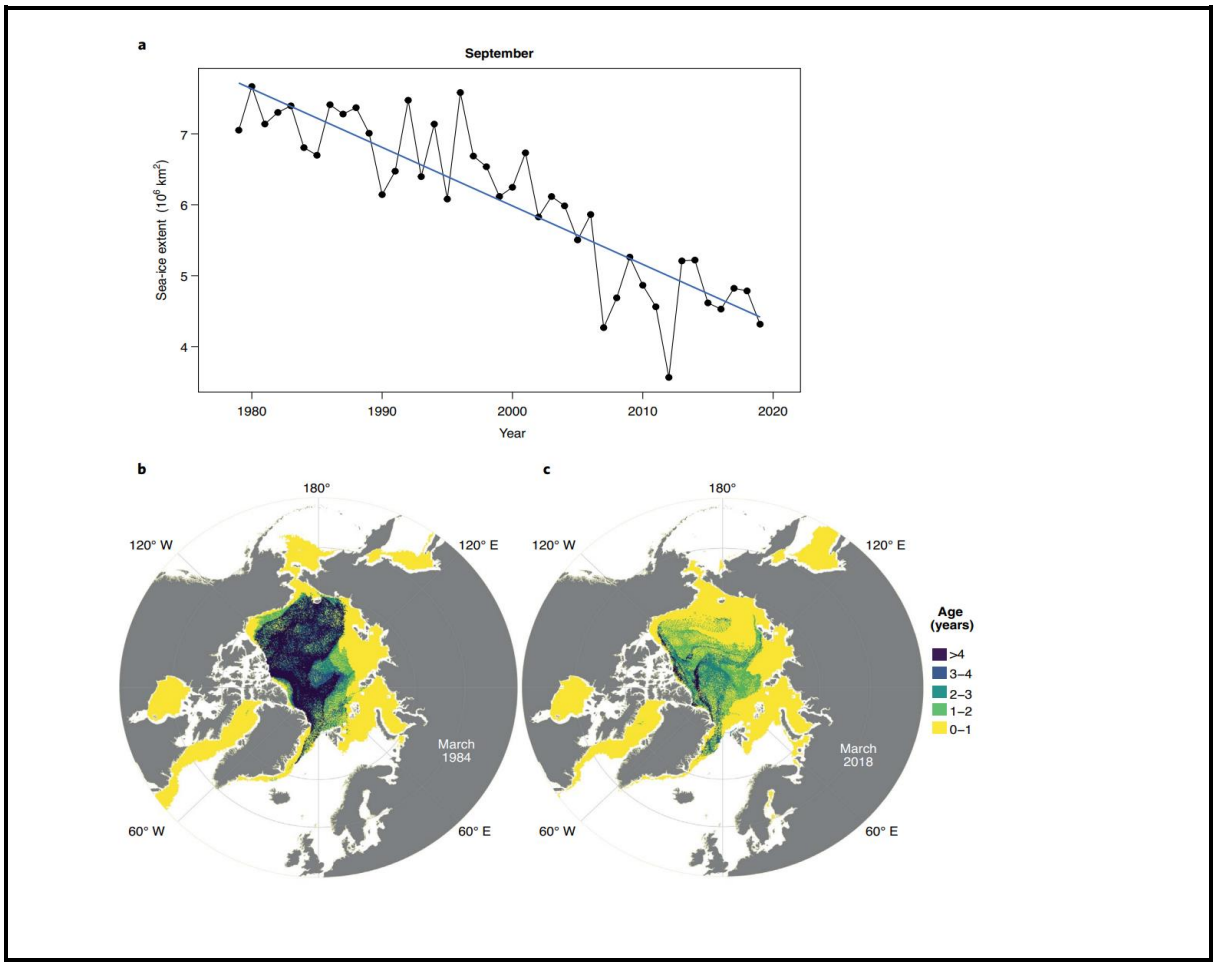


Figure 2. Changing sea-ice age and extend. **a**, *A time series depicting the extent of summer sea ice (in 10 km) from 1979 to 2019.* **b**, **c**, *Maps that compare the age of Arctic Sea ice during winter in 1984 (***b***) and 2018 (***c***). The data used is satellite-derived sea ice data from the National Snow and Ice Data Centre (NSIDC, Adryna et al., 2020). This figure is taken from the review article "Phytoplankton dynamics in a changing Arctic Ocean" by Adryna et al., 2020.*

According to the AMAP (Arctic Monitoring and Assessment Programme) assessment (2019), increased human activity, such as shipping and oil and gas exploration, in the Arctic can have a negative impact on phytoplankton species diversity through the introduction of invasive species, pollution, and other factors.

A decrease in the number and variety of phytoplankton species present in the Arctic Ocean could result in a domino effect throughout the entire marine ecosystem, as phytoplankton is a vital food source for many marine animals (Arrigo at al., 2014).

1.3 Aims and questions

The aim of this thesis is to investigate the identity, taxonomy, phylogeny, and geographical distribution of micro algal species collected from the Barents Sea free water masses (pelagic) and sea ice (sympagic). During an expedition in the Arctic in August 2019, algae were isolated, and cultures started. The cultures were characterised using a combination of molecular lab work and data analysis, including DNA sequencing of 18S small ribosomal subunit (SSU) and 28S large ribosomal subunit (LSU) ribosomal RNA genes, as well as light microscopy. The main focus of the analysis was molecular phylogenetic analysis and taxonomic placement. The 18S rRNA gene sequences were then used to determine the global biogeographic distribution of the studied strains by searching for these sequences in the metaPR2 (metabarcoding database by Daniel Vaulot) to determine under which environmental factors these algae have been found in the past (Vaulot at al., 2022).

The main questions that this thesis aims to answer are:

The overall aim is divided into six complementary questions. These are:

1. What are the taxonomy and phylogenetic placement of the algal strains isolated into culture during the Nansen Legacy cruise in August 2019?
2. At which station, habitat and water depth were the different isolated genotypes found during this cruise?
3. Are any of the species isolated new to science?
4. How are the biogeographical distribution of selected genotypes based on metaPR2?
5. Are there any species unique to the Arctic (endemic species)?
6. What is the global distribution, including habitat types, of the described species or genotypes in the past?

1.4 The hypotheses that will be tested in the thesis are:

H1: In our material there are ice algae or phytoplankton species novel to science.

H0: In our material there are no ice algae or phytoplankton species novel to science.

H1: In our material there are ice algae or phytoplankton species that are unique to the Arctic.

H0: In our material there are no ice algae or phytoplankton species that are unique to the Arctic, they have previously been found in other geographical regions as well.

2 Materials and Methods

2.1 Sampling and strain isolation

Done by Bente Edvardsen and Karoline Saubrekka.

Seawater samples were collected from the Northern Barents Sea. Sampling was conducted in conjunction with the Nansen Legacy ("Arven etter Nansen") project, on board of the icebreaker F/F Crown Prince Haakon between August 5-27, 2019. Water samples were collected at various depths and locations including open-water areas on the shelf, sea-ice regions on the slope region, and deep in the Nansen Basin. Detailed information about each sampling station is presented in Table 1 and Figure 3.

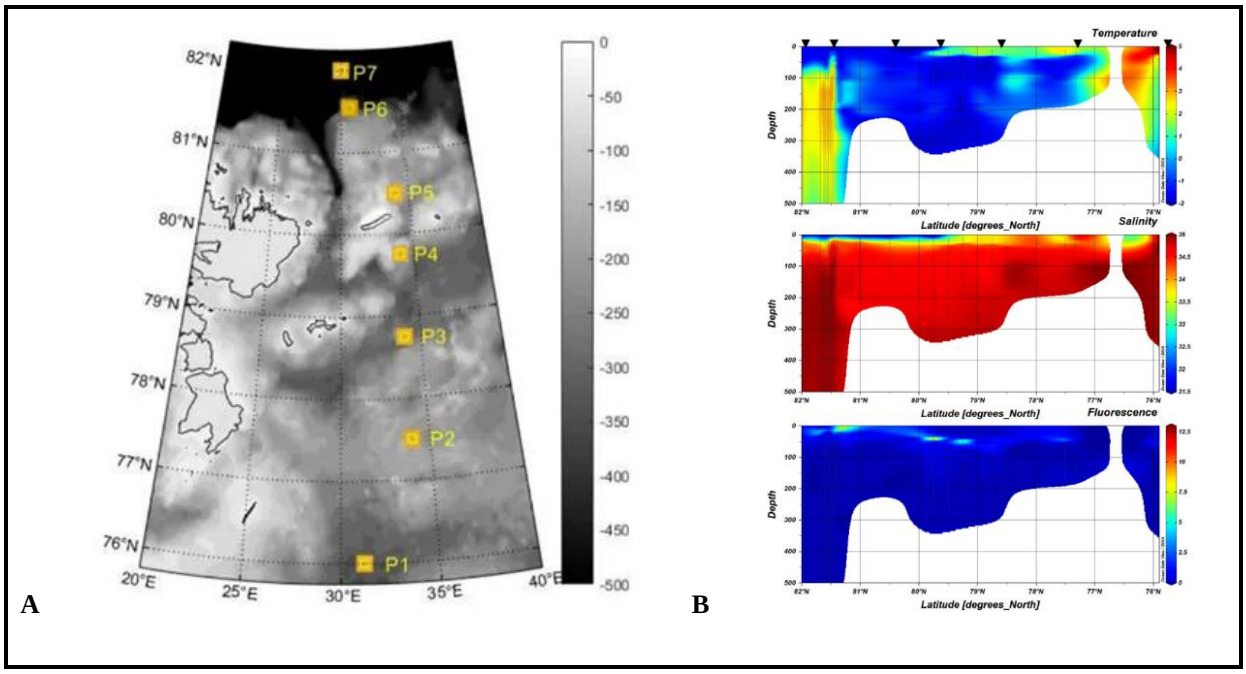


Figure 3. This figure illustrates the sampling stations (P1-P7, stations P6-ICE and P7-ICE are not shown) of the studied strains during the Nansen LEGACY cruise in panel A and a representation of the temperature, salinity, and fluorescence distribution in panel B. The locations of the process stations (P1-P7. Stations P6-ICE and P7-ICE are not shown) are indicated by inverted triangles above slide with temperature indicator. Figure is taken from the «Nansen Legacy cruise Q3 report» (Reigstad et al. 2022).

Pelagic seawater samples were collected using 10L Niskin bottles mounted on a conductivity-temperature-depth (CTD) instrument at various depths (5/10m, deep chlorophyll maximum (DCM), 200m, bottom). Net samples were collected from a vertical net haul (0-50 or 100 m) with a 10 µm mesh size. Sympagic (ice associated) samples were collected either from a hole made in the ice, using a Niskin Bottle descended just below the sea surface (0.5m), or from a melt pond, where water was collected from the surface using a bucket, or from the bottom 0-10 cm of an ice core obtained by drilling a hole through the ice. The locations, sampling depth, water salinity and temperature where monoalgal strains were obtained are presented in Table 1.

Table 1. This table presents the station's name, latitude and longitude, depth of the sampling, water salinity, temperature of the water and the sampling date.

Station	Latitude	Longitude	Depth of sampling (m)	Salinity (PSU)	Temperature (C°)	Date
P1	76.0000	31.2194	45	35	5	2019-08-08
P2	77.5006	34.0012	10, 50	34.4	3, -1	2019-08-11
P3	78.7498	34.0008	10, 75	34.2, 34.6	2, -1	2019-08-13
P4	79.7494	33.9971	40	34.2	-1	2019-08-14
P5	80.4966	33.9898	20	34	2	2019-08-16
P6	81.5726	31.2134	0, 0.5	34	0, -1	2019-08-18
P6-ICE	81.5327	30.9684	0, 0.5	34	0, -1	2019-08-17
P7	81.9184	29.1151	0, 0.5	34	-1	2019-08-20
P7-ICE	81.9861	29.9975	0, 0.5	34	-1	2019-08-20

Further information regarding the specific characteristics of each station can be found in «Nansen Legacy cruise Q3 report» (Reigstad et al., 2022).

Microalgal strains were isolated on board the vessel and later in the laboratory at the University of Oslo (UiO). Monoalgal cultures were obtained through a combination of serial dilution series and capillary isolation techniques. Cultures were eventually cultivated in the laboratory at UiO in individual cultivation borosilicate tubes with 15 ml of growth medium (IMR ½ medium with salinity 30) (Eppley et al., 1967). Cultures were maintained on board in a climate room at 4ºC, under light-dark conditions (12:12 light-dark cycle). The light on board was provided by LED illumination with full light spectrum (LEDlife Pro-Grow 2.0, 4W, Claes Ohlsson).

Overview of research methods and materials

To provide a clear understanding of the research methods and materials, a brief schematic overview is presented in Figure 4. before going into the detailed description of the procedures employed in this project.

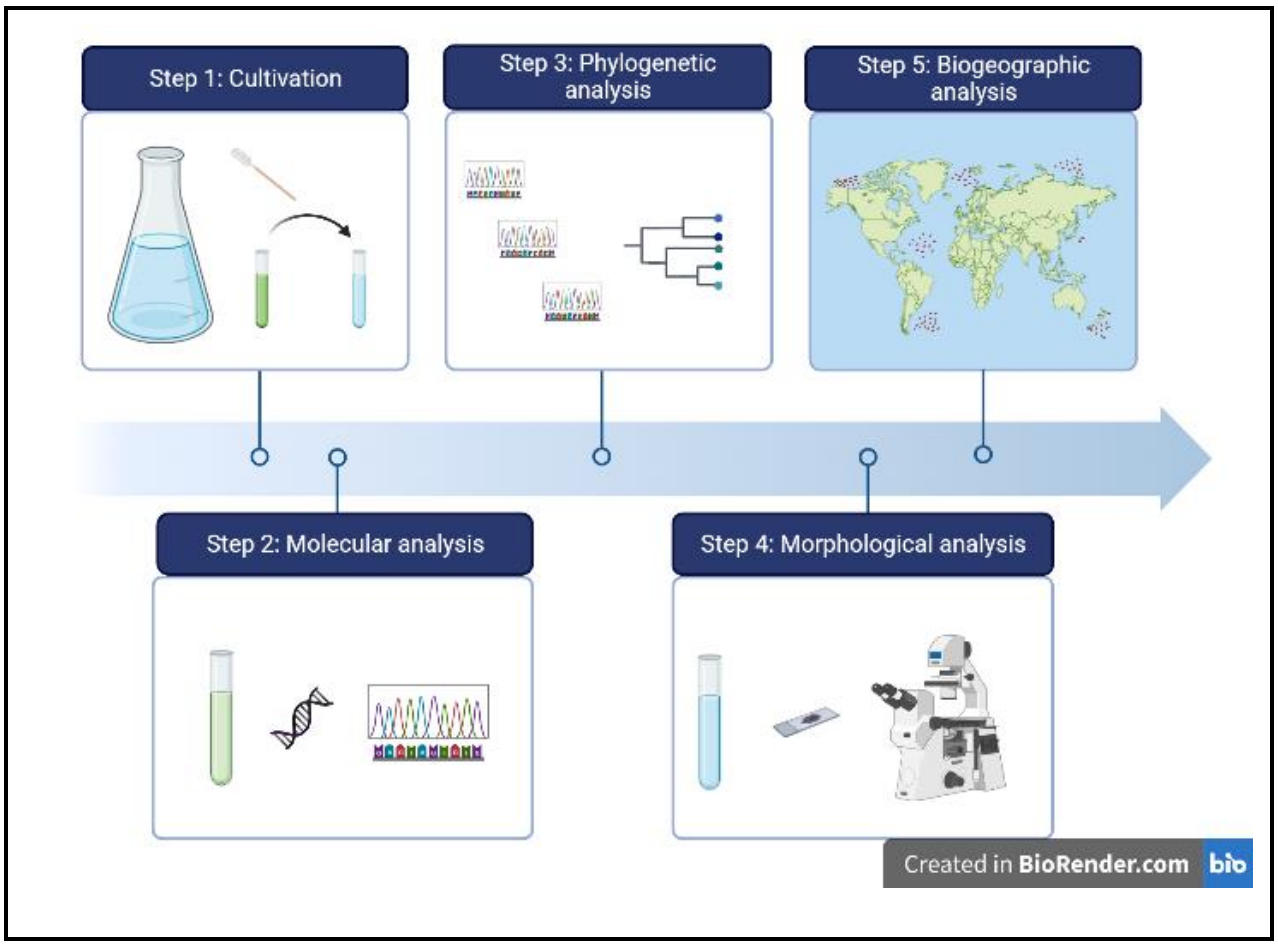


Figure 4. This schematic illustration summarises the methods and analyses used in this thesis, which involved the cultivation of the studied strains, microscopic examination, isolation, amplification, and sequencing of DNA, BLAST analysis, and the construction of a phylogenetic tree. Additionally, biogeographic analysis was conducted to understand the geographical distribution of the strains. Figure created with Biorender.com.

First, the isolated strains were examined under a microscope to ensure their maintenance, including cultivation. Next, a set of analytical tasks were carried out, including molecular analysis, phylogenetic analysis, morphological analysis, and biogeographical analysis (see Figure 4).

2.2 Cultivation

The isolated algal strains were maintained at the AQUA facilities of the Department of Biosciences, University of Oslo, and cultured in climate rooms and incubators at 4°C under a 12:12 light-dark cycle. The light was provided by LED illumination with full light spectrum and partly by natural light. To ensure optimal growth conditions, all strains were transferred to new growth medium every 6 weeks.

The medium used for culturing the micro algae was a half-defined algae medium IMR ½ + Si with a PSU (practical salinity unit) of 30 (Eppley et al., 1967). As a base for the medium, natural seawater was collected from the Oslofjord near Drøbak at a depth of 40 meters. The seawater was filtered using Whatman Glass Microfiber Filters Grade GF/C (Cytiva Lifesciences, USA), and the salinity was measured using a handheld refractometer (N-8, Atago CO LTD., Japan). The salinity was then adjusted by adding distilled water to achieve the desired salinity.

The stock solutions of the IMR ½ medium contained all the required nutrients and vitamins for algae growth, including phosphate, nitrate, and selenite solutions. To ensure the wellbeing of diatoms, silicate solutions were also added. Finally, a trace metals solution was added, which not only contained trace metals but also chelators to complex inhibitory heavy metals.

The IMR ½ medium was autoclaved (HS 6610EC-1 Autoclave, Getinge, Sweden) at 121°C for 20 minutes. After autoclaving, the algae growth medium was stored in a temperature-controlled room at 4°C.

2.3 Molecular analysis

All molecular analyses were conducted at the AQUA DNA laboratories, using equipment and facilities provided by the Department of Biosciences at the University of Oslo (UiO).

2.3.1 DNA Extraction

For DNA extraction, 2 ml aliquot of high-density cell culture was transferred to a 2 ml Eppendorf tube and centrifuged at 8,000 rpm for 8 minutes (Centrifuge 5424 R, Eppendorf, Germany). The resulting pellet was processed with three different types of kits: NucleoSpin II Plant, NucleoSpin II Fungi and NucleoSpin® Soil. The highest yield of DNA was obtained using the NucleoSpin® Soil kit and therefore used in this work. The manufacturer's protocol was followed with minor adjustments. The spin column protocol involves the following general steps: lysing the cell walls to make DNA available, degrading proteins, binding the lysate (DNA) to a silica membrane in a spin column, washing away contaminants, and eluting DNA. The extracted DNA was stored frozen at -20 ºC in labelled Eppendorf® Safe-Lock 1.5 ml microcentrifuge tubes.

2.3.2 Measurement of DNA concentration

After DNA extraction, the quantity of DNA was measured to ensure that the amount of DNA was sufficient for PCR and sequencing. The quantity of DNA was evaluated using the Qubit® 3.0 Fluorometer (Thermo Fisher Scientific, USA) according to the manufacturer's protocol. The Qubit dsDNA BR (Broad Range) Assay Kit was applied as it measures samples with concentrations between 2 and 1000 ng/µL. This corresponds with previous DNA estimation in

similar algae cultures. The optimal DNA concentration for PCR is 5-20 ng/µl. Based on the results obtained, some samples were diluted or used in a higher volume quantity for the PCR. The detailed DNA quantification description can be found in an appendix (Appendix 1).

2.3.3 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) is a technique for amplifying a specific region of DNA. In this study, the PCR was used to amplify two genes 18S rRNA and 28S rRNA in order to determine phylogeny and taxonomy of isolated strains and ultimately characterise some of the diversity of microalgae in the Arctic Ocean in the deep basin north of the Barents Sea.

The analysis in this study used the entire 18S rRNA sequence along with a partial sequence of the 28S rRNA gene, specifically targeting the D1 and D2 domains, which spanned approximately 650 base pairs. These genes are widely used in molecular studies due to their high level of conservation and their ability to provide phylogenetic information for a broad range of eukaryotic organisms. This approach is widely accepted in the field and has been used in many previous studies to infer the evolutionary relationships of diverse taxa (Guillou et al., 2013; Medlin et al., 1988).

The PCR was performed separately for 18S rRNA gene and for 28S rRNA gene. For the amplification of the 18S rRNA gene, two different primer sets were used: SSUA, SSUB (Theriot et al., 2015) and 1F, 1528R (Medlin et al., 1988). The PCRs were performed for each primer set separately. For amplification of the partial 28S rRNA gene the primers D1C-F and D2C-R were used (Scholin et al., 1994). Sequences of the used primers for PCR are presented in Table 2.

Table 2. Sequences of the used primers for PCR and sequencing (modified from Edvardsen et al. 2003). Sd = synthesis direction: F = forward; R = reverse.

Code	PCR	Sequencing	Sd	Nucleotide sequence 5’ to 3’
1F	SSU	SSU	F	AACCTGGTTGATCCTGCCAGT
1528R	SSU	SSU	R	TGATCCTTCTGCAGGTTCACCTAC
DIR	LSU	LSU	F	ACCCGCTGAATTTAAGCATA
D2C	LSU	LSU	R	CCTTGGTCCGTGTTTCAAGA
SSUB	SSU	SSU	R	CCTTCTGCAGGTTCACCT AC
850+		SSU	F	GGGACAGTTGGGGGTATTCGTA
1147-		SSU	R	AGTTTCAGCCTTGCGACCATAC

The reaction mix (see Table 3 below) was prepared in a 1.5 mL Eppendorf tube for all PCR reactions, with one extra set of the master mix components per ten reactions. The concentration of the primers used was 5 µM. An amount of DNA template for amplifying the rRNA genes was 1-5 µl, depending on the DNA concentration.

The PCR program was also adjusted depending on the gene and primers used. The program consisted of an initial denaturation step at 95°C for 2-3 minutes, followed by 35 cycles of denaturation at 95°C for 45 seconds, annealing as a rule at 55ºC (or 50ºC for SSU when needed) for 45 seconds, and extension at 72°C for 2 minutes for 18S rRNA or 1 minute for 28S rRNA (1 minutes per 1000 bp). A final extension step was performed at 72°C for 5 minutes. The detailed 28S rRNA gene (LSU) PCR program and the 18S rRNA gene (SSU) PCR program can be found in appendix (Appendix 2).

Table 3. Reaction (master) mix (µL per sample). The contents of the PCR master mix consist of MilliQ water, primers, and polymerase.

N	Reagent	1 reaction (µL)	X reactions (µL)
1	PCR water	7.5 (4.5-8.5)	
2	Forward primer (5 µM)	1.5	
3	Reserve primer (5 µM)	1.5	
4	GoTaq Green Master mix	12.5	
5	DNA template (ca 20 ng μL^{-1})	2 (1-5 µL)	
	Total	**25**	

2.3.4 Agarose Gel Electrophoresis

The electrophoresis of the PCR products was performed using agarose gel electrophoresis. This method uses an electric current to separate the DNA molecules based on their size, with larger molecules migrating more slowly than smaller ones. The agarose gel was made by dissolving 0.8 g of agarose powder in 100 ml of 1X TAE buffer and adding a nucleotide stain called GelRed for visualization under UV light. When the gel was solidified a DNA- size marker (Lambda DNA/EcoRI+HindIII, Thermo Scientific™ SM0192) was added in the first well and the samples accordingly. The gel was then run at 80V for 40 minutes (Electrophoresis Power Supply – EPS 301, Amersham Pharmacia Biotech, USA), and the separated DNA molecules were visualised using a UV detector cabinet (Gene Genious Bioimaging System, Syngene, Cambridge, England).

2.3.5 Purification of PCR-product and preparation for sequencing

Before sequencing the PCR products need to be cleaned up. To inactivate excessive primers, unincorporated nucleotides, and other enzymes in PCR products, special enzymes (PCR Product Cleanup Reagent by Exosap-IT) were used. To Eppendorf tube with 12.5 µl PCR-product was added 5 µl Exosap-IT. This mixture was incubated for 15 minutes at 37 °C, thereafter 15 minutes at 80 °C (Grant QBD2 Block Heater, Grant Instruments, United Kingdom).

To prepare for sequencing, 5 µl of sequencing primer and 5 µl of purified PCR product were added to pre-labelled 1.5 mL Eppendorf tubes for each sequencing reaction. These tubes were sent to DNA sequencing at EuroFins Genomics (Germany).

2.3.6 Processing of the sequences, BLAST

The partial sequences of approximately 700 base pairs (bp) were obtained for the 28S rRNA gene. The sequencing of the 18S rRNA gene, which is approximately 1800 bp in length, was also conducted. All used in the study sequences can be found in Appendix 4.

The sequencing data was analysed with Geneious Prime software. First, each sequence was assessed for quality by examining the corresponding chromatogram. Regions of low quality were removed from the beginning and end of the final sequences. The De Novo Assemble settings were used to align the sequences.

Then sequences were subjected to a similarity search against the sequences available in GenBank using the Basic Local Alignment Search Tool (BLAST) algorithm. The algorithm is hosted by the National Center for Biotechnology Information (NCBI), which is a part of the National Institute for Health in the United States. The BLAST analysis was performed within the Geneious Prime software.

To construct a phylogenetic tree, a reference sequences from the database with the highest similarity were chosen.

2.4 Phylogenetic analysis

Phylogenetic trees were constructed to describe the evolutionary relationships among the strains and classify them taxonomically.

Based on the initial BLAST results, the studied sequences classified into four distinct taxonomic groups: Chlorophyta (green algae), Miozoa (dinoflagellates), Ochrophyta (dictyochophytes), and Bacillariophyta (diatoms). Separate phylogenetic trees were constructed for each taxonomic group using 18S (see section 3. Results, Figures 5, 7, 9, 11) and 28S (see section 3. Results, Figures 6, 8, 10, 12) rRNA gene sequences. An alignment was created by incorporating sequences of the studied strains, reference sequences and an outgroup. The sequenses were aligned using the L-INS-I algorithm in MAFFT within Geneious Prime software, with five DNA reference sequences per genotype available from the NCBI database. *Bolidomonas pacifica* (phylum Ochrophyta, class Bolidophyceae) was used as an outgroup in all phylogenetic trees. The newly taxonomically accepted name of this species is *Triparma pacifica* (Ichinomiya et al., 2016). Accession number HQ912557 for the 18S rRNA phylogenetic trees and AB430658 for the 28S rRNA phylogenetic trees.The maximum-likelihood phylogenetic trees was generated in RAxML program using GTRGAMMA I model with 1000 bootstrap replications. The bootstrap values, represented as

numerical values between the nodes, indicate how frequently the nodes are recovered through resampling and thus represent relative support values of the observed branching pattern.

2.5 Morphological analysis

The morphological characteristics of the studied strains were analysed using a light microscope (Zeiss Axio Vert.A1, Carl Zeiss AG, Germany) equipped with differential interference contrast (DIC) and phase contrast objectives at x40 magnification. Micrographs were captured using an attached camera (Leica MC170 HD, Leica Biosystems, Germany) and the imaging program (LAS EZ Digital Imaging System, Leica Biosystems, Germany). The strains were identified based on their morphological characteristics, and in some cases, their motility patterns were also taken into consideration.

Table 4. List of the strain isolated during the Nansen Legacy cruise Q3 and used in the present study. Table contains information on the algal strains characterised in this study including information on their code, station name where the particular strain was isolated, the collection date, corresponded latitude and longitude, type of habitat and the result of the morphological identification by light microscopy (LM). ND - no data.

Strain code	Station name	Collection date	Latitude	Longitude	Type of habitat, depth	Identification by LM
BE_AeN706-3	P2	2019-08-11	77.4987	34.0012	pelagic, 50 m	*Pyramimonas* sp.
BE_AeN706-5	P2	2019-08-11	77.4987	34.0012	pelagic, 50 m	ND
BE_AeN706-6	P2	2019-08-12	77.5006	33.9865	pelagic, 10 m	*Pyramimonas* sp.
BE_AeN706-8	P2	2019-08-12	77.5006	33.9865	pelagic, 10 m	*Pyramimonas* sp.
BE_AeN706-9	P2	2019-08-12	77.5006	33.9865	pelagic, 10 m	*Pyramimonas* sp.
BE_AeN706-10	P2	2019-08-12	77.5006	33.9865	pelagic, 10 m	*Pyramimonas* sp.
BE_AeN706-11	P2	2019-08-12	77.5006	33.9865	pelagic, 50 m	*Pennate diatom*
BE_AeN706-12	P2	2019-08-12	77.5006	33.9865	pelagic, 50 m	*Prorocentrum* sp.
BE_AeN706-13	P6-ICE	2019-08-17	81.5327	30.9684	melt pond, 0 m	ND
BE_AeN706-15	P6-ICE	2019-08-17	81.5327	30.9684	melt pond, 0 m	*Micromonas* sp.
BE_AeN706-19	P6-ICE	2019-08-17	81.5720	31.2128	under ice water, 0.5 m	*Micromonas* sp.
BE_AeN706-21	P6-ICE	2019-08-17	81.5720	31.2128	under ice water, 0.5 m	*Chaetoceros neogracilis*
BE_AeN706-22	P6-ICE	2019-08-17	81.5720	31.2128	under ice water, 0.5 m	*Chaetoceros neogracilis*
BE_AeN706-23	P6-ICE	2019-08-17	81.5720	31.2128	under ice water, 0.5 m	*Thalassiosira gravida*
BE_AeN706-24	P6-ICE	2019-08-17	81.5720	31.2128	under ice water, 0.5 m	*Chaetoceros neogracilis*
BE_AeN706-25	P6-ICE	2019-08-17	81.5327	30.9684	melt pond, 0 m	*Thalassiosira gravida*
BE_AeN706-26	P6-ICE	2019-08-17	81.5327	30.9684	melt pond, 0 m	*Thalassiosira gravida*
BE_AeN706-27	P6-ICE	2019-08-17	81.5327	30.9684	melt pond, 0 m	*Bacterosira* sp.
BE_AeN706-30	P1	2019-08-08	76.0000	31.2198	pelagic, 45 m	*Chaetoceros neogracilis*
BE_AeN706-31	P1	2019-08-08	76.0000	31.2198	pelagic, 45 m	*Chaetoceros neogracilis*
BE_AeN706-34	P4	2019-08-14	79.7494	33.9971	pelagic, 40 m	*Chaetoceros neogracilis*
BE_AeN706-35	P4	2019-08-14	79.7494	33.9971	pelagic, 40 m	*Chaetoceros neogracilis*
K-AeN706-5	P3	2019-08-13	78.7498	34.0008	pelagic, 75 m	*Thalassiosira gravida*
K-AeN706-6	P3	2019-08-13	78.7498	34.0008	pelagic, 75 m	*Thalassiosira gravida*
K-AeN706-7	P3	2019-08-13	78.7498	34.0008	pelagic, 75 m	*Chaetoceros cinctus*
K-AeN706-11	P3	2019-08-13	78.7498	34.0008	pelagic, 10 m	*Chatoceros neogracilis*
K-AeN706-14	P3	2019-08-13	78.7498	34.0008	pelagic, 0-100m	*Thalassiosira gravida*
K-AeN706-15	P6-ICE	2019-08-17	81.5327	30.9684	under ice water, 0.5 m	*Atteya septentrionalis*
K-AeN706-17	P7-ICE	2019-08-20	81.9861	29.9975	under ice, 0-5 m	*Pedinellaceae* sp.
K-AeN706-18	P6	2019-08-18	81.5720	31.2128	pelagic, 15 m	*Chaetoceros* cf. *neogracilis*
K-AeN706-19	P6	2019-08-18	81.5720	31.2128	pelagic, 15 m	*Neodenticula seminae*
K-AeN706-22	P7-ICE	2019-08-20	81.9861	29.9975	ice core, 0-10 cm	*Nitzschia* sp.
K-AeN706-26	P7	2019-08-21	81.9184	29.1151	pelagic, 0-100 m	*Shionodiscus bioculatus*
K-AeN706-28	P7-ICE	2019-08-20	81.9861	29.9975	ice core, 0-10 cm	*Chatoceros neogracilis*
K-AeN706-30	P7-ICE	2019-08-20	81.9861	29.9975	ice core, 0-10 cm	*Pennate diatom*
K-AeN706-31	P7-ICE	2019-08-20	81.9861	29.9975	ice core, 0-10 cm	cf. *Navicula*
K-AeN706-32	P7-ICE	2019-08-20	81.9861	29.9975	ice core, 0-10 cm	*Pennat diatom*
K-AeN706-33	P7-ICE	2019-08-20	81.9861	29.9975	ice core, 0-10 cm	*Pennat diatom*
K-AeN706-34	P7-ICE	2019-08-20	81.9861	29.9975	ice core, 0-10 cm	ND
K-AeN706-35	P7-ICE	2019-08-20	81.9861	29.9975	ice core, 0-10 cm	*Pennat diatom*
K-AeN706-36	P6	2019-08-18	81.5720	31.2128	pelagic, 15 m	*Atteya septentrionalis*
K-AeN706-37	P6	2019-08-18	81.5720	31.2128	pelagic, 15 m	*Atteya septentrionalis*
K-AeN706-38	P6	2019-08-18	81.5720	31.2128	pelagic, 15 m	cf. *Chaetoceros neogracilis*
K-AeN706-45	P5	2019-08-15	80.4966	33.9898	pelagic, 20 m	*Pennate diatom*
K-AeN706-52	P6	2019-08-18	81.5720	31.2128	pelagic, 15 m	*Pennate diatom*
K-AeN706-56	P6	2019-08-18	81.5762	31.3259	pelagic, 0-50 m	*Thalassiosira* sp.

2.6 Biogeographic analysis

To study the biogeography of the algal strains, the metaPR2 interactive website was used, which includes an 18S rRNA gene metabarcoding database (Vaulot et al., 2022). This database comprises 41 metabarcoding datasets that correspond to more than 4,000 samples and 90,000 amplicon sequence variants (ASVs). The 18S rRNA gene sequence of each studied strain was compared with the ASVs found in the compiled dataset using the local BLAST tool. The distribution of the ASVs, which were identical (99-100% pairwise identity) to the queried sequence, was shown using the local mapping tool (ASV information was retrieved from the database available at https://shiny.metapr2.org/metapr2/). Insights into the geographical distribution of studied strains and an understanding of their biogeographic patterns were gained through the utilisation of this website and database.

3 Results

3.1 Identification by DNA sequencing

A total of 174 DNA sequences were obtained (LSU: 39 from BE-cultures and 42 from K-cultures; SSU: 47 from BE-cultures and 46 from K-cultures) and needed to be processed. The results from BLAST analyses of the obtained 28S rRNA nuclear gene sequences and 18S rRNA nuclear gene sequences of the studied strains, along with their best matching DNA sequences from the NCBI nucleotide database, are presented in Table 5 (for 18S rRNA sequences) and Table 6 (for 28S rRNA sequences). The highest match between the obtained sequence and a reference sequence is determined by two parameters: pairwise identity and query cover. Pairwise identity refers to the percentage of identical nucleotide or amino acid residues between two sequences in a pairwise alignment, such as the sequenced strain and its best matching DNA sequence from the NCBI database. The query cover provides information on how much of the obtained sequence overlaps with the reference sequence.

The BLAST results for both genes were compared, and a common identification was assigned to each strain (see Table 7 below). If the BLAST results for both genes matched and provided a detailed species-level identification, the strain was given an identical name. If the BLAST results for both genes differed, the strain was identified up to the first matching taxonomic rank.

The DNA sequences of the 18S and 28S RNA genes were analysed twice by BLAST. The first analysis involved a BLAST search for similar sequences in the NCBI database, and the result are presented in Tables 5, 6, and 7. In the second BLAST analysis aligned strain sequences with reference sequences were used, and any questionable base pairs were manually corrected after alignment. The results of the second BLAST were used to construct phylogenetic trees (see section 3.2. Phylogenetic analysis, Figures 5-12).

Table 5. Strains identification based on 18S rRNA gene sequencing (BLAST). Table contains the code of the strain, type of the analysed sequence, accession number of the highest match in NCBI database, species name of the best match and submission year to the NCBI database, pairwise similarity (%) and the query cover (%). ND - no data.

Strain code	SSU (18S sequencing)	Best match (BLAST) accession number	Best match (BLAST) species name (submission year in NCBI database)	Pairwise identity (%)	Query cover (%)
BE_AeN706-3	Consensus	KY980369	*Pyramimonas australis* (2017)	99.9	95.91
BE_AeN706-5	Consensus	ON888455	*Micromonas polaris* (2022)	100.0	99.88
BE_AeN706-6	Consensus	FN562443	*Pyramimonas parkeae* (2009)	97.7	99.61
BE_AeN706-8	Consensus	FN562443	*Pyramimonas parkeae* (2009)	97.6	99.44
BE_AeN706-9	ND	ND	ND	ND	ND
BE_AeN706-10	Consensus	FN562443	*Pyramimonas parkeae* (2009)	97.7	100
BE_AeN706-11	Consensus	LR812489	*Fragilariopsis kerguelensis* (2020)	99.5	100
BE_AeN706-12	Consensus	MZ593908	*Prorocentrum shikokuense* (2021)	99.6	100
BE_AeN706-13	Consensus	ON888455	*Micromonas polaris* (2022)	100	99.89
BE_AeN706-15	Consensus	ON888455	*Micromonas polaris* (2022)	99.9	100
BE_AeN706-19	Consensus	DQ025753	*Micromonas pusilla* (2005)	100	100.0
BE_AeN706-21	Consensus	EU090014	*Chaetoceros* sp. (2007)	99.4	99.36
BE_AeN706-22	Consensus	EU090014	*Chaetoceros* sp. (2007)	99.4	99
BE_AeN706-23	ND	ND	ND	ND	ND
BE_AeN706-24	Consensus	EU090014	*Chaetoceros* sp. (2007)	98.9	100
BE_AeN706-25	ND	ND	ND	ND	ND
BE_AeN706-26	Consensus	KT860983	*Thalassiosira rotula* (2015)	100	100
BE_AeN706-27	Consensus	KC771209	*Uncultured marine eukaryote* (2013)	99.7	100
BE_AeN706-30	850F+	KC771204	*Uncultured marine eukaryote* (2013)	99.9	100
BE_AeN706-31	850F+	KC771204	*Uncultured marine eukaryote* (2013)	99.9	100
BE_AeN706-34	850F+	KC771204	*Uncultured marine eukaryote* (2013)	99.9	100
BE_AeN706-35	850F+	KC771204	*Uncultured marine eukaryote* (2013)	99.8	100
K-AeN706-5	ND	ND	ND	ND	ND
K-AeN706-6	Consensus	MW205690	*Thalassiosira rotula* (2022)	99.83	100
K-AeN706-7	Consensus	KC771202	*Uncultured marine eukaryote* (2013)	99.9	99.71
K-AeN706-11	850F+	KC771204	*Uncultured marine eukaryote* (2013)	99.8	100
K-AeN706-15	Consensus	JX401230	*Attheya longicornis* (2013)	99.9	100.0
K-AeN706-17	Consensus	EU247836	Pedinellales (2008)	100	97
K-AeN706-18	Consensus	KC771204	*Uncultured marine eukaryote* (2013)	99.8	100
K-AeN706-19	Consensus	LR812489	*Fragilariopsis kerguelensis* (2020)	99.5	100
K-AeN706-22	Consensus	KC771161	*Uncultured marine eukaryote* (2013)	99.8	100
K-AeN706-26	Consensus	X85401	*Actinocyclus curvatulus* (2002)	99.5	100
K-AeN706-28	Consensus	KC771204	*Uncultured marine eukaryote* (2013)	99.8	100
K-AeN706-30	Consensus	AY485458	*Nitzschia thermalis* (2007)	99.7	100
K-AeN706-31	Consensus	JQ240485	*Amphora sp*. (2012)	99.3	99.4
K-AeN706-32	850F+	KC771155	*Uncultured marine eukaryote* (2013)	99.5	98.08
K-AeN706-33	Consensus	KY320391	*Nitzschia aequorea* (2017)	99.4	100
K-AeN706-34	Consensus	KC771204	*Uncultured marine eukaryote* (2013)	99.8	99.66
K-AeN706-35	Consensus	EU090019	*Fragilaria* sp. (2007)	99.7	100
K-AeN706-36	Consensus	AY485450	*Attheya longicornis* (2007)	99.7	99.61
K-AeN706-37	Consensus	JX401230	*Attheya longicornis* (2013)	99.9	99
K-AeN706-38	Consensus	KC771204	*Uncultured marine eukaryote* (2013)	99.9	100.0
K-AeN706-45	Consensus	GU373969	*Pseudo-nitzschia seriata* (2010)	99.4	99.94
K-AeN706-52	ND	ND	ND	ND	ND
K-AeN706-56	Consensus	EU371262	*Uncultured marine eukaryote* (2009)	97.5	100.0

Table 6. Strains identification based on 28S rRNA gene sequencing (BLAST). Table contains the code of the strain, type of the analysed sequence, accession number of the highest match in NCBI database, species name of the best match and submission year to the NCBI database, pairwise similarity (%) and the query cover (%).

Strain code	LSU (28S sequencing)	Best match (BLAST) accession number	Best match (BLAST) species name and submission year to NCBI database	Pairwise identity (%)	Query covered (%)
BE_AeN706-3	Consensus	HE610152	*Pyramimonas tetrarhynchus* (2012)	95.9	99.16
BE_AeN706-5	Consensus	OM688886	*Micromonas polaris* (2022)	100.0	99.24
BE_AeN706-6	Consensus	HE610152	*Pyramimonas tetrarhynchus* (2012)	95.9	99.66
BE_AeN706-8	Consensus	HE610152	*Pyramimonas tetrarhynchus* (2012)	95.7	100.0
BE_AeN706-9	Consensus	HE610152	*Pyramimonas tetrarhynchus* (2012)	95.7	99.67
BE_AeN706-10	Consensus	HE610152	*Pyramimonas tetrarhynchus* (2012)	95.7	99.67
BE_AeN706-11	Consensus	OK147701	*Fragilariopsis* sp. (2022)	99.6	99.3
BE_AeN706-12	Consensus	MT831988	*Prorocentrum* sp. (2020)	99.4	99.71
BE_AeN706-13	Consensus	OM688886	*Micromonas polaris* (2022)	99.4	99.27
BE_AeN706-15	Consensus	OM688886	*Micromonas polaris* (2022)	100.0	98.31
BE_AeN706-16	D1R-F	OM688886.1	*Micromonas polaris* (2022)	98.82	85
BE_AeN706-19	D1R-F	OM688886.1	*Micromonas polaris* (2022)	98.64	87
BE_AeN706-21	Consensus	OK147711	*Chaetoceros neogracilis* (2022)	100.0	100.0
BE_AeN706-22	D1R-F	OK147711	*Chaetoceros neogracilis* (2022)	99.8	100.0
BE_AeN706-23	Consensus	OK147689	*Thalassiosira gravida* (2022)	99.8	100.0
BE_AeN706-24	D1R-F	OK147711	*Chaetoceros neogracilis* (2022)	99.8	100.0
BE_AeN706-25	Consensus	OK147689	*Thalassiosira gravida* (2022)	100.0	99.84
BE_AeN706-26	Consensus	OK147689	*Thalassiosira gravida* (2022)	99.7	99.22
BE_AeN706-27	Consensus	MH843510	*Bacterosira* sp. (2018)	98.1	100.0
BE_AeN706-30	D2C-R	KT884484	*Chaetoceros* cf. *neogracilis* (2016)	100.0	100.0
BE_AeN706-31	D2C-R	KT884485	*Chaetoceros* cf. *neogracilis* (2016)	99.8	99.82
BE_AeN706-34	D2C-R	OK147711	*Chaetoceros neogracilis* (2022)	100.0	100.0
BE_AeN706-35	D2C-R	OK147711	*Chaetoceros neogracilis* (2022)	100.0	100.0
K-AeN706-5	D2C-R	OK147689	*Thalassiosira gravida* (2022)	99.8	100.0
K-AeN706-6	D2C-R	OK147689	*Thalassiosira gravida* (2022)	99.3	100.0
K-AeN706-7	Consensus	KY852287	*Chaetoceros cinctus* (2017)	99.7	99.83
K-AeN706-11	D2C-R	OK147711	*Chaetoceros neogracilis* (2022)	99.5	99.66
K-AeN706-15	D1R-F	JQ995405	*Attheya septentrionalis* (2018)	100.0	100.0
K-AeN706-17	Consensus	AF289045	*Apedinella radians* (2000)	88.5	99.81
K-AeN706-18	D2C-R	OK147711	*Chaetoceros neogracilis* (2022)	99.8	100.0
K-AeN706-19	Consensus	GU734797	*Neodenticula seminae* (2010)	99.1	98.91
K-AeN706-22	Consensus	AF417667	*Nitzschia lecointei* (2009)	96.7	74.66
K-AeN706-26	Consensus	MW176068	*Planktoniella tubulata* (2021)	88.2	26.48
K-AeN706-28	Consensus	OK147711	*Chaetoceros neogracilis* (2022)	99.5	100.0
K-AeN706-30	Consensus	MN725812	*Nitzschia* cf. *palea* (2021)	98.5	99.64
K-AeN706-32	D2C-R	KU898815	*Uncultured eukaryote clone Billefjorden_Sea_Ice14* (2016)	97.7	99.43
K-AeN706-33	Consensus	MN725812	*Nitzschia* cf. *palea* (2021)	98.5	99.46
K-AeN706-34	Consensus	OK147711	*Chaetoceros neogracilis* (2022)	99.8	99.31
K-AeN706-35	Consensus	AF417685	*Synedropsis hyperboreoides* (2009)	98.6	99.4
K-AeN706-36	Consensus	MH020639	*Attheya septentrionalis* (2018)	99.7	99.68
K-AeN706-37	Consensus	MH020639	*Attheya septentrionalis* (2018)	99.8	98.58
K-AeN706-38	Consensus	OK147711	*Chaetoceros neogracilis* (2022)	99.8	99.82
K-AeN706-45	Consensus	OK147699	*Pseudo-nitzschia granii* (2022)	99.6	99.09
K-AeN706-52	D2C-R	MH843510	*Bacterosira* sp. (2018)	97.5	100.0
K-AeN706-56	Consensus	JQ995464.1	*Thalassiosira hispida* (2018)	100	97

The BLAST search of the 18S rRNA sequences identified *Pyramimonas australis* as the best match for the BE-3 strain. For strains BE-6, BE-8, and BE-10, the best matching DNA sequence was *Pyramimonas parkeae*, specifically belonging to the CCMP 726 strain (Marin et al., 2010).

However, the BLAST analysis of the 28S rRNA sequences for the BE-3, BE-6, BE-8, BE-9, and BE-10 strains revealed the same best matching DNA sequence from the NCBI database, which is *Pyramimonas tetrarhynchus*. This best matching sequence corresponds to the strain SCCAP K-0002 (Marin, 2012).

The 18S rRNA sequences of the BE-5, BE-13, and BE-15 strains all share the same best match, which is *Micromonas polaris* .The best match for the BE-19 strain 18S rRNA sequence is *Micromonas pusilla*. *Micromonas pusilla* has, however, been divided into several species, including *Micromonas polaris*.

Micromonas pusilla is the type species (holotypes) of the genus *Micromonas* (Guiry, 2023), but has been suggested to be renamed *Micrinomonas pusilla* (R.W.Butcher) Doweld (Guiry, 2023). The two names are both valid and are synonyms.

The 28S rRNA sequences of the strains BE-5, BE-13, BE-15, BE-16 and BE-19 showed the same best matching DNA sequence from the NCBI database, which is *Micromonas polaris*.

The 18S rRNA sequences of the BE-11 and K19 strains share the same best match sequence, which is *Fragilariopsis kerguelensis*. In the case of the 28S rRNA sequence BLAST search, the strain BE-11 corresponds to *Fragilariopsis* sp., while for K-19, it corresponds to *Neodenticula seminae*.

The 18S rRNA sequence blast search for the BE-12 strain identified *Prorocentrum shikokuense* as the best match. However, in the BLAST analysis of the 28S rRNA sequence, the best match for this strain was identified as *Prorocentrum* sp.

The 18S rRNA sequence BLAST search for strains BE-21, BE-22, and BE-24 shows the best match as *Chaetoceros* sp. However, for strains BE-34, BE-35, K-11, K-18, K-28, K-34, and K-38, the same best matching DNA sequence from the NCBI database is identified as an uncultured marine eukaryote (accession number KC771204). It is worth noting that the second-best match in the NCBI database for these strains, with a pairwise identity value of 100% and a query covered value of 99%, is *Chaetoceros* sp. (accession number EU090014) (Choi et al., 2008).

In the case of the 28S rRNA sequence BLAST search, the strains BE-21, BE-22, BE-24, BE-34, BE-35, K-11, K-18, K-28, K-34, and K-38 all have the same best matching DNA sequence from the NCBI database, which is identified as *Chaetoceros neogracilis*. However, for strains BE-30 and BE-31, the best match is identified as *Chaetoceros* cf. *neogracilis*.

The 18S rRNA sequence BLAST search for strains BE-26 and K-6 shows that the best match is *Thalassiosira rotula*. However, for the 28S rRNA sequence of strains BE-23, BE-25, BE-26, K-5, and K-6, the BLAST search indicates that the best match DNA sequence is *Thalassiosira gravida*. It is important to mention that the 18S rRNA sequence is not available for strains BE-23, BE-25, and K-5.

The 18S rRNA sequence BLAST search for strain BE-27 shows that the best match is an uncultured marine eukaryote. The second-best match in the NCBI database for this strain, with a pairwise identity value of 99.49% and a query covered value of 100%, is *Bacterosira constricta* (accession number KT692951.1) (Park et al., 2016). However, for the 28S rRNA sequence of strain BE-27, the BLAST search indicates that the best match DNA sequence is *Bacterosira* sp.

The BLAST analysis for the 18S rRNA sequences of strain K-7 resulted in an uncultured marine eukaryote as the best match. The second-best match in the NCBI database for this strain, with a pairwise identity value of 99.77% and a query covered value of 98%, is *Chaetoceros cinctus* (accession number KY852266.1) (Gaonkar et al., 2017). The BLAST search for the 28S rRNA sequence also showed *Chaetoceros cinctus* as the best match for strain K-7.

For the strains K-15, K-36, and K-37 the 18S rRNA sequence shows the best matching sequence through a BLAST search as *Attheya longicornis*. However, it is important to mention that *Attheya septentrionalis* (accession number HQ912618.1), with a pairwise identity value of 100% and a query covered value of 98%, is also among the best match sequences in the NCBI database (Theriot et al., 2010). The BLAST search for the 28S rRNA sequences of strains K-15, K-36 and K-37 showed *Attheya septentrionalis* as the best match.

The K-17 strain shows the best match to Pedinellales sequence for the 18S rRNA sequence and to *Apedinella radians* for the 28S rRNA sequence in the NCBI database, as determined via a BLAST search.

The BLAST analysis for the 18S rRNA sequences of strain K-22 revealed an uncultured marine eukaryote as the best match. *Nitzschia longissima* (accession number MT259195) showed a pairwise identity value of 98.96% and a query coverage of 100%. It is considered one of the best match sequences in the NCBI database as well. As for the 28S rRNA, the best match was found to be *Nitzschia lecointei*.

The K-26 strain shows the best match to *Actinocyclus curvatulus* for the 18S rRNA sequence and to *Planktoniella tubulata* for the 28S rRNA sequence in the NCBI database, as determined via a BLAST search.

The K-30 strain shows the best match to *Nitzschia thermalis* for the 18S rRNA sequence and to *Nitzschia* cf. *palea* for the 28S rRNA sequence in the NCBI database, as determined via a BLAST search.

The BLAST analysis for the 18S rRNA sequences of strain K-32 identified an uncultured marine eukaryote as the best match. However, the second best match was *Bacillaria* cf. *paxillifer* (accession number HM805020), with a pairwise identity value of 98.39% and a query coverage of 100% (Pniewski et al., 2010). Regarding the 28S rRNA sequence, the best match was an uncultured eukaryote. The second best match was *Nitzschia lecointei* (accession number AF417667), with a relatively lower pairwise identity value of 93.04% and a query coverage of 99%.

For the K-35 strain, the BLAST analysis for the 18S rRNA sequence revealed *Fragilaria* sp. as the best match, while *Synedropsis hyperboreoides* was the best match for the 28S rRNA sequence in the NCBI database, as determined via a BLAST search.

Similarly, for the K-45 strain, the best match for the 18S rRNA sequence was *Pseudo-nitzschia seriata*, and for the 28S rRNA sequence, it was *Pseudo-nitzschia granii* in the NCBI database, as determined via a BLAST search.

The BLAST analysis for the 28S rRNA sequences of strain K-52 identified *Bacterosira* sp. as the best match. The 18S rRNA sequence for this strain is not available.

Regarding the K-56 strain, the BLAST analysis for the 18S rRNA sequence identified an uncultured marine eukaryote as the best match. Additionally, one of the best matches was *Minidiscus trioculatus* (accession number FJ590769.1), with a pairwise identity value of 98.67% and a query coverage of 100% (Kaczmarska et al., 2009). As for the 28S rRNA sequence, the best match was *Thalassiosira hispida*.

Table 7. Comparison of BLAST analysis results for 18S and 28S genes of the studied strains. Table includes the strain code, the highest match for the strain based on BLAST analysis results for 18S gene sequence, the highest match for the strain based on BLAST analysis results for 28S gene sequence and a concluding identification based on both results. ND - no data.

Strain code	Identification by DNA sequencing, SSU (BLAST)	Identification by DNA sequencing, LSU (BLAST)	Identification total
BE_AeN706-3	*Pyramimonas australis*	*Pyramimonas tetrarhynchus*	*Pyramimonas* sp.
BE_AeN706-5	*Micromonas polaris*	*Micromonas polaris*	*Micromonas polaris*
BE_AeN706-6	*Pyramimonas parkeae*	*Pyramimonas tetrarhynchus*	*Pyramimonas* sp.
BE_AeN706-8	*Pyramimonas parkeae*	*Pyramimonas tetrarhynchus*	*Pyramimonas* sp.
BE_AeN706-9	ND	*Pyramimonas tetrarhynchus*	*Pyramimonas* sp.
BE_AeN706-10	*Pyramimonas parkeae*	*Pyramimonas tetrarhynchus*	*Pyramimonas* sp.
BE_AeN706-11	*Fragilariopsis kerguelensis*	*Fragilariopsis* sp.	*Fragilariopsis* sp.
BE_AeN706-12	*Prorocentrum shikokuense*	*Prorocentrum* sp.	*Prorocentrum* sp.
BE_AeN706-13	*Micromonas polaris*	*Micromonas polaris*	*Micromonas polaris*
BE_AeN706-15	*Micromonas polaris*	*Micromonas polaris*	*Micromonas polaris*
BE_AeN706-16	ND	*Micromonas polaris*	*Micromonas* sp.
BE_AeN706-19	*Micromonas pusilla*	*Micromonas polaris*	*Micromonas* sp.
BE_AeN706-21	*Chaetoceros* sp.	*Chaetoceros neogracilis*	*Chaetoceros* cf. *neogracilis*
BE_AeN706-22	*Chaetoceros* sp.	*Chaetoceros neogracilis*	*Chaetoceros* cf. *neogracilis*
BE_AeN706-23	ND	*Thalassiosira gravida*	*Thalassiosira gravida*
BE_AeN706-24	*Chaetoceros* sp.	*Chaetoceros neogracilis*	*Chaetoceros* cf. *neogracilis*
BE_AeN706-25	ND	*Thalassiosira gravida*	*Thalassiosira gravida*
BE_AeN706-26	*Thalassiosira rotula*	*Thalassiosira gravida*	*Thalassiosira gravida*
BE_AeN706-27	*Uncultured marine eukaryote*	*Bacterosira* sp.	*Bacterosira* sp.
BE_AeN706-30	*Uncultured marine eukaryote*	*Chaetoceros* cf. *neogracilis*	*Chaetoceros* cf. *neogracilis*
BE_AeN706-31	*Uncultured marine eukaryote*	*Chaetoceros* cf. *neogracilis*	*Chaetoceros* cf. *neogracilis*
BE_AeN706-34	*Uncultured marine eukaryote*	*Chaetoceros neogracilis*	*Chaetoceros* cf. *neogracilis*
BE_AeN706-35	*Uncultured marine eukaryote*	*Chaetoceros neogracilis*	*Chaetoceros* cf. *neogracilis*
K-AeN706-5	ND	*Thalassiosira gravida*	*Thalassiosira gravida*
K-AeN706-6	*Thalassiosira rotula*	*Thalassiosira gravida*	*Thalassiosira gravida*
K-AeN706-7	*Uncultured marine eukaryote*	*Chaetoceros cinctus*	*Chaetoceros* cf. *cinctus*
K-AeN706-11	*Uncultured marine eukaryote*	*Chaetoceros neogracilis*	*Chaetoceros* cf. *neogracilis*
K-AeN706-15	*Attheya longicornis*	*Attheya septentrionalis*	*Attheya septentrionalis*
K-AeN706-17	Pedinellales	*Apedinella radians*	Pedinellales
K-AeN706-18	*Uncultured marine eukaryote*	*Chaetoceros neogracilis*	*Chaetoceros* cf. *neogracilis*
K-AeN706-19	*Fragilariopsis kerguelensis*	*Neodenticula seminae*	*Fragilariopsis* sp.
K-AeN706-22	*Uncultured marine eukaryote*	*Nitzschia lecointei*	*Nitzschia* sp.
K-AeN706-26	*Actinocyclus curvatulus*	*Planktoniella tubulata*	*Actinocyclus* sp.
K-AeN706-28	*Uncultured marine eukaryot*	*Chaetoceros neogracilis*	*Chaetoceros* cf. *neogracilis*
K-AeN706-30	*Nitzschia thermalis*	*Nitzschia* cf. *palea*	*Nitzschia* sp.
K-AeN706-31	*Amphora* sp.	ND	*Amphora* sp.
K-AeN706-32	*Uncultured marine eukaryote*	*Uncultured eukaryote*	*Uncultured eukaryote*
K-AeN706-33	*Nitzschia aequorea*	*Nitzschia* cf. *palea*	*Nitzschia* sp.
K-AeN706-34	*Uncultured marine eukaryote*	*Chaetoceros neogracilis*	*Chaetoceros* cf. *neogracilis*
K-AeN706-35	*Fragilaria* sp.	*Synedropsis hyperboreoides*	*Fragilaria* sp.
K-AeN706-36	*Attheya longicornis*	*Attheya septentrionalis*	*Attheya longicornis* cf. *septentrionalis*
K-AeN706-37	*Attheya longicornis*	*Attheya septentrionalis*	*Attheya longicornis* cf. *septentrionalis*
K-AeN706-38	*Uncultured marine eukaryote*	*Chaetoceros neogracilis*	*Chaetoceros* cf. *neogracilis*
K-AeN706-45	*Pseudo-nitzschia seriata*	*Pseudo-nitzschia granii*	*Pseudo-nitzschia* sp.
K-AeN706-52	ND	*Bacterosira* sp.	*Bacterosira* sp.
K-AeN706-56	*Uncultured marine eukaryote*	*Thalassiosira hispida*	*Thalassiosira* cf. *hispida*

3.2 Phylogenetic analysis

In total, we constructed eight phylogenetic trees (see Figures 5-12 below) to illustrate the phylogenetic relationships and to obtain a more precise taxonomic placement of the cultured strains.

Only bootstrap values above 70 are shown. The scale bar of each phylogenetic tree shows the estimated number of nucleotide substitutions per site (number of substitutions divided by the total number of characters in the sequence alignment). By comparing the branch lengths of the tree to the scale bar, one can gain a better understanding of the evolutionary distances between the different branches and the degree of divergence between the sequences. Together with the branch order (tree topology) his information can be used to infer relationships between the organisms being studied and to generate hypotheses about their evolutionary history.

The sequences obtained in this study lack accession numbers since they have not yet been submitted to a genetic sequence database. However, these sequences will be deposited into the GenBank (NCBI) database at a later time. The algal strains have been submitted to and will be available from the Norwegian culture collection (NORCCA). This process of depositing the sequences into a publicly accessible database will allow for others in the scientific community to access and utilise the data, thereby promoting further research and collaboration in the field.

Chlorophyta

The 18S rRNA gene phylogenetic tree of the phylum Chlorophyta can be seen in Figure 5. The 28S rRNA gene phylogenetic tree of the phylum Chlorophyta can be seen in Figure 6 below.

Of the studied strains, nine representatives belong to the phylum Chlorophyta and are classified into two distinct classes, Pyramimonadophyceae and Mamiellophyceae.

Class Pyramimonadophyceae

***Pyramimonas* sp.**

In the SSU rRNA gene region strains BE-3, BE-6, BE-8, BE-9, BE-10 were identical. The sequence of strain BE-9 is shorter than the others (853 bp) and has a mismatch with the sequences of BE-3, BE-6, BE-8, BE-10 by 1 bp in the corresponding part of the gene. All five strains are most similar to the *Pyramimonas australis* (accession number KY980369), but not identical. We can identify these strains as *Pyramimonas* sp.

In the LSU rRNA gene region strains BE-3, BE-6, BE-8, BE-9, BE-10 were identical. The sequences formed a sister group to sequences of *Pyramimonas tetrarhynchus* and *Pyramimonas parkeae.* The difference between analysed sequences and reference sequence of

Pyramimonas tetrarhynchus was in a few positions of the partial 28S gene. *Pyramimonas parkeae* had even more nucleotide mismatches. These strains will thus be identified to *Pyramimonas* sp.

Class Mamiellophyceae

Micromonas polaris

In the SSU rRNA gene region strains BE-19, BE-15, BE-5, and BE-13 have only few base pair difference of their sequences. These strains are most similar to the reference sequence of *Micromonas polaris* (accession number ON888444*)*, but the reference sequences are shorter than our analysed sequences.

In the LSU rRNA gene region strains BE-5, BE-15, BE-13 have some difference in length of the sequences and are identical in the overlapping region. The closest reference sequence is *Micromonas polaris* (OM688886) sequence with 1 bp mismatch. Among the reference sequences the is also one identified to the species *Micromonas pusilla. Micromonas polaris* and *Micrinomonas pusilla* sequences in the overlapping region are identical. These strains will be referred to as *Micromonas polaris*.

The partial 28S rRNA gene of strains BE-19 and BE-16 was sequenced using forward and reverse primers, but we obtained a useful sequence only with the forward primer (D1R-F). These are thus sequences based on only one strand, but of good quality and rather long. The sequence of strain BE-19 is 664 bp long, and the sequence of strain BE-16 is 670 bp long. In the area of overlap, the two sequences are identical. Most similar sequence is *Micromonas polaris* (accession number OM688882). The BE-19 and BE-16 strains sequences differ from BE-5, BE-15, and BE-13 strains sequences in 8 positions, suggesting that these sequences belong to a different species than *Micromonas polaris*. The strains BE-19 and BE-16 will be referred to as *Micromonas* sp.

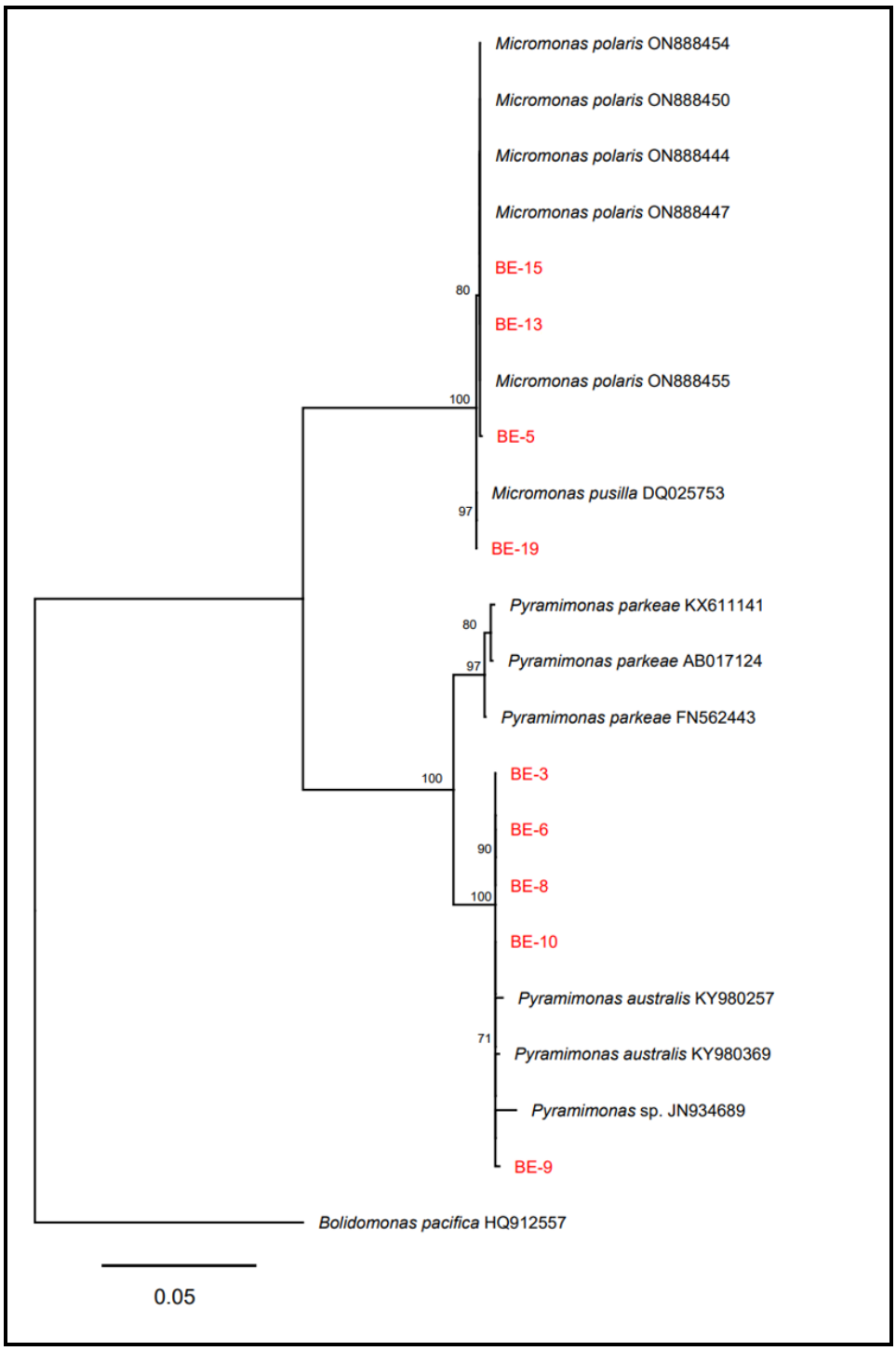


Figure 5. The 18S rRNA gene phylogenetic tree of cultured strains of the taxonomic group Chlorophyta. Maximum-likelihood was generated in RAxML using GTR GAMMAI model with 1000 bootstrap replications. Only bootstrap values above 70 are shown. The scale-bar of the tree is a genetic distance of 0.05 base pairs. The cultured strains are marked in red and reference sequences are marked in black. The DNA sequence BE-9 is represented by a single sequence provided by forward primer 850F. Other sequences are represented by the consensus sequence of both the forward and reverse strand.

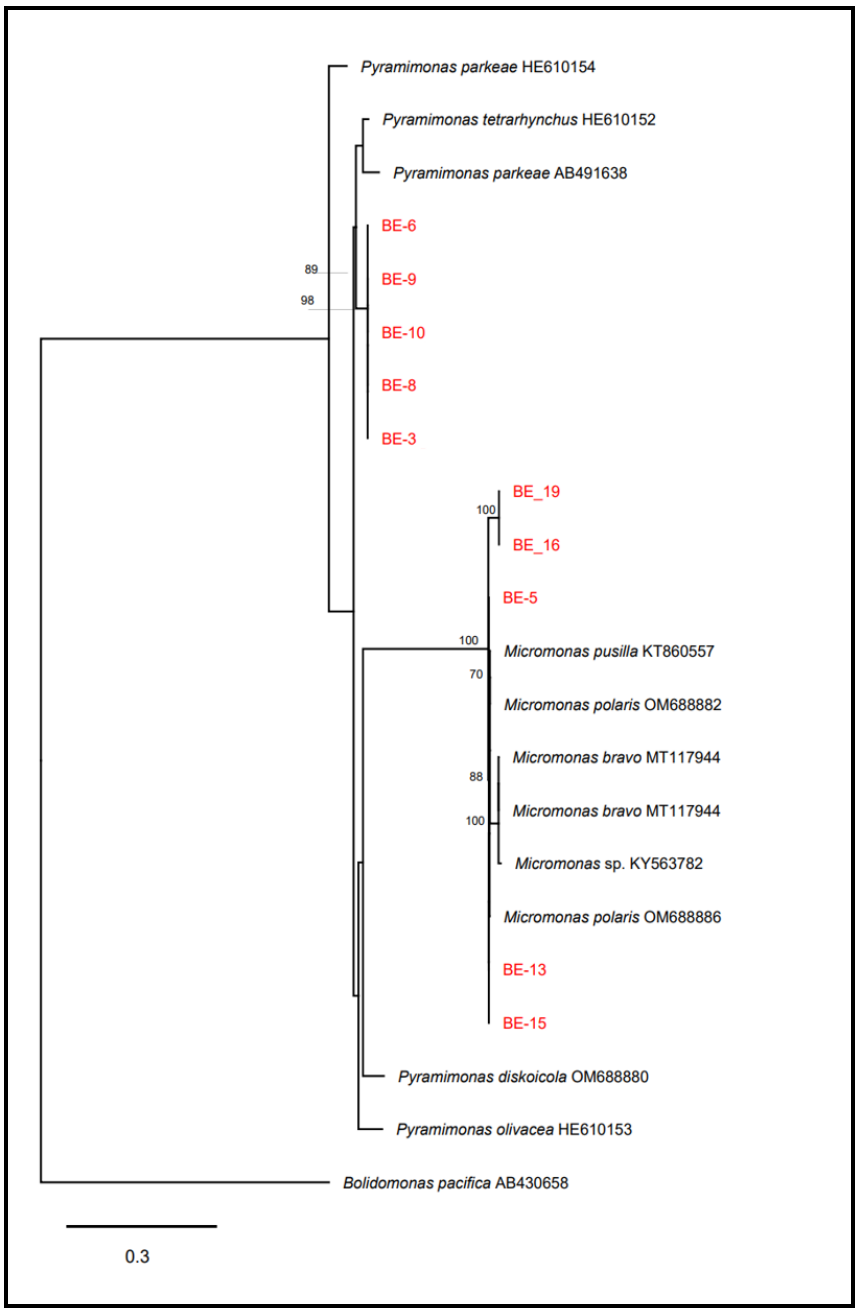


Figure 6. The 28S rRNA phylogenetic tree of cultured strains of the taxonomic group Chlorophyta. Maximum-likelihood was generated in RAxML using GTR GAMMAI model with 1000 bootstrap replications. Only bootstrap values above 70 are shown. The scale-bar of the tree is a genetic distance of 0.3 base pairs. The cultured strains are marked in red and reference sequences are marked in black. Strain BE_19 and BE_16 represented by single sequence provided by forward primer. Other strains are presented by the consensus sequence.

Miozoa

Class Dinophyceae

18S rRNA gene sequence of the BE-12 strain (Figure 7 below) is most similar to *Prorocentrum minimum* (accession number JX402086). The alignment has a few base pair differences. According to the World Register of Marine Species (WoRMS), *Prorocentrum minimum* has changed its name to *Prorocentrum cordatum* (Ostenfeld) J.D.Dodge, 1976 (Guiry, 2023).

The 28S rRNA gene sequence of the BE-12 strain (Figure 8 below) is most similar to *Prorocentrum* sp. sequence (accession number MT831988), but not identical. The alignment shows three base pair differences between BE-12 and MT831988 sequence. This strain will be referred to as *Prorocentrum* sp.

However, a new alignment for the 18S rRNA gene of the BE-12 sequence with *Prorocentrum* reference sequences was completed by Bente Edvardsen after the article about *Prorocentrum pervagatum* (Tillmann at al., 2023) was published. A *Prorocentrum* tree which include BE-12 18S rRNA sequence placed it together with a *Prorocentrum pervagatum* sequence (Appendix 3, Figure A3). BE-12 18S rRNA sequence is totally identical to *Prorocentrum pervagatum* (accession number OP094113).

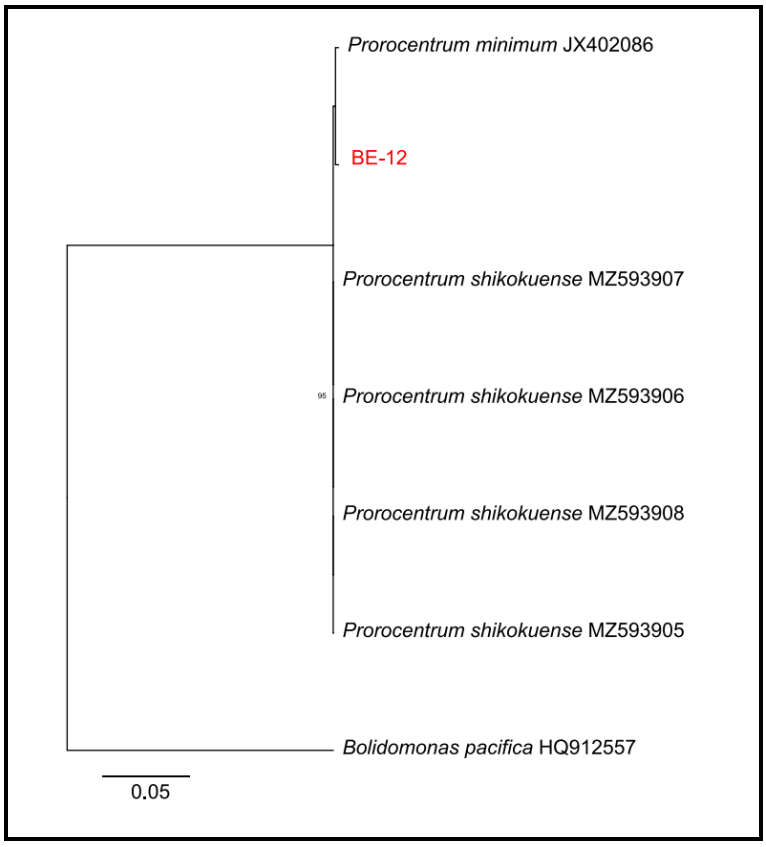


Figure 7. The 18S rRNA phylogenetic tree of BE-12 (UIO 573) strain, dinoflagellate taxonomic group. Maximum-likelihood was generated in RAxML using GTR GAMMAI model with 1000 bootstrap replications. Only bootstrap values above 70 are shown. The scale-bar of the tree is a genetic distance of 0.05 base pairs. The cultured strain sequenced here is labelled in red and reference sequences are labelled in black.

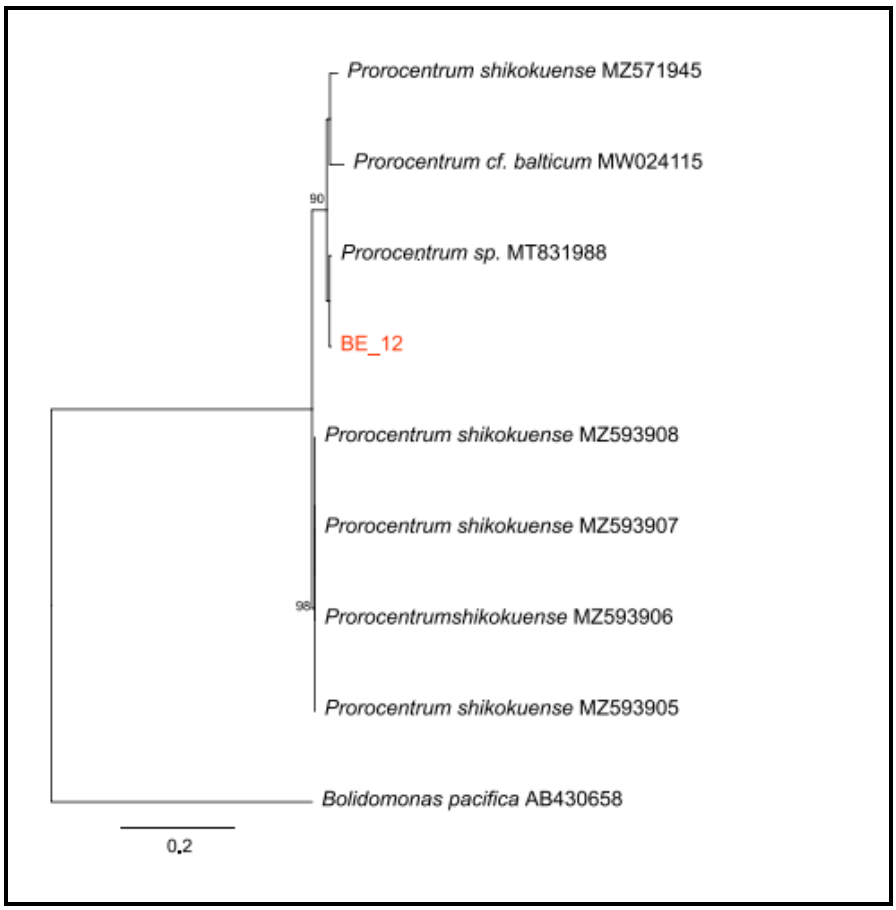


Figure 8. The 28S rRNA phylogenetic tree of BE-12 strain in the dinoflagellate taxonomic group. Maximum-likelihood was generated in RAxML using GTR GAMMAI model with 1000 bootstrap replications. Only bootstrap values above 70 are shown. The scale-bar of the tree is a genetic distance of 0.2 base pairs. The cultured strain sequenced here is labelled in red and reference sequences are labelled in black.

Ochrophyta

Class Dictyochophyceae

Strain K-17

The SSU rRNA gene sequence of the K-17 strain (Figure 9 below) is identical with sequence belonging to the order Pedinellales (accession nmber JN934682) which is of the RCC2301 strain isolated in Arctic Ocean (Canada) (Balzano et al. 2012a). Also, strain K-17 has an identical sequence to Pedinellales CCMP2098 strain (accession number EU247836), isolated in the Arctic Ocean (Canada) (Hamilton et al., 2008).

The LSU rRNA gene sequence of the K-17 strain (Figure 10 below) does not have an identical reference sequence, which means that there is no matching species in the NCBI database described earlier. However, this sequence forms a sister clade with *Pteridomonas danica* which is consistent with previous findings (Balzano et al., 2012a). This strain will be referred to as Pedinellales.

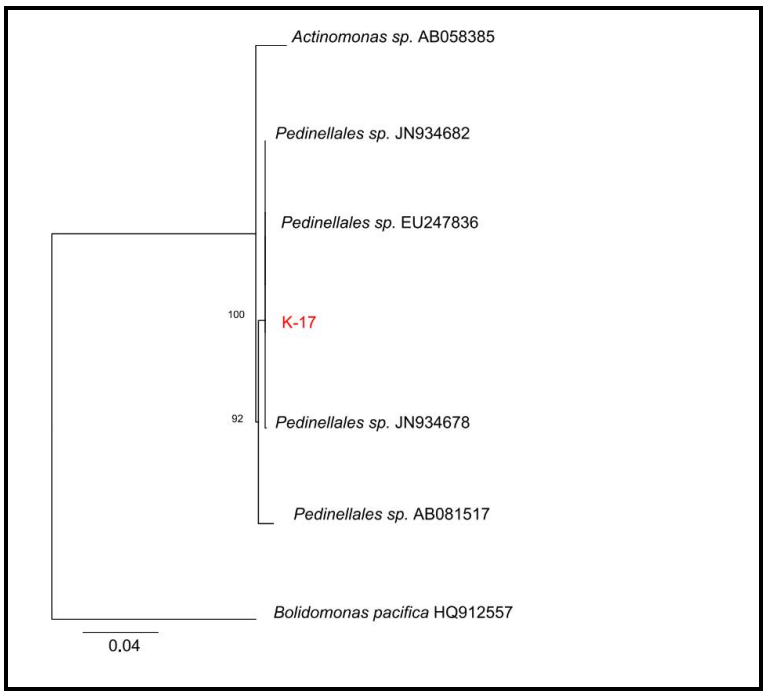


Figure 9. The 18S rRNA phylogenetic tree of K-17 strain of the dictyochophyte taxonomic group. Maximum-likelihood was generated in RAxML using GTR GAMMAI model with 1000 bootstrap replications. Only bootstrap values above 70 are shown. The scale-bar of the tree is a genetic distance of 0.04 base pairs. The cultured strain sequenced here is labelled in red and reference sequences are labelled in black.

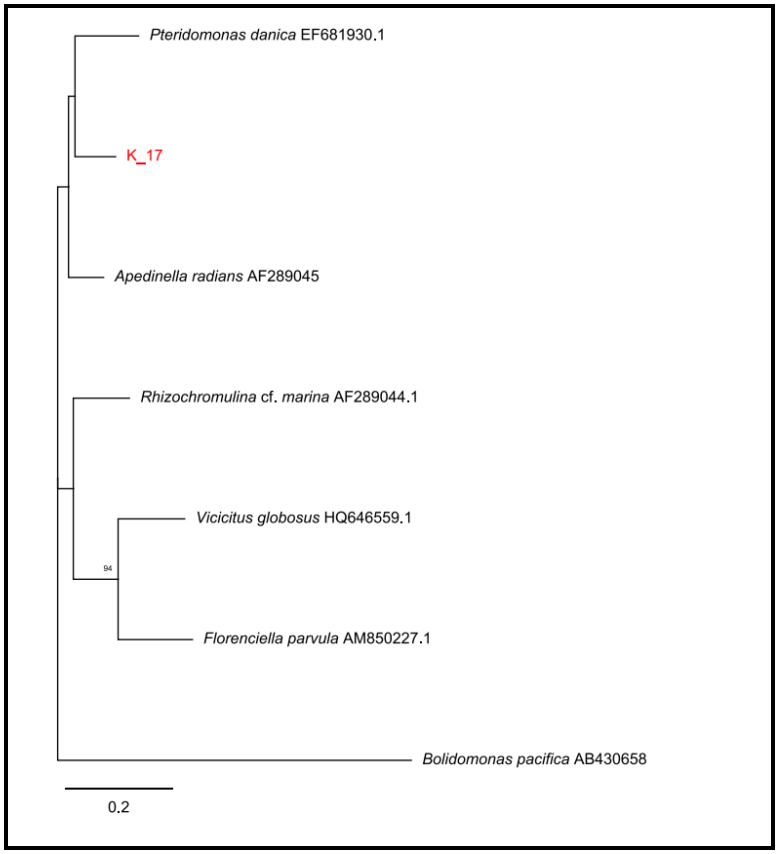


Figure 10. The 28S rRNA phylogenetic tree of K-17 strain, dictyochophyte taxonomic group. Maximum-likelihood was generated in RAxML using GTR GAMMAI model with 1000 bootstrap replications. Only bootstrap values above 70 are shown. The scale-bar of the tree is a genetic distance of 0.2 base pairs. The cultured strain is marked in red and reference sequences are marked in black.

Bacillariophyta Diatoms

Phylogenetic trees are shown in Figures 11 and 12.

Class Bacillariophyceae

***Fragilaria* sp.**

The 18S rRNA gene sequence of the K-35 is most similar to *Fragilaria* cf. *striatula* (accession number is AJ971377) but not identical. The 28S rRNA gene sequence of the K-35 is most similar to *Fragilariaceae* (accession number is JQ995460). This strain will be referred to as *Fragilaria* sp.

***Nitzschia* sp.**

The strains K-30 and K-33 are identical and are grouped together in 18S and in 28S rRNA phylogenetic trees. However, identical reference sequences are missing. In 18S rRNA phylogenetic tree most similar sequence belongs to species *Nitzschia aequorea* (accession number KY320391). In 28S rRNA phylogenetic tree most similar sequence belongs to species *Nitzschia* cf. *palea* (accession number MN725812). These strains will be referred to as *Nitzschia* sp.

The strain K-22 has no identical sequences in either of the phylogenetic trees. The closest sequence belongs to *Nitzschia longissima* (accession number AY881968) in 18S rRNA phylogenetic tree. The closest sequence belongs to *Nitzschia frustulum* (accession number KX839245) in 28S rRNA phylogenetic tree. This strain will be referred to as *Nitzschia* sp.

Pseudo-nitzschia

The sequences of strains BE-11 and K-19 are identical and appear together in both the 18S and 28S rRNA phylogenetic trees. However, determining their exact evolutionary relationship is challenging. These strains belong to a sister clade that includes distinct sequences from the genus Pseudo-nitzschia in both trees. Additionally, in the 18S rRNA tree, there is a close relation to the genus Fragilariopsis, while in the 28S rRNA tree, the relation to the family Fragilariaceae is more distant. These strains will be referred to as cf. *Pseudo-nitzschia.*

The strain K-45 is most similar to the species *Pseudo-nitzschia subcurvata (*accession number is KX253952*)* in 18S rRNA phylogenetic tree and is identical with the species *Pseudo-nitzschia granii* (accession number is OK147699) in 28S rRNA phylogenetic tree. This strain will be referred to as *Pseudo-nitzschia* sp.

Amphora sp.

The 18S rRNA gene sequence of the strain K-31 is very similar to *Amphora* sp. sequence (accession number JQ240485) but not identical. The 28S rRNA gene sequence of the strain K-31 is not available. This strain will be referred to as *Amphora* sp.

***Bacillaria* sp.**

The strain K-32 is most similar to the species *Bacillaria* cf. *paxillifer (*accession number is HM805020*)* in 18S rRNA phylogenetic tree. In the 28S rRNA phylogenetic tree this strain does not have identical sequences and is most similar to the uncultured eukaryote clone sequence (accession number is KU898815.1). This strain will be referred to as *Bacillaria* sp.

Class Mediophyceae

Thalassiosira gravida

The strains K-6 and BE-23 in 18S rRNA and in 28S rRNA phylogenetic tree are identical. Strains BE-25, BE-26, K-6, and K-5 are identical and are grouped together in 18S as well but are not available in the 28S rRNA phylogenetic tree. Identical reference sequence in the 18S rRNA phylogenetic tree is *Thalassiosira gravida* (accession number OK147688). Identical reference sequence in the 28S rRNA phylogenetic tree is *Thalassiosira gravida* (accession number OK147675). These strains will be referred to as *Thalassiosira gravida.*

The strain K-56 has no identical sequences in the 18S rRNA phylogenetic tree. The closest sequence is *Thalassiosira anguste-lineata* (accession number AJ810854). In the 28S rRNA phylogenetic tree this strain is identical with *Thalassiosira hispida* (accession number JQ995464), strain RCC2521 (Balzano et al., 2017). This strain will be referred to as *Thalassiosira* cf. *hispida.*

Attheya septentrionalis

The strain K-15 in 18S rRNA phylogenetic tree is identical with species *Attheya septentrionalis* (accession number is HQ912618), strain CCMP2084 (Theriot et al., 2010). In 28S rRNA phylogenetic tree this strain was identical with the same species (JQ995404). The K-15 strain will be referred to as *Attheya septentrionalis.*

Attheya longicornis

The strains K-36 and K-37 in 18S rRNA phylogenetic tree are identical and have an identical reference sequence, species *Attheya longicornis* (accession number is AY485450). In 28S rRNA phylogenetic tree these strains are identical and have an identical reference sequence, species *Attheya longicornis* (accession number is GQ219677) (Sorhannus et al., 2012). However, one of the identical reference sequences of these strains is *Attheya septentrionalis* (accession number is JQ995433) strain RCC2042 as well (Balzano et al., 2017). As this result is confirmed by both phylogenetic trees, the appropriate designation for these strains will be *Attheya longicornis*.

***Bacterosira* sp.**

The strain BE-27 in 18S rRNA phylogenetic tree is most similar to the *Bacterosira bathyomphala* (accession number is DQ514894). In 28S rRNA phylogenetic tree strain BE-27

is identical with strain K-52 (the 18S rRNA gene sequence is not available). The closest reference sequence is *Bacterosira* sp. (accession number is MH843510). These strains will be referred to as *Bacterosira* sp.

Chaetoceros neogracilis

The strains BE-21, BE-30, BE-31, BE-34, BE-35, K-11, K-18, K-28, K-34, K-38 are identical in both phylogenetic trees. The identical reference 18S rRNA sequence is *Chaetoceros neogracilis* (accession number is OM688898). The identical reference 18S rRNA sequence is *Chaetoceros neogracilis* (accession number is OK147711). These strains will be referred to as *Chaetoceros neogracilis.* The strains BE-22, BE-24 are placed together in *Chaetoceros neogracilis* group but with few bases pair difference in 18S rRNA phylogenetic tree. In 28S rRNA phylogenetic trees are identical with other reference sequences of species *Chaetoceros neogracilis.* These strains will be referred as *Chaetoceros* cf. *neogracilis.*

Chaetoceros cinctus

The strain K-7 in 18S rRNA phylogenetic tree has no identical reference sequences. Most similar to the species *Chaetoceros cinctus* (accession number is KY852268). In 28S rRNA phylogenetic tree this strain is identical with species *Chaetoceros cinctus* (accession number is KY852282). These strains will be referred to as *Chaetoceros* cf. *cinctus.*

Class Coscinodiscophyceae

***Actinocyclus* sp.**

Strain K-26 in18S rRNA phylogenetic tree is most similar to the *Actinocyclus* sp. sequence (accession number is KC309522) (Ashworth at al.,2013). In 28S rRNA phylogenetic tree this strain has no similar reference sequences. These strains will be referred to as *Actinocyclus* sp.

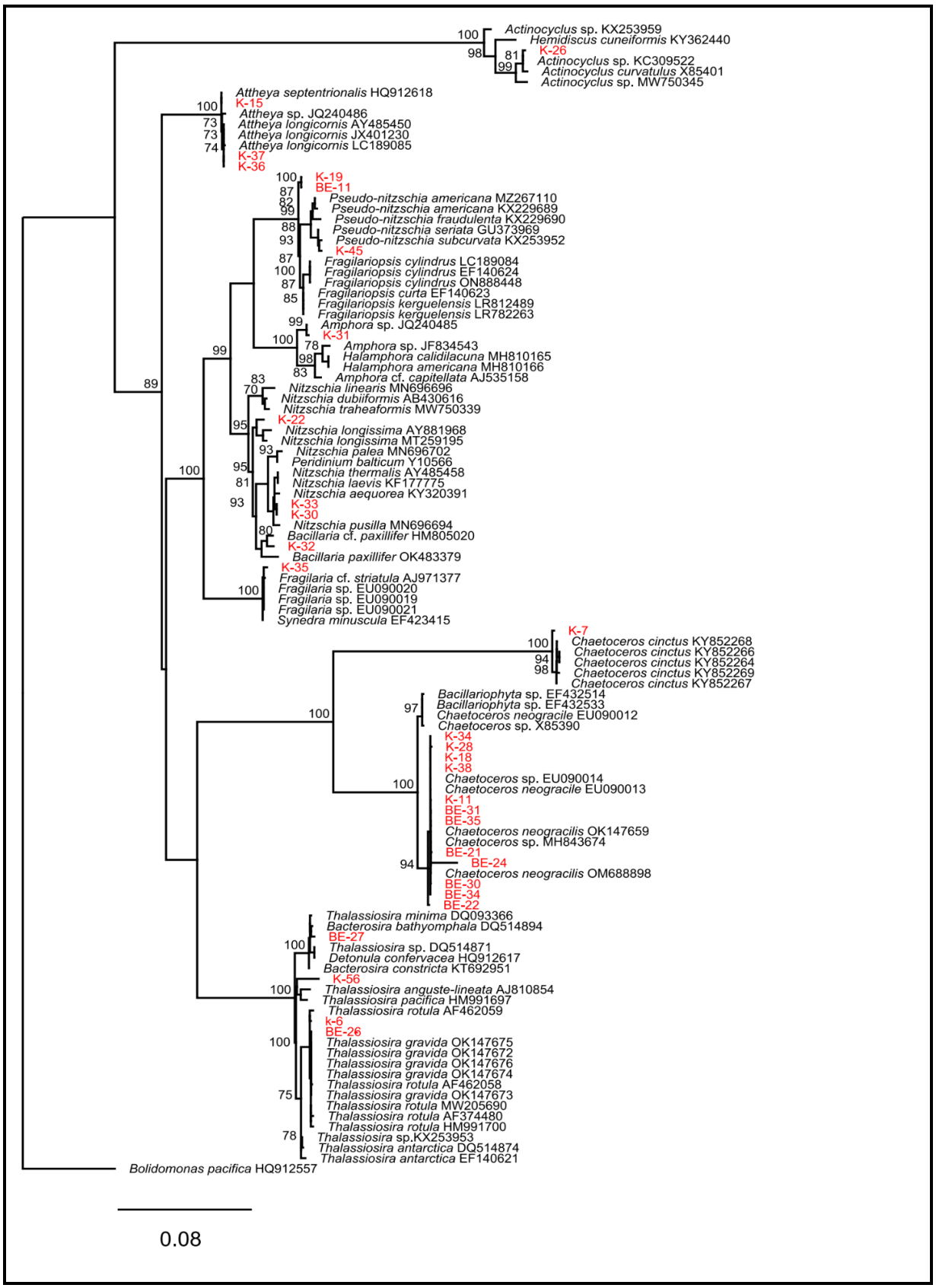


Figure 11. The 18S rRNA RaXML phylogenetic tree of cultured strains, taxonomic group, diatoms, GTR GAMMAI model with 1000 bootstrap replications. The scale-bar of the tree is a genetic distance of 0.08 base pairs. The cultured strains are marked in red and reference sequences are marked in black. Strain BE-30, BE-31, BE-34, BE-35 and K-11, K-31 represented by single sequence provided by forward primer 850F. Other strains are presented by consensus sequence.

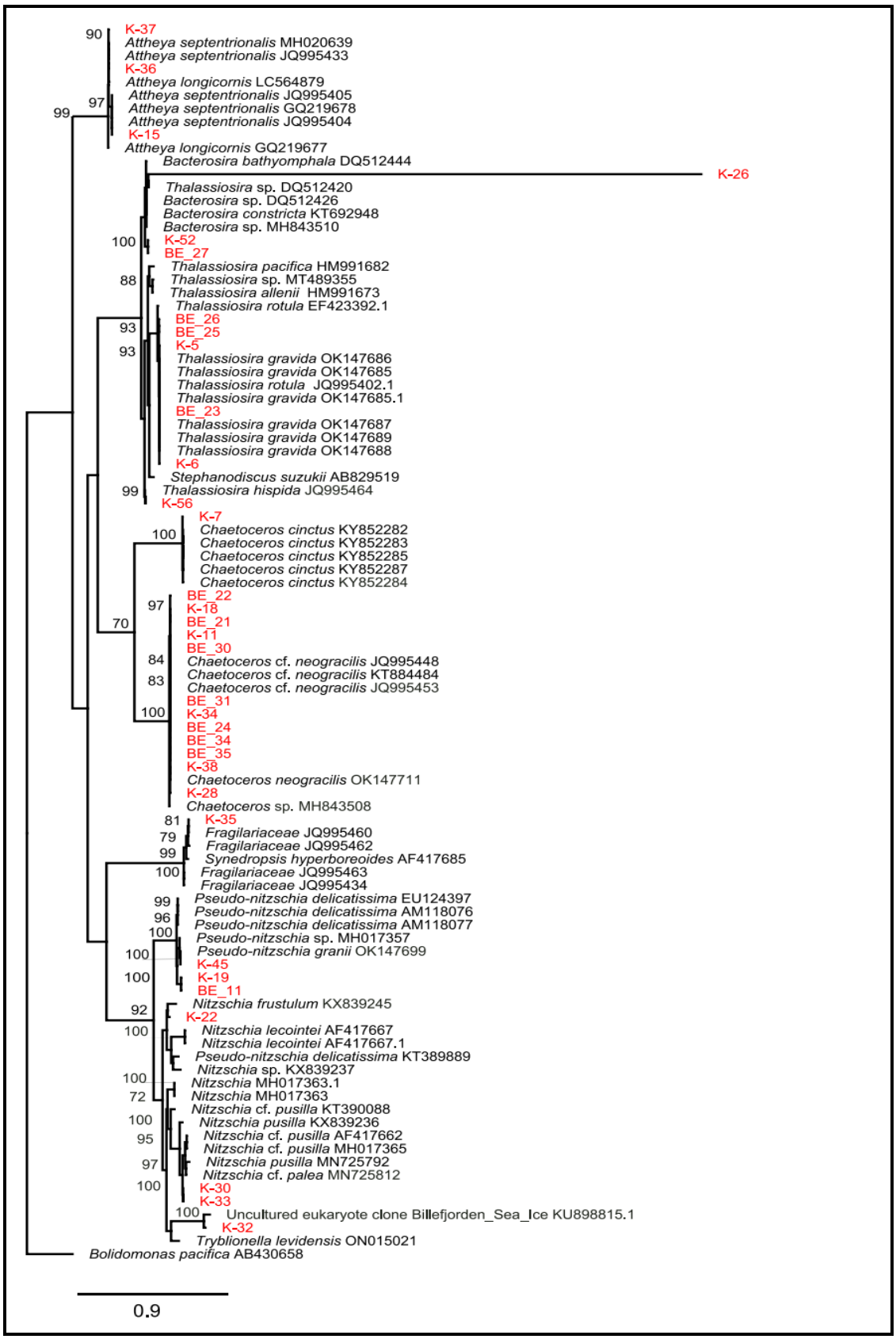


Figure 12. The 28S rRNA RaXML phylogenetic tree of cultured strains, taxonomic group, diatoms, GTR GAMMAI model with 1000 bootstrap replications. The scale-bar of the tree is a genetic distance of 0.9 base pairs. The cultured strains are marked in red and reference sequences are marked in black. Strain BE-22, BE-24, BE-30, BE-31, BE-34, BE-35 and K-5, K-6, K-11, K-15, K-52 represented by single sequence. Other strains are presented by consensus sequence.

Table 8. Strain identification based on phylogeny. Taxonomic classification of the identified algal strains was verified by checking in the online database Algae Base. ND – no data.

Strain code	Identification (phylogeny)	Phylum	Class	Genus	Species
BE_AeN706-3	*Pyramimonas* sp.	Chlorophyta	Pyramimonadophyceae	*Pyramimonas*	sp.
BE_AeN706-5	*Micromonas polaris*	Chlorophyta	Mamiellophyceae	*Micromonas*	*polaris*
BE_AeN706-6	*Pyramimonas* sp.	Chlorophyta	Pyramimonadophyceae	*Pyramimonas*	sp.
BE_AeN706-8	*Pyramimonas* sp.	Chlorophyta	Pyramimonadophyceae	*Pyramimonas*	sp.
BE_AeN706-9	*Pyramimonas* sp.	Chlorophyta	Pyramimonadophyceae	*Pyramimonas*	sp.
BE_AeN706-10	*Pyramimonas* sp.	Chlorophyta	Pyramimonadophyceae	*Pyramimonas*	sp.
BE_AeN706-11	cf. *Pseudo-nitzschia*	Bacillariophyta	Bacillariophyceae	*Fragilariopsis*	sp.
BE_AeN706-12	*Prorocentrum* sp.	Miozoa	Dinophyceae	*Prorocentrum*	sp.
BE_AeN706-13	*Micromonas polaris*	Chlorophyta	Mamiellophyceae	*Micromonas*	*polaris*
BE_AeN706-15	*Micromonas polaris*	Chlorophyta	Mamiellophyceae	*Micromonas*	*polaris*
BE_AeN706-16	*Micromonas* sp.	Chlorophyta	Mamiellophyceae	*Micromonas*	sp.
BE_AeN706-19	*Micromonas* sp.	Chlorophyta	Mamiellophyceae	*Micromonas*	sp.
BE_AeN706-21	*Chaetoceros neogracilis*	Bacillariophyta	Mediophyceae	*Chaetoceros*	*neogracilis*
BE_AeN706-22	*Chaetoceros neogracilis*	Bacillariophyta	Mediophyceae	*Chaetoceros*	cf. *neogracilis*
BE_AeN706-23	*Thalassiosira gravida*	Bacillariophyta	Mediophyceae	*Thalassiosira*	*gravida*
BE_AeN706-24	*Chaetoceros neogracilis*	Bacillariophyta	Mediophyceae	*Chaetoceros*	cf. *neogracilis*
BE_AeN706-25	*Thalassiosira gravida*	Bacillariophyta	Mediophyceae	*Thalassiosira*	*gravida*
BE_AeN706-26	*Thalassiosira gravida*	Bacillariophyta	Mediophyceae	*Thalassiosira*	*gravida*
BE_AeN706-27	*Bacterosira* sp.	Bacillariophyta	Mediophyceae	*Bacterosira*	sp.
BE_AeN706-30	*Chaetoceros neogracilis*	Bacillariophyta	Mediophyceae	*Chaetoceros*	*neogracilis*
BE_AeN706-31	*Chaetoceros neogracilis*	Bacillariophyta	Mediophyceae	*Chaetoceros*	*neogracilis*
BE_AeN706-34	*Chaetoceros neogracilis*	Bacillariophyta	Mediophyceae	*Chaetoceros*	*neogracile*
BE_AeN70635	*Chaetoceros neogracilis*	Bacillariophyta	Mediophyceae	*Chaetoceros*	*neogracile*
K-AeN706-5	*Thalassiosira gravida*	Bacillariophyta	Mediophyceae	*Thalassiosira*	*gravida*
K-AeN706-6	*Thalassiosira gravida*	Bacillariophyta	Mediophyceae	*Thalassiosira*	*gravida*
K-AeN706-7	*Chaetoceros cinctus*	Bacillariophyta	Mediophyceae	*Chaetoceros*	*cinctus*
K-AeN706-11	*Chaetoceros neogracilis*	Bacillariophyta	Mediophyceae	*Chaetoceros*	*neogracilis*
K-AeN706-14	*Thalassiosira gravida*	Bacillariophyta	Mediophyceae	*Thalassiosira*	*gravida*
K-AeN706-15	*Attheya septentrionalis*	Bacillariophyta	Mediophyceae	*Attheya*	*septentrionalis*
K-AeN706-17	Pedinellales	Ochrophyta	Dictyochophyceae	Pedinellales	ND
K-AeN706-18	*Chaetoceros neogracilis*	Bacillariophyta	Mediophyceae	*Chaetoceros*	*neogracilis*
K-AeN706-19	cf. *Pseudo-nitzschia*	Bacillariophyta	Bacillariophyceae	*Fragilariopsis*	sp.
K-AeN706-22	*Nitzschia* sp.	Bacillariophyta	Bacillariophyceae	*Nitzschia*	sp.
K-AeN706-26	*Actinocyclus* sp.	Bacillariophyta	Coscinodiscophyceae	*Actinocyclus*	sp.
K-AeN706-28	*Chaetoceros neogracilis*	Bacillariophyta	Mediophyceae	*Chaetoceros*	*neogracilis*
K-AeN706-30	*Nitzschia* sp.	Bacillariophyta	Bacillariophyceae	*Nitzschia*	sp.
K-AeN706-31	*Amphora* sp.	Bacillariophyta	Bacillariophyceae	*Amphora*	sp.
K-AeN706-32	*Bacillaria* sp.	Bacillariophyta	Bacillariophyceae	*Bacillaria*	sp.
K-AeN706-33	*Nitzschia* sp.	Bacillariophyta	Bacillariophyceae	*Nitzschia*	sp.
K-AeN706-34	*Chaetoceros neogracilis*	Bacillariophyta	Mediophyceae	*Chaetoceros*	*neogracilis*
K-AeN706-35	*Fragilaria* sp.	Bacillariophyta	Bacillariophyceae	*Fragilaria*	sp.
K-AeN706-36	*Attheya longicornis*	Bacillariophyta	Mediophyceae	*Attheya*	*longicornis*
K-AeN706-37	*Attheya longicornis*	Bacillariophyta	Mediophyceae	*Attheya*	*longicornis*
K-AeN706-38	*Chaetoceros neogracilis*	Bacillariophyta	Mediophyceae	*Chaetoceros*	*neogracilis*
K-AeN706-45	*Pseudo-nitzschia* sp.	Bacillariophyta	Bacillariophyceae	*Pseudo-nitzschia*	sp.
K-AeN706-52	*Bacterosira* sp.	Bacillariophyta	Mediophyceae	*Bacterosira*	sp.
K-AeN706-56	*Thalassiosira* cf.*hispida*	Bacillariophyta	Mediophyceae	*Thalassiosira*	cf. *hispida*

3.3 Morphological analysis

A total of 46 algal strains were characterised, with 26 strains isolated from pelagic habitats and 20 strains from the sympagic habitat of the Northern Barents Sea and the Arctic Ocean (see section 2.5 “Morphological analysis”, Table 4). For morphological identification, the keys and species descriptions by Throndsen et al. (2003) were used.

In this project, morphological analysis played a supplementary role in identification. Its purpose was to validate the results obtained through the sequencing of the 18S rRNA and 28S rRNA genes. Nonetheless, all strains underwent examination under a light microscope and a selection were documented by a light micrograph shown in Figures 13, 14, 15 and 16 below.

The identification of microalgae by microscopy was found to be quite challenging. In many cases the microalgal species could not be identified. In certain instances, only the class could be determined. The lack of detailed morphology prevented the analysis and identification of microalgae at the species level by light microscopy.

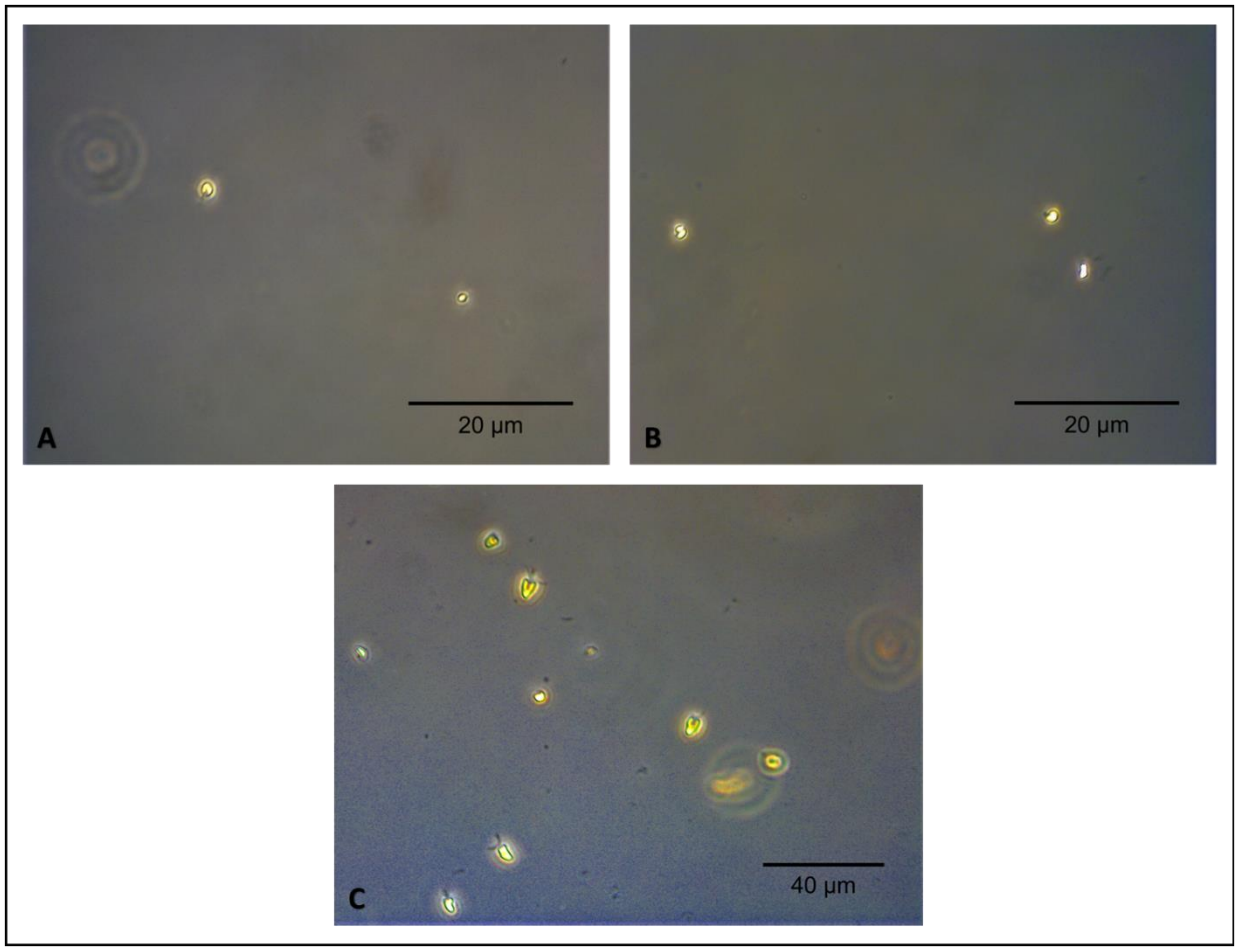


Figure 13. Light microscopy of the class Mamiellophyceae and Pyramimonadophyceae. Strain BE-5 Micromonas polaris (A), BE-19 Micromonas sp. *(B), BE-10 Pyramimonas* sp. *(C).*

Some micrographs illustrate the typical morphological characteristics of the studied strains. For instance, in Figure 13 (A and B), strains BE-5 and BE-19 exhibit the typical size of the

genus *Micromonas*. The coma-form of the cell is slightly recognisable in Figure 13 (B). Additionally, Figure 13 (C) shows that some cells of the strain BE-10 (*Pyramimonas* sp.) possess four flagella with an apical axis measuring 6 µm and a transapical axis of 4 µm.

Strain K-17 shown in Figure 14 (D, E) presented by a single cell. The cell has radial symmetry and has a typical flagellum which pulls the cell.

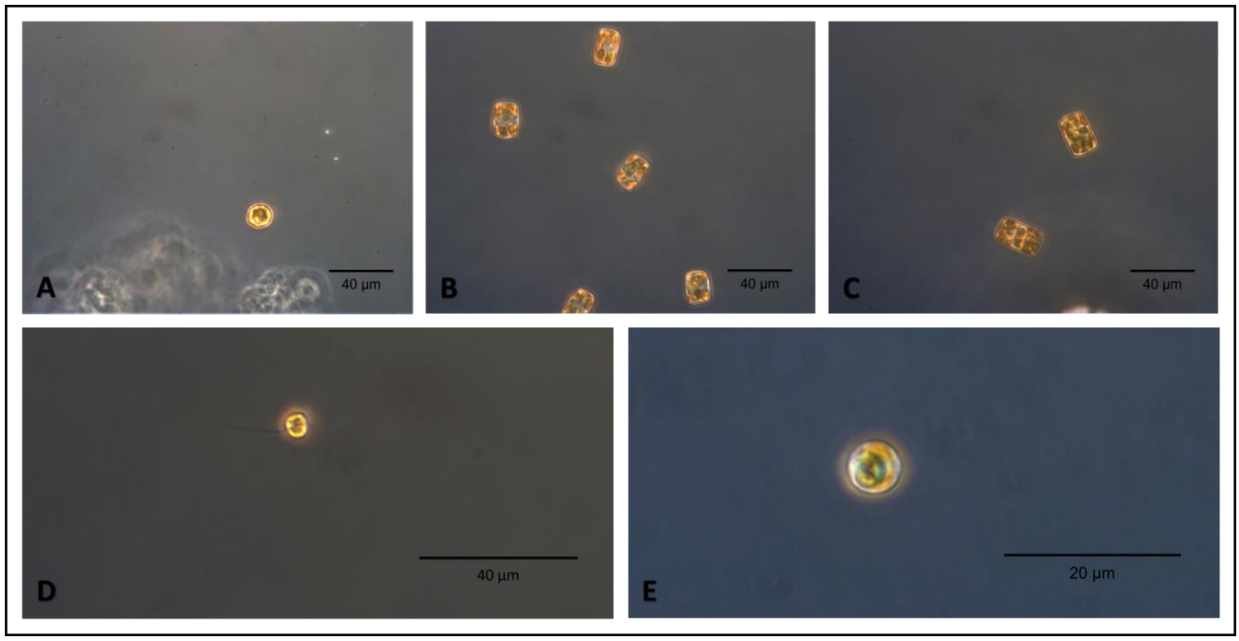


Figure 14. Light microscopy of the class Dinophyceae, Coscinodiscophyceae and Dictyochophyceae. Strain BE-12 Prorocentrum sp. *(A); Strain K-26 Actinocyclus* sp. *(B, C); strain K-17 Pedinellales (D, E).*

Strain BE-12 was identified as *Prorocentrum* sp. through molecular analysis, and the observed round cells with a diameter of 15 µm are consistent with the morphological description of *Prorocentrum* sp.

Micrographs of the strain K-26, *Actinocyclus* sp., were taken only from the girdle view (see Figure 14 (B and C) above), and the valve view images are missing. The valve morphology of *Actinocyclus* is important for identification purposes. The cells appear to be 37 µm in apical axis and 17 µm in transapical axis with many chloroplasts. This is consistent with the morphological descriptions of *Actinocyclus* sp.

However, microscopy revealed some noteworthy details as well. Specifically, strains BE-11 and K-19 (identical in the sequenced regions of the both genes) exhibited pronounced phenotypic differences (see Figure 15 (A, D) below). While all analytical methods used in this study identified strains K-22, K-30, and K-33 as Nitzschia sp., microscopy uncovered clear morphological distinctions among these strains (see Figure 15 (B, C, E, J and K) below). These distinctions encompass variations in size, cell grouping, and cell shape. Moreover, Figure 15 (B, C) below illustrates that strain K-22 displayed a rather atypical shape of the cells, with a bend present in the central part of the valve, suggesting the possibility of it being a new species.

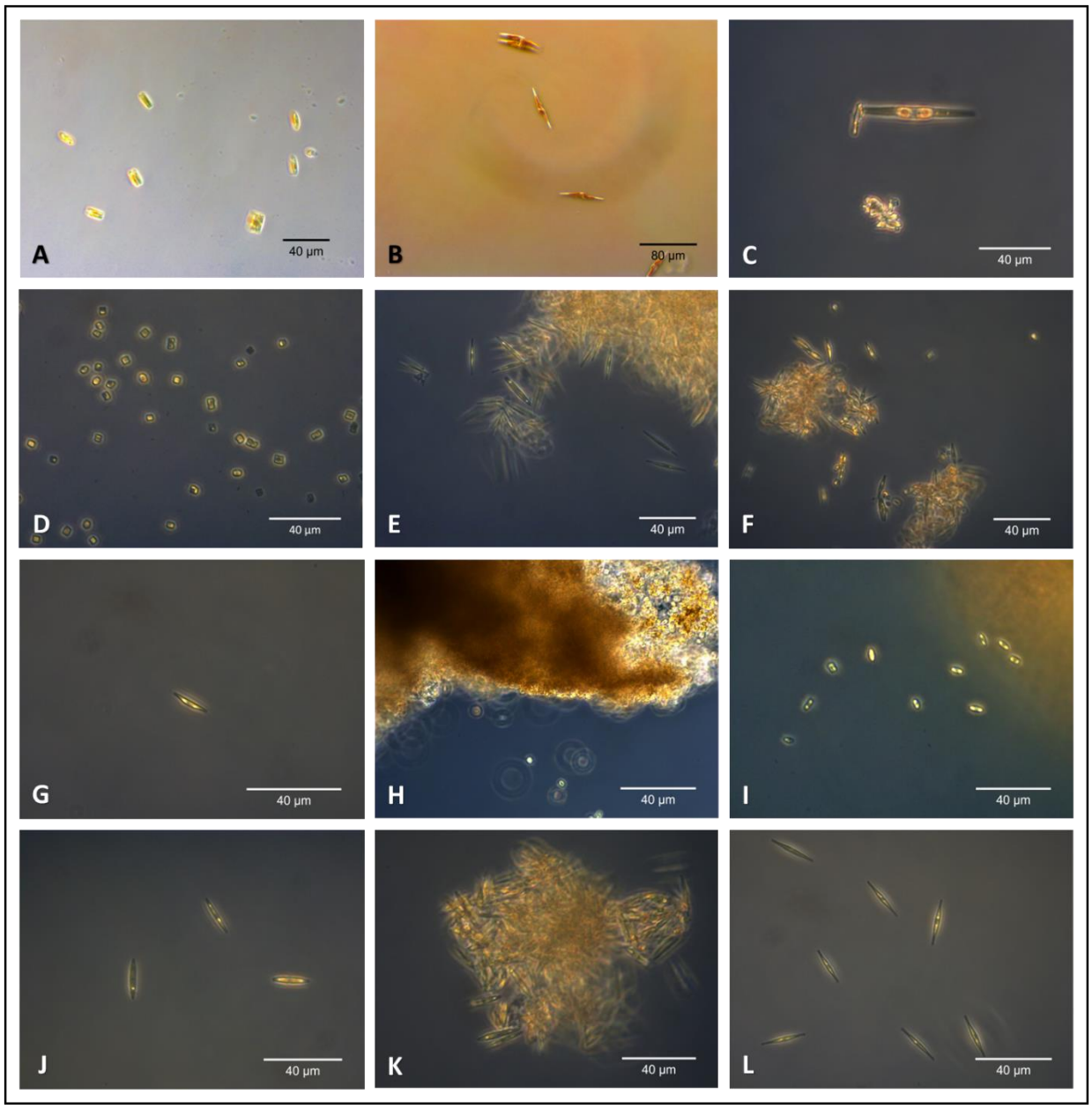


Figure 15. Light microscopy of the class Bacillariophyceae. Strain BE-11 Fragilariopsis sp. *(A); K-22 Nitzschia* sp. *(B Lugol-fixed cell, C living cell); K-19 Fragilariopsis* sp. *(D); K-30 Nitzschia* sp. *(E); K-31 Amphora* sp. *(F, G); K-32* cf. *Bacillaria* sp. *(H, I); K-33 Nitzschia* sp. *(J, K); K-35 Fragilaria* sp. *(L).*

The strain K-35 identified through molecular analysis as *Fragilaria* sp., exhibits morphological characteristics more similar to *Synedropsis* cf. *hyperborean*, with an apical axis measuring 23-25 µm and a transapical axis of 3 µm.

Light microscopy revealed significant inconsistencies with the molecular identification of strain K-32. The suggested identification as *Bacillaria* sp., based on phylogenetic analysis, cannot be supported by microscopy observations. The cells of strain K-32 appear to be only 6 µm in apical axis (instead of the expected size of at least 70 µm) and 2 µm in transapical axis (see Figure 15 (I) above). Furthermore, the cell formation (see figure 15 (H) above) does not resemble the structured colony characteristic of *Bacillaria* sp.

Strains BE-23, BE-25, *BE-26*, K-5, and K-6 are identified as *Thalassiosira gravida*. These strains exhibit variations in their morphological characteristics, including cell size, shape, and the distance between cells in chain formation (see Figure 16 (I, J, K, M, N and O) below). Strains K-15, K-36, and K-37 were identified as *Attheya longicornis* cf. *septentrionalis* (see Figure 16 (F, G and H) below). Light microscopy revealed that strain K-15 might be *Attheya septentrionalis* due to the length of the horns of the cells, which are approximately three times as long as the apical axis. Strains K-36 and K-37 are more likely to be identified as *Attheya longicornis*, based on the observation that some of their horns are ten times longer than their apical axis. This can be seen in Figure 16 (G and H), where arrows point to the elongated horns of the *Attheya* cells. This characteristic aligns with the morphological description of *Attheya* longicornis and provides supporting evidence for their identification.

The micrograph of the strain BE-27, *Bacterosira* sp. (see Figure 16 (L) below), shows single cells with a diameter of 17 µm. The tight chain formation, which is typical for *Bacterosira* sp., is not observable.

Strain K-7, identified as *Chaetoceros cinctus*, forms a curved chain (see Figure 16 (E) below). From the girdle view, the cells appear as elongated rectangles with an apical axis of 15 µm and a transapical axis of 6 µm. The setae are long (35 µm) and diverge at variable angles. These characteristics are consistent with the morphological descriptions of *Chaetoceros cinctus* (Gaonkar et al., 2017).

The micrographs of *Chaetoceros neogracilis* are presented in Figure 16 (A, B, C, and D). Solitary cells appear to have a rectangular form with an apical axis of 10 µm and a transapical axis of 5 µm. The micrographs are not as clear, so it is difficult to say anything particular about the cells' setae. In general, the morphological descriptions coincide with the previously given morphological characteristics of the *Chaetoceros neogracilis* (Balzano et al., 2017).

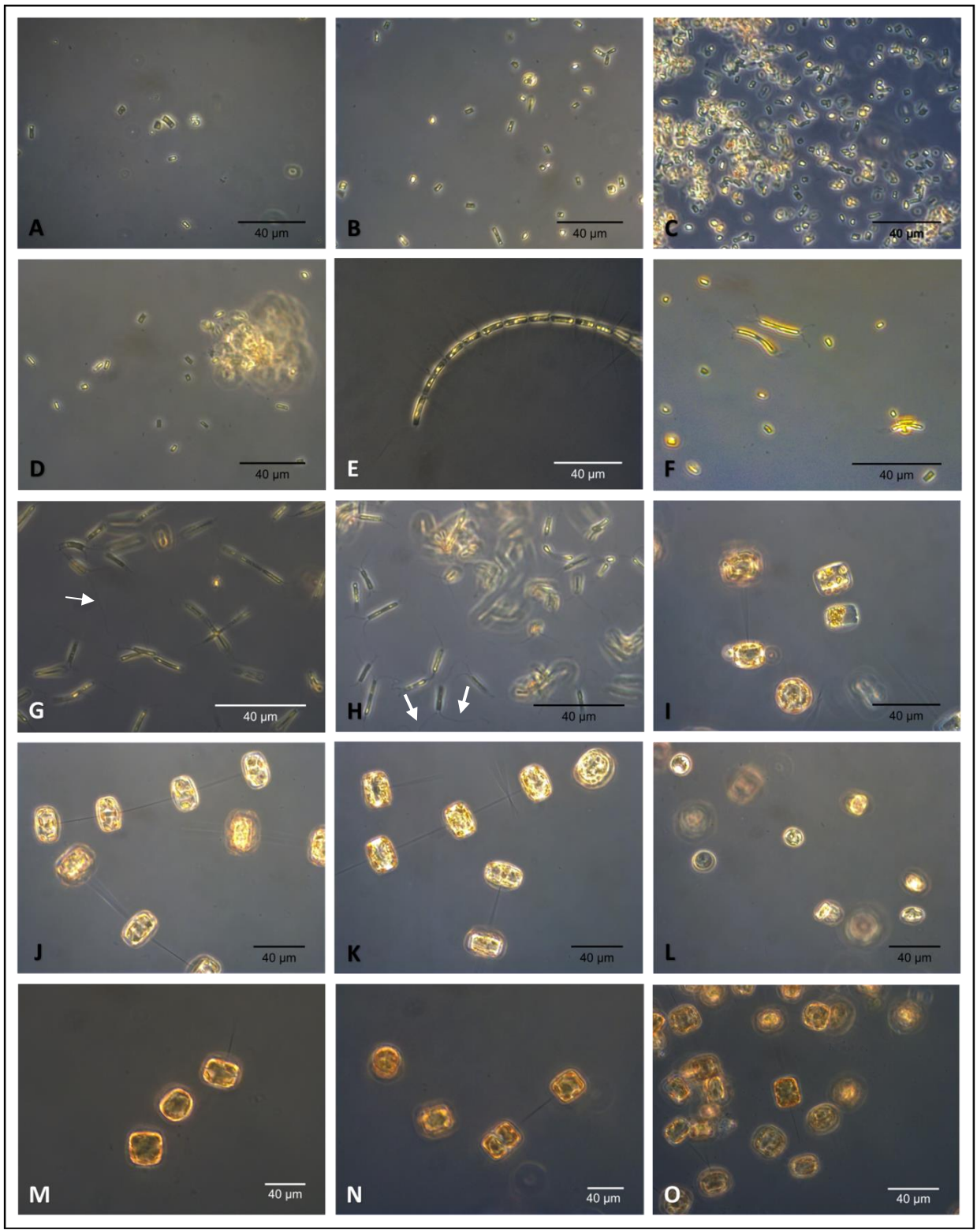


Figure 16. Light microscopy of the class Mediophyceae. Strains BE-21 (A), BE-30 (B), BE-34 (C), BE-38 Chaetoceros neogracilis; K-7 Chaetoceros cinctus (E); strains K-15 (F), K-36 (G), K-37 (H) Attheya longicornis cf. *septentrionalis; strain BE-27 Bacterosira* sp. *(L); strains BE-26 (I), BE-23 (J), BE-25 (K), K-5(M), K-6(N, O) Thalassiosira gravida.*

The results of the various analyses used in this study are not always consistent. The gathered results from the BLAST analysis, phylogenetic analysis, and light microscopy of the studied strains, along with the overall identification, are presented in Table 9 below.

Tabell 9. Total identification based on BLAST analysis, phylogeny, and microscopy of the studied strains. ND - no data.

Strain code	Identification (BLAST)	Identification (phylogeny)	Identification (morphology)	Total
BE_AeN706-3	*Pyramimonas* sp.	*Pyramimonas* sp.	*Pyramimonas* sp.	*Pyramimonas* sp.
BE_AeN706-5	*Micromonas polaris*	*Micromonas polaris*	*Micromonas* sp.	*Micromonas polaris*
BE_AeN706-6	*Pyramimonas* sp.	*Pyramimonas* sp.	*Pyramimonas* sp.	*Pyramimonas* sp.
BE_AeN706-8	*Pyramimonas* sp.	*Pyramimonas* sp.	*Pyramimonas* sp.	*Pyramimonas* sp.
BE_AeN706-9	*Pyramimonas* sp.	*Pyramimonas* sp.	*Pyramimonas* sp.	*Pyramimonas* sp.
BE_AeN706-10	*Pyramimonas* sp.	*Pyramimonas* sp.	*Pyramimonas* sp.	*Pyramimonas* sp.
BE_AeN706-11	*Fragilariopsis* sp.	cf. *Pseudo-nitzschia*	*Fragilariopsis* sp.	*Fragilariopsis* sp.
BE_AeN706-12	*Prorocentrum* sp.	*Prorocentrum* sp.	*Prorocentrum* sp.	*Prorocentrum* sp.
BE_AeN706-13	*Micromonas polaris*	*Micromonas polaris*	*Micromonas* sp.	*Micromonas polaris*
BE_AeN706-15	*Micromonas polaris*	*Micromonas polaris*	*Micromonas* sp.	*Micromonas polaris*
BE_AeN706-16	*Micromonas* sp.	*Micromonas* sp.	*Micromonas* sp.	*Micromonas* sp.
BE_AeN706-19	*Micromonas* sp.	*Micromonas* sp.	*Micromonas* sp.	*Micromonas* sp.
BE_AeN706-21	*Chaetoceros* cf. *neogracilis*	*Chaetoceros neogracilis*	*Pennat diatom*	*Chaetoceros neogracilis*
BE_AeN706-22	*Chaetoceros* cf. *neogracilis*	*Chaetoceros neogracilis*	*Pennat diatom*	*Chaetoceros neogracilis*
BE_AeN706-23	*Thalassiosira gravida*	*Thalassiosira gravida*	*Thalassiosira*	*Thalassiosira gravida*
BE_AeN706-24	*Chaetoceros* cf. *neogracilis*	*Chaetoceros neogracilis*	*Pennat diatom*	*Chaetoceros neogracilis*
BE_AeN706-25	*Thalassiosira gravida*	*Thalassiosira gravida*	*Thalassiosira*	*Thalassiosira gravida*
BE_AeN706-26	*Thalassiosira gravida*	*Thalassiosira gravida*	*Thalassiosira*	*Thalassiosira gravida*
BE_AeN706-27	*Bacterosira* sp.	*Bacterosira* sp.	*Bacterosira* sp.	*Bacterosira* sp.
BE_AeN706-30	*Chaetoceros* cf. *neogracilis*	*Chaetoceros neogracilis*	*Pennat diatom*	*Chaetoceros neogracilis*
BE_AeN706-31	*Chaetoceros* cf. *neogracilis*	*Chaetoceros neogracilis*	*Pennat diatom*	*Chaetoceros neogracilis*
BE_AeN706-34	*Chaetoceros* cf. *neogracilis*	*Chaetoceros neogracilis*	*Pennat diatom*	*Chaetoceros neogracilis*
BE_AeN706-35	*Chaetoceros* cf. *neogracilis*	*Chaetoceros neogracilis*	*Pennat diatom*	*Chaetoceros neogracilis*
K-AeN706-5	*Thalassiosira gravida*	*Thalassiosira gravida*	*Thalassiosira* sp.	*Thalassiosira gravida*
K-AeN706-6	*Thalassiosira gravida*	*Thalassiosira gravida*	*Thalassiosira* sp.	*Thalassiosira gravida*
K-AeN706-7	*Chaetoceros* cf. *cinctus*	*Chaetoceros cinctus*	*Chaetoceros* sp.	*Chaetoceros cinctus*
K-AeN706-11	*Chaetoceros* cf. *neogracilis*	*Chaetoceros neogracilis*	*Pennat diatom*	*Chaetoceros neogracilis*
K-AeN706-15	*Attheya septentrionalis*	*Attheya septentrionalis*	*Attheya septentrionalis*	*Attheya septentrionalis*
K-AeN706-17	Pedinellales	Pedinellales	Pedinellales	Pedinellales
K-AeN706-18	*Chaetoceros* cf. *neogracilis*	*Chaetoceros neogracilis*	*Pennat diatom*	*Chaetoceros neogracilis*
K-AeN706-19	*Fragilariopsis* sp.	cf. *Pseudo-nitzschia*	*Fragilariopsis* sp.	*Fragilariopsis* sp.
K-AeN706-22	*Nitzschia* sp.	*Nitzschia* sp.	*Nitzschia* sp.	*Nitzschia* sp.
K-AeN706-26	*Actinocyclus* sp.	*Actinocyclus* sp.	*Shionodiscus* sp.	*Actinocyclus* sp.
K-AeN706-28	*Chaetoceros* cf. *neogracilis*	*Chaetoceros neogracilis*	*Pennat diatom*	*Chaetoceros neogracilis*
K-AeN706-30	*Nitzschia* sp.	*Nitzschia* sp.	*Nitzschia* sp.	*Nitzschia* sp.
K-AeN706-31	*Amphora* sp.	*Amphora* sp.	*Amphora* sp.	*Amphora* sp.
K-AeN706-32	*Uncultured eukaryote*	*Bacillaria* sp.	*Pennat diatom*	*Bacillaria* sp.
K-AeN706-33	*Nitzschia* sp.	*Nitzschia* sp.	*Nitzschia* sp.	*Nitzschia* sp.
K-AeN706-34	*Chaetoceros* cf. *neogracilis*	*Chaetoceros neogracilis*	*Pennat diatom*	*Chaetoceros neogracilis*
K-AeN706-35	*Fragilaria* cf. *Synedropsis*	*Fragilaria* sp.	*Synedropsis* sp.	*Synedropsis* sp.
K-AeN706-36	*Attheya longicornis* cf. *septentrionalis*	*Attheya longicornis*	*Attheya longicornis*	*Attheya longicornis*
K-AeN706-37	*Attheya longicornis* cf. *septentrionalis*	*Attheya longicornis*	*Attheya longicornis*	*Attheya longicornis*
K-AeN706-38	*Chaetoceros* cf. *neogracilis*	*Chaetoceros neogracilis*	*Pennat diatom*	*Chaetoceros neogracilis*
K-AeN706-45	*Pseudo-nitzschia* sp.	*Pseudo-nitzschia* sp.	*Pennat diatom*	*Pseudo-nitzschia* sp.
K-AeN706-52	*Bacterosira* sp.	*Bacterosira* sp.	*Pennat diatom*	*Bacterosira* sp.
K-AeN706-56	*Thalassiosira* cf.*hispida*	*Thalassiosira* cf.*hispida*	*Thalassiorira* sp.	*Thalassiosira* cf.*hispida*

3.4 Biogeographic analysis

For the biogeographic analysis of the studied strains, the 18S rRNA gene sequences were used. Consequently, distribution maps were not presented for all strains. The provided maps illustrate the geographic distribution of identical ASVs (Amplicon Sequence Variants) identified in the metaPR2 metabarcoding datasets (Vaulot et al. 2022).

Distribution patterns

After producing and examining distribution maps of the studied strains, the following biogeographic distribution types were identified (see Table 10): Arctic (Figure 17, Arctic-temperate (Figures 18 and 19 (A) below), cosmopolitan (Figure 19 (B and C) below) and polar (Figure 19 (D) below).

Table 10. Description of biogeographic distribution types as defined in this study.

Biogeographic distribution type	Description
Arctic	Genotype has a biogeographic distribution mostly restricted to the Arctic circle (above 66ºN; Figure 17).
Arctic-temperate	Genotype is present in the Arctic and at temperate latitudes (30ºN - 66ºN and 30ºS - 60ºS; Figures 18 and 19(A)).
Cosmopolitan	Genotype is present in the Arctic and has a cosmopolitan distribution (Figure 19(B, C)).
Polar	Genotype has a biogeographic distribution mostly restricted to the Arctic and Antarctic circle (above 66ºN and 66ºS; Figure 19 (D)).

The template for Table 10. is taken from the article “Diversity and biogeography of planktonic diatoms in Svalbard fjords: The role of dispersal and Arctic endemism in phytoplankton community structuring.” by Supraha et al. (2022).

Genotypes with Arctic biogeographic distribution.

An Arctic distribution, shown in Figure 17, was found for *Micromonas polaris* and *Micromonas* sp. (strains BE-5, BE-13, BE-15, BE-19); *Chaetoceros neogracilis* (BE-21, K - 28); *Chaetoceros cinctus* (K-7); *Bacterosira* sp. (BE-27); *Synedropsis hyperboreoides* (K-35); *Attheya septentrionalis* (K-15); *Thalassiosira* sp. (K-56); *Fragilariopsis kerguelensis* (K-19).

Genotypes with Arctic-temperate biogeographic distribution

According to the metabarcoding metaPR2 database there are following genotypes with Arctic-temperate biogeographic distribution (Figure 18 and 19, A below): *Chaetoceros* cf. *neogracilis* (BE-22, BE-24, K-18, K-28, K-34, K-38); *Thalassiosira gravida* (K-6, BE-26); *Bacillaria* cf. *paxillifer* (K-32); *Attheya longicornis* (K-36, K-37); Pedinellales (K-17); *Pyramimonas* sp. (BE-3, BE-6, BE-8, BE-10); *Actinocyclus curvatulus* (K-26); *Nitzschia thermalis* (K-30); *Nitzschia aequorea (*K-33); *Amphora* sp. (K-31).

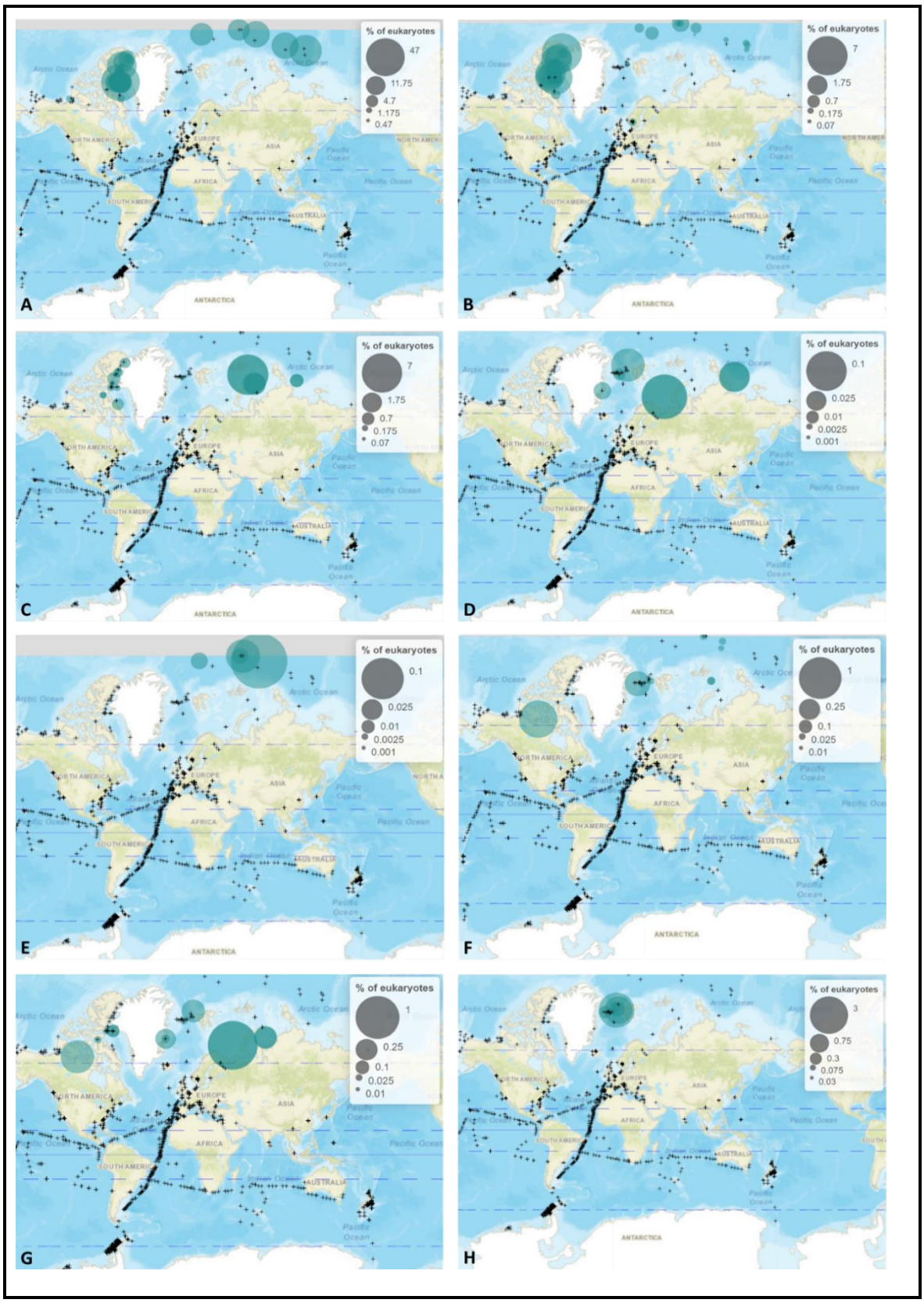


Figure 17. Genotypes with Arctic biogeographic distribution: A- Micromonas polaris and Micromonas sp. *(BE-5, BE-13, BE-15, BE-19); B – Chaetoceros neogracilis (BE-21, K -28); C –Chaetoceros cinctus (K-7); D – Bacterosira* sp. *(BE-27); E – Fragilaria* sp. *(K-35); F – Attheya septentrionalis (K-15); G – Thalassiosira* cf.*hispida (K-56); H – Fragilariopsis* sp. *(K-19). Spherical symbols indicate the relative abundance of the genotype compared to the total eukaryotic reads within each sample of the compiled metabarcoding dataset. A cross symbol (+) is used to indicate samples where the genotypes were not detected.*

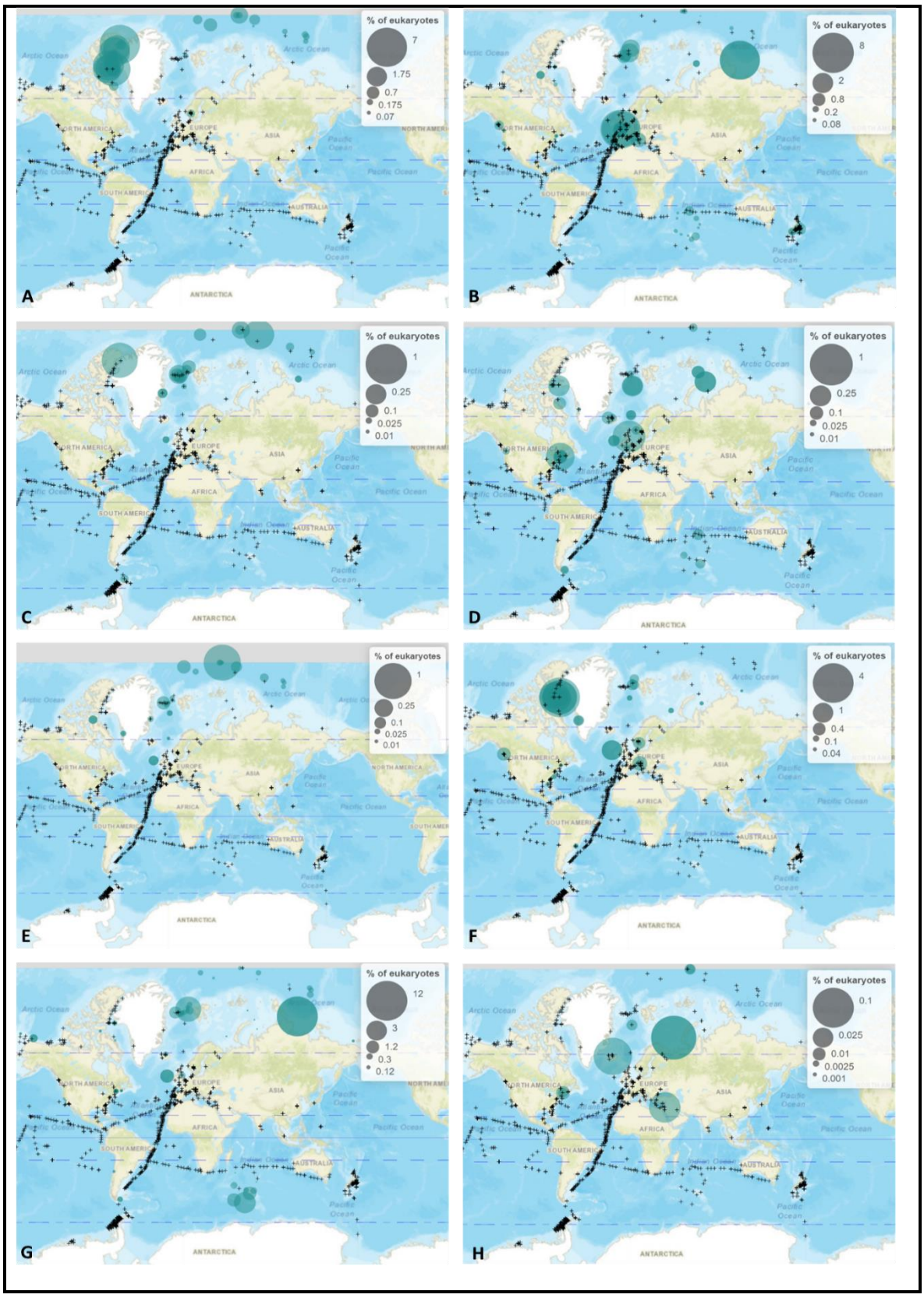


Figure 18. Genotypes with Arctic-temperate biogeographic distribution. A – Chaetoceros cf. *neogracilis (BE-22, BE-24, K-18, K-28, K-34, K-38); B – Thalassiosira gravida (K-6, BE-26); C – Bacillaria* cf. *paxillifer (K-32); D – Attheya longicornis (K-36, K-37); E – Pedinellales (K-17); F – Pyramimonas* sp. *(BE-3, BE-6, BE-8, BE-10); G – Actinocyclus* sp. *(K-26); Nitzschia* sp. *(K-30). Spherical symbols indicate the relative abundance of the genotype compared to the total eukaryotic reads within each sample of the compiled metabarcoding dataset. A cross symbol (+) is used to indicate samples where the genotypes were not detected.*

Genotypes with a cosmopolitan and polar biogeographic distribution

Among the studied strains there are two genotypes with a cosmopolitan biogeographic distribution (Figure 19 (B, C) below): *Pseudo-nitzschia seriata* (K-45); *Nitzschia* sp. (K-22)). There is one genotype with polar biogeographic distribution (Figure 19, D below): *Prorocentrum* sp. (BE-12).

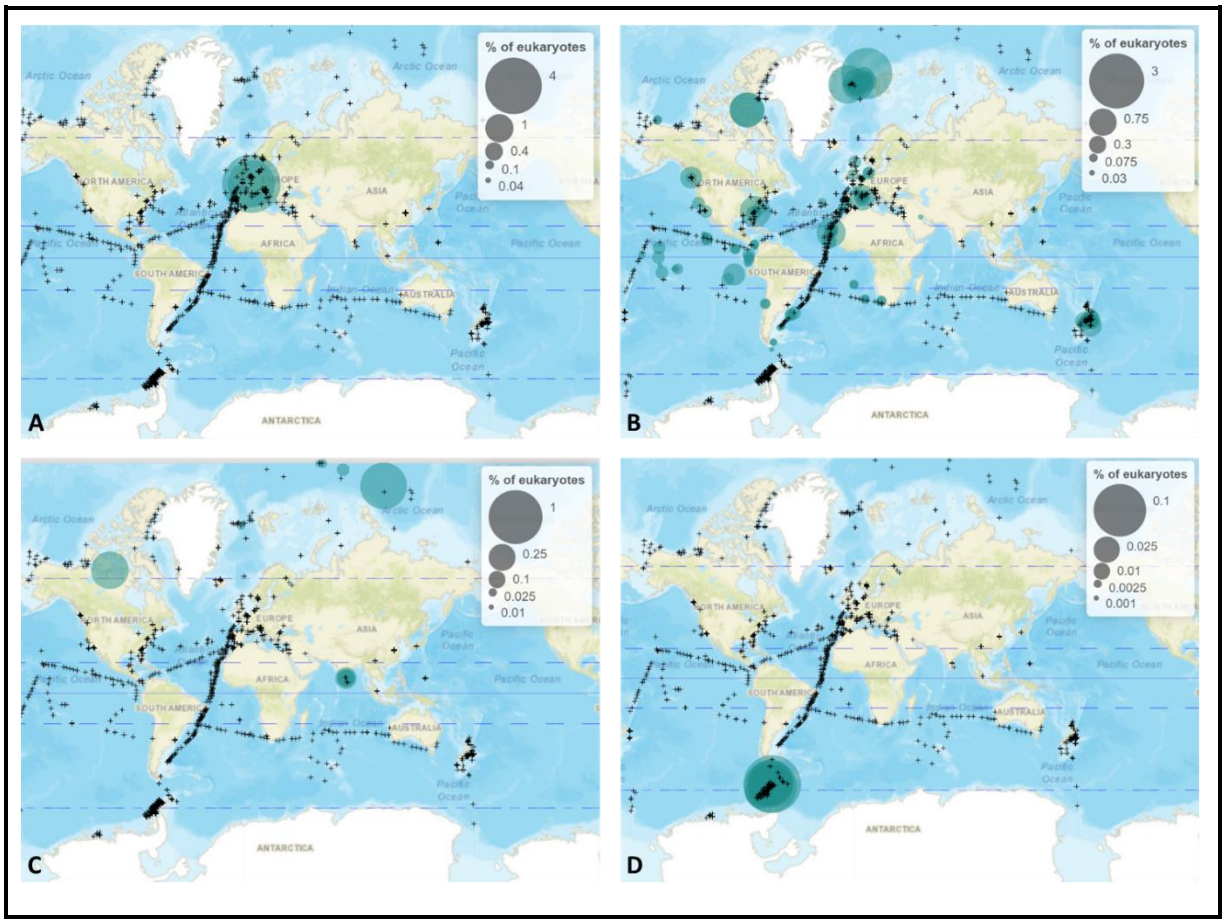


Figure 19. Genotypes with an Arctic-temperate biogeographic distribution (A – Amphora sp. *(K-31)). Genotypes with a cosmopolitan biogeographic distribution (B - Pseudo-nitzschia* sp. *(K-45), C – Nitzschia* sp. *(K-22)). Genotypes with a Polar distribution (D – Prorocentrum* sp. *(BE-12)). Spherical symbols indicate the relative abundance of the genotype compared to the total eukaryotic reads within each sample of the compiled metabarcoding dataset. A cross symbol (+) is used to indicate samples where the genotypes were not detected.*

The content of Table 11 provides information about the taxonomy of the strains, including their distribution type, the name of the isolation station, the type of habitat, and the sampled depth. However, no clear correlation can be observed between the distribution type of a particular genotype, the sampled station, and the habitat of the genotype.

The following genotypes potentially represent endemic Arctic species: BE-5, BE-13, BE-15, B-19 (*Micromonas polaris*); K-7 (*Chaetoceros cinctus*); BE-25 (*Bacterosira* sp.); K-35 (*Synedropsis* sp.); K-15 (*Attheya septentrionalis*); K-56 (*Thalassiosira* sp.); BE-11 and K-19 (*Fragilariopsis* sp.); BE-21 and K-28 (*Chaetoceros neogracilis*).

Table 11. Gathered information about the taxonomy of the studied strains, their distribution type, name of the isolation station and a type of the habitat. ND - no data.

Strain code	Taxon	Distribution type	Station name	Type of habitat, sampled depth
BE_AeN706-3	*Pyramimonas* sp.	Arctic-temperate	P2	pelagic, 50 m
BE_AeN706-5	*Micromonas polaris*	Arctic	P2	pelagic, 50 m
BE_AeN706-6	*Pyramimonas* sp.	Arctic-temperate	P2	pelagic, 10 m
BE_AeN706-8	*Pyramimonas* sp.	Arctic-temperate	P2	pelagic, 10 m
BE_AeN706-9	*Pyramimonas* sp.	ND	P2	pelagic, 10 m
BE_AeN706-10	*Pyramimonas* sp.	Arctic-temperate	P2	pelagic, 10 m
BE_AeN706-11	*Fragilariopsis* sp.	Arctic	P2	pelagic, 50 m
BE_AeN706-12	*Prorocentrum* sp.	Polar	P2	pelagic, 50 m
BE_AeN706-13	*Micromonas polaris*	Arctic	P6-ICE	melt pond, 0 m
BE_AeN706-15	*Micromonas polaris*	Arctic	P6-ICE	melt pond, 0 m
BE_AeN706-16	*Micromonas* sp.	ND	P6-ICE	under ice water, 0.5 m
BE_AeN706-19	*Micromonas* sp.	Arctic	P6-ICE	under ice water, 0.5 m
BE_AeN706-21	*Chaetoceros neogracilis*	Arctic	P6-ICE	under ice water, 0.5 m
BE_AeN706-22	*Chaetoceros neogracilis*	Arctic-temperate	P6-ICE	under ice water, 0.5 m
BE_AeN706-23	*Thalassiosira gravida*	ND	P6-ICE	under ice water, 0.5 m
BE_AeN706-24	*Chaetoceros neogracilis*	Arctic-temperate	P6-ICE	melt pond, 0 m
BE_AeN706-25	*Thalassiosira gravida*	ND	P6-ICE	melt pond, 0 m
BE_AeN706-26	*Thalassiosira gravida*	Arctic-temperate	P6-ICE	melt pond, 0 m
BE_AeN706-27	*Bacterosira* sp.	Arctic	P1	pelagic, 45 m
BE_AeN706-30	*Chaetoceros neogracilis*	Arctic-temperate	P1	pelagic, 45 m
BE_AeN706-31	*Chaetoceros neogracilis*	Arctic-temperate	P4	pelagic, 40 m
BE_AeN706-34	*Chaetoceros neogracilis*	Arctic-temperate	P4	pelagic, 40 m
BE_AeN706-35	*Chaetoceros neogracilis*	Arctic-temperate	P3	pelagic, 75 m
K-AeN706-5	*Thalassiosira gravida*	ND	P3	pelagic, 75 m
K-AeN706-6	*Thalassiosira gravida*	Arctic-temperate	P3	pelagic, 75 m
K-AeN706-7	*Chaetoceros cinctus*	Arctic	P3	pelagic, 10 m
K-AeN706-11	*Chaetoceros neogracilis*	Arctic-temperate	P3	pelagic, 0-100m
K-AeN706-15	*Attheya septentrionalis*	Arctic	P6-ICE	under ice water, 0.5 m
K-AeN706-17	Pedinellales	Arctic-temperate	P7-ICE	under ice, 0-5 m
K-AeN706-18	*Chaetoceros neogracilis*	Arctic-temperate	P6	pelagic, 15 m
K-AeN706-19	*Fragilariopsis* sp.	*Arctic*	P6	pelagic, 15 m
K-AeN706-22	*Nitzschia* sp.	Cosmopolitan	P7-ICE	ice core, 0-10 cm
K-AeN706-26	*Actinocyclus* sp.	Arctic-temperate	P7	pelagic, 0-100 m
K-AeN706-28	*Chaetoceros neogracilis*	Arctic/Arctic-temperate	P7-ICE	ice core, 0-10 cm
K-AeN706-30	*Nitzschia* sp.	Arctic-temperate	P7-ICE	ice core, 0-10 cm
K-AeN706-31	*Amphora* sp.	Arctic-temperate	P7-ICE	ice core, 0-10 cm
K-AeN706-32	*Bacillaria* sp.	Arctic-temperate	P7-ICE	ice core, 0-10 cm
K-AeN706-33	*Nitzschia* sp.	Arctic-temperate	P7-ICE	ice core, 0-10 cm
K-AeN706-34	*Chaetoceros neogracilis*	Arctic-temperate	P7-ICE	ice core, 0-10 cm
K-AeN706-35	*Synedropsis* sp.	Arctic	P7-ICE	ice core, 0-10 cm
K-AeN706-36	*Attheya longicornis*	Arctic-temperate	P6	pelagic, 15 m
K-AeN706-37	*Attheya longicornis*	Arctic-temperate	P6	pelagic, 15 m
K-AeN706-38	*Chaetoceros neogracilis*	Arctic-temperate	P6	pelagic, 15 m
K-AeN706-45	*Pseudo-nitzschia* sp.	Cosmopolitan	P5	pelagic, 20 m
K-AeN706-52	*Bacterosira* sp.	ND	P6	pelagic, 15 m
K-AeN706-56	*Thalassiosira* cf.*hispida*	Arctic	P6	pelagic, 0-50 m

4 Discussion

In this section of the thesis, the obtained results will be discussed, with a separate analysis conducted for each analysis method employed, namely BLAST analysis, phylogenetic analysis, light microscopy of arctic phytoplankton, and ice algae. The goal of such analysis is to gain insight into the genetic relatedness, taxonomy, and morphological characteristics of the studied algae. An overall assessment of the analysis methods will be provided, with evaluation of their strengths and limitations. Subsequently, the thesis questions will be addressed to determine whether the initial hypothesis is supported or contradicted by the findings. Finally, an assessment of the obtained results will be presented, with the key findings summarised and followed by a concluding statement.

4.1 BLAST Analysis

Several significant findings were revealed in our study through the BLAST analysis. Its main advantage lies in its ability to identify sequence similarities and provide insights into the genetic relatedness of organisms. Potential matches were identified, and evolutionary relationships were inferred by comparing our sequences with existing databases. Potentially novel species were uncovered, and their presence in the arctic phytoplankton and ice algal samples was assessed through this method. However, it is important to note that BLAST analysis has some limitations. The accuracy of the results heavily depends on the completeness and quality of the reference databases. Additionally, challenges may be encountered in cases where genetic diversity is high or when dealing with poorly characterised organisms.

In the analysis of the obtained BLAST results, it was observed that a majority of genotypes lacked sequences that matched with 100% pairwise identities and query coverage. Among the 18S rRNA gene sequences, only the B-19 *Micromonas pusilla* and BE-26 *Thalassiosira gravida* strains showed 100% similarity with their best match sequences in the NCBI database. Among the 28S rRNA gene sequences, BE-21, BE-30, BE-34, BE-35, and K-15 had 100% matches in the GenBank. However, certain genotypes had no satisfactory matching sequences in database. Namely, both strain K-26 and strain K-22 exhibited remarkably distant analogues in the database with a pairwise identity value of 88% and a query covered value of 26% for strain K-26 and with pairwise identity value of 96% and a query covered value of 74% for strain K-22. The absence of reference sequences for these strains in GenBank may indicate potentially novel or undiscovered species in the scientific community.

The use of two markers, 18S rRNA and 28S rRNA genes, proved beneficial as they provided complementary data. In some cases, the identification was confirmed. For example, the BLAST search of the 18S and 28S rRNA sequences of the strains BE-5 and BE-15 revealed the same species, *Micromonas polaris*. This supports the accuracy and reliability of using multiple markers for species identification. In other cases, one of the markers offered more

specific identification of the strain. For example, the BLAST search of the 18S rRNA sequences of the strains BE-11 resulted in the identification of *Fragilariopsis kerguelensis*, while the BLAST search of the 28S rRNA sequence only identified it as *Fragilariopsis* sp. A possible explanation for such discrepancy could be the absence of 28S rRNA reference sequences of *Fragilariopsis kerguelensis* in the NCBI database. Additionally, the partial sequence of the 28S rRNA gene is shorter and can produce poorer phylogenetic trees then 18S rRNA phylogenetic trees. In certain instances, the results of the BLAST search for the two gene sequences differed in their identification at the species or genus level. For example, the BLAST search of the 18S rRNA gene sequence of the BE-6 strain resulted in the identification of *Pyramimonas parkeae*, while the BLAST search of the 28S rRNA gene sequence identified it as *Pyramimonas tetrarhynchus*. This discrepancy suggests the possibility of genetic variation or sequence divergence between these closely related species. It highlights the importance of considering multiple markers and conducting further investigation to determine the accurate taxonomic classification of the strain. Again the reason could be in lack of reference sequences in GenBank. This means that the unidentified microalgal species may not have closely related sequences available for comparison and identification.

It is worth mentioning that both the 18S and 28S rRNA genes identified numerous genotypes as uncultured marine eukaryotes, which did not provide precise identification of the strains. This occurrence indicates the presence of previously uncharacterised or undiscovered organisms in the studied samples. The identification of these genotypes as uncultured marine eukaryotes underscores the need for algal cultures for further research to unravel the diversity and taxonomy of these organisms, such as this study.

4.2 Phylogenetic Analysis

The phylogenetic analysis is common in determining the taxonomy of the studied strains and understanding the evolutionary relationships among the arctic phytoplankton and ice algal species (Jeyapandi et al., 2018). This analysis provided a valuable opportunity to visually compare the level of similarity between algal strains and their reference sequences. By constructing phylogenetic trees based on 18S and 28S rRNA genetic markers, the phylogenetic positions and evolutionary histories of the studied strains were revealed with higher accuracy. This method offered a visual representation of the relatedness between organisms and facilitated the identification of common ancestors. However, it is important to acknowledge potential limitations, such as the selection of genetic markers and the accuracy of the alignment algorithms utilised. Moreover, the availability of representative sequences in public databases can impact the resolution and reliability of the phylogenetic analysis.

Phylogenetic analysis uncovered discrepancies in the placement of certain strains. Specifically, in the 18S rRNA phylogenetic tree, strain BE-11 and K19 were observed to be placed separately from other reference sequences, with the *Pseudo-nitzschia* clade identified as their sister group. However, in the 28S rRNA phylogenetic tree, these strains were grouped

together within the clade of *Pseudo-nitzschia*. In the case of these strains, the combined use of BLAST search and microscopy proved to be more valuable in accurately identifying them as *Fragilariopsis* sp. These inconsistencies highlight the complexity of the evolutionary relationships among the studied strains and may indicate potential taxonomic challenges or genetic divergence within these lineages.

4.3 Light Microscopy of Arctic Phytoplankton and Ice Algae

Essential insights into the morphological characteristics and abundance of arctic phytoplankton and ice algae were provided through the application of light microscopy. Direct observation of the specimens enabled the identification and quantification of different taxa through this traditional method. Accessibility, ease of use, and cost-effectiveness are advantages of light microscopy. Valuable information about cell size, shape, pigmentation, and other morphological features was obtained. It is important to acknowledge that light microscopy has limitations, particularly in identifying cryptic or morphologically similar species. Accurate identification of some taxa may require more advanced techniques, such as electron microscopy or molecular approaches. Despite the secondary role of microscopy in this project, this traditional method of analysis has provided valuable insights. For instance, it revealed some notable observations, such as the atypical cell form of strain K-22 identified as *Nitzschia* sp. Additionally, there is a clear discrepancy between the morphological characteristics of *Bacillaria* sp. and the strain K-32. In particular, light microscopy has been instrumental in distinguishing between two species, namely *Attheya longicornis* and *Attheya septentrionalis*, represented by strains K-36, K-37, and K-15.

A significant role was played by each analysis method employed in this study, enhancing our understanding of the arctic phytoplankton and ice algal communities. The BLAST analysis identified genetic similarities and led to the discovery of novel species. The phylogenetic analysis provided insights into the evolutionary relationships among the studied strains. Light microscopy allowed for the direct examination of morphological characteristics, colony formation, and even motility patterns in certain cases. By integrating these methods, a reliable identification of the arctic phytoplankton and ice algal representatives was achieved. However, it is crucial to recognise that no single method is without limitations. Hence, a combination of different approaches is essential for a more accurate and holistic understanding of studied strains.

4.4 Taxonomy and phylogenetic placement

The primary objective was to ascertain the taxonomy and phylogenetic placement of the algal species isolated into culture during the Nansen Legacy cruise in August 2019.A total of 46 strains were characterised during the present study. The following taxa were revealed: 10 strains belonging to the phylum Chlorophyta (green algae), represented by the genera

Pyramimonas and *Micromonas*; one strain belonging to the phylum Miozoa (dinoflagellates), genus *Prorocentrum*; one strain belonging to the phylum Ochrophyta (dictyochophytes), order Pedinellales; and 34 strains belonging to the phylum Bacillariophyta (diatoms). The phylum Bacillariophyta was represented by the following genera: *Chaetoceros, Thalassiosira, Bacterosira, Nitzschia, Pseudo-nitzschia, Actinocyclus, Amphora, Bacillaria, Fragilaria, Fragilariopsis,* and *Atheya*. The analysis of the studied representatives of the Arctic phytoplankton and ice algae reveals a strong dominance of diatoms among the taxa. Specifically, *Chaetoceros neogracilis* was represented by 12 strains, accounting for one-third of all the studied strains. *Thalassiosira gravida* had six representatives among the characterised strains. These findings align with the results obtained in a similar study on Arctic phytoplankton diversity in the North Water (Lovejoy et al., 2002), in the Beaufort Sea (Balzano et al., 2017) and in the Baffin Bay (Ribeiro et al., 2020). The genera *Micromonas* and *Pyramimonas* were well-described in research on phytoplankton diversity in the Northeast Pacific and Arctic Oceans (Balzano et al., 2012a). The presence of a limited range of dinoflagellate diversity was observed, suggesting a potential requirement for alternative methods of isolation and cultivation to capture their full diversity in the Arctic (Okolodkov et al., 1996; Van de Schootbrugge et al., 2020).

4.5 Station, habitat, and water depth of the isolated genotypes

The second aspect under investigation aimed to identify the station, habitat, and water depth associated with the different isolated genotypes encountered during this cruise. The studied genotypes were isolated from various habitat types, including melt ponds, ice cores, under ice water, and pelagic waters. It is worth noting that identical genotypes were found in different habitat types. Specifically, *Micromonas* sp. strains BE-15 and BE-19 were present in both melt ponds and under ice water. Similarly, *Attheya septentrionalis* and *Attheya longicornis* strains K-15, K-36, and K-37 were isolated from both pelagic waters and under ice water. Additionally, *Chaetoceros neogracilis* strains BE-21, K-28, and K-38 were found in under ice water, ice cores, and the pelagic habitat. Establishing a definitive correlation between genotypes and their respective habitats poses a significant challenge. However, when considering algal strains inhabiting solid ice environments, diatoms were found to dominate, further affirming their contribution as key players in sympagic assemblages (Mundy et al., 2011).

4.6 Species novel to science

The third question is whether any of the isolated species are novel discoveries in the field of science. One of the main objectives of this thesis was to investigate whether any of the studied strains could potentially be represented by novel species that have not yet been formally described or genetically characterised using the 18S or 28S rRNA genes. Several

genotypes were identified that exhibited distinct positions in the DNA trees and displayed differences from the reference sequences available in the NCBI database. Genotypes, such as BE-12 (*Prorocentrum* sp.), BE-11 and K-19 (*Fragilariopsis* sp.), K-17 (order Pedinellales), K-22 (*Nitzschia* sp.), K-26 (*Actinocyclis* sp.), and K-32 (*Bacillaria* sp.) were of particular interest. As mentioned previously, the 18S rRNA sequence of the BE-12 strain is identical to that of a recently described species, *Prorocentrum pervagatum* sp. nov., which was isolated from the North Atlantic waters (Tillmann et al., 2023). A thorough morphological analysis by electron microscopy may be necessary to ensure precise identification and determination of the potential novel species. Additional gene markers may also be needed. Overall, the results obtained from detailed molecular data, phylogenetic analysis, and morphological examination provide compelling evidence for the existence of previously uncharacterised microalgal species. It is likely that these species have been described solely based on their morphological characteristics and have not been subjected to cultivation or sequencing. Among the microalgae examined, strains mentioned show promising potential as novel species. However, it is challenging to conclusively determine if any of the strains represent novel species based on the data collected during the thesis project. Therefore, a definitive answer to this question is currently unavailable. Gathering additional data on the questionable strains is necessary to further investigate and address this inquiry.

4.7 Biogeographical distribution based on metaPR2

The fourth inquiry aims to explore the biogeographical distribution of selected genotypes based on metaPR2.The biogeographical distribution of algal genotypes in this study was determined using data from the metaPR2 database. The metaPR2 database is a compiled dataset that includes various sources such as raw sequence files, online metabarcoding datasets, and direct submissions of the sequencing data from researchers. By comparing the 18S rRNA DNA sequences of the studied strains with those in the metaPR2 database, it is possible to infer the biogeographical distribution patterns of the identified genotypes. It provides metabarcoding data from different regions, including coastal areas and oceanic expeditions, enabling analysis of the presence and distribution of algal genotypes across diverse geographical locations. The metaPR2 has sufficient coverage of the Arctic region with planktonic and sea-ice habitats. However, it is important to acknowledge that the database may not cover all geographical regions, and there could be limitations in the coverage of certain areas. It was demonstrated by the biogeographic analysis that four general biogeographic distribution types exist for arctic phytoplankton and ice algae: polar, arctic, arctic-temperate, and cosmopolitan. Genotypes with arctic-temperate distribution was most common among the characterised strains, and genotypes endemic to the arctic was also found. Only two genotypes with cosmopolitan distribution and one genotype with polar distribution were discovered. The results highlight that the phytoplankton and ice algal communities in the Barents Sea and Arctic Ocean consist of genotypes endemic to the Arctic and genotypes with broader biogeographic distributions.

4.8 Unique to the Arctic species (endemic species)

The fifth query focuses on whether there are any species unique to the Arctic (endemic species). According to the results obtained during this study, the following strains have been identified as endemic arctic algae: BE-5, BE-13, BE-15, B-19 (*Micromonas polaris*); K-7 (*Chaetoceros cinctus*); BE-25 (*Bacterosira* sp.); K-35 (*Synedropsis* sp.); K-15 (*Attheya septentrionalis*); K-56 (*Thalassiosira* sp.); BE-11 and K-19 (*Fragilariopsis* sp.); and BE-21 and K-28 (*Chaetoceros neogracilis*). These strains exhibit a biogeographic distribution mostly restricted to the Arctic Circle (above 66ºN) based on the information provided by the metaPR2 database. It is important to acknowledge that this statement is not definitive. The biogeographic data obtained in this study are solely based on metabarcoding data sourced from the metaPR2 database. It is worth noting that numerous microalgal species observed through microscopy in the Arctic are not included in the database due to the lack of 18S gene sequencing. Additionally, it is important to consider that the metaPR2 database may not encompass all geographical regions.

4.9 Alternative data on the global distribution

The sixth aspect under investigation aimed to identify the global distribution, including habitat types, of the species or genotypes in question in the past. Characterised algae have been found in various marine environments worldwide. For example, *Chaetoceros neogracilis* has been documented in both polar and temperate regions, including the Arctic (Choi et al., 2008; Balzano et al., 2017), Antarctic (Trimborn et al., 2017), and Baltic Sea (Majaneva et al., 2012). These algae can thrive in cold, nutrient-rich conditions and are found in coastal and open ocean waters. *Fragilariopsis* sp. is commonly associated with sea ice habitats in the Arctic and Antarctic regions (Moreno et al., 2020; Morin et al., 2020).

Attheya longicornis is frequently found in northern cold water regions, such as the Barents Sea (Artamonova et al., 2017) and the Sea of Japan (Stonik et al., 2006). Similarly, *Attheya septentrionalis* is typically found in cold, polar, and subpolar regions, with a particular emphasis on the Arctic and sub-Arctic (Throndsen et al., 2003; Balzano et al., 2017). This species is known to be planktonic, and it is also known as an epiphyte.

Bacterosira sp., represented by the strain K-52, is a genus containing two known species: *Bacterosira consticta* and *Bacterosira bathyomphala* (Guiry, 2023). These species are commonly found in Arctic waters (Throndsen et al., 2003; Guiry, 2023).

Thalassiosira gravida is commonly found in northern and southern cold waters (Whittaker et al., 2012), including the Arctic region (Balzano et al., 2017). *Thalassiosira hispida* inhabits a range of northern cold water regions, including the Arctic (Luddington et al., 2016; Balzano et al., 2017) and extends into temperate regions (Throndsen et al., 2003).

Synedropsis hyperborea is a common species in northern cold water regions, including the Arctic (Throndsen et al., 2003; Balzano et al., 2017). *Micromonas polaris* is well-known for its presence in polar regions, particularly the Arctic Ocean (Balzano et al., 2012a) and the Southern Ocean surrounding Antarctica (Trefault et al., 2021).

It is challenging to provide detailed distribution information for genera such as *Pseudo-nitzschia*, *Nitzschia*, *Amphora*, *Pyramimonas*, and *Prorocentrum*, as they are widely distributed in marine environments. These genera can be found in various regions worldwide, including northern cold water, temperate, and even tropical areas.

4.10 The rejection of hypotheses

Both hypotheses were rejected, as there was insufficient evidence to support them.

In this study, three types of identification were used to understand the diversity and biogeography of Arctic phytoplankton and ice algae. Both morphological and molecular characterisations of studied cultures were used to link taxonomic identification with metabarcoding datasets. It was revealed by our findings that the current morphological and molecular databases do not fully represent Arctic phytoplankton diversity, as several strains that could not be identified using traditional methods were discovered.

To analyse the phytoplankton and ice algae diversity in the Barents Sea, the distribution of strains genotypes at the genus level was examined. Distinct biogeographic patterns were shown by our results, indicating that Arctic algal communities include both endemic and cosmopolitan genotypes. However, limitations in the interpretative power of metabarcoding were also observed due to the representation of taxa in reference libraries. Only about half of the strains could be identified to the species level, and this level of identification was mainly possible for well-studied genera. It became evident that Arctic diatoms are underrepresented in reference databases, particularly for diatoms.

To improve the interpretation of metabarcoding data, the need for large-scale cultivation efforts to link morphological and molecular information was identified. However, the cultivation of strains enabled the identification of previously unknown genotypes. Two common genetic markers, the 18S and 28S rRNA gene sequences, were used for species-level identification and phylogenetic placement. The 18S rRNA marker showed sufficient resolution for most species-level identifications, and its global datasets allowed for biogeographic mapping. However, in some cases, a better resolution and reveal of more genotypes could be provided by additional markers such as 28S rRNA.

Four general distribution types were found when analysing the biogeographic patterns of Arctic phytoplankton and ice algae: polar, arctic, arctic-temperate, and cosmopolitan. No clear phylogenetic relationship was associated with these distribution types, and even within well-studied genera, diatoms exhibited a wide range of biogeographic patterns. Consistency with earlier morphology-based studies was observed in our findings, which revealed both

similarities and differences compared to those studies. The use of genotype-level approaches allowed the detection of potential endemic species and a better understanding of the biogeography of Arctic phytoplankton and ice algae.

4.11 Conclusion

The study integrated morphological and molecular approaches to investigate Arctic phytoplankton and ice algae diversity. By utilising two genetic markers, 18S and 28S rRNA gene sequences, strain identification at the genus or species level was achieved for more accurate phylogenetic placement. However, it should be noted that the current morphological and molecular databases do not fully represent Arctic phytoplankton diversity, as traditional methods failed to identify certain strains. Sequencing also had limitations in data interpretation due to incomplete representation of taxa in reference databases, resulting in only about half of the strains being identified to the species level. Large-scale cultivation efforts are needed to establish a link between morphological and molecular information and to improve the interpretation of metabarcoding data.

The analysis of microalgae biodiversity isolated during the Nansen Legacy Project cruise in August 2019 (AeN706) confirmed the dominance of diatoms among Arctic phytoplankton and ice algae, with *Chaetoceros neogracilis*, *Thalassiosira gravida*, and *Attheya septentrionalis* being the most abundant diatomic species (Lovejoy et al., 2002; Wiedmann et al., 2020). Among the Chlorophyta (green algae), the genera *Micromonas* and *Pyramimonas* were found to be dominant, consistent with previous research (Balzano et al., 2012b; Ribeiro et al., 2020). Furthermore, the study identified potentially novel or previously unsequenced algal strains, including a genotype classified under the order Pedinellales and another genotype belonging to the genus *Prorocentrum*. However, it is important to note that further research is required to validate and corroborate these conclusions.

The biogeographic analysis revealed four distribution types among the studied strains: polar, arctic, arctic-temperate, and cosmopolitan. The variety of distribution patterns indicate the presence of both endemic and cosmopolitan genotypes in Arctic algal communities. A better understanding of the diversity and distribution of Arctic algae will be achieved through further research using additional markers and expanding reference databases.

References

AMAP (Arctic Monitoring and Assessment Programme). (2019). *Arctic ocean acidification assessment: 2018 summary for policymakers*. Retrieved from https://www.amap.no/documents/doc/arctic-ocean-acidification-assessment-2018-summary-for-policymakers/1542

Ardyna, M., & Arrigo, K. R. (2020). Phytoplankton dynamics in a changing Arctic Ocean. *Nature Climate Change*, *10*(10), 892–903. https://doi.org/10.1038/s41558-020-0905-y

Arrigo, K. R. (2014). Sea Ice Ecosystems. *Annual Review of Marine Science*, *6*(1), 439–467. https://doi.org/10.1146/annurev-marine-010213-135103

Artamonova, E. Y., Vasskog, T., & Eilertsen, H. C. (2017). Lipid content and fatty acid composition of Porosira glacialis and Attheya longicornis in response to carbon dioxide (CO2) aeration. PloS One, 12(5), e0177703–e0177703. https://doi.org/10.1371/journal.pone.0177703

Ashworth, M. P., Nakov, T., Theriot, E. C., & Wetherbee, R. (2013). Revisiting Ross and Sims (1971): toward a molecular phylogeny of the Biddulphiaceae and Eupodiscaceae (Bacillariophyceae). Journal of Phycology, 49(6), 1207–1222. https://doi.org/10.1111/jpy.12131

Balzano, S., Gourvil, P., Siano, R., Chanoine, M., Marie, D., Lessard, S., Sarno, D., & Vaulot, D. (2012a). 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

Balzano, S., Marie, D., Gourvil, P., & Vaulot, D. (2012b). Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples. The ISME Journal, 6(8), 1480–1498. https://doi.org/10.1038/ismej.2011.213

Basu, S., & Mackey, K. R. M. (2018). Phytoplankton as key mediators of the biological carbon pump: Their responses to a changing climate. *Sustainability*, *10*(3), 869. https://doi.org/10.3390/su10030869

Behrenfeld, M. J., & Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography*, *42*(1), 1–20. https://doi.org/10.4319/lo.1997.42.1.0001

Bråte, J., Logares, R., Berney, C., Ree, D. K., Klaveness, D., Jakobsen, K. S., & Shalchian-Tabrizi, K. (2010). Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. The ISME Journal, 4(9), 1144–1153. https://doi.org/10.1038/ismej.2010.39

Choi, H.-G., Joo, H. M., Jung, W., Hong, S. S., Kang, J.-S., & Kang, S.-H. (2008). Morphology and phylogenetic relationships of some psychrophilic polar diatoms (Bacillariophyta). Nova Hedwigia, 7–30.

Das, S., Mohapatra, S. C., & Hsu, J. T. (1999). Studies on primer-dimer formation in polymerase chain reaction (PCR). *Biotechnology Techniques*, *13*(10), 643-646.

Dodge, J. (1975). Prorocentrales (Dinophyceae). II. Revision of the taxonomy within the genus Prorocentrum. Botanical Journal of the Linnean Society, 71(2), 103–125. https://doi.org/10.1111/j.1095-8339.1975.tb02449.x

Duarte, C. M., Agustí, S., Wassmann, P., Arrieta, J. M., Alcaraz, M., Coello, A., Marbà, N., Hendriks, I. E., Holding, J., García-Zarandona, I., Kritzberg, E., & Vaqué, D. (2012). Tipping Elements in the Arctic Marine Ecosystem. *Ambio*, *41*(1), 44–55. https://doi.org/10.1007/s13280-011-0224-7

Edvardsen, B., Shalchian-Tabrizi, K., Jakobsen, K. S., Medlin, L. K., Dahl, E., Brubak, S., & Paasche, E. (2003). Genetic variability and molecular phylogeny of Dinophysis species (Dinophyceae) from Norwegian waters inferred from single cell analyses of rDNA 1. *Journal of Phycology*, *39*(2), 395-408. https://doi.org/10.1046/j.1529-8817.2003.02103.x

Eppley, R. W., Holmes, R. W., & Strickland, J. D. H. (1967). Sinking rates of marine phytoplankton measured with a fluorometer. *Journal of Experimental Marine Biology and Ecology*, *1*(2), 191–208. https://doi.org/10.1016/0022-0981(67)90014-7

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science (American Association for the Advancement of Science), 281(5374), 237–240. https://doi.org/10.1126/science.281.5374.237

Gaonkar, C. C., Kooistra, W. H. C. F., Lange, C. B., Montresor, M., Sarno, D., & Wetherbee, R. (2017). Two new species in the Chaetoceros socialis complex (Bacillariophyta): C. sporotruncatus and C. dichatoensis, and characterization of its relatives, C. radicans and C. cinctus. Journal of Phycology, 53(4), 889–907. https://doi.org/10.1111/jpy.12554

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. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., … Christen, R. (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. In *Nucleic acids research* (Vol. *41*, Issue Database issue, pp. D597–D604). Oxford University Press. https://doi.org/10.1093/nar/gks1160

Guiry, G.M. (2023, February 28). AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. Retrieved May 21, 2023, from https://www.algaebase.org

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., & Troedsson, C. (2014). Characterization of the 18S rRNA gene for designing universal eukaryote-specific primers. *PloS One*, *9*(2), e87624. https://doi.org/10.1371/journal.pone.0087624

Haines, A. M., Tobe, S. S., Kobus, H. J., & Linacre, A. (2015). Properties of nucleic acid staining dyes used in gel electrophoresis. *Electrophoresis*, *36*(6), 941–944. https://doi.org/10.1002/elps.201400496

Hamilton, A. K., Lovejoy, C., Galand, P. E., & Ingram, R. G. (2008). Water masses and biogeography of picoeukaryote assemblages in a cold hydrographically complex system. *Limnology and Oceanography*, *53*(3), 922-935.

Ichinomiya, M., Dos Santos, A. L., Gourvil, P., Yoshikawa, S., Kamiya, M., Ohki, K., Audic, S., De Vargas, C., Noël, M.-H., Vaulot, D., & Kuwata, A. (2016). Diversity and oceanic distribution of the Parmales (Bolidophyceae), a picoplanktonic group closely related to diatoms. The ISME Journal, 10(10), 2419–2434. https://doi.org/10.1038/ismej.2016.38

Jeyapandi, S., Rajthilak, C., Santhanam, P., Begum, A., Ravikumar, S., Pachiappan, P., & Ahmed, S. U. (2018). Study on Molecular Taxonomy and Phylogenetic Analysis of Phytoplankton. In Basic and Applied Phytoplankton Biology (pp. 63–75). Springer Singapore. https://doi.org/10.1007/978-981-10-7938-2_3

Kaczmarska, I., Lovejoy, C., Potvin, M., & Macgillivary, M. (2009). Morphological and molecular characteristics of selected species of Minidiscus (Bacillariophyta, Thalassiosiraceae). European Journal of Phycology, 44(4), 461–475. https://doi.org/10.1080/09670260902855873

Krumhardt, K. M., Lovenduski, N. S., Long, M. C., Luo, J. Y., Lindsay, K., Yeager, S., & Harrison, C. (2020). Potential Predictability of Net Primary Production in the Ocean. *Global Biogeochemical Cycles*, *34*(6), n/a–n/a. https://doi.org/10.1029/2020GB006531

Langbehn, T. J., & Varpe, Ø. (2017). Sea-ice loss boosts visual search: Fish foraging and changing pelagic interactions in polar oceans. *Global Change Biology*, *23*(11), 4556–4567. https://doi.org/10.1111/gcb.13797

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J.-M., Decelle, J., Dolan, J., Dunthorn, M., Edvardsen, B., Gobet, A., Kooistra, W. C. F., Mahé, F., Not, F., Ogata, H., Pawlowski, J., Pernice, M., Romac, S., … Massana, R. (2014). Patterns of Rare and Abundant Marine Microbial Eukaryotes. Current Biology, 24(8), 813–821. https://doi.org/10.1016/j.cub.2014.02.050

Lovejoy, C., Legendre, L., Martineau, M.-J., Bâcle, J., & von Quillfeldt, C. H. (2002). Distribution of phytoplankton and other protists in the North Water. Deep-Sea Research. Part II, Topical Studies in Oceanography, 49(22), 5027–5047. https://doi.org/10.1016/S0967-0645(02)00176-5

Luddington, I. A., Lovejoy, C., Kaczmarska, I., & Moisander, P. (2016). Species-rich meta-communities of the diatom order Thalassiosirales in the Arctic and northern Atlantic Ocean. Journal of Plankton Research, 38(4), 781–797. https://doi.org/10.1093/plankt/fbw030

Majaneva, M., Rintala, J.-M., Piisilä, M., Fewer, D. P., & Blomster, J. (2012). Comparison of wintertime eukaryotic community from sea ice and open water in the Baltic Sea, based on sequencing of the 18S rRNA gene. Polar Biology, 35(6), 875–889. https://doi.org/10.1007/s00300-011-1132-9

Marin, B., Melkonian, M. (2010). Molecular Phylogeny and Classification of the Mamiellophyceae class. nov. (Chlorophyta) based on Sequence Comparisons of the Nuclear- and Plastid-encoded rRNA Operons. Protist, 161(2), 304–336. https://doi.org/10.1016/j.protis.2009.10.002

Marin, B. (2012). Nested in the Chlorellales or Independent Class? Phylogeny and Classification of the Pedinophyceae (Viridiplantae) Revealed by Molecular Phylogenetic Analyses of Complete Nuclear and Plastid-encoded rRNA Operons. Protist, 163(5), 778–805. https://doi.org/10.1016/j.protis.2011.11.004

Medlin, L., Elwood, H. J., Stickel, S., & Sogin, M. L. (1988). The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene, 71*(2), 491–499. https://doi.org/10.1016/0378-1119(88)90066-2

Moreno, C. M., Gong, W., Cohen, N. R., DeLong, K., & Marchetti, A. (2020). Interactive effects of iron and light limitation on the molecular physiology of the Southern Ocean diatom Fragilariopsis kerguelensis. Limnology and Oceanography, 65(7), 1511–1531. https://doi.org/10.1002/lno.11404

Morin, P., Lacour, T., Grondin, P., Bruyant, F., Ferland, J., Forget, M., Massicotte, P., Donaher, N., Campbell, D. A., Lavaud, J., & Babin, M. (2020). Response of the sea-ice diatom Fragilariopsis cylindrus to simulated polar night darkness and return to light. Limnology and Oceanography, 65(5), 1041–1060. https://doi.org/10.1002/lno.11368

Mousing, E. A., Ribeiro, S., Chisholm, C., Kuijpers, A., Moros, M., & Ellegaard, M. (2017). Size differences of Arctic marine protists between two climate periods—using the paleoecological record to assess the importance of within-species trait variation. *Ecology and Evolution, 7*(1), 3–13. https://doi.org/10.1002/ece3.2592

Mundy, C. J., Gosselin, M., Ehn, J. K., Belzile, C., Poulin, M., Alou, E., Roy, S., Hop, H., Lessard, S., Papakyriakou, T. N., Barber, D. G., & Stewart, J. (2011). Characteristics of two distinct high-light acclimated algal communities during advanced stages of sea ice melt. Polar Biology, 34(12), 1869–1886. https://doi.org/10.1007/s00300-011-0998-x

Notz, D., & SIMIP Community. (2020). Arctic sea ice in CMIP6. *Geophysical Research Letters*, *47*(10), e2019GL086749. https://doi.org/10.1029/2019GL086749

Okolodkov, Y. B., & Dodge, J. D. (1996). Biodiversity and biogeography of planktonic dinoflagellates in the Arctic Ocean. Journal of Experimental Marine Biology and Ecology, 202(1), 19–27. https://doi.org/10.1016/0022-0981(96)00028-7

Park, J. S., Alverson, A. J., & Lee, J. H. (2016). A phylogenetic re-definition of the diatom genus Bacterosira (Thalassiosirales, Bacillariophyta), with the transfer of Thalassiosira constricta based on morphological and molecular characters. Phytotaxa, 245(1), 1–16. https://doi.org/10.11646/phytotaxa.245.1.1

Pniewski, F., Friedl, T., & Latała, A. (2010). Identification of diatom isolates from the Gulf of Gdańsk: testing of species identifications using morphology, 18S rDNA sequencing and DNA barcodes of strains from the Culture Collection of Baltic Algae (CCBA). Oceanological and Hydrobiological Studies, 39(3), 3–20. https://doi.org/10.2478/v10009-010-0031-7

Polyak, L., Alley, R. B., Andrews, J. T., Brigham-Grette, J., Cronin, T. M., Darby, D. A., Dyke, A. S., Fitzpatrick, J. J., Funder, S., Holland, M., Jennings, A. E., Miller, G. H., O'Regan, M., Savelle, J., Serreze, M., St. John, K., White, J. W. C., & Wolff, E. (2010). History of sea ice in the Arctic. *Quaternary Science Reviews*, *29*(15), 1757–1778. https://doi.org/10.1016/j.quascirev.2010.02.010

Poulin, M., Daugbjerg, N., Gradinger, R., Ilyash, L., Ratkova, T., & von Quillfeldt, C. (2011). The pan-Arctic biodiversity of marine pelagic and sea-ice unicellular eukaryotes: a first-attempt assessment. *Marine Biodiversity*, *41*(1), 13–28. https://doi.org/10.1007/s12526-010-0058-8

Reigstad, M., Gabrielsen, T., Amargant, M., Amundsen, R., Bluhm, B., Bodur, Y., ... & Yadetie, F. (2022). Seasonal Cruise Q3: Cruise Report. *The Nansen Legacy Report Series*, *27*. https://doi.org/10.7557/nlrs.6407)

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. Elementa (Washington, D.C.), 8(1), 6. https://doi.org/10.1525/elementa.401

Riebesell, U., & Gattuso, J.-P. (2015). Lessons learned from ocean acidification research. *Nature Climate Change*, *5*(1), 12–14. https://doi.org/10.1038/nclimate2456

Scholin, C. A., Herzog, M., Sogin, M., & Anderson, D. M. (1994). Identification of group- and strain-specific genetic markers for globally distributed Alexandrium (Dinophyceae). II. Sequence analysis of a fragment of the LSU rRNA gene. *Journal of Phycology*, *30*(6), 999–1011. https://doi.org/10.1111/j.0022-3646.1994.00999.x

Shimada, C., & Tanimura, Y. (2006). Spatial variability in valve morphology of Neodenticula seminae, an oceanic diatom in the subarctic North Pacific and the Bering Sea. *Paleontological Research*, *10*(1), 79–89. https://doi.org/10.2517/prpsj.10.79

Simon, N., Cras, A.-L., Foulon, E., & Lemée, R. (2009). Diversity and evolution of marine phytoplankton. *Comptes Rendus. Biologies*, *332*(2), 159–170. https://doi.org/10.1016/j.crvi.2008.09.009

Sorhannus, U., & Fox, M. G. (2012). Phylogenetic Analyses of a Combined Data Set Suggest that the Attheya Lineage is the Closest Living Relative of the Pennate Diatoms (Bacillariophyceae). Protist, 163(2), 252–262. https://doi.org/10.1016/j.protis.2011.04.005

Stonik, I., Orlova, T., & Aizdaicher, N. (2006). Diatoms of the genus Attheya west, 1860 from the Sea of Japan. Russian Journal of Marine Biology, 32(2), 123–126. https://doi.org/10.1134/S1063074006020088

Supraha, L., Klemm, K., Gran-Stadniczenko, S., Horstmann, C., Vaulot, D., Edvardsen, B., & John, U. (2022). Diversity and biogeography of planktonic diatoms in Svalbard fjords: The role of dispersal and Arctic endemism in phytoplankton community structuring. *Elementa (Washington, D.C.)*, *10*(1). https://doi.org/10.1525/elementa.2021.00117

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

Theriot, E. C., Ashworth, M. P., Nakov, T., Ruck, E., & Jansen, R. K. (2015). Dissecting signal and noise in diatom chloroplast protein encoding genes with phylogenetic information profiling. Molecular Phylogenetics and Evolution, 89, 28–36. https://doi.org/10.1016/j.ympev.2015.03.012

Throndsen, J., Tangen, K., & Hasle, G. R. (2003). Norsk kystplanktonflora (p. 341). Almater forl.

Thyrring, J., Tremblay, R., & Sejr, M. K. (2017). Importance of ice algae and pelagic phytoplankton as food sources revealed by fatty acid trophic markers in a keystone species (Mytilus trossulus) from the High Arctic. Marine Ecology. Progress Series (Halstenbek), 572, 155–164. https://doi.org/10.3354/meps12143

Tillmann, U., Wietkamp, S., Gottschling, M., & Hoppenrath, M. (2023). Prorocentrum pervagatum sp. nov. (Prorocentrales, Dinophyceae): A new, small, planktonic species with a global distribution. *Phycological Research*, *71*(1), 56–71. https://doi.org/10.1111/pre.12502

Trefault, N., De la Iglesia, R., Moreno-Pino, M., Lopes dos Santos, A., Gérikas Ribeiro, C., Parada-Pozo, G., Cristi, A., Marie, D., & Vaulot, D. (2021). Annual phytoplankton dynamics in coastal waters from Fildes Bay, Western Antarctic Peninsula. Scientific Reports, 11(1), 1368–1368. https://doi.org/10.1038/s41598-020-80568-8

Trimborn, S., Brenneis, T., Hoppe, C. J. M., Laglera, L. M., Norman, L., Santos-Echeandía, J., Völkner, C., Wolf-Gladrow, D., & Hassler, C. S. (2017). Iron sources alter the response of

Southern Ocean phytoplankton to ocean acidification. Marine Ecology. Progress Series (Halstenbek), 578, 35–50. https://doi.org/10.3354/meps12250

Van de Schootbrugge, B., Houben, A. J. P., Ercan, F. E. Z., Verreussel, R., Kerstholt, S., Janssen, N. M. M., Nikitenko, B., & Suan, G. (2020). Enhanced Arctic-Tethys connectivity ended the Toarcian Oceanic Anoxic Event in NW Europe. Geological Magazine, 157(10), 1593–1611. https://doi.org/10.1017/S0016756819001262

Vaquer-Sunyer, R., Duarte, C. M., Santiago, R., Wassmann, P., & Reigstad, M. (2010). Experimental evaluation of planktonic respiration response to warming in the European Arctic Sector. *Polar Biology*, *33*(12), 1661–1671. https://doi.org/10.1007/s00300-010-0788-x

Vaulot, D., Sim, C. W. H., Ong, D., Teo, B., Biwer, C., Jamy, M., & Lopes dos Santos, A. (2022). metaPR2: A database of eukaryotic 18S rRNA metabarcodes with an emphasis on protists. *Molecular Ecology Resources*, *22*(8), 3188–3201. https://doi.org/10.1111/1755-0998.13674

Verde, C., Giordano, D., Bellas, C. M., di Prisco, G., & Anesio, A. M. (2016). Polar Marine Microorganisms and Climate Change. *Advances in Microbial Physiology*, *69*, 187–215. https://doi.org/10.1016/bs.ampbs.2016.07.002

Wiedmann, I., Ceballos-Romero, E., Villa-Alfageme, M., Renner, A. H. H., Dybwad, C., Jagt, H., Svensen, C., Assmy, P., Wiktor, J. M., Tatarek, A., Różańska-Pluta, M., & Iversen, M. H. (2020). Arctic Observations Identify Phytoplankton Community Composition as Driver of Carbon Flux Attenuation. Geophysical Research Letters, 47(14). https://doi.org/10.1029/2020GL087465

Whittaker, K. A., Rignanese, D. R., Olson, R. J., & Rynearson, T. A. (2012). Molecular subdivision of the marine diatom Thalassiosira rotula in relation to geographic distribution, genome size, and physiology. BMC Evolutionary Biology, 12(1), 209–209. https://doi.org/10.1186/1471-2148-12-209

Appendix 1.

DNA quantification description.

The Qubit working solution was prepared by mixing 199 µL of dilution buffer and 1 µL of concentrated Qubit reagent (fluorophore) for each DNA sample. The volume of the Qubit Master Mix (working solution) varied depending on the number of samples (Table A1). Before pipetting (Table A2, A3), samples (DNA solutions) were defrosted on ice and then left to reach room temperature. The result is presented in ng/µL.

Table A1. Reagents of the Qubit working solution per DNA sample/standard

Component	Volume
Dilution buffer	199 µl
Qubit reagent	1 µl
Final volume	200 µl

Table A2. Pipetting for DNA quantification in Qubit standards (calibration)

Component	Volume
Qubit working solution	190 µl
Standard (S1 and S2)	10 µl
Final volume	200 µl

Table A3. Pipetting for DNA quantification in Qubit standards (calibration)

Component	Volume
Qubit working solution	198 µl
DNA solution	2 µl
Final volume	200 µl

The principle of this method is based on the binding of a fluorochrome reagent to DNA in the sample (Figure A1). The emitted signal from the fluorochrome-DNA complexes is measured, and the amount of DNA present in the sample is estimated using a fluorometer. The result is displayed on the screen of the Qubit device.

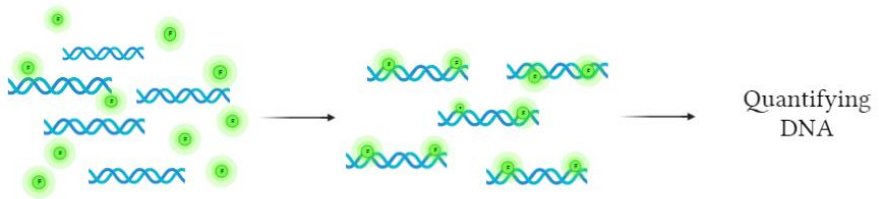


Figure A1. Illustration of the DNA quantification process by Qubit. The green dots represent the fluorochrome, and the double helix represents the DNA. The fluorochrome binds to the DNA, and the result is obtained using the Qubit device measuring the fluorescence emitted. The image was created using BioRender.com.

Appendix 2.

The PCR program was also adjusted depending on the gene and primers used.

Tabell A4. The PCR program for amplification of the 18S rRNA gene.

Temperature	Time	PCR step
95°C	2 min	Denaturation
95°C	45 s	Denaturation
50°C	45s	Primer annealing
72°C	2 min	Extension
72°C	5 min	Extension
8°C		Soak

Tabell A5. The PCR program for amplification of the 28S rRNA gene.

Temperature	Time	PCR step
95°C	3 min	Denaturation
95°C	45 s	Denaturation
55°C	45s	Primer annealing
72°C	1 min	Extension
72°C	5 min	Extension
8°C		Soak

Date:

	Template: (Strain/sample)	Extracted date
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		

Gene:
Primers:
Taq polymerase:

Reaction mix (µL per sample)

	Reagent	**1 reaction (µL)**	**X reactions (µL)**
1	PCR water	7.5	
2	primer 1 (5 µM->1 µM)	1.5	
3	primer 2 (5 µM->1 µM)	1.5	
4	GoTaq Green Master mix	12.5	
5	DNA template (ca 20 ng μL^{-1})	2	
	Total	25	

PCR Program
1: Denaturation at 95°C for 3 min)
2. DNA synthesis in (30 cycles consisting of:
2.1: Denaturation at 95°C for 45s
2.2: Primer annealing at 55°C for 45s
2.3: Extension at 72°C for 1-2 min (or 1 min per 1000 bp)
3. Extension at 72°C for 5 min
4. Soak at 8 °C

Figure A2. Scheme used for PCR to calculate the components of the reaction and to account for the strains used.

Appendix 3.

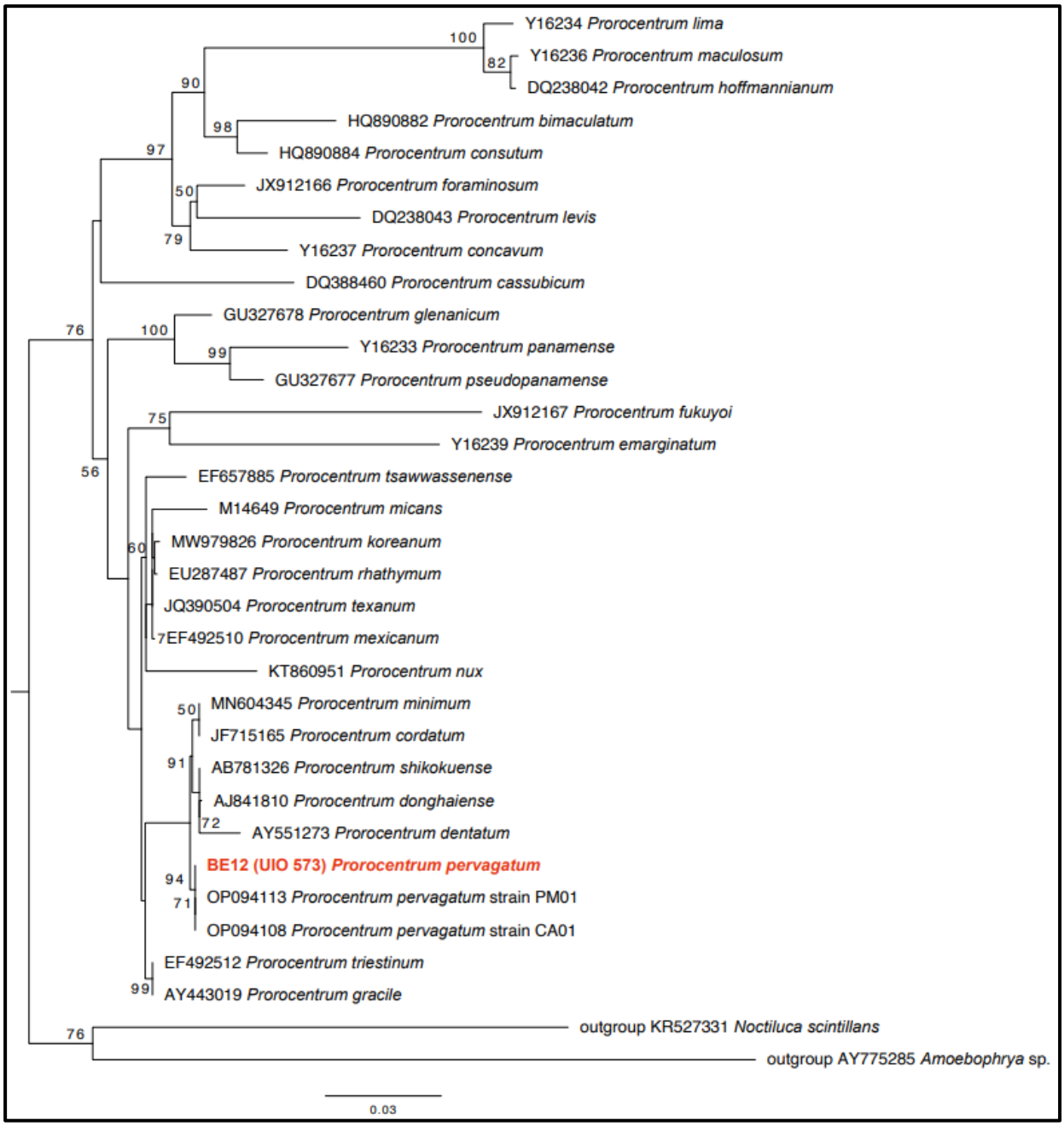


Figure A3. A new alignment of BE-12 sequence with Prorocentrum reference sequences one sequence per species for SSU. A Prorocentrum tree (made by Bente Edvardsen) placed BE-12 together with Prorocentrum pervagatum (Tillmann at al., 2023). Phylogenetic tree made with PHYML in Geneious with bootstrap values SH-like. The partial sequence of 28S ribosomal gene has also been sequenced. The article was published in 2023.

Appendix 4.

18S rRNA sequences:

BE_AeN706-3

CCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTATAAGCTGATTATACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTATTTGATGG
TA--CTACTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCG-
CAACTCCCGACTTCTGGAAGGGACGTATTTATTAGATAAAAGACCAGCCGCCCTCGGGCGTTTTGTGGTGAATCATGATAACTTGTCGGATCGCATGAGCTTGTCTCGGCG
ACGTTTCATTCAAATTTCTGCCCTATCAACTTTCGATGGTAGGATAGAGGCCTACCATGGTGGTAACGGGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGA
GAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACACAGGGAGGTAGTGACAATAAATAACAATACCGGGCTTTTTCAAGTCTGGTAA
TTGGAATGAGAACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAG
TTAAAAAGCTCGTAGTTGGATTTCGGGTGAGGATGACCGGTCTGCCGTTGCGGTACGCACTGGACGTTTTCATCTTGTTGTCGGGGACGCGCTTCTGGGCTTCACTGTTCG
GGACGCGGAGTCGGCGCTGTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCAATCGCTCTGAATACATTAGCATGGAATAACGCTATAGGACTCTGGTCCTATTGTGT
TGGTCTTCGGGACCGGAGTAATGATTAAGAGGGACAGTTGGGGGCATTCGTATTTCATTGTCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACTTCTGCGAAAGCAT
TTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGAT--
GCCGACTAGGGATTGGCGGATGTTATATCGATGACTTCGCCAGCACCTTATGAGAAATCAAAGTTTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTAA-
AGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGATTGAGAGCT
CTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCA-GCCTGCTAAATAGTC-
TCGCAAATTCTGAATTTGCGCTTGGCTTCTTAGAGGGACTATTGGCGTATAGCCAATGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCGCAC
GCGCGCTACACTGATGAATTCAACGAGCTCCTAACCTAGGCTGAAGAGTCCGGGTAATCTTTGAAA-
TTTCATCGTGATGGGGATAGATTATTGCAATTATTAATCTTCAACGAGGAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGATTACGTCCCTGCCCTTTGTACACACCG
CCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAATGTTCGGACTGTTGCGTGGCGAACGGTC-CGCCGTCTGCTTCGCGATGGGAAGTTCAT-
TAAACCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCG

BE_AeN706-5

TAAGCCATGCATGTCTAAGTATAAGC--
GTTATACTGKGAAACTGCGAATGGCTCATTAAATCAGCAATAGTTTCTTTGGTGGTGACCTACTACATGGATAACCGTAGTAATTCTAGAGCTAATACATGCG-
TAAATCTCGACTCACG-AAGAGACGTATTTATTAGATAAAGACCGACCT---------
CGTTCTGCGGTGAATCATAATAACTTCACGGACCGCATGGCTTTATGCCGGCGGTGTTCCATTCAAATTTCTGCCCTATCAACTTTCGACGGTAGGATAGAGGCCTACCGT
GGTGTTCACGGGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACA
CAGGGAGGTAGTGACAATAAATAACAATATCGGGGTTTTTCAACTCTGATAATTGGAATGAGAACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGTCTGGTG
CCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATTTCGGTTGAGAACGGCCGGTCCGCCGTT-
TGGTGTGCACTGGCTGGTCTCAACTTCCTGTAGAGGACGCGCTCTGGGTTAACGCTC----
GGACGCGGAGTCTACGTGGTTACTTTGAAAAAATTAGAGTGTTCAAAGCGGGCTTACGCT-
TGAATATTTCAGCATGGAATAACACTATAGGACTCCTGTCCTATTTCGTTGGTCT-
CGGGACGGGAGTAATGATTAAGAGGAACAGTTGGGGGCATTCGTATTTCATTGTCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACTTCTGCGAAAGCATTTGCCAA
GGATGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGATGATTAGATACCATCCTAGTCTCAACCATAAACGAT--GCCGACTAGGGATTGCAGGATGTTA-
ATTGATGACTCCTGCAGCACCTTATGAGAAATCAAAGTTTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTAA-
AGGAATTGACGGAAGGGCACCACCAGGCGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGATTGATAGCT
CTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCA-GCCTGCTAAATAGTT-
GTACACTACTCTT-AGTGCAGCAA-
CTTCTTAGAGGGACTATGTGCGTTTAGCACATGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACACTGACGAATGCAAC
GAGCTTATAACCTTGGCCGAAAGGTCTGGGTAATCTCCAAA--
TTTCGTCGTGATGGGGATAGATTATTGCAATTATTAATCTTCAACGAGGAATGCCTAGTAAGCGCAAGTCATCAGCTTGCGTTGATTACGTCCCTGCCCTTTGTACACACC
GCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAGCGTTCGGACCGCTTCCTC-TGGACGGGC-AACCGTCTG-TTGGTCGTGGGAAGTTCGT-
TAAACCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGA----------G

BE_AeN706-6

CCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTATAAGCTGATTATACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTATTTGATGG
TA--CTACTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCG-
CAACTCCCGACTTCTGGAAGGGACGTATTTATTAGATAAAAGACCAGCCGCCCTCGGGCGTTTTGTGGTGAATCATGATAACTTGTCGGATCGCATGAGCTTGTCTCGGCG
ACGTTTCATTCAAATTTCTGCCCTATCAACTTTCGATGGTAGGATAGAGGCCTACCATGGTGGTAACGGGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGA
GAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACACAGGGAGGTAGTGACAATAAATAACAATACCGGGCTTTTTCAAGTCTGGTAA
TTGGAATGAGAACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAG
TTAAAAAGCTCGTAGTTGGATTTCGGGTGAGGATGACCGGTCTGCCGTTGCGGTACGCACTGGACGTTTTCATCTTGTTGTCGGGGACGCGCTTCTGGGCTTCACTGTTCG
GGACGCGGAGTCGGCGCTGTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCAATCGCTCTGAATACATTAGCATGGAATAACGCTATAGGACTCTGGTCCTATTGTGT
TGGTCTTCGGGACCGGAGTAATGATTAAGAGGGACAGTTGGGGGCATTCGTATTTCATTGTCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACTTCTGCGAAAGCAT
TTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGAT--
GCCGACTAGGGATTGGCGGATGTTATATCGATGACTTCGCCAGCACCTTATGAGAAATCAAAGTTTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTAA-
AGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGATTGAGAGCT
CTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCA-GCCTGCTAAATAGTC-
TCGCAAATTCTGAATTTGCGCTTGGCTTCTTAGAGGGACTATTGGCGTATAGCCAATGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCGCAC
GCGCGCTACACTGATGAATTCAACGAGCTCCTAACCTAGGCTGAAGAGTCCGGGTAATCTTTGAAA-
TTTCATCGTGATGGGGATAGATTATTGCAATTATTAATCTTCAACGAGGAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGATTACGTCCCTGCCCTTTGTACACACCG
CCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAATGTTCGGACTGTTGCGTGGCGAACGGTC-CGCCGTCTGCTTCGCGATGGGAAGTTCAT-
TAAACCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAAC

BE_AeN706-8

CCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTATAAGCTGATTATACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTATTTGATGG
TA--CTACTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCG-
CAACTCCCGACTTCTGGAAGGGACGTATTTATTAGATAAAAGACCAGCCGCCCTCGGGCGTTTTGTGGTGAATCATGATAACTTGTCGGATCGCATGAGCTTGTCTCGGCG
ACGTTTCATTCAAATTTCTGCCCTATCAACTTTCGATGGTAGGATAGAGGCCTACCATGGTGGTAACGGGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGA
GAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACACAGGGAGGTAGTGACAATAAATAACAATACCGGGCTTTTTCAAGTCTGGTAA
TTGGAATGAGAACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAG
TTAAAAAGCTCGTAGTTGGATTTCGGGTGAGGATGACCGGTCTGCCGTTGCGGTACGCACTGGACGTTTTCATCTTGTTGTCGGGGACGCGCTTCTGGGCTTCACTGTTCG
GGACGCGGAGTCGGCGCTGTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCAATCGCTCTGAATACATTAGCATGGAATAACGCTATAGGACTCTGGTCCTATTGTGT
TGGTCTTCGGGACCGGAGTAATGATTAAGAGGGACAGTTGGGGGCATTCGTATTTCATTGTCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACTTCTGCGAAAGCAT
TTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGAT--
GCCGACTAGGGATTGGCGGATGTTATATCGATGACTTCGCCAGCACCTTATGAGAAATCAAAGTTTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTAA-

AGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGATTGAGAGCT
CTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCA-GCCTGCTAAATAGTC-
TCGCAAATTCTGAATTTGCGCTTGGCTTCTTAGAGGGACTATTGGCGTATAGCCAATGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCGCAC
GCGCGCTACACTGATGAATTCAACGAGCTCCTAACCTAGGCTGAAGAGTCCGGGTAATCTTTGAAA-
TTTCATCGTGATGGGGATAGATTATTGCAATTATTAATCTTCAACGAGGAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGATTACGTCCCTGCCCTTTGTACACACCG
CCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAATGTTCGGACTGTTGCGTGGCGAACGGTC-CGCCGTCTGCTTCGCGATGGGAAGTTCAT-
TAAACCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAAC

BE_AeN706-9

GCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGACGATYAGATACCGTCCTAGTCTCAACCATAAACGAT--
GCCGACTAGGGATTGGCGGATGTTATATCGATGACTTCTCCAGCACCTTATGAGAAATCAAAGTTTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTAA-
AGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGATTGAGAGCT
CTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCA-GCCTGCTAAATAGTC-
TCGCAAATTCTGAATTTGCGCTTGGCTTCTTAGAGGGACTATTGGCGTATAGCCAATGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCGCAC
GCGCGCTACACTGATGAATTCAACGAGCTCCTAACCTAGGCTGAAGAGTCCGGGTAATCTTTGAAA-
TTTCATCGTGATGGGGATAGATTATTGCAATTATTAATCTTCAACGAGGAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGATTACGTCCCTGCCCTTTGTACACACCG
CCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAATGTTCGGACTGTTGCGTGGCGAACGGTC-CGCCGTCTGCTTCGCGATGGGAAGTTCAT-
TAAACCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAA

BE_AeN706-10

GTCTCAAAGATTAAGCCATGCATGTCTAAGTATAAGCTGATTATACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTATTTGATGGTA--
CTACTACTCGGATAACCGTAGTAATTCTAGAGCTAATACGTGCG-
CAACTCCCGACTTCTGGAAGGGACGTATTTATTAGATAAAAGACCAGCCGCCCTCGGGCGTTTTGTGGTGAATCATGATAACTTGTCGGATCGCATGAGCTTGTCTCGGCG
ACGTTTCATTCAAATTTCTGCCCTATCAACTTTCGATGGTAGGATAGAGGCCTACCATGGTGGTAACGGGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGA
GAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACACAGGGAGGTAGTGACAATAAATAACAATACCGGGCTTTTTCAAGTCTGGTAA
TTGGAATGAGAACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAG
TTAAAAAGCTCGTAGTTGGATTTCGGGTGAGGATGACCGGTCTGCCGTTGCGGTACGCACTGGACGTTTTCATCTTGTTGTCGGGGACGCGCTTCTGGGCTTCACTGTTCG
GGACGCGGAGTCGGCGCTGTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCAATCGCTCTGAATACATTAGCATGGAATAACGCTATAGGACTCTGGTCCTATTGTGT
TGGTCTTCGGGACCGGAGTAATGATTAAGAGGGACAGTTGGGGGCATTCGTATTTCATTGTCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACTTCTGCGAAAGCAT
TTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGAT--
GCCGACTAGGGATTGGCGGATGTTATATCGATGACTTCGCCAGCACCTTATGAGAAATCAAAGTTTTTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTAA-
AGGAATTGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGATTGAGAGCT
CTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCA-GCCTGCTAAATAGTC-
TCGCAAATTCTGAATTTGCGCTTGGCTTCTTAGAGGGACTATTGGCGTATAGCCAATGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCGCAC
GCGCGCTACACTGATGAATTCAACGAGCTCCTAACCTAGGCTGAAGAGTCCGGGTAATCTTTGAAA-
TTTCATCGTGATGGGGATAGATTATTGCAATTATTAATCTTCAACGAGGAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGATTACGTCCCTGCCCTTTGTACACACCG
CCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAATGTTCGGACTGTTGCGTGGCGAACGGTC-CGCCGTCTGCTTCGCGATGGGAAGTTCAT-
TAAACCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAG

BE_AeN706-11

CTACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTTAGACAAGTTCTCGCGACTAACTTCCAATAAAGGAAACTAACCACAATCCCGAG-
GCTTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACGA
TGCAAGTTCAAAAGATTGCCCAGGCCTCTCGGCCAAGGT-
AAGAACTCGTTGAATGCATCAGTGTAGCGCGCGTGCGGCCCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCTATCTTCCTGCGTCTAATAGAACGCACGTCCCT
CTAAGAAGCAC-AA-TGCAGTG--
ATAAACACTACACGTGCTATTTAGCAGGCAGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAAT
CAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTTATGATTTCTCACAAGGTGCTGACGGAGACGAAACGAAACTCCGCCAATCCCTTGTCGGCATAGTTTATGGTTAAGACTACGATGGTATC
TAATCATCTTCGATCCCCTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTTTCGCAGTAGTTTGTCTTCCAAAAATCCAAGAATTTCACCTCTGACAATGG
AATACAAATACCCCCAACTGTCCCTCTTAATCATTACCTCAGTGCGCAAACCAACAAAATAGTACTAAGGTCCTATATTATTATTCCATGCTAATATATTCAA--
GGCATAAGCCTGCTTTGAACACTCTAATTTTTTCACAGTAAACGATGGGCATCCCCTGC-ACGACAACTTAATGCCACACAGATTCCACCCAAACATGG-C--AGACCAGT-
TCAGCACTATCACTCAAAGAGCAAAGGCCGACTGGACACACCACAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATACTCGATAAGGGGTTTAAATTGTTCTCATTCCAATTGCCAGACCTAAGAAGGCCCGGCATTGTTATTTATTGTCA
CTACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGGAATCAAACCCTAATTTCCCGTTACCCG
TTAAAGCCATGGTAGGCCAATACCCTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-GCC--
AGAGGCATGCGATCCGCAAGTTTATTATGAATCACCACA-TCACCCC--GA--AGGGGTTGGTTTCAATCTAATAAATACTACCC--------
CAGAAGGGTATTGACGCATGTATTAGCTCTAGAATTACTACGGGTATCCAAGTAGTAAGGGACTATCAAATAAACTATAACTGATATAATGAGCCGTTCGCAG-
TTTCAAAG--T-AAA-TATTTATACTTAGACATGCATGGCTTAATCTT

BE_AeN706-12

TCACCTACGGAAACCTTGTTACGACTTCTCCTTCCTCTAAGTGATAAGGTTCACTAAACTTTCCGCAAGCAGGTCCAGG-
AACTGGACACCGCAGCAGTCCGAATTATTCACCGGATCACTCAATCGGTAGGAGCGACGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAGCACGAGCTGATGACTCG
CGCTTACTAGGAATTCCTCGTTGAAGATTAATAATTGCAATAATCTATCCCCATCACGATGCGATTTAAAAAGATTACCCAACCTTTCCAGGCAAGGTCATAAACTCGTTG
AACGCATCAGTGTAGCGCGCGTGCAGCCCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCTCAAACTTCCTTGCGTTAGACACGCAAAGTCCCTCTAAGAAGTTGC
CCACGTAACCGAAGTTACGTG---
TAACTATTTAGCAGGTTAAGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAATCAAGAAAGAGCTA
TCAATCTGTCAATCCTTACTATGTCTGGACCTGGTAAGTTTCCCCGTGTTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTCCTTTAAGTTT
CAGCCTTGCGACCATACTCCCCCCGGAACCCAAAGACTTTGATTTCTCATAAGGTGCTGAAGGAGTCGTATAAATAACGACCTCCAATCTCTAGTCGGCATGGTTTATGGT
TAAGACTAGGACGGTATCTGATCGTCTTCGATCCCCTAACTTTCGTTCTTGATCAATGAAAACATCCTTGGCAAATGCTTTCGCAGTAGTCCGTCTTTAACAAATCCAAGA
ATTTCACCTCTGACAGTTAAATACGAATGCCCCCAACTATCCCTATTAATCATTACCTCAGCTCTAGAAACCAACAAAATAGAACCGAGGTCCTATCTTATTATTCCATGCT
AATGTATTCAA-GGCGTAAGCCTGCTTGAAACACTCTAATTTCCTCAAAGTAAAAG------
TCCTGGATACCGCACCACACAGTCAAGTGCAGATACGTTCTCCAAGAAGATGCCCAGGCCGAGCCAGATACTCACCCAGAGGGCGGACCGGTCGTCCTCGGCAGAAATC
CAACTACGAGCTTTTTAACCGCAACAACTTTAATATACGCTATTGGAGCTGGAATTACCGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGGTACTCGTAAAGGGATTTA
AATTCTACTCATTCCAATTACAAGAC--
ATATATGCCCTGTATTGTTATTTCTTGTCACTACCTCCCTGTGTCAGGATTGGGTAATTTGCGCGCCTGCTGCCTTCCTTAGATGTGGTAGCCGTTTCTCAGGCTCCCTCTCC
GGAATCGAACCCTAATTCTCCGTTACCCGTCATTGCCACGGTAGGCCAATACCCTACCGTCGGAAGCTGATAGGTCAGAAACTTGAATGATTCATCGCCAG--
CGAAGCTATGCGATTCGTTCGGTTATTATGAATCACCAAAGGACCAGGCGGAGCCTGGGCTGGTTCTGTAACTAATAAACACAACCCTTC--
CTAGAAGTCGGGTTTGGACGCATGTATTAGCTCTACAATTAGCACAGTTATCCATGTAAAGAATGACCATCAAATAAACTATAACTGTTTTAATGAGCCATTCGCAGTTTC
GCCGTATAGAA-GCTTATACTGAGACATGCATGGCTTAATCTT

BE_AeN706-13

CAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTATAAGC--
GTTATACTGTGAAACTGCGAATGGCTCATTAAATCAGCAATAGTTTCTTTGGTGGTGACCTACTACATGGATAACCGTAGTAATTCTAGAGCTAATACATGCG-
TAAATCTCGACTCACG-AAGAGACGTATTTATTAGATAAAGACCGACCT---------
CGTTCTGCGGTGAATCATAATAACTTCACGGACCGCATGGCTTTATGCCGGCGGTGTTCCATTCAAATTTCTGCCCTATCAACTTTCGACGGTAGGATAGAGGCCTACCGT
GGTGTTCACGGGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACA
CAGGGAGGTAGTGACAATAAATAACAATATCGGGGTTTTTCAACTCTGATAATTGGAATGAGAACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGTCTGGTG
CCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATTTCGGTTGAGAACGGCCGGTCCGCCGTT-
TGGTGTGCACTGGCTGGTCTCAACTTCCTGTAGAGGACGCGCTCTGGGTTAACGCTC----
GGACGCGGAGTCTACGTGGTTACTTTGAAAAAATTAGAGTGTTCAAAGCGGGCTTACGCT-
TGAATATTTCAGCATGGAATAACACTATAGGACTCCTGTCCTATTTCGTTGGTCT-
CGGGACGGGAGTAATGATTAAGAGGAACAGTTGGGGGCATTCGTATTTCATTGTCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACTTCTGCGAAAGCATTTGCCAA
GGATGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGATGATTAGATACCATCCTAGTCTCAACCATAAACGAT--GCCGACTAGGGATTGCAGGATGTTA-
ATTGATGACTCCTGCAGCACCTTATGAGAAATCAAAGTTTTTGGGTTCCGGGGGGGAGTATGGTCGCAAGGCTGAAACTTAA-
AGGAATTGACGGAAGGGCACCACCAGGCGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGATTGATAGCT
CTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCA-GCCTGCTAAATAGTT-
GTACACTACTCTT-AGTGCAGCAA-
CTTCTTAGAGGGACTATGTGCGTTTAGCACATGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACACTGACGAATGCAAC
GAGCTTATAACCTTGGCCGAAAGGTCTGGGTAATCTCCAAA--
TTTCGTCGTGATGGGGATAGATTATTGCAATTATTAATCTTCAACGAGGAATGCCTAGTAAGCGCAAGTCATCAGCTTGCGTTGATTACGTCCCTGCCCTTTGTACACACC
GCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAGCGTTCGGACCGCTTCCTC-TGGACGGGC-AACCGTCTG-TTGGTCGTGGGAAGTTCGT-
TAAACCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACTG

BE_AeN706-15

TGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTATAAGC--
GTTATACTGTGAAACTGCGAATGGCTCATTAAATCAGCAATAGTTTCTTTGGTGGTGACCTACTACATGGATAACCGTAGTAATTCTAGAGCTAATACATGCG-
TAAATCTCGACTCACG-AAGAGACGTATTTATTAGATAAAGACCGACCT---------
CGTTCTGCGGTGAATCATAATAACTTCACGGACCGCATGGCTTTATGCCGGCGGTGTTCCATTCAAATTTCTGCCCTATCAACTTTCGACGGTAGGATAGAGGCCTACCGT
GGTGTTCACGGGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACA
CAGGGAGGTAGTGACAATAAATAACAATATCGGGGTTTTTCAACTCTGATAATTGGAATGAGAACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGTCTGGTG
CCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATTTCGGTTGAGAACGGCCGGTCCGCCGTT-
TGGTGTGCACTGGCTGGTCTCAACTTCCTGTAGAGGACGCGCTCTGGGTTAACGCTC----
GGACGCGGAGTCTACGTGGTTACTTTGAAAAAATTAGAGTGTTCAAAGCGGGCTTACGCT-
TGAATATTTCAGCATGGAATAACACTATAGGACTCCTGTCCTATTTCGTTGGTCT-
CGGGACGGGAGTAATGATTAAGAGGAACAGTTGGGGGCATTCGTATTTCATTGTCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACTTCTGCGAAAGCATTTGCCAA
GGATGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGATGATTAGATACCATCCTAGTCTCAACCATAAACGAT--GCCGACTAGGGATTGCAGGATGTTA-
ATTGATGACTCCTGCAGCACCTTATGAGAAATCAAAGTTTTTGGGTTCCGGGGGGGAGTATGGTCGCAAGGCTGAAACTTAA-
AGGAATTGACGGAAGGGCACCACCAGGCGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGATTGATAGCT
CTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCA-GCCTGCTAAATAGTT-
GTACACTACTCTT-AGTGCAGCAA-
CTTCTTAGAGGGACTATGTGCGTTTAGCACATGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACACTGACGAATGCAAC
GAGCTTATAACCTTGGCCGAAAGGTCTGGGTAATCTCCAAA--
TTTCGTCGTGATGGGGATAGATTATTGCAATTATTAATCTTCAACGAGGAATGCCTAGTAAGCGCAAGTCATCAGCTTGCGTTGATTACGTCCCTGCCCTTTGTACACACC
GCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAGCGTTCGGACCGCTTCCTC-TGGACGGGC-AACCGTCTG-TTGGTCGTGGGAAGTTCGT-
TAAACCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAG

BE_AeN706-19

TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTATAAGC--
GTTATACTGTGAAACTGCGAATGGCTCATTAAATCAGCAATAGTTTCTTTGGTGGTGACCTACTACATGGATAACCGTAGTAATTCTAGAGCTAATACATGCG-
TAAATCTCGACTCACG-AAGAGACGTATTTATTAGATAAAGACCGACCT---------
CGTTCTGCGGTGAATCATAATAACTTCACGGACCGCATGGCTTTATGCCGGCGGTGTTCCATTCAAATTTCTGCCCTATCAACTTTCGACGGTAGGATAGAGGCCTACCGT
GGTGTTCACGGGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTGACA
CAGGGAGGTAGTGACAATAAATAACAATATCGGGGTTTTTCAACTCTGATAATTGGAATGAGAACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGTCTGGTG
CCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATTTCGGTTGAGAACGGCCGGTCCGCCGTT-
TGGTGTGCACTGGCTGGTCTCAACTTCCTGTAGAGGACGCGCTCTGGGTTAACGCTC----
GGACGCGGAGTCTACGTGGTTACTTTGAAAAAATTAGAGTGTTCAAAGCGGGCTTACGCT-
TGAATATTTCAGCATGGAATAACACTATAGGACTCCTGTCCTATTTCGTTGGTCT-
CGGGACGGGAGTAATGATTAAGAGGAACAGTTGGGGGCATTCGTATTTCATTGTCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACTTCTGCGAAAGCATTTGCCAA
GGATGTTTTCATTAATCAAGAACGAAAGTTGGGGGCTCGAAGATGATTAGATACCATCCTAGTCTCAACCATAAACGAT--GCCGACTAGGGATTGCAGGATGTTA-
ATTGATGACTCCTGCAGCACCTTATGAGAAATCAAAGTTTTTGGGTTCCGGGGGGGAGTATGGTCGCAAGGCTGAAACTTAA-
AGGAATTGACGGAAGGGCACCACCAGGCGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAAACTTACCAGGTCCAGACATAGTAAGGATTGACAGATTGATAGCT
CTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCA-GCCTGCTAAATAGTT-
GTACACTACTCTT-AGTGCAGCAA-
CTTCTTAGAGGGACTATGTGCGTTTAGCACATGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACACTGACGAATGCAAC
GAGCTTATAACCTTGGCCGAAAGGTCTGGGTAATCTCCAAA--
TTTCGTCGTGATGGGGATAGATTATTGCAATTATTAATCTTCAACGAGGAATGCCTAGTAAGCGCAAGTCATCAGCTTGCGTTGATTACGTCCCTGCCCTTTGTACACACC
GCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAGCGTTCGGACCGCTTCCTC-TGGACGGGC-AACCGTCTG-TTGGTCGTGGGAAGTTCGT-
TAAACCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCCAGAAGGATCA

BE_AeN706-21

ACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAGTGAT-AAGGTTTAGACAGGTTCTCGTGGACAAACTGCAATAAAGTAGATT-GCCACAATCCCGAG-
TCTTCACCGGACCACTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCATGATTAATAATTGCAATAATCAATCCCTATCACGA
TGCATGCTAAAAAGATTTCCCAGGCCTCTCGGCCAAGGT--
GAAGCTCGTTGAATGCATCAGTGTAACACGCGTGCGGTTCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCAATCTTCCTGTAGTTT-
TATACTACATGTCCCTCTAAGAAACCA-TC-GTCAATG-
AAAAATCATTAGCGTGATTATTTAGCAGGCTGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
ATAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCAGACGAGGCAAAAATGCA--
CCCGCCCATCCTGAGTCGGCATAGTTTATGGTATAGACTACGATGGTATCTAATCATCTTCGATCCCCATACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTCCGGAAATCTAAGAATTTCACCTCTGACAACTGAATACGAATACCCCCAACTGTCCCTATTGATCATTACCTTGGTTCTCAAACCAACAAAATA
GTACCAAAGTCCTATCTTATTATTCCATGCTAATATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTCTCACAGTAAACGATGAACGGCTCTCACCCCGACAACTTAATGCCAGGATGCACC-ATTCAAGGACGG-
CATGAGAGGAC-CAAGTACTGAC-----

TAGAAGTCAGACTGATCCCTCACGCCGGAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATCCTCGATGTGGATTCTAAAATTCTCTCATTCCAAAAGCCAGACCCGAAAGGGCCCGGCATTGTTATTTATTGTCA
CTACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGGGATCAAACCCTAATCCTCCGTTACCCG
TTAAAGCCACGGTAGTCCAATACACTACCGTC--TAGCTGATAGGGCAGAAACTTGAATGGTCCATCGCC-GGCA-
GAAGCCATGCGATTCGCAAAGTTATTATGAATCACCATT-ACATCGCCGAA--ACGAGTTGGTTTCAATCTAATAAATACCACTCTTCCAA---
AAGTCGAGTGTTTATGCATGTATTAGCTCTAGAATTACTACGGGTATCCAAATAGTAAGGGACTACCAAATAAACTATAACTGATATAATGAGCCATTCGCAG-
TTTCAAAG--TAAAA-GA-TTATACTTACACATGCATGGCTTAATCTTTA

BE_AeN706-22

GAAACCTTGTTA-CGACTTCACCTTCCTMTAAGTGAT-AAGGTTTAGACAGGTTCTCGTGGACAAACTGCAATAAAGTAGATT-GCCACAATYCCGAG-
TCTTCACCGGACCACTCAATCGGTAGGTRCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCATGATTAATAATTGCAATAATCAATCCCTATCACGA
TGCATGCTAAAAAGATTTCCCAGGCCTCTCGGCCAAGGT--
GAAGCTCGTTGAATGCATCAGTGTAACACGCGTGCGGTTCMGAACMTCTAAGGGCATCACMGRCCTGTTATTGCCCCAATCTTCCTGTAGTTT-
TATACTACATGTCCCTCTAAGAAACCA-TC-GTCAATG-
AAAAATCATTAGCGTGATTATTTAGCAGGCTGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
ATAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCAGACGAGGCAAAAATGCA--
CCCGCCCATCCTGAGTCGGCATAGTTTATGGTATAGACTACGATGGTATCTAATCATCTTCGATCCCCATACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTCCGGAAATCTAAGAATTTCACCTCTGACAACTGAATACGAATACCCCCAACTGTCCCTATTGATCATTACCTTGGTTCTCAAACCAACAAAATA
GTACCAAAGTCCTATCTTATTATTCCATGCTAATATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTCTCACAGTAAACGATGAACGGCTATCACCCCGACAACTTAATGCCAGGATGCACC-ATTCAAGGACGG-
CATGAGAGGAT-CAAGTACTGAC-----
TAGAAGTCAGACTGATCCCTCACGCCGGAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATCCTCGATGTGGATTCTAAAATTCTCTCATTCCAAAAGCCAGACCCGAAAGGGCCCGGCATTGTTATTTATTGTCA
CTACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGGGATCAAACCCTAATCCTCCGTTACCCG
TTAAAGCCACGGTAGTCCAATACACTACCGTC--TAGCTGATAGGGCAGAAACTTGAATGGTCCATCGCC-GGCA-
GAAGCCATGCGATTCGCAAAGTTATTATGAATCACCATT-ACATCGCCGAA--ACGAGTTGGTTTCAATCTAATAAATACCACTCTTCCAA---
AAGTCGAGTGTTTATGCATGTATTAGCTCTAGAATTACTACGGGTATCCAAATAGTAAGGGACTACCAAATAGACTATAACTGATATAATGAGCCATTCGCAG-
TTTCAAAG--TAAAA-GATTTATACTTACACATGCATGGCTTAATCTTTGAGAC

BE_AeN706-24

CCTTCCTTTAAAGTGAT-AAGGTTTAGACAGTTTTTGGTGGACAAACGGCAATAAAGTAGATT-
GCCCCAATCCCGAGTTTTTCMCCGGACCACTCAATCGGTAGGTGCGACCGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAG
GAATTCCTCGTTCATGATTAATAATTGCAATAATCAATCCCTATCACGATGCATGCTAAAAAGATTTCCCAGGCCTCTCGGCCAAGGT--
GAAGCTCGTTGAATGCATCAGTGTAACACGCGTGCGGTTCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCAATCTTCCTGTAGTTT-
TATACTACATGTCCCTCTAAGAAACCA-TC-GTCAATG-
AAAAATCATTAGCGTGATTATTTAGCAGGCTGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
ATAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCAGACGAGGCAAAAATGCA--
CCCGCCCATCCTGAGTCGGCATAGTTTATGGTATAGACTACGATGGTATCTAATCATCTTCGATCCCCATACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTCCGGAAATCTAAGAATTTCACCTCTGACAACTGAATACGAATACCCCCAACTGTCCCTATTGATCATTACCTTGGTTCTCAAACCAACAAAATA
GTACCAAAGTCCTATCTTATTATTCCATGCTAATATATTCAA-CGGCATAAGCCT

BE_AeN706-26

CACCAACTAAGAACGGCCATGCACCACCACCCATAGAATCAAG-AAAGAACTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
----------GTAAGTTTTCCC--
GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-CCCCCCGGAACCCAAAGACTT-
TGATTTCTCATACGGTGCCGAAGGAGTCAAAAAACA--
ACCTCCGATCCCGAGTCGGCATAGTTTATGGTTAAGACTACGATGGTATCTAATCATCTTCGATCCCCTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTTCGGAAATCCAAGAATTTCACCTCTGACAACGAAATACGAATACCCCCAACTGTCCCTATTAATCATTACTTCGTAACGCAAACCAACAAAATA
GTTCCGAAGTCCTATCTTATTATTCCATGCTAATATATTCAA-CGGCATAAGCCTGCTTTAAACACTCTAATTTTTTCACAGTAAACGATGGGTATCCCCTGC-
CCGACAACTTAATGCCAAACAGGATCTCCCCAAGGATGGCC-AGAGACAACACAAGTTCACGCAC---
AGAGTGCGTGACCGGTCACTCCTGCCAGAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATACTCGATAAGGGATTTAAATTGTTCTCATTCCAATTGCCAGACCTGTAAAGGCCCGGCATTGTTATTTATTGTCAC
TACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGAAATCTTGCCCTAACAATTCGTTACCCGT
TAAAGCCATGGTAGTCCAATACACTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-
GGCATGGAGCCATGCGATTCGAAAAGTTATTATGAATCACCAAA-GCACCTCCGAA—
GAGGGTTGGTCTATACCTAATAAATACTACCCTTCCGCGAACAGTTGGGTATTGATGCATGT

BE_AeN706-27

GGAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTTGGACAAGTTCTCGCGGTCAGGCCCCAATGAAGGAGCCAAACCACAATCCCGAG-
TCCTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCACTGCAGTTTGATGAACTGCGATTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACG
ATGCATGTTAACAAGATTACCCAGGCCTCTCGGCCAAGGTTATATGCTCGTTGAGTGCATCAGTGTAACGCGCGTGCGGCCCAGGACATCTAAGGGCATCACAGACCTGT
TATTGCCGCCATCTTCCTTCATCTTGTAGAATGAACGTCCCTCTAAGAAGCTC-TT-GCCAATG-
AAAAACCATTAGCAGAACTATTTAGCAGGCGGCGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAACTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAAGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCCGATGCTGTCAAAAAACG--
AGCACCGATCCTGAGTCGGCATAGTTTATGGTTAAGACTACGATGGTATCTAATCATCTTCGATCCCCTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTGCAGAAATCCAAGAATTTCACCTCTGACAAYGMAATACGAATACCCCCAACTGTCCCTATTAATCATTACTTCGTAACGCAAACCAACAAAAT
AGTTCCGAAGTCCTATCTTATTATTCCATGCTAATATATTCAA-CGGCATAAGCCTGCTTTAAACACTCTAATTTTTTCACAGTAAACGATGGGTATCCCCTGC-
CCGACAACTTAATGCCAAACAGGATCTCCCCAAGGATGGCC-AGAGACAACACAAGTTCACGCAC---
AGAGTGYGTGACCGGTCRCTCCTGCCAGAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATACTCGATAAGGGATTTAAATTGTTCTCATTCCAATTGCCAGACCTGTAAAGGCCCGGCATTGTTATTTATTGTCAC
TACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGAAATCTTGCCCTAACAATTCGTTACCCGT
TAAAGCCATGGTAGTCCAATACACTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-
GGCATGGAGCCRTGCGATTCGAAAAGTTATTATGAATCACCAAA-GCACCTCCGAA--
GAGGGTTGGTCTATACCTAATAAATACTACCCTTCCGCGAACAGTTGGGTATTGATGCATGTATTAGCTCTAGAATTACTACGGGTATCCAAGTAGTAAGGGACTATCAAA
GGAACTATAACTGATATAATGAGCCGTTCGCAGTTTTCAAAG--TAAAA-GAGTTATACTTAGACATGCATGGCTTAATCTTTGAGACGAGCGTATGA

BE_AeN706-30

CGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAGTGAT-AAGGTTTAGACAGGTTCTCGTGGACAAACTGCAATAAAGTAGATT-GCCACAATCCCGAG-
TCTTCACCGGACCACTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCATGATTAATAATTGCAATAATCAATCCCTATCACGA
TGCATGCTAAAAAGATTTCCCAGGCCTCTCGGCCAAGGT--
GAAGCTCGTTGAATGCATCAGTGTAACACGCGTGCGGTTCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCAATCTTCCTGTAGTTT-
TATACTACATGTCCCTCTAAGAAACCA-TC-GTCAATG-
AAAAATCATTAGCGTGATTATTTAGCAGGCTGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
ATAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCAGACGAGGCAAAAATGCA—
CCCGCCCATCCTGAGTCGGCATAGTTTATGGTATAGACTACGATGGTATCTAATCATCTTCGATCCCCATACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTCCGGAAATCTAAGAATT

BE_AeN706-31

TACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAGTGAT-AAGGTTTAGACAGGTTCTCGTGGACAAACTGCAATAAAGTAGATT-GCCACAATCCCGAG-
TCTTCACCGGACCACTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCATGATTAATAATTGCAATAATCAATCCCTATCACGA
TGCATGCTAAAAAGATTTCCCAGGCCTCTCGGCCAAGGT--
GAAGCTCGTTGAATGCATCAGTGTAACACGCGTGCGGTTCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCAATCTTCCTGTAGTTT-
TATACTACATGTCCCTCTAAGAAACCA-TC-GTCAATG-
AAAAATCATTAGCGTGATTATTTAGCAGGCTGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
ATAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCAGACGAGGCAAAAATGCA—
CCCGCCCATCCTGAGTCGGCATAGTTTATGGTATAGACTACGATGGTATCTAATCATCTTCGATCCCCATACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTCCGGAA

BE_AeN706-34

ACCTACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAGTGAT-AAGGTTTAGACAGGTTCTCGTGGACAAACTGCAATAAAGTAGATT-GCCACAATCCCGAG-
TCTTCACCGGACCACTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCATGATTAATAATTGCAATAATCAATCCCTATCACGA
TGCATGCTAAAAAGATTTCCCAGGCCTCTCGGCCAAGGT--
GAAGCTCGTTGAATGCATCAGTGTAACACGCGTGCGGTTCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCAATCTTCCTGTAGTTT-
TATACTACATGTCCCTCTAAGAAACCA-TC-GTCAATG-
AAAAATCATTAGCGTGATTATTTAGCAGGCTGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
ATAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCAGACGAGGCAAAAATGCA—
CCCGCCCATCCTGAGTCGGCATAGTTTATGGTATAGACTACGATGGTATCTAATCATCTTCGATCCCCATACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTCCGGAAATCTAAGAATT

BE_AeN706-35

GGAAACCTTGTTA-CGACTTCACCTTCCTCTAAGTGAT-AAGGTTTAGACAGGTTCTCGTGGACAAACTGCAATAAAGTAGATT-GCCACAATCCCGAG-
TCTTCACCGGACCACTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCATGATTAATAATTGCAATAATCAATCCCTATCACGA
TGCATGCTAAAAAGATTTCCCAGGCCTCTCGGCCAAGGT--
GAAGCTCGTTGAATGCATCAGTGTAACACGCGTGCGGTTCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCAATCTTCCTGTAGTTT-
TATACTACATGTCCCTCTAAGAAACCA-TC-GTCAATG-
AAAAATCATTAGCGTGATTATTTAGCAGGCTGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
ATAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCAGACGAGGCAAAAATGCA—
CCCGCCCATCCTGAGTCGGCATAGTTTATGGTATAGACTACGATGGTATCTAATCATCTTCGATCCCCATACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTCCGGAAAT

K-AeN706-6

CTACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTCGGACAAGTTCTCGCGGTCAGGCCCCAATGAAGGAGCCAAACCACAATCCCGAG-
TCCTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCATTGCGGTTTGATGAACCGCGATTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACG
ATGCAAGTTAACAAGATTACCCAGGCCTCTCGGCCAAGGTTATATGCTCGTTGAGTGCATCAGTGTAACGCGCGTGCGGCCCAGGACATCAAAGGGCATCACAGACCTGT
TATTGCCGCCATCTTCCTTCATCTTGTAGAATGAACGTCCCTCTAAGAAGCTC-TTAACCAATC-
AAAAACGATTAGTAGAACTATTTAGCAGGCGGCGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAACTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACGGTGCCGAAGGAGTCAAAAAACA--
ACCTCCGATCCCGAGTCGGCATAGTTTATGGTTAAGACTACGATGGTATCTAATCATCTTCGATCCCCTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTTCGGAAATCCAAGAATTTCACCTCTGACAACGAAATACGAATACCCCCAACTGTCCCTATTAATCATTACTTCGTAACGCAAACCAACAAAATA
GTTCCGAAGTCCTATCTTATTATTCCATGCTAATATATTCAA-CGGCATAAGCCTGCTTTAAACACTCTAATTTTTTCACAGTAAACGATGGGTATCCCCTGC-
CCGACAACTTAATGCCAAACAGGATCTCCCCAAGGATGGCC-AGAGACAACACAAGTTCACGCAC---
AGAGTGCGTGACCGGTCACTCCTGCCAGAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATACTCGATAAGGGATTTAAATTGTTCTCATTCCAATTGCCAGACCTGTAAAGGCCCGGCATTGTTATTTATTGTCAC
TACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGAAATCTTGCCCTAACAATTCGTTACCCGT
TAAAGCCATGGTAGTCCAATACACTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-
GGCATGGAGCCATGCGATTCGAAAAGTTATTATGAATCACCAAA-GCACCTCCGAA--
GAGGGTTGGTCTATACCTAATAAATACTACCCTTCCGCGAACAGTTGGGTATTGATGCATGTATTAGCTCTAGAATTACTACGGGTATCCAAGTAGTAAGGGACTATCAAA
TAAACTATAACTGATATAATGAGCCGTTCGCAGTTTTCAAAG—TAAAATTAGTTATACTTAGACATGCATGGCTTAATCTTTGAGACGAGC

K-AeN706-7

CCTACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAATCGGT-AAAGTTCGGCGAGATTCACTCGATTGAAATCTAATAAAAGAAATCAATTAAGCTCCTGAC-
GCCTCACTTTCACGATCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCACATTGATGATGTGCGTTTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACGA

CGCAATTTAACAAGATTGCCCATACCTCTCGGTAAAGGTTATATGCTCGTTGAGTGCGTCAGTGTAACACGCGTGCGGTTCAGAACATCTAAGGGCATCACAGACCTGTTA
TTGCCGCAATCTTCCTGTAGTTT-TATACTACATGTCCCTCTAAGAAACCA-TC-GTCAATG-
AAGAATCATTAGCGTGATTATTTAGCAGGCTGCGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCAAAGAA
TTAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTTT-GTCTGGACCTG----------
ATAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCAGACGAGGTCATACAACA--
CCCGCCCATCCTGAGTCGGCATAGTTTATGGTATAGACTACGATGGTATCTAATCATCTTCGATCCCCATACTTTCGTTCTTGATCAATTAAAACATCCTTGGTAAATGCTT
TCGCAATTGTTTGTCTTTCCAAAATCTAAGAATTTCACCTCTGACAAGAAAATACAAATACCCCCAACTGTCCATGTTCATCATTACTTTGGTGCGCAAACCAACAAAATA
GTACCGAAGTCTTATCTTATTATTCCATGCTAATACATTCAA-
CGGCCTAAGCCTGCTTTGAACACTCTAATTTTCTCACAGTAAACGATGAGCGGCTAACACTCCGACAACTTAATGCCAGAATGGACC-ACTCAAAGATGG-
AACAGAAGAGT-CGAATACCAAC-----
CTAAGGCTGGATTGACTTTCCGTTCCAGAAATGCGACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTACTGAAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTTCAGTTGATCCTCGATAGGGGATTTAAATATTTCTCATTCCAATTGCCAGACCTAGAAAGGCCCGGCACTGTTATTTATTGTCAC
TACCTCCCTGTGTCAGGATTGGGTAATTTGCGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTTTCTAAGGCTCCCTCTCCGGAATCAAACCCTAATCCTCCGTTACCCGT
TAAAGCCATGGTAGTCCAATACACTACCATCC-AAGCTGATAGGGCAGAATCTCGAACGATCTATCGCC-GGCA-
AAAGCCATGCGATTCGCAAAGTTATTATGAATCACCATT-ACACTGCCGAA--ACAGGTTGGTTTCAATCTAATAAATACCACTCTTCCAA---
AGGTTGAGTGTTTACGCATGTATTAGCTCTAGAATTACTACGGATATCCAAGTAGTAAGAGACTATCGAATAGACTAAAACTGATATAATGAGCCGTTCGCAG-
TTTCATGG--TAAAA-TATTTTTACTTAGACATGCATGGCTTAATCTTT

K-AeN706-11

CTACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAGTGAT-AAGGTTTAGACAGGTTCTCGTGGACAAACTGCAATAAAGTAGATT-GCCACAATCCCGAG-
TCTTCACCGGACCACTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCATGATTAATAATTGCAATAATCAATCCCTATCACGA
TGCATGCTAAAAAGATTTCCCAGGCCTCTCGGCCAAGGT--
GAAGCTCGTTGAATGCATCAGTGTAACACGCGTGCGGTTCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCAATCTTCCTGTAGTTT-
TATACTACATGTCCCTCTAAGAAACCA-TC-GTCAATG-
AAAAATCATTAGCGTGATTATTTAGCAGGCTGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
ATAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCAGACGAGGCAAAAATGCA—
CCCGCCCATCCTGAGTCGGCATAGTTTATGGTATAGACTACGATGGTATCTAATCATCTTCGATCCCCATACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTCCGGAA

K-AeN706-15

TACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTTAGACAAGTTCTCACGGTACAATCCCAATAAAGGAACCAAACCACAATCCCGAG-
GCTTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACGA
TGCACATTCAAAAGATTGCCCAGGCCTCTCGGCCAAGGTTATAAACTCGTTGAATGCATCAGTGTAGCGCGCGTGCGGCCCAGGACATCTAAGGGCATCACAGACCTGTT
ATTGCCCCTATCCTCCTCCAAGTTG-ATCTTGGACGTCCCTCTAAGAAACCC-TA-
CCCAATGCAAAAAGCATTCGGCGGATTATTTAGCAGGCAGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCAC
CCATAGAATCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
----------GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGACTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-
TGATTTCTCATAAGGTGCTCATGGGGTCAAAAATAACACCCACAAATCCCTAGTCGGCATAGTTTATGGTTAAGACTACGATGGTATCTAATCATCTTCGATCCCCTAACT
TTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTTTCGCAGTAGTTCGTCTTCCAGAAATCCAAGAATTTCACCTCTGACAATTGAATACGAATACCCCCAACTGTCC
CTCTTAATCATTACCTCGGTGCGCAAACCAACAAAATAGTACCAAGGTCCTATCTTATTATTCCATGCTAATATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTTTCACAGTAAACGATGGGCAGCCACTGC-CCGACAACTAAATGCCAGACAGCTTAGACCCAAGGATGG-
CAAGAGACAGG-AAAGTACCAACAC---
AAAGTGCCGGACCACCCGCTCCCGCCAGAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATCCTCGTTAAGGGATTTAAATTGTTCTCATTCCAATTGCTAGACCTAGAAAGGCCCGGCATTGTTATTTATTGTCAC
TACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCATGCTCCCTCTCCGGAATCAAACCCTAATTTTCCGTTACCCGTT
AAAGCCATGGTAGGCCAATACCCTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-
GGCACAAAGCCATGCGATTCGCGAAGTTATTATGAATCACCACA-TCACCTCCGAA--AAGGGTTGGTTTCTATCTAATAAATACTACCCGTCCACAA-
GAGTTGGGTATTGATGCATGTATTAGCTCTAGAATTACTACGGTTATCCAAGTAGTAAGGGACTATCAAATAAACTATAACTGATATAATGAGCCATTCGCAG-
TTTCAAAG--TAAAA-GATTTATACTTAGACATGCATGGCTTAATCTTTGAGACK-18_UIO 595: GAAACCTTGTTA-CGACTTCACCTTCCTCTAAGTGAT-
AAGGTTTAGACAGGTTCTCGTGGACAAACTGCAATAAAGTAGATT-GCCACAATCCCGAG-TCTTCACCGGACCACTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCATGATTAATAATTGCAATAATCAATCCCTATCACGA
TGCATGCTAAAAAGATTTCCCAGGCCTCTCGGCCAAGGT--
GAAGCTCGTTGAATGCATCAGTGTAACACGCGTGCGGTTCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCAATCTTCCTGTAGTTT-
TATACTACATGTCCCTCTAAGAAACCA-TC-GTCAATG-
AAAAATCATTAGCGTGATTATTTAGCAGGCTGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
ATAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCAGACGAGGCAAAAATGCA--
CCCGCCCATCCTGAGTCGGCATAGTTTATGGTATAGACTACGATGGTATCTAATCATCTTCGATCCCCATACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTCCGGAAATCTAAGAATTTCACCTCTGACAACTGAATACGAATACCCCCAACTGTCCCTATTGATCATTACCTTGGTTCTCAAACCAACAAAATA
GTACCAAAGTCCTATCTTATTATTCCATGCTAATATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTCTCACAGTAAACGATGAACGGCTCTCACCCCGACAACTTAATGCCAGGATGCACC-ATTCAAGGACGG-
CATGAGAGGAC-CAAGTACTGAC-----
TAGAAGTCAGACTGATCCCTCACGCCGGAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATCCTCGATGTGGATTCTAAAATTCTCTCATTCCAAAAGCCAGACCCGAAAGGGCCCGGCATTGTTATTTATTGTCA
CTACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGGGATCAAACCCTAATCCTCCGTTACCCG
TTAAAGCCACGGTAGTCCAATACACTACCGTC--TAGCTGATAGGGCAGAAACTTGAATGGTCCATCGCC-GGCA-
GAAGCCATGCGATTCGCAAAGTTATTATGAATCACCATT-ACATCGCCGAA--ACGAGTTGGTTTCAATCTAATAAATACCACTCTTCCAA---
AAGTCGAGTGTTTATGCATGTATTAGCTCTAGAATTACTACGGGTATCCAAATAGTAAGGGACTACCAAATAAACTATAACTGATATAATGAGCCATTCGCAG-
TTTCAAAG--TAAAA-GATTTATACTTACACATGCATGGCTTAATCTTTGAGACGAGCGATGACTACTG

K-AeN706-19

GGAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTTAGACAAGTTCTCGCGACTAACTTCCAATAAAGGAAACTAACCACAATCCCGAG-
GCTTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACGA
TGCAAGTTCAAAAGATTGCCCAGGCCTCTCGGCCAAGGT-
AAGAACTCGTTGAATGCATCAGTGTAGCGCGCGTGCGGCCCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCTATCTTCCTGCGTCTAATAGAACGCACGTCCCT
CTAAGAAGCAC-AA-TGCAGTG--

ATAAACACTACACGTGCTATTTAGCAGGCAGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAAT
CAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTTATGATTTCTCACAAGGTGCTGACGGAGACGAAACGAAACTCCGCCAATCCCTTGTCGGCATAGTTTATGGTTAAGACTACGATGGTATC
TAATCATCTTCGATCCCCTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTTTCGCAGTAGTTTGTCTTCCAAAAATCCAAGAATTTCACCTCTGACAATGG
AATACAAATACCCCCAACTGTCCCTCTTAATCATTACCTCAGTGCGCAAACCAACAAAATAGTACTAAGGTCCTATATTATTATTCCATGCTAATATATTCAA--
GGCATAAGCCTGCTTTGAACACTCTAATTTTTTCACAGTAAACGATGGGCATCCCCTGC-ACGACAACTTAATGCCACACAGATTCCACCCAAACATGG-C--AGACCAGT-
TCAGCACTATCACTCAAAGAGCAAAGGCCGACTGGACACACCACAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATACTCGATAAGGGGTTTAAATTGTTCTCATTCCAATTGCCAGACCTAAGAAGGCCCGGCATTGTTATTTATTGTCA
CTACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGGAATCAAACCCTAATTTCCCGTTACCCG
TTAAAGCCATGGTAGGCCAATACCCTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-GCC--
AGAGGCATGCGATCCGCAAGTTTATTATGAATCACCACA-TCACCCC--GA--AGGGGTTGGTTTCAATCTAATAAATACTACCC---------
CAGAAGGGTATTGACGCATGTATTAGCTCTAGAATTACTACGGGTATCCAAGTAGTAAGGGACTATCAAATAAACTATAACTGATATAATGAGCCGTTCGCAG-
TTTCAAAG--TAAAA-TATTTATACTTAGACATGCATGGCTTAAT

K-AeN706-22

GGAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTTAGACAAGTTCTCGCGACAAACTCCCAATAAAGGGAACTCGCCACAATCCCGAG-
GCTTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACGA
TGCACGTTCAAAAGATTGCCCAGGCCTCTCGGCCAAGGA--
AAAACTCGTTGAATGCATCAGTGTAGCGCGCGTGCGGCCCAGGACATCTAAGGGCATCACAGACCTGTTATTGCCCCTATCTTCCTGCATTTAA-
AGAATGCACGTCCCTCTAAGAAGTTT-CA-CCCAGTA-
AAAAATCACTCGGCAAACTATTTAGCAGGCAGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTTGTGATTTCTCATAAGGTGCTGACGGAGACAAAACGAGACTCCGCCAATCCCTTGTCGGCATAGTTTATGGTTAAGACTACGATGGTATCT
AATCATCTTCGATCCCCTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTTTCGCAGTAGTTCGTCTTCCAAAAATCCAAGAATTTCACCTCTGACAATGGA
ATACGAATACCCCCAACTGTCCCTATTAATCATTACCTTGGTGCGCAAACCAACAAAATAGTACCAAGGTCCTATCTTATTATTCCATGCTAATATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTTTCACAGTAAACGATGGGCATCCCCTGC-ACGACAACCTAATGCCACACAGGTTCCACCCAAGGATGG-C--GACGCAAG-
CAAACACTAACAC---TAGGTGTCAGGCCGCACGCGACAGCCACAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATACTCGATAAGGGGTTTAAATTGTTCTCATTCCAATTGCCAGACCTATAAAGGCCCGGCATTGTTATTTATTGTCAC
TACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGGAATCAAACCCTAATTTCCCGTTACCCGT
TAAAGCCATGGTAGGCCAATACCCTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-GGCA-
AAAGCCATGCGATCCGCAAGCTTATTATGAATCACCACA-TCACCCC--GA--AGGGGTTGGTTTCAATCTAATAAATACTACCC---------
CAGAAGGGTATTGACGCATGTATTAGCTCTAGAATTACTACGGTTATCCAAGTAGTAAGGGACTATCAAATAAACTATAACTGATATAATGAGCCGTTCGCAG-
TTTCAAAG--TAAGA-GATTTATACTTAGACATGCATGGCTTAATCTT-GAGAC

K-AeN706-26

CCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTTAAACGAGTTCTACTAAACACATACCTAAAACAGCACATATCTAATATCCCAGG-
TTTTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAGCACGCTTTGATGAAGCGCACTTACTAGATATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACG
ACGCACATTTATAAGATTACCCAAGCCTCTCGGCTAAGGTTATAAACTTGTTGAATGCATCAGTGTAGCGCGCGTGCCGTCCAGAACATCTAAGGGCATCACAGACCTGTT
ATTGCCCCTAACTTCCTTTCGTTAGTCACG-AAAAGTTCCTCTAAGAAGCTA-AA-AACAATG-
AACAAACATTATATTAGCTATTTAGCAGGCAGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCAATCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAACTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCATGCTCCACATCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGACTTGCGACCATACT-
CCCCCCGGAACCCAAAAACTT-TGATTTCTCGGAAGGTGCGGAAGGTGTTGTTT-
GCACAACCACCCATCCCTAGTTGGCATAGTTTATGGTTAAGACTACGATGGTATCTAATCATCTTCGATCCCTTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAA
TGCTTTCGCAGTTGTTCGTCTTCCAAAAATCCAAGAATTTCACCTCTAACATTGGAATACGAATACCCCCAACTATCCCTATTAATCATTACCTAGGTGCATAAACCAACA
AAGTAGTACCTAAGTCCTATTTCATTATTCCATGCTAATACATTCAA-AGCTAAATGCTTACTTTGAACGTTCTGATTTTTTCACAGTAAAAGATGAATATCTTTAGT-
TCGACAACTTAATGCCAAACTAAATCTCTTCAAGGATGG-TTGAAGATAGCAACTTCATAGGTC----
AAAGACCCAAGGTCGCTACTTCTACCAGAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATCCTCGATATGAGATTTAAATGGTTCTCATTCCAATTGCCAGAGC-
AAGATGCCCCAGCATTGTTATTTATCGTCACTACCTCCCTGTATCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTTTCTCAGGCTCCCTCTC
CGGGGTCAAGCCCTAACCCTCCGTCACCCGTAAAAGCCATGGTAGGCCAATACCCTACCATCCAAAGCTGATAGGGCGGAAACATGAATGATTTGTCGCT-GGTA-
AAAACCATGCGATCCATCAAGTTATTATGAATCACCAGACATACTTCCGAA--GAAGGTTGGTTTGTTTCTAATAAATACAACCGTTCCCTA--
AAGTCCGGTTTATATGCATGTATTAGCTCTAGAATTACTACGGTTATCCAAGTAGTAAGTAATTATCAAATAAGTTAAAGCCGATATAATGAGCCATTCGCAG-
TTTTAGAG--TAAAA-TTCTTGTACTTAGACATGCATGGCTTAATCTTTGA

K-AeN706-28

ACCTTGTTA-CGACTTCACCTTCCTCTAAGTGAT-AAGGTTTAGACAGGTTCTCGTGGACAAACTGCAATAAAGTAGATT-GCCACAATCCCGAG-
TCTTCACCGGACCACTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCATGATTAATAATTGCAATAATCAATCCCTATCACGA
TGCATGCTAAAAAGATTTCCCAGGCCTCTCGGCCAAGGT--
GAAGCTCGTTGAATGCATCAGTGTAACACGCGTGCGGTTCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCAATCTTCCTGTAGTTT-
TATACTACATGTCCCTCTAAGAAACCA-TC-GTCAATG-
AAAAATCATTAGCGTGATTATTTAGCAGGCTGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
ATAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCAGACGAGGCAAAAATGCA--
CCCGCCCATCCTGAGTCGGCATAGTTTATGGTATAGACTACGATGGTATCTAATCATCTTCGATCCCCATACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTCCGGAAATCTAAGAATTTCACCTCTGACAACTGAATACGAATACCCCCAACTGTCCCTATTGATCATTACCTTGGTTCTCAAACCAACAAAATA
GTACCAAAGTCCTATCTTATTATTCCATGCTAATATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTCTCACAGTAAACGATGAACGGCTCTCACCCCGACAACTTAATGCCAGGATGCACC-ATTCAAGGACGG-
CATGAGAGGAC-CAAGTACTGAC-----
TAGAAGTCAGACTGATCCCTCACGCCGGAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATCCTCGATGTGGATTCTAAAATTCTCTCATTCCAAAAGCCAGACCCGAAAGGGCCCGGCATTGTTATTTATTGTCA
CTACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGGGATCAAACCCTAATCCTCCGTTACCCG
TTAAAGCCACGGTAGTCCAATACACTACCGTC--TAGCTGATAGGGCAGAAACTTGAATGGTCCATCGCC-GGCA-
GAAGCCATGCGATTCGCAAAGTTATTATGAATCACCATT-ACATCGCCGAA--ACGAGTTGGTTTCAATCTAATAAATACCACTCTTCCAA---
AAGTCGAGTGTTTATGCATGTATTAGCTCTAGAATTACTACGGGTATCCAAATAGTAAGGGACTACCAAATAAACTATAACTGATATAATGAGCCATTCGCAG-
TTTCAAAG--TAAAA-GA-TTATACTTACCCATGCATGGCTTAATCTTTGAGA

K-AeN706-30

CTACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTTAGACAAGTTCTCGCGATCAACTCCCAATAAAGGGAACTCACCACAATCCCGAG-
GCTTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACGA
TGCACGTTCAAAAGATTGCCCAGGCCTCTCGGCCAAGGA--
AAAACTCGTTGAATACATCAGTGTAGCGCGCGTGCGGCCCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCTATCTTCCTGCGTCTAATAGAACGCACGTCCCTC
TAAGAAGCTT-CA-CCCAGTG-
AAAATTCACTCGGCAAGCTATTTAGCAGGCAGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTTGTGATTTCTCATAAGGTGCTGACGGAGACAAAACGAGACTCCGCCAATCCCTTGTCGGCATAGTTTATGGTTAAGACTACGATGGTATCT
AATCATCTTCGATCCCCTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTTTCGCAGTAGTTCGTCTTCCAGAAATCCAAGAATTTCACCTCTGACAATGGA
ATACGAATACCCCCAACTGTCCCTATTAATCATTACCTTGGTGCGCAAACCAACAAAATAGTACCAAGGTCCTATCTTATTATTCCATGCTAATATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTTTCACAGTAAACGATGGGCATCCCCTGC-ACGACAACCTAATGCCACACAGGTTCCACCCAAGGATGG-C--GACGCTAG-
CAAGCACCGGCAC----GAATGCCGGGCCGCACGCGACTGCCACAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATACTCGATAAGGGGTTTAAATTGTTCTCATTCCAATTGCCAGACCTATAAAGGCCCGGCATTGTTATTTATTGTCAC
TACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGGAATCAAACCCTAATTTCCCGTTACCCGT
TAAAGCCATGGTAGGCCAATACCCTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-GGCA--
AAGCCATGCGATCCGCAAGCTTATTATGAATCACCACA-TCACCCC--GA--AGGGGTTGGTTTCAATCTAATAAATACTACCC---------
CAGAAGGGTATTGACGCATGTATTAGCTCTAGAATTACTACGGTTATCCAAGTAGTAAGGGACTATCAAATAAACTATAACTGATATAATGAGCCGTTCGCAG-
TTTCAAAG--TAAAA-GATTTATACTTAGACATGCATGGCTTAA

K-AeN706-31

TTCACCTACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAGATGAT-AAGGTTTAGACAAGTTCTCGCAACCAA-CACCAATAAAGGCA-CTGGTCACAATCCCGAG-
GCTTCACCGGATCATCCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACGA
TACACGTTCAAAAGATTGCCCAGGCCTCTCGGCCAAGG----
AAACTCGTTGAATGCATCAGTGTAGCGCGCGTGCGGCCCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCTATCTTCCTGCATCTAATAGAATGCACGTCCCTCT
AAGAAGCTC-GA-AGCAGTG--
AAAATCACTCCAAGAGCTATTTAGCAGGCAGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAAT
CAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-
GTCTGGACCTGAAGGAATACAAAAGGCGGGAAGGATAGAGAGTTAGATGCGTGTTGGGGAAGTGAAAGCTGTTTGGAMAGGGTGTTCCTTCCTTCAGTAGTAGAGCAGA
ATAGAATAGATGCYATGCTTACCTGGTAAGTTTTCCC--
GTGTTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAAACTTATGATTTCTCATAAGGTGCTGATGGAGACGTAACGAAACCCCACCAATCCCTTGTCGGCATAGTTTATGGTTAAGACTACGATGGTATCT
AATCATCTTCGATCCCCTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTTTCGCATTAGTTCGTCTTCCAGAAATCCAAGAAT

K-AeN706-32

GAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTTAGACAAGTTCTCGCGACAAACTCCCAATAAAGGAAACTCGCCACAATCCCGAG-
GCTTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACGA
TGCACGTTCAAAAGATTGCCCAGGCCTCTCGGCCAAGG----
AAACTCGTTGAATGCATCAGTGTAGCGCGCGTGCGGCCCAGGACATCTAAGGGCATCACAGACCTGTTATTGCCCCTATCTTCCTGCSTCTAATAGAAMGCACGTCCCTCT
AAGAAGCWT-AW-CCCMGTG-
AACATTCACTCGGCAAGCTATTTAGCAGGCAGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTTGTGATTTCTCATAAGGTGCTGACGGAGACAAAACGAAACTCCGCCAATCCCTTGTCGGCATAGTTTATGGTTAAGACTACGATGGTATCT
AATCATCTTCGATCCCCTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTTTCGCAGTAGTTCGTCTTCCAAAAATCCAAGAATTTCACCTCTGACAATGGA
ATACGAATACCCCCAACTGTCCCTATTAATCATTACCTTGGTGCGCAAACCAACAAAATAGTACCAAGGTCCTATCTTATTATTCCATGCTAACATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTTTCACAGTAAACGATGGGCATCCCCTGC-ACGACAACCTAATGCCACACAGGWTCCACCCAAGGATGG-C--
GACGCAAG-CAAACACCAACAC---
TAGGKGTCGGGTCGCACGCGACAGCCACAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAKASTTGCCCTCCWRTTGATCCTCGATAAGGGATTTAARTTGTTCTCMTTCCWRTTGCCASACCTATAAAGGCCCGGYRTTGTTATTTATTGTCA
CTACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAKGCTCCCTCTCCGGAATCAAACCCTAATTTCCCGTTACCCG
TTAAAGCCATGGTAGGCCAATACCCTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-GGCG-
TGAGCCATGCGATCCGCAAGCTTATTATGAATCACCACA-TCACCCC--GA--AGGGGTTGGTTTCAATCTAATAAATACTACCC---------
CAGAAGGGTATTGACGCATGTATTAGCTCTAGAATTACTACGGTTATCCAAGTAGTAAGGGACTATCAAATAAACTATAACTGATATAATGAGCCGTTCGCAG-
TTTCAAAG--TAAAA-GATTTATACTTAGACATGCATGGCTTAATCTTTGAGACGAGCGTATGACT

K-AeN706-33

TACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTTAGACAAGTTCTCGCGATCAACTCCCAATAAAGGGAACTCACCACAATCCCGAG-
GCTTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACGA
TGCACGTTCAAAAGATTGCCCAGGCCTCTCGGCCAAGGA--
AAAACTCGTTGAATACATCAGTGTAGCGCGCGTGCGGCCCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCTATCTTCCTGCGTCTAATAGAACGCACGTCCCTC
TAAGAAGCTT-CA-CCCAGTG-
AAAATTCACTCGGCAAGCTATTTAGCAGGCAGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTTGTGATTTCTCATAAGGTGCTGACGGAGACAAAACGAGACTCCGCCAATCCCTTGTCGGCATAGTTTATGGTTAAGACTACGATGGTATCT
AATCATCTTCGATCCCCTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTTTCGCAGTAGTTCGTCTTCCAGAAATCCAAGAATTTCACCTCTGACAATGGA
ATACGAATACCCCCAACTGTCCCTATTAATCATTACCTTGGTGCGCAAACCAACAAAATAGTACCAAGGTCCTATCTTATTATTCCATGCTAATATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTTTCACAGTAAACGATGGGCATCCCCTGC-ACGACAACCTAATGCCACACAGGTTCCACCCAAGGATGG-C--GACGCTAG-
CAAGCACCGGCAC----GAATGCCGGGCCGCACGCGACTGCCACAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATACTCGATAAGGGGTTTAAATTGTTCTCATTCCAATTGCCAGACCTATAAAGGCCCGGCATTGTTATTTATTGTCAC
TACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGGAATCAAACCCTAATTTCCCGTTACCCGT
TAAAGCCATGGTAGGCCAATACCCTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-GGCA--
AAGCCATGCGATCCGCAAGCTTATTATGAATCACCACA-TCACCCC--GA--AGGGGTTGGTTTCAATCTAATAAATACTACCC---------
CAGAAGGGTATTGACGCATGTATTAGCTCTAGAATTACTACGGTTATCCAAGTAGTAAGGGACTATCAAATAAACTATAACTGATATAATGAGCCGTTCGCAG-
TTTCAAAG--TAAAA-GATTTATACTTAGACATGCATGGCTTAATCTTTGAGACGAGCG

K-AeN706-34

CCTACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAGTGAT-AAGGTTTAGACAGGTTCTCGTGGACAAACTGCAATAAAGTAGATT-GCCACAATCCCGAG-
TCTTCACCGGACCACTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCATGATTAATAATTGCAATAATCAATCCCTATCACGA
TGCATGCTAAAAAGATTTCCCAGGCCTCTCGGCCAAGGT--
GAAGCTCGTTGAATGCATCAGTGTAACACGCGTGCGGTTCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCAATCTTCCTGTAGTTT-
TATACTACATGTCCCTCTAAGAAACCA-TC-GTCAATG-
AAAAATCATTAGCGTGATTATTTAGCAGGCTGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
ATAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCAGACGAGGCAAAAATGCA--
CCCGCCCATCCTGAGTCGGCATAGTTTATGGTATAGACTACGATGGTATCTAATCATCTTCGATCCCCATACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTCCGGAAATCTAAGAATTTCACCTCTGACAACTGAATACGAATACCCCCAACTGTCCCTATTGATCATTACCTTGGTTCTCAAACCAACAAAATA
GTACCAAAGTCCTATCTTATTATTCCATGCTAATATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTCTCACAGTAAACGATGAACGGCTCTCACCCCGACAACTTAATGCCAGGATGCACC-ATTCAAGGACGG-
CATGAGAGGAC-CAAGTACTGAC-----
TAGAAGTCAGACTGATCCCTCACGCCGGAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATCCTCGATGTGGATTCTAAAATTCTCTCATTCCAAAAGCCAGACCCGAAAGGGCCCGGCATTGTTATTTATTGTCA
CTACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGGGATCAAACCCTAATCCTCCGTTACCCG
TTAAAGCCACGGTAGTCCAATACACTACCGTC--TAGCTGATAGGGCAGAAACTTGAATGGTCCATCGCC-GGCA-
GAAGCCATGCGATTCGCAAAGTTATTATGAATCACCATT-ACATCGCCGAA--ACGAGTTGGTTTCAATCTAATAAATACCACTCTTCCAA---
AAGTCGAGTGTTTATGCATGTATTAGCTCTAGAATTACTACGGGTATCCAAATAGTAAGGGACTACCAAATAAACTATAACTGATATAATGAGCCATTCGCAG-
TTTCAAAG--TAAAA-GATTTATACTTACACATGCATGGCTTAATCTTTGAGACGAGCG

K-AeN706-35

CTACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTTAGACAAGTTCTCGCGACAAATCCCCAATAAAGGAAACTAATCACAATCCCGAG-
GCTTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGACTGATGATCTGCGTTTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACG
ATGCACGTTC-ACAGATTTCCCAGGCCTCTCGGCCAAGGT---
ACACTCGTTGCATGCATCAGTGTAGCGCGCGTGCGGCCCAGGACATCTAAGGGCATCACAGACCTGTTTTTGCCCGTATCTTCCTGCGTTTTATAAAACGCACGTCCCTCT
AAGAAGTTG-TA-GCCAATG-
AAAATTCATTAACCCAACTATTTAGCAGGCACGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTTGTGATTTCTCATAAGGTGCTGACGGAGTCAATACAACG-
ACCGCCAATCCCTTGTCGGCATAGTTTATGGTTAAGACTACGATGGTATCTAATCATCTTCGATCCCCTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTCCAAAAATCCAAGAATTTCACCTCTGACAATGGAATACGAATACCCCCAACTGTCCCTATTAATCATTACCTCGGTGCGCAAACCAACAAAATA
GTACCAAGGTCCTATCTCATTATTCCATGCTAATATATTCAA-CGGCATAAGCCTGCTTTGAACACTCTAATTTTTTCACAGTAAACGATGGGCATCCCCTGC-
ACGACAACTTAATGCCACACAGATTCCACCCAAGGATGGAC--GAGTCCAT-CAAGTACCCGCAC---
AATGTGCCGGACCGACGAACCCAACCACAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGGTCCTCGATAAGGGGTTTAAATTGTTCTCATTCCAATTGCCAGACCTTTAAAGGCCCGGCATTGTTATTTATTGTCAC
TACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGGAATCAAACCCTAATTTCCCGTTACCCGT
TAAAGCCATGGTAGGCCAATACCCTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-
AGCATAAAGCCGTGCGATTCGCAAACTTATTATGATTCACCACA-TCACCCCGCAAGCGGAGGTTGGTTTCATTCTAATAAATACCACCC----------
AAAGGTGTTGACGCATGTATTAGCTCTAGAATTACTACGGTTATCCAAGTAGTAAGGGACTATCAAATAAACTATAACTGATATAATGAGCCGTTCGCAG-TTTCAAAG--
TAAAA-GATTTATACTTAGACATGCATGGCTTAATCTTTGAGACGAGCGTATGACTACTGG

K-AeN706-36

TTCACCTACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTTAGACAAGTTCTCACGGTACAATCCCAATAAAGGAACCAAACCACAATCCCGAG-
GCTTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACGA
TGCACATTCAAAAGATTGCCCAGGCCTCTCGGCCAAGGTTATAAACTCGTTGAATGCATCAGTGTAGCGCGCGTGCGGCCCAGGACATCTAAGGGCATCACAGACCTGTT
ATTGCCCCTATCCTCCTCCAAGTTG-ATCTTGGACGTCCCTCTAAGAAACCC-TA-
CCCAATGCAAAAAGCATTCGGCGGATTATTTAGCAGGCAGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCAC
CCATAGAATCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
----------GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGACTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-
TGATTTCTCATAAGGTGCTCATGGGGTCATAAATAACACCCACAAATCCCTAGTCGGCATAGTTTATGGTTAAGACTACGATGGTATCTAATCATCTTCGATCCCCTAACTT
TCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTTTCGCAGTAGTTCGTCTTCCAGAAATCCAAGAATTTCACCTCTGACAATTGAATACGAATACCCCCAACTGTCCC
TCTTAATCATTACCTCGGTGCGCAAACCAACAAAATAGTACCAAGGTCCTATCTTATTATTCCATGCTAATATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTTTCACAGTAAACGATGGGCAGCCACTGC-CCGACAACTAAATGCCAGACAGCTTAGACCCAAGGATGG-
CAAGAGACAGG-AAAGTACCAACAC---
TAAGTGCCGGACCACCCGCTCCCGCCAGAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATCCTCGTTAAGGGATTTAAATTGTTCTCATTCCAATTGCTAGACCTAGAAAGGCCCGGCATTGTTATTTATTGTCAC
TACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCATGCTCCCTCTCCGGAATCAAACCCTAATTTTCCGTTACCCGTT
AAAGCCATGGTAGGCCAATACCCTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-
GGCACAAAGCCATGCGATTCGCGAAGTTATTATGAATCACCACA-TCACCTCCGAA--AAGGGTTGGTTTCTATCTAATAAATACTACCCGTCCACAA-
GAGTTGGGTATTGATGCATGTATTAGCTCTAGAATTACTACGGTTATCCAAGTAGTAAGGGACTATCAAATAAACTATAACTGATATAATGAGCCATTCGCAG-
TTTCAAAG--TAAAA-GATTTATACTTAGACATGCATGGCTTAATCTTTGAGACGAGCGTATGACTACTG

K-AeN706-37

CTACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTTAGACAAGTTCTCACGGTACAATCCCAATAAAGGAACCAAACCACAATCCCGAG-
GCTTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACGA
TGCACATTCAAAAGATTGCCCAGGCCTCTCGGCCAAGGTTATAAACTCGTTGAATGCATCAGTGTAGCGCGCGTGCGGCCCAGGACATCTAAGGGCATCACAGACCTGTT
ATTGCCCCTATCCTCCTCCAAGTTG-ATCTTGGACGTCCCTCTAAGAAACCC-TA-
CCCAATGCAAAAAGCATTCGGCGGATTATTTAGCAGGCAGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCAC
CCATAGAATCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
----------GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGACTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-
TGATTTCTCATAAGGTGCTCATGGGGTCATAAATAACACCCACAAATCCCTAGTCGGCATAGTTTATGGTTAAGACTACGATGGTATCTAATCATCTTCGATCCCCTAACTT
TCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTTTCGCAGTAGTTCGTCTTCCAGAAATCCAAGAATTTCACCTCTGACAATTGAATACGAATACCCCCAACTGTCCC
TCTTAATCATTACCTCGGTGCGCAAACCAACAAAATAGTACCAAGGTCCTATCTTATTATTCCATGCTAATATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTTTCACAGTAAACGATGGGCAGCCACTGC-CCGACAACTAAATGCCAGACAGCTTAGACCCAAGGATGG-
CAAGAGACAGG-AAAGTACCAACAC---

TAAGTGCCGGACCACCCGCTCCCGCCAGAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATCCTCGTTAAGGGATTTAAATTGTTCTCATTCCAATTGCTAGACCTAGAAAGGCCCGGCATTGTTATTTATTGTCAC
TACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCATGCTCCCTCTCCGGAATCAAACCCTAATTTTCCGTTACCCGTT
AAAGCCATGGTAGGCCAATACCCTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-
GGCACAAAGCCATGCGATTCGCGAAGTTATTATGAATCACCACA-TCACCTCCGAA--AAGGGTTGGTTTCTATCTAATAAATACTACCCGTCCACAA-
GAGTTGGGTATTGATGCATGTATTAGCTCTAGAATTACTACGGTTATCCAAGTAGTAAGGGACTATCAAATAAACTATAACTGATATAATGAGCCATTCGCAG-
TTTCAAAG--TAAAA-GA-TTATACTTAGACATGCATGGCTTAATCTTTGAGACGAGCGTATGACTACTGG----------A

K-AeN706-38

CCTACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAGTGAT-AAGGTTTAGACAGGTTCTCGTGGACAAACTGCAATAAAGTAGATT-GCCACAATCCCGAG-
TCTTCACCGGACCACTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCATGATTAATAATTGCAATAATCAATCCCTATCACGA
TGCATGCTAAAAAGATTTCCCAGGCCTCTCGGCCAAGGT--
GAAGCTCGTTGAATGCATCAGTGTAACACGCGTGCGGTTCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCAATCTTCCTGTAGTTT-
TATACTACATGTCCCTCTAAGAAACCA-TC-GTCAATG-
AAAAATCATTAGCGTGATTATTTAGCAGGCTGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
ATAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACAGTGCAGACGAGGCAAAAATGCA--
CCCGCCCATCCTGAGTCGGCATAGTTTATGGTATAGACTACGATGGTATCTAATCATCTTCGATCCCCATACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTCCGGAAATCTAAGAATTTCACCTCTGACAACTGAATACGAATACCCCCAACTGTCCCTATTGATCATTACCTTGGTTCTCAAACCAACAAAATA
GTACCAAAGTCCTATCTTATTATTCCATGCTAATATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTCTCACAGTAAACGATGAACGGCTCTCACCCCGACAACTTAATGCCAGGATGCACC-ATTCAAGGACGG-
CATGAGAGGAC-CAAGTACTGAC-----
TAGAAGTCAGACTGATCCCTCACGCCGGAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATCCTCGATGTGGATTCTAAAATTCTCTCATTCCAAAAGCCAGACCCGAAAGGGCCCGGCATTGTTATTTATTGTCA
CTACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGGGATCAAACCCTAATCCTCCGTTACCCG
TTAAAGCCACGGTAGTCCAATACACTACCGTC--TAGCTGATAGGGCAGAAACTTGAATGGTCCATCGCC-GGCA-
GAAGCCATGCGATTCGCAAAGTTATTATGAATCACCATT-ACATCGCCGAA--ACGAGTTGGTTTCAATCTAATAAATACCACTCTTCCAA---
AAGTCGAGTGTTTATGCATGTATTAGCTCTAGAATTACTACGGGTATCCAAATAGTAAGGGACTACCAAATAAACTATAACTGATATAATGAGCCATTCGCAG-
TTTCAAAG--TAAAA-GATTTATACTTACACATGCATGGCTTAATCTTTGAGACGAGCG

K-AeN706-45

CTACGGAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTTAGACAGGTTCTCGTGGTCAGATTCCAATAAAGGAAACCAACCACAATCCCGAG-
GCTTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAATGCAGATTGATGATCTGCGTTTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACGA
TGCAGGTTCAAAAGATTGCCCAGGCCTCTCGGCCAAGGT--
AGAACTCGTTGAATGCATCAGTGTAGCGCGCGTGCGGCCCAGAACATCTAAGGGCATCACAGACCTGTTATTGCCCCTATCTTCCTGCGTCTAATAGAACGCACGTCCCTC
TAAGAAGCAC-TA-CACAGTG--
ATAAACACTATGCGTGCTATTTAGCAGGCAGGGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAAT
CAAG-AAAGAGCTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTTATGATTTCTCACAAGGTGCTGACGGAGACGAAACGAGACTCCGCCAATCCCTTGTCGGCATAGTTTATGGTTAAGACTACGATGGTATC
TAATCATCTTCGATCCCCTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTTTCGCAGTAGTTTGTCTTCCAAAAATCCAAGAATTTCACCTCTGACAATGG
AATACAAATACCCCCAACTGTCCCTCTTAATCATTACCTTAGTGCGCAAACCAACAAAATAGTACCAAGGTCCTATATCATTATTCCATGCTAATATATTCAA-
CGGCATAAGCCTGCTTTGAACACTCTAATTTTTTCACAGTAAACGATGGGCATCCCCTGC-ACGACAACCTAATGCCACACAGATTCCACCCAAACATGG-C--AGACCAGT-
ACAGCACAATCACTCAAAGAGCAAAGGCCGACTGGACACACCACAAATCCAACTACGAGCTTTTTAACTGCAACAACTTTAATATACGCTATTGGAGCTGGAATTAC-
CGCGGCTGCTGGCACCAGACTTGCCCTCCAATTGATACTCGATAAGGGGTTTAAATTGTTCTCATTCCAATTGCCAGACCTAAGAAGGCCCGGCATTGTTATTTATTGTCA
CTACCTCCCTGTGTCAGGATTGGGTAATTTACGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTCTCTCAGGCTCCCTCTCCGGAATCAAACCCTAATTTCCCGTTACCCG
TTAAAGCCATGGTAGGCCAATACCCTACCATCCAAAGCTGATAGGGCAGAAACTTGAATGATCCATCGCC-GCC--
AGAGGCATGCGATCCGCAAGCTTATTATGAATCACCACA-TCACCCC--GA--AGGGGTTGGTTTCAATCTAATAAATACTACCC---------
CAGAAGGGTATTGACGCATGTATTAGCTCTAGAATTACTACGGGTATCCAAGTAGTAAGGGACTATCAAATAAACTATAACTGATATAATGAGCCGTTCGCAG-
TTTCAAAG--TAAAA-AATTTATACTTAGACATGCATGGCTTAATCTTTGAGACGAGCGTATGACTACTG

K-AeN706-56

GAAACCTTGTTA-CGACTTCACCTTCCTCTAAATGAT-AAGGTTCGGACAAGTTCCCGCAGTCTGGCCCCAATGAAGGAGCCAAACCACAGTCCCGAG-
TCCTCACCGGACCATTCAATCGGTAGGTGCGA-
CGGGCGGTGTGTACAAAGGGCAGGGACGTAATCATTGCAGTTTGATGAACTGCGATTACTAGGAATTCCTCGTTCAAGATTAATAATTGCAATAATCTATCCCTATCACGA
TGCATGTTAACAAGATTACCCAGGCCTCTCGGCCAAGGTTATATGCTCGTTGAGTGCATCAGTGTAACGCGCGTGCGGCCCAGGACATCTAAGGGCATCACAGACCTGTT
ATTGCCGCCATCTTCCTTCATCTTGTAGAATGAACGTCCCTCTAAGAAGCTC-TT-ACCAATG-
AAAAACCATTAGTAGAGCTATTTAGCAGGCGGCGGTCTCGTTCGTTAACGGAATTAACCAGACAAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAA
TCAAG-AAAGAACTCTCAATCTGTCAATCCTCACTAT-GTCTGGACCTG----------
GTAAGTTTTCCC--GTGTTGAGTCAAATTAAGCCGCAGGTTCCACTCCTGGTGGTGCCCTTCCGTCAATTTCTTTAAGTTTCAGCCTTGCGACCATACT-
CCCCCCGGAACCCAAAGACTT-TGATTTCTCATACGGTGCTGATGGAGTCAAAAAACA--
ACCACCAATCCCGAGTCGGCATAGTTTATGGTTAAGACTACGATGGTATCTAATCATCTTCGATCCCCTAACTTTCGTTCTTGATTAATGAAAACATCCTTGGTAAATGCTT
TCGCAGTAGTTCGTCTTTCGGAAATCCAAGAATTTCACCTCTGACAACGAAATACGAATACCCCCAACTGTCCCTATTAATCATTACTTTGTAACGCAAACCAACAAAATA
GTTCCAAAGTCCTATCTTATTATTCCATGCTAATATATTCAA-CGGCATAAGCCTGCTTTAAACACTCTAATTTTTTCACAGTAAACGATGGGTATCCCCCGC-
CCGACAACTTAATGCCAAACGGGATCTCCCCAAGGATGGCC-AGAGACAACACAAGTTCACACGC---ATAGCGTGCGACCGGTCGCTCC

28S rRNA sequences:

BE_AeN706-3
TAAGCATATTACTAAGCGGAGGAAAAGAAACTAACCAGGATTCCCCTAGTAACGGCGAGTGAACCGGGAGAAGCCCAACTTGTAAATCTGGCGGCG-----
TGGCCGTCCGAATTGTAGTCTAGAGAAGCGCCCTCTGCGGCGGACGAAGCCCAAGTTCCTTGGAATGGGACGCTCGAGAGGGTGAGAGCCCCGTCGGTTTCAGTC-
TGTTCGCTCCACGAGGCGCTGTCGAAGAGTCGGGTTGTTTGGGAATGCAGCCCAAAGCAGGTGGTAAATTCCATCTAAGGCTAAATACTGGCGAGAGACCGATAGCGAAC
AAGTACCGTGAGGGAAAGATGAAAAGAACTTTGAAAAGAGAGTTAAACAGTGCTTGAAATTGTTGAGGGGGAAGCGATTAGAAGT-
ACGGGTGCGCCTCGGCTTTGCTTCGGCAGTGCG--GGGAGCTGGTCAGCGTGGGTTTGCCGGGCGGGAGAAACGCG-----
GGGATTGTTCCCCTGCCGATGCCGTCTGGTGGACCGAGGAGTGAAGTGCGCTCTC-
TGAGACCTTCGGGATCTGCGCACTCAAGACGCTGGCGGAATGCTTCTAATCGGCCCGTCTTGAA

BE_AeN706-5
AGCATATTACTAAGCGGAGGAAAAGAAAATAACAATGATTCCCCTAGTAACGGCGAGCGAACCGGGAAGAGCCCAACTTGGAAATCTGGAGGCT-----
TGTCCTTCCGAATTGTAGTCTAGAGTAGTGCCATCTGCGGCTGACCGGTCCCAAGTTCCCTGGAAAGGGACGTCAGAGAGGGTGAGAGCCCCGTCGAGACCGGTT-
TTGTCGCTCCACGATGCGCTTACGAGGAGTCGGGTTGTTTGGGAATGCAGCCCTAAGCGGGAGGTAAATTCCTTCTAAGGCTAAATATTGGCGAGAGACCGATAGCGAAC
AAGTACCGCGAGGGAAAGATGAAAAGACCTTTGAAAAGAGAGTTAAAAAGTACCTGAAATTGTTGAGGGGGAAGCGATTAAAACCAATGGGTGCTCCCTAGGTAGGCG
CAAGTCACACGCTAGGGGCTGGTCAACGTGGGTTAGCCGGGTGAAA----------TATTACACTTGACTGACCGAGGACT----------
CGACGTTGGCGGAATGGTTTTAATCGGCCCGTCTTGA

BE_AeN706-6
TAAGCATATTACTAAGCGGAGGAAAAGAAACTAACCAGGATTCCCCTAGTAACGGCGAGTGAACCGGGAGAAGCCCAACTTGTAAATCTGGCGGCG-----
TGGCCGTCCGAATTGTAGTCTAGAGAAGCGCCCTCTGCGGCGGACGAAGCCCAAGTTCCTTGGAATGGGACGCTCGAGAGGGTGAGAGCCCCGTCGGTTTCAGTC-
TGTTCGCTCCACGAGGCGCTGTCGAAGAGTCGGGTTGTTTGGGAATGCAGCCCAAAGCAGGTGGTAAATTCCATCTAAGGCTAAATACTGGCGAGAGACCGATAGCGAAC
AAGTACCGTGAGGGAAAGATGAAAAGAACTTTGAAAAGAGAGTTAAACAGTGCTTGAAATTGTTGAGGGGGAAGCGATTAGAAGT-
ACGGGTGCGCCTCGGCTTTGCTTCGGCAGTGCG--GGGAGCTGGTCAGCGTGGGTTTGCCGGGCGGGAGAAACGCG-----
GGGATTGTTCCCCTGCCGATGCCGTCTGGTGGACCGAGGAGTGAAGTGCGCTCTC-
TGAGACCTTCGGGATCTGCGCACTCAAGACGCTGGCGGAATGCTTCTAATCGGCCCGTCTTGAAACACGG

BE_AeN706-8
TTAAGCATATTACTAAGCGGAGGAAAAGAAACTAACCAGGATTCCCCTAGTAACGGCGAGTGAACCGGGAGAAGCCCAACTTGTAAATCTGGCGGCG-----
TGGCCGTCCGAATTGTAGTCTAGAGAAGCGCCCTCTGCGGCGGACGAAGCCCAAGTTCCTTGGAATGGGACGCTCGAGAGGGTGAGAGCCCCGTCGGTTTCAGTC-
TGTTCGCTCCACGAGGCGCTGTCGAAGAGTCGGGTTGTTTGGGAATGCAGCCCAAAGCAGGTGGTAAATTCCATCTAAGGCTAAATACTGGCGAGAGACCGATAGCGAAC
AAGTACCGTGAGGGAAAGATGAAAAGAACTTTGAAAAGAGAGTTAAACAGTGCTTGAAATTGTTGAGGGGGAAGCGATTAGAAGT-
ACGGGTGCGCCTCGGCTTTGCTTCGGCAGTGCG--GGGAGCTGGTCAGCGTGGGTTTGCCGGGCGGGAGAAACGCG-----
GGGATTGTTCCCCTGCCGATGCCGTCTGGTGGACCGAGGAGTGAAGTGCGCTCTC-
TGAGACCTTCGGGATCTGCGCACTCAAGACGCTGGCGGAATGCTTCTAATCGGCCCGTCTTGAAACACGG

BE_AeN706-9
AATTAAGCATATTACTAAGCGGAGGAAAAGAAACTAACCAGGATTCCCCTAGTAACGGCGAGTGAACCGGGAGAAGCCCAACTTGTAAATCTGGCGGCG-----
TGGCCGTCCGAATTGTAGTCTAGAGAAGCGCCCTCTGCGGCGGACGAAGCCCAAGTTCCTTGGAATGGGACGCTCGAGAGGGTGAGAGCCCCGTCGGTTTCAGTC-
TGTTCGCTCCACGAGGCGCTGTCGAAGAGTCGGGTTGTTTGGGAATGCAGCCCAAAGCAGGTGGTAAATTCCATCTAAGGCTAAATACTGGCGAGAGACCGATAGCGAAC
AAGTACCGTGAGGGAAAGATGAAAAGAACTTTGAAAAGAGAGTTAAACAGTGCTTGAAATTGTTGAGGGGGAAGCGATTAGAAGT-
ACGGGTGCGCCTCGGCTTTGCTTCGGCAGTGCG--GGGAGCTGGTCAGCGTGGGTTTGCCGGGCGGGAGAAACGCG-----
GGGATTGTTCCCCTGCCGATGCCGTCTGGTGGACCGAGGAGTGAAGTGCGCTCTC-
TGAGACCTTCGGGATCTGCGCACTCAAGACGCTGGCGGAATGCTTCTAATCGGCCCGTCTTGAAACACGG

BE_AeN706-10
AATTAAGCATATTACTAAGCGGAGGAAAAGAAACTAACCAGGATTCCCCTAGTAACGGCGAGTGAACCGGGAGAAGCCCAACTTGTAAATCTGGCGGCG-----
TGGCCGTCCGAATTGTAGTCTAGAGAAGCGCCCTCTGCGGCGGACGAAGCCCAAGTTCCTTGGAATGGGACGCTCGAGAGGGTGAGAGCCCCGTCGGTTTCAGTC-
TGTTCGCTCCACGAGGCGCTGTCGAAGAGTCGGGTTGTTTGGGAATGCAGCCCAAAGCAGGTGGTAAATTCCATCTAAGGCTAAATACTGGCGAGAGACCGATAGCGAAC
AAGTACCGTGAGGGAAAGATGAAAAGAACTTTGAAAAGAGAGTTAAACAGTGCTTGAAATTGTTGAGGGGGAAGCGATTAGAAGT-
ACGGGTGCGCCTCGGCTTTGCTTCGGCAGTGCG--GGGAGCTGGTCAGCGTGGGTTTGCCGGGCGGGAGAAACGCG-----
GGGATTGTTCCCCTGCCGATGCCGTCTGGTGGACCGAGGAGTGAAGTGCGCTCTC-
TGAGACCTTCGGGATCTGCGCACTCAAGACGCTGGCGGAATGCTTCTAATCGGCCCGTCTTGAAACACGG

BE_AeN706-11
AATAGCATATAATTAAGCGGAGGAAAAGAAACTAACTAGGATTCCCTCAGTAACGGCGAGTGAAGCGGGACTAGCTCAGGATGTGAATCTGCGCTTTTATG----------
----------GCG--CCGAATTGTGGTCT-GTAG----ACCTTGACGTTA-TTTGCCGGGCCAAGTTCCTTGGAAAAGGACAGCTGAGAGGGTGAGACTCCCGT---
CCGCCTGGCAGAGTGAGTCATGTG---------
TCAACGAGTCGAGTTGTTTGGGATTGCAGCTCTAATTTGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAGACCGATAGCGTACAAGTACCGTGAGGGAAAGATG
CAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAACGGAAGCGA-AGGAAACCAGTGTT--
TGTTGGTTCATATTTCCCTAGCCACTTGTGGTTTGGGCGCTGTGAGCTTGCGTGAGTTTGGGTTGGTTGATCCCTTTGGAAGAGCGCAGACAGAGTTGATGT--
CTGTTGCTA----------GCACTGGGATTGACTGAT----------GCAGACGAAATGGTTTTCTTTACCCCGTCTTGAAACACGGACCAAGGA

BE_AeN706-12
T---------AGCATATAAGTAAGCGGAGGATAAGAAACTAAATAGGATTCCCTTAGTAATGGCGAATGAACAGGGATCAGCTCAGCATG-GAAATTGGGGCCCTCGGC------
CTTGAATTGTAGTCTCGAGATGTATCGCCAACGGAGGCGCAGATGTAAGCCTCTTGGAAAGGAGCATCAACGAGGGTGAGAGTCCCGTT--
TGTCATCTGCAGTCCCCCGTGCACGGCATACCTTCTAAGAGTCGCGTTCCTCGGAATTGGAGCGTAAATTGGGTGGTAAATTTCATCTCAAGCTAAATATTGGTTCGAGAC
CGATAGCAAACAAGTACCATGAGGGAAAGATGAAAAGGACTTTGAAAAGAGAGTTAAA-
AGTGCCTGAAATTGCTGAAAGGGAAGCGAATGGAACCAGTGTTGCTTGGATGTGATTGGCGGGTGCATGACGCGCTCGTTGTATTTCCTTCCTTGTGTGTCAACGCCAGTT
CGCGATCGAGGAAAACTCCAGGGTCATGGTAGCTTGTCTCAGGGTGAGTGAATAGCCTTGGCAGAACTCATTTGCGGACTGTTTCTTTCGTGTCTGGTTGCAGTGTCTTTG
GCACCCTGGCGATCTTGGCTTGGCT-
TTGTGCCACAGGCGAGTGACCCAGGTGACTCGGGCATTTCCCTGACACAAAGACGATGACTAAATGGTTCTATTCGACCCGTCTTGAAACA

BE_AeN706-13
AGCATATTACTAAGCGGAGGAAAAGAAAATAACAATGATTCCCCTAGTAACGGCGAGCGAACCGGGAAGAGCCCAACTTGGAAATCTGGAGGCT-----
TGTCCTTCCGAATTGTAGTCTAGAGTAGTGCCATCTGCGGCTGACCGGTCCCAAGTTCCCTGGAAAGGGACGTCAGAGAGGGTGAGAGCCCCGTCGAGACCGGTT-
TTGTCGCTCCACGATGCGCTTACGAGGAGTCGGGTTGTTTGGGAATGCAGCCCTAAGCGGGAGGTAAATTCCTTCTAAGGCTAAATATTGGCGAGAGACCGATAGCGAAC

AAGTACCGCGAGGGAAAGATGAAAAGACCTTTGAAAAGAGAGTTAAAAAGTACCTGAAATTGTTGAGGGGGAAGCGATTAAAACCAATGGGTGCTCCCTAGGTAGGCG
CAAGTCACACGCTAGGGGCTGGTCAACGTGGGTTAGCCGGGTGAAA----------TATTACACTTGACTGACCGAGGACT----------
CGACGTTGGCGGAATGGTTTTAATCGGCCCGTCTTGAAACACGG

BE_AeN706-15

GGAATTAAGCATATTACTAAGCGGAGGAAAAGAAAATAACAATGATTCCCCTAGTAACGGCGAGCGAACCGGGAAGAGCCCAACTTGGAAATCTGGAGGCT-----
TGTCCTTCCGAATTGTAGTCTAGAGTAGTGCCATCTGCGGCTGACCGGTCCCAAGTTCCCTGGAAAGGGACGTCAGAGAGGGTGAGAGCCCCGTCGAGACCGGTT-
TTGTCGCTCCACGATGCGCTTACGAGGAGTCGGGTTGTTTGGGAATGCAGCCCTAAGCGGGAGGTAAATTCCTTCTAAGGCTAAATATTGGCGAGAGACCGATAGCGAAC
AAGTACCGCGAGGGAAAGATGAAAAGACCTTTGAAAAGAGAGTTAAAAAGTACCTGAAATTGTTGAGGGGGAAGCGATTAAAACCAATGGGTGCTCCCTAGGTAGGCG
CAAGTCACACGCTAGGGGCTGGTCAACGTGGGTTAGCCGGGTGAAA----------TATTACACTTGACTGACCGAGGACT----------
CGACGTTGGCGGAATGGTTTTAATCGGCCCGTCTTGA

BE_AeN706-16

ACATGATTCCCCTAGTAACGGCGAGCGAACCGGGAAGAGCCCAACTTGGAAATCTGGAGGCT-----
TGTCCTTCCGAATTGTAATCTAGAGTAGTGCCATCTGCGGCTGACCGGTCCCAAGTTCCCTGGAAAGGGACGTCGGAGAGGGTGAGAGCCCCGTCGAGACCGGTT-
TTGTCGCTCCACGATGCGCTTACCAGGAGTCGGGTTGTTTGGGAATGCAGCCCTAAGCGGGAGGTAAATTCCTTCTAAGGCTAAATATTGGCGAGAGACCGATAGCGAAC
AAGTACCGCGAGGGAAAGATGAAAAGACCTTTGAAAAGAGAGTTAAAAAGTACCTGAAATTGTTGAGGGGGAAGCGATTAAAACCAATGGGTGCTCCCTAGGTAGGCG
CAAGTCACACGCTAGGGGCTGGTCAACGTGGGTTAGCCGGGTGAAA----------TATTACACTTGACTGACCGAGGACT----------
CGACGTTGGCGGAATGGTTTTAATCGGCCCGTCTTGAAACCGGAACCAAGGACCTGGGTCGGGCTGAGGTCAGTCACTCTGTGCTCGTGATGCTGGCAAAATGGTTTTCTT
TACCCCGTCTTGAAACAGGGACCAA

BE_AeN706-19

ATAACATGATTCCCCTAGTAACGGCGAGCGAACCGGGAAGAGCCCAACTTGGAAATCTGGAGGCT-----
TGTCCTTCCGAATTGTAATCTAGAGTAGTGCCATCTGCGGCTGACCGGTCCCAAGTTCCCTGGAAAGGGACGTCGGAGAGGGTGAGAGCCCCGTCGAGACCGGTT-
TTGTCGCTCCACGATGCGCTTACCAGGAGTCGGGTTGTTTGGGAATGCAGCCCTAAGCGGGAGGTAAATTCCTTCTAAGGCTAAATATTGGCGAGAGACCGATAGCGAAC
AAGTACCGCGAGGGAAAGATGAAAAGACCTTTGAAAAGAGAGTTAAAAAGTACCTGAAATTGTTGAGGGGGAAGCGATTAAAACCAATGGGTGCTCCCTAGGTAGGCG
CAAGTCACACGCTAGGGGCTGGTCAACGTGGGTTAGCCGGGTGAAA----------TATTACACTTGACTGACCGAGGACT----------
CGACGTTGGCGGAATGGTTTTAATCGGCCCGTCTTGAAACCGGAACCAAGGACCTGGGTCGGGCTGAGGTCAGTCACTCTGTGCTCGTGATGCTGGCAAAATGGTTTTCTT
TACCCCGTCTTGAAAC

BE_AeN706-21

AGCATATAATTAAGCGGAGGAAAAGAAAATAACTATGATTCCCTCAGTAAGGGCGACTGAAGCGGGAACAGCTCATTCTGTGAATCTCTGCTTCG----------
------GCA--GCGAGTTGTGG-------AAAGGCATGTCAGCTAA-GCTACTGGTCTAAGTCTCTTGGAAAAGAGCGCTTGAAAGGGTGACAGCCCCGT-
CTTTGACTGGTTAGCTGTGGCTTTAGCACATGTTTTATACGAGTCGAGTTGCTTGGGATTGCAGCTCTAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAC
ACCGATAGCGTACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCGA-AGGAAACCAGTGCC-
ATTCTTCACTATTTTTCTGTTCCTACTTGTGGGAATAGCGCTGTGGTGTTGAATAGGCCAGCGTGAGTTTGTACCGGGGGATAACA--CTCTTCAGAAGGCAG--
CTAGATTTCCTAGCGATTGCTGTTGAGCTGTGCTCTGGTACAGACTGAGG--CTAGTCACTT---GTGCTCGTGACGCTGGCGAAATGGTTTTCTTCACCCCGTCTT

BE_AeN706-22

CTATGATTCCCTCAGTAAGGGCGACTGAAGCGGGAACAGCTCATTCTGTGAATCTCTGCTTCG----------GCA--GCGAGTTGTGG-------
AAAGGCATGTCAGCTAA-GCTACTGGTCTAAGTCTCTTGGAAAAGAGCGCTTGAAAGGGTGACAGCCCCGT-
CTTTGACTGGTTAACTGTGGCTTTAGCACATGTTTTATACGAGTCGAGTTGCTTGGGATTGCAGCTCTAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAC
ACCGATAGCGTACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCGA-AGGAAACCAGTGCC-
ATTCTTCACTATTTTTCTGTTCCTACTTGTGGGAATAGCGCTGTGGTGTTGAATAGGCCAGCGTGAGTTTGTACCGGGGGATAACA--CTCTTCAGAAGGCAG--
CTAGATTTCCTAGCGATTGCTGTTGAGCTGTGCTCTGGTACAGACTGAGG--CTAGTCACTT---
GTGCTCGTGACGCTGGCGAAATGGTTTTCTTCACCCCGTCTTGAAACACGG

BE_AeN706-24

ATGATTCCCTCAGTAAGGGCGACTGAAGCGGGAACAGCTCATTCTGTGAATCTCTGCTTCG----------GCA--GCGAGTTGTGG-------
AAAGGCATGTCAGCTAA-GCTACTGGTCTAAGTCTCTTGGAAAAGAGCGCTTGAAAGGGTGACAGCCCCGT-
CTTTGACTGGTTAGCTGTGGCTTTAGCACATGTTTTATACGAGTCGAGTTGCTTGGGATTGCAGCTCTAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAC
ACCGATAGCGTACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCGA-AGGAAACCAGTGCC-
ATTCTTCACTATTTTTCTGTTCCTACTTGTGGGAATAGCGCTGTGGTGTTGAATAGGCCAGCGTGAGTTTGTACCGGGGGATAACA--CTCTTCAGAAGGCAG--
CTAGATTTCCTAGCGATTGCTGTTGAGCTGTGCTCTGGTACAGACTGAGG--CTAGTCACTT---GTGCTCGTGACGCTGGCGAAATGGTTTTCTTCACCCCGTCTTGA

BE_AeN706-25

TAGCATATAATTAAGCGGAGGAAAAGAAACTAACAAGGATTCCCCTAGTAACGGCGAGTGAAGCGGGATGAGCTCACCATGTGAATCTGTGTAACCGGAAACGGT---------
----------GCA--CCGAATTGTGGTCT-GGAGAAGTACTGTCGGCCGT-GTTCCCGGGCCAAGTCTCTTGGAAAAGGGCAGCTGAGAGGGTGAAACTCCCGT-
TCTTGCCTGGGAACATTGCGCTTTGGCATATACTTTCTACGAGTCGAGTTGCTTGGGATTGCAGCTCAAATTTGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAT
ACCGATAGTGCACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGTTGAAAGGGAAGCGA-
AGGAAACCAGTGCCGAAGCTTAGTCATACTTCTCTTACTACTTGTGGTAAGGGCGCTGTGGCTTTGCGTGGGTCAGCATCGGCTCTTTGCCTGGGGTATATCTTCAGTAGG
TAGACGA--CCTCTTCGGAG-GTGAGTGCCTATTTTTGTTATTCCGGGTTGGGCTGAGG--TCAGTCACTCT—
GTGCTCGTGATGCTGGCAAAATGGTTTTCTTTACCCCGTCTTG

BE_AeN706-26

TAGCATATAATTAAGCGGAGGAAAAGAAACTAACAAGGATTCCCCTAGTAACGGCGAGTGAAGCGGGATGAGCTCACCATGTGAATCTGTGTAACCGGAAACGGT---------
----------GCA--CCGAATTGTGGTCT-GGAGAAGTACTGTCGGCCGT-GTTCCCGGGCCAAGTCTCTTGGAAAAGGGCAGCTGAGAGGGTGAAACTCCCGT-
TCTTGCCTGGGAACATTGCGCTTTGGCATATACTTTCTACGAGTCGAGTTGCTTGGGATTGCAGCTCAAATTTGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAT
ACCGATAGTGCACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGTTGAAAGGGAAGCGA-
AGGAAACCAGTGCCGAAGCTTAGTCATACTTCTCTTACTACTTGTGGTAAGGGCGCTGTGGCTTTGCGTGGGTCAGCATCGGCTCTTTGCCTGGGGTATATCTTCAGTAGG
TAGACGA--CCTCTTCGGAG-GTGAGTGCCTATTTTTGTTATTCCGGGTTGGGCTGAGG--TCAGTCACTCT—
GTGCTCGTGATGCTGGCAAAATGGTTTTCTTTACCCCGTCTTGAAACACGG

BE_AeN706-27

AGCATATAATTAAGCGGAGGAAAAGAAACTAACAAGGATTCCCCTAGTAACGGCGAGTGAAGCGGGAAGAGCTCACCATGTGAATCTGTGTAACCCGCAAGGGT---------
----------GCA--CCGAATTGTGGTCT-GGAGAAGTACTGTCGGCCAT-GTTTCCGGGCCAAGTCTCTTGGAAAAGAGCAGCTGAGAGGGTGAAACTCCCGT-
TCTTGCCTGGAAATGTTGCGCCATGGCATATGCTTTCTATGAGTCGAGTTGCTTGGGATTGCAGCTCAAATTTGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAT
ACCGATAGTGCACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGTTGAAAGGGAAGCGA-
AGGAAACCAGTGCCGAAGCTTAGTCATATTTCTCTTGCTACTTGTGGCAAGGGCGCTGTGGCTTTGCATGGGTCAGCATCGGCTCTTTGCCTGGGGTAAATCTTCGGTTGG

TAGACGA--CCCTCTTATGGGGTGAGTGCCTTCCGTTGCTATCCTGGGTTGGGCCGAGG--TCAGTCACTCT—
GTGCTCGTGATGCTGGCAAAATGGTTTTCTTTACCCCGTCTTGA

BE_AeN706-30

TCCCTCAGTAAGGGCGACTGAAGCGGGAACAGCTCATTCTGTGAATCTCTGCTTCG----------GCA--GCGAGTTGTGG-------AAAGGCATGTCAGCTAA-
GCTACTGGTCTAAGTCTCTTGGAAAAGAGCGCTTGAAAGGGTGACAGCCCCGT-
CTTTGACTGGTTAGCTGTGGCTTTAGCACATGTTTTATACGAGTCGAGTTGCTTGGGATTGCAGCTCTAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAC
ACCGATAGCGTACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCGA-AGGAAACCAGTGCC-
ATTCTTCACTATTTTTCTGTTCCTACTTGTGGGAATAGCGCTGTGGTGTTGAATAGGCCAGCGTGAGTTTGTACCGGGGGATAACA--CTCTTCAGAAGGCAG--
CTAGATTTCCTAGCGATTGCTGTTGAGCTGTGCTCTGGTACAGACTGAGG--CTAGTCACTT---GTGCTCGTGACGCTGGCGAAATGGTTTTCTTCACCCCGTCT

BE_AeN706-31

GGGCGACTG-AGCGGGAACAGCTCATTCTGTGAATCTCTGCTTCG----------GCA--GCGAGTTGTGG-------AAAGGCATGTCAGCTAA-
GCTACTGGTCTAAGTCTCTTGGAAAAGAGCGCTTGAAAGGGTGACAGCCCCGT-
CTTTGACTGGTTAGCTGTGGCTTTAGCACATGTTTTATACGAGTCGAGTTGCTTGGGATTGCAGCTCTAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAC
ACCGATAGCGTACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCGA-AGGAAACCAGTGCC-
ATTCTTCACTATTTTTCTGTTCCTACTTGTGGGAATAGCGCTGTGGTGTTGAATAGGCCAGCGTGAGTTTGTACCGGGGGATAACA--CTCTTCAGAAGGCAG--
CTAGATTTCCTAGCGATTGCTGTTGAGCTGTGCTCTGGTACAGACTGAGG--CTAGTCACTT---GTGCTCGTGACGCTGGCGAAATGGTTTTCTTCACCCCGTCTTGAAA

BE_AeN706-34

CTATGATTCCCTCAGTAAGGGCGACTGAAGCGGGAACAGCTCATTCTGTGAATCTCTGCTTCG----------GCA--GCGAGTTGTGG-------
AAAGGCATGTCAGCTAA-GCTACTGGTCTAAGTCTCTTGGAAAAGAGCGCTTGAAAGGGTGACAGCCCCGT-
CTTTGACTGGTTAGCTGTGGCTTTAGCACATGTTTTATACGAGTCGAGTTGCTTGGGATTGCAGCTCTAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAC
ACCGATAGCGTACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCGA-AGGAAACCAGTGCC-
ATTCTTCACTATTTTTCTGTTCCTACTTGTGGGAATAGCGCTGTGGTGTTGAATAGGCCAGCGTGAGTTTGTACCGGGGGATAACA--CTCTTCAGAAGGCAG--
CTAGATTTCCTAGCGATTGCTGTTGAGCTGTGCTCTGGTACAGACTGAGG--CTAGTCACTT---GTGCTCGTGACGCTGGCGAAATGGTTTTCTTCACCCCGTC

BE_AeN706-35

TATGATTCCCTCAGTAAGGGCGACTGAAGCGGGAACAGCTCATTCTGTGAATCTCTGCTTCG----------GCA--GCGAGTTGTGG-------
AAAGGCATGTCAGCTAA-GCTACTGGTCTAAGTCTCTTGGAAAAGAGCGCTTGAAAGGGTGACAGCCCCGT-
CTTTGACTGGTTAGCTGTGGCTTTAGCACATGTTTTATACGAGTCGAGTTGCTTGGGATTGCAGCTCTAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAC
ACCGATAGCGTACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCGA-AGGAAACCAGTGCC-
ATTCTTCACTATTTTTCTGTTCCTACTTGTGGGAATAGCGCTGTGGTGTTGAATAGGCCAGCGTGAGTTTGTACCGGGGGATAACA--CTCTTCAGAAGGCAG--
CTAGATTTCCTAGCGATTGCTGTTGAGCTGTGCTCTGGTACAGACTGAGG--CTAGTCACTT---GTGCTCGTGACGCTGGCGAAATGGTTTTCTTCACCCCGTCTT

K-AeN706-5

TAACAAGGATTCCCCTAGTAACGGCGAGTGAAGCGGGATGAGCTCACCATGTGAATCTGTGTAACCGGAAACGGT----------GCA--CCGAATTGTGGTCT-
GGAGAAGTACTGTCGGCCGT-GTTCCCGGGCCAAGTCTCTTGGAAAAGGGCAGCTGAGAGGGTGAAACTCCCGT-
TCTTGCCTGGGAACATTGCGCTTTGGCATATACTTTCTACGAGTCGAGTTGCTTGGGATTGCAGCTCAAATTTGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAT
ACCGATAGTGCACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGTTGAAAGGGAAGCGA-
AGGAAACCAGTGCCGAAGCTTAGTCATACTTCTCTTACTACTTGTGGTAAGGGCGCTGTGGCTTTGCGTGGGTCAGCATCGGCTCTTTGCCTGGGGTATATCTTCAGTAGG
TAGACGA--CCTCTTCGGAG-GTGAGTGCCTATTTTTGTTATTCCGGGTTGGGCTGAGG--TCAGTCACTCT--
GTGCTCGTGATGCTGGCAAAATGGTTTTCTTTACCCCGTCTTGA

K-AeN706-6

AACTAACAAGGATTCCCCTAGTAACGGCGAGTGAAGCGGGATGAGCTCACCATGTGAATCTGTGTAACCGGAAACGGT----------GCA--CCGAATTGTGGTCT-
GGAGAAGTACTGTCGGCCGT-GTTCCCGGGCCAAGTCTCTTGGAAAAGGGCAGCTGAGAGGGTGAAACTCCCGT-
TCTTGCCTGGGAACATTGCGCTTTGGCATATACTTTCTACGAGTCGAGTTGCTTGGGATTGCAGCTCAAATTTGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAT
ACCGATAGTGCACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGTTGAAAGGGAAGCGA-
AGGAAACCAGTGCCGAAGCTTAGTCATACTTCTCTTACTACTTGTGGTAAGGGCGCTGTGGCTTTGCGTGGGTCAGCATCGGCTCTTTGCCTGGGGTATATCTTCAGTAGG
TAGACGA--CCTCTTCGGAG-GTGAGTGCCTATTTTTGTTATTCCGGGTTGGGCTGAGG--TCAGTCACTCT—GTGCTCGTGATGCTGGCAAAATGGTTTTCTTTACCCCGT

K-AeN706-7

AATAACTATGATTCCCTCAGTAAGGGCGACTGAAGCGGGAAGAGCTCATTCTGTGAATCTGAATACTTTTTTG----------GTATTCCGAATTGTGAATT-
TAAGAAGCA-CATCAGCGAG-AGTTCTGGTCTAAGTCTCTTGGAAAAGAGCAGCTGAGAGGGTGACACTCCCGTATTTTGGCTAG--
AACTCAAACTTTAGCATGTGGTTTCGACGAGTCGAGTTGCTTGGGATTGCAGCTCAAAGTTGGTGGTAAATTCCATCTAAAGCTAAATATTGGAAAATGACCGATAGCGT
ACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAAAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCAA-ACGGATCCAGTGCC-
ATGCCTTATCATTTTTCTGTTCTTGCTTGCGAGAATAGCGCTGTGATCTGGTGTAGGTCAACATGGGTTTGAACCGAAGGAAGCAG--CTATTTAGGAGGCAG--
CACAATTTCTGTGCGATTACTATTTAGCTATGCTTTGGTTCAGACCGAGG--ATAGTCACTTTGTGCTTTCGTGATGTTGACAAAATGGATCTGTTTACCCCGTCTTGAA

K-AeN706-11

GGAAAAGAAAATAACTATGATTCCCTCAGTAAGGGCGACTGAAGCGGGAACAGCTCATTCTGTGAATCTCTGCTTCG----------GCA--
GCGAGTTGTGG-------AAAGGCATGTCAGCTAA-GCTACTGGTCTAAGTCTCTTGGAAAAGAGCGCTTGAAAGGGTGACAGCCCCGT-
CTTTGACTGGTTAGCTGTGGCTTTAGCACATGTTTTATACGAGTCGAGTTGCTTGGGATTGCAGCTCTAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAC
ACCGATAGCGTACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCGA-AGGAAACCAGTGCC-
ATTCTTCACTATTTTTCTGTTCCTACTTGTGGGAATAGCGCTGTGGTGTTGAATAGGCCAGCGTGAGTTTGTACCGGGGGATAACA--CTCTTCAGAAGGCAG--
CTAGATTTCCTAGCGATTGCTGTTGAGCTGTGCTCTGGTACAGACTGAGG--CTAGTCACTT---GTGCTCGTGACGCTGGCGAAATGGTTTTCTTCACCCCGTCTTG

K-AeN706-15

CTCAGTAAGGGCGACTGAAGCGGGAAGAGCTCACCGTGTGAATCTGTGCGGGTGTTTGTTGTT-TTTT------GAATTCAATAAACCCAGCA--CCGAATTGTGGTCT-
GTAGAAGCATTGTCGGCCGC-GGTCCCGGGCCAAGTCCCTTGGAAAAGGGCAGCATAGAGGGTGAGACTCCCGT-
TCTTGCCTGGGACGTGTTGGCCATGGCACATGTATTCGACGAGTCGAGTTGCTTGGGATTGCAGCTCCAAGTGGGTGGTAAATTCCATCTAAAGCTAAATACCGGTGGGA
GACCGATAGCGAACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAAGGGAAGCGA-AGGAAACCAGTGTC-
GTGGCTGTGCATATTTCCCGTGCCGCTTGCGGTGTGGGCGCTGTGCACGGTCTTGGGCCAGCATGGGTTTGTATTTCATGGCAAATGGTGGTTTGGGAGGTAGGTTCCACT
TGTGGTTCCGATACCAAGTCATTTTGTGGTGGATGAGACCGAGGGAGTCGTCT-TTTTGACGCTCGTGATGCTGGCGAAATGGTTTTCTTTACCCCGTCTTGA

K-AeN706-17

ATTAAGCATATAAGTAAGCGGAGGAAAAGAAACTAACAAGGATTCCCTCAGTAAGGGCGACTGAAGCGGGAAGAGCCCATGTTGCCAA-------
ACCTTCGGGTGTTGTGCCATGCGTGTCTTTTGTCTGGCCA----------GGAGTAAATCCCTTGGAACAGGGTATCAGAGAGGGTGACAATCCCGTTCGCCTCTTGG---
CTTTGGCAATT--GTAGACCGTG-
TAACGGAGTCGAGTTGCTTGGGATTGCAGCTCAAAGCGGGTGGTAAATTCCATCTAAGGCTAAATACGGGTGGGAGACCGATAGCGAACAAGTACTGTGAAGGAAAGAT
GAAAAGAACTTTGAAAAGAGAGTTAAAGAGTGCCTGAAATTGCTGAAAGGGAACCGTAGGTCATCAGTGTTGGCTTCGTGCATGTTTGGCTCTCCTACCTGGGGTGGT--
CTTTGTGCACGCG---G----------CCTACTCATCGTCACCATTGCCCGGAGGACAGCTT---GC----------
GCTCTGGGTGGTGTTGGGGCTGAGAAAATGGTGCTCTACACCCCGTCTTG

K-AeN706-18

AGCATATAATTAAGCGGAGGAAAAGAAAATAACTATGATTCCCTCAGTAAGGGCGACTGAAGCGGGAACAGCTCATTCTGTGAATCTCTGCTTCG----------
------GCA--GCGAGTTGTGG-------AAAGGCATGTCAGCTAA-GCTACTGGTCTAAGTCTCTTGGAAAAGAGCGCTTGAAAGGGTGACAGCCCCGT-
CTTTGACTGGTTAGCTGTGGCTTTAGCACATGTTTTATACGAGTCGAGTTGCTTGGGATTGCAGCTCTAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAC
ACCGATAGCGTACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCGA-AGGAAACCAGTGCC-
ATTCTTCACTATTTTTCTGTTCCTACTTGTGGGAATAGCGCTGTGGTGTTGAATAGGCCAGCGTGAGTTTGTACCGGGGGATAACA--CTCTTCAGAAGGCAG--
CTAGATTTCCTAGCGATTGCTGTTGAGCTGTGCTCTGGTACAGACTGAGG--CTAGTCACTT---GTGCTCGTGACGCTGGCGAAATGGTTTTCTTCACCCCGTCTTG

K-AeN706-19

AAACTAACTAGGATTCCCTCAGTAACGGCGAGTGAAGCGGGACTAGCTCAGGATGTGAATCTGCGCTTTTATG----------GCG--CCGAATTGTGGTCT-
GTAG----ACCTTGACGTTA-TTTGCCGGGCCAAGTTCCTTGGAAAAGGACAGCTGAGAGGGTGAGACTCCCGT---CCGCCTGGCAGAGTGAGTCATGTG---------
TCAACGAGTCGAGTTGTTTGGGATTGCAGCTCTAATTTGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAGACCGATAGCGTACAAGTACCGTGAGGGAAAGATG
CAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAACGGAAGCGA-AGGAAACCAGTGTT--
TGTTGGTTCATATTTCCCTAGCCACTTGTGGTTTGGGCGCTGTGAGCTTGCGTGAGTTTGGGTTGGTTGATCCCTTTGGAAGAGCGCAGACAGAGTTGATGT--
CTGTTGCTA----------GCACTGGGATTGACTGAT----------GCAGACGAAATGGTTTTCTTTACCCCGTCTTGAAACACGGACCAAGG

K-AeN706-22

AATAGCATATAATTAAGCGGAGGAAAAGAAACTAACCAGGATTCCCTCAGTAAGGGCGACTGAAGCGGGAAGTGCTCAGGATGTGAATCTGCGC----TAT----------
----------GCG--CCGAATTGTGGTCT-GTAG----ACGGTGGCGCTA-CTAGCCGGGCCAAGTCCCTTGGAAAAGGGCAGCTGAGAGGGTGAGACTCCCGT---
CCGCCCGGCTGGGTGAGCTACTAG---------
TCAACGAGTCGAGTTGTTTGGGATTGCAGCTCAAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAGACCGATAGCGTACAAGTACCGTGAGGGAAAGATG
CAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAAGGGAAGCGA-AGGAAACCAGTGTT--
TGCTCGGTCATATTTCCCTGGCCGCTTGCGGCTTGGGCGCTGTGTCCG-GCTTGTGTTGTTCTTGGTTGGGACCCTTGGAAGAGCGCAGAGGGAGTTGACCT--CTGTTGCTA-
----------GCATTGGTCYTGACTGAG----------GAGGACGAAATGGTTTTCTTTACCCCGTCTTG

K-AeN706-26

AGGATTCCCCAAGTAACGGCGAGTGAACAGGGAGAAGATCTATATGTAAATCAGTGTTTCTG----------ACA--TTGAATTGTGAATT-
ATAGAAATTGCATAGATTGGCGTAAATAAATCGAATCTCTTGGAAAAGAGTAGAATATAGGGTGATACTCCCAT-ATGTGTTTATTTTACAACGATCATA-
TGTGCAATCTTATAGAGTCCGGTTGCTTGGGATTGCAGCCGAAAGTGGGAGGTGAATTCCTTCTAAATCTAAATATTTGTAAGATACCAATAGATGAAAAGTACCGTGAG
GGACTGATGAAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAACTATAGAGATGTTACGGATAAGAAAGAATTGTCTAGAGTTGCTCGTCTATCT-
TCTTCCCTTGTGGTTGAAGTCTAATGGCAGCTCCATAGAGTTATTTGATATAATCATTGGTTTATAT----------TTATTCCTT----------
TTTTTATGTCTCTTAAAT----------AACAAAATTCATTTTTATCTCCCGTCTTGAAAC

K-AeN706-28

TATAATTAAGCGGAGGAAAAGAAAATAACTATGATTCCCTCAGTAAGGGCGACTGAAGCGGGAACAGCTCATTCTGTGAATCTCTGCTTCG----------
GCA--GCGAGTTGTGG-------AAAGGCATGTCAGCTAA-GCTACTGGTCTAAGTCTCTTGGAAAAGAGCGCTTGAAAGGGTGACAGCCCCGT-
CTTTGACTGGTTAGCTGTGGCTTTAGCACATGTTTTATACGAGTCGAGTTGCTTGGGATTGCAGCTCTAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAC
ACCGATAGCGTACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCGA-AGGAAACCAGTGCC-
ATTCTTCACTATTTTTCTGTTCCTACTTGTGGGAATAGCGCTGTGGTGTTGAATAGGCCAGCGTGAGTTTGTACCGGGGGATAACA--CTCTTCAGAAGGCAG--
CTAGATTTCCTAGCGATTGCTGTTGAGCTGTGCTCTGGTACAGACTGAGG--CTAGTCACTT---GTGCTCGTGACGCTGGCGAAATGGTTTTCTTCACCCCGTCTTGAA

K-AeN706-30

AGCATATAATTAAGCGGAGGAAAAGAAACTAACCAGGATTCCCTCAGTAAGGGCGACTGAAGCGGGAAGTGCTCAGGATGTGAATCTGCGCCTT-GTT----------
---------GCG--CCGAATTGTGGTCT-GACG----ACTGTGGCATTA-CTGGCCGGGCCAAGTCCCTTGGAACAGGGCGGCTGAGAGGGTGAGACCCCCGT---
TCGCCTGGCTGGGTGAGCCGCTAG---------
TCTACGAGTCGAGTTGTTTGGGATTGCAGCTCAAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAGACCGATAGCGTACAAGTACCGTGAGGGAAAGATG
CAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAAGGGAAGCGA-AGGAAACCAGTGTC--
GGCTTGGTCATATTTCCCTGTCCGCTTGCGGTTTGGGCGCTGTGTCCG-GCCTTGGTTGTTCTTGGTTGGGTCTGCTGGAAGAGCGCGGTGAGAGTTGATCT--CCGTTGCTA--
----------GCAGCGGTCCTGACTGAG----------GAGGACGAAATGGTTTTCTTTACCCCGTCTT

K-AeN706-32

TAGCATATTATTAAGCGGAGGAAAAGAAACTAACCAGGATTCCCTCAGTAAGGGCGACTGAAGCGGGAAGTGCTCAGGATGTGAATCTGCGC----TAT----------
---------GCG--CCGAATTGTGGTCT-GTAG----ACGGTGGCATTA-CCAGCCGGGCCAAGTCCCTTGGAACAGGGCAGCTGAGAGGGTGAGACTCCCGT---
TCGCCCGGCAGGATGAGCTACTAG---------
TCAACGAGTCGAGTTGTTTGGGATTGCAGCTCAAAGCGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAGACCGATAGCGTACAAGTACCGTGAGGGAAAGAT
GCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATCGCTGAAAGGGAAGCGA-AGGAAACCAGTGTT--
TGCGTAGTTATACTTCCCTAGCCGCTTGCGGTTGGGGCGCTGTGGCTGCGCTTGTGTTTTGCTTAGTTGGAACGTGCGGAAGAGCGGTGAGGGAGTTGACCT--
CATTTGCTA----------GCGTTCGTTTCGACCGAG----------GAGGACGAAATGGTTTTCTTTACCCCGTCTTGA

K-AeN706-33

AATAGCATATAATTAAGCGGAGGAAAAGAAACTAACCAGGATTCCCTCAGTAAGGGCGACTGAAGCGGGAAGTGCTCAGGATGTGAATCTGCGCCTT-GTT----------
----------GCG--CCGAATTGTGGTCT-GACG----ACTGTGGCATTA-CTGGCCGGGCCAAGTCCCTTGGAACAGGGCGGCTGAGAGGGTGAGACCCCCGT---
TCGCCTGGCTGGGTGAGCCGCTAG---------
TCTACGAGTCGAGTTGTTTGGGATTGCAGCTCAAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAGACCGATAGCGTACAAGTACCGTGAGGGAAAGATG
CAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAAGGGAAGCGA-AGGAAACCAGTGTC--
GGCTTGGTCATATTTCCCTGTCCGCTTGCGGTTTGGGCGCTGTGTCCG-GCCTTGGTTGTTCTTGGTTGGGTCTGCTGGAAGAGCGCGGTGAGAGTTGATCT--CCGTTGCTA--
----------GCAGCGGTCCTGACTGAG----------GAGGACGAAATGGTTTTCTTTACCCCGTCTTG

K-AeN706-34

ATGATTCCCTCAGTAAGGGCGACTGAAGCGGGAACAGCTCATTCTGTGAATCTCTGCTTCG----------GCA--GCGAGTTGTGG-------
AAAGGCATGTCAGCTAA-GCTACTGGTCTAAGTCTCTTGGAAAAGAGCGCTTGAAAGGGTGACAGCCCCGT-

CTTTGACTGGTTAGCTGTGGCTTTAGCACATGTTTTATACGAGTCGAGTTGCTTGGGATTGCAGCTCTAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAC
ACCGATAGCGTACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCGA-AGGAAACCAGTGCC-
ATTCTTCACTATTTTTCTGTTCCTACTTGTGGGAATAGCGCTGTGGTGTTGAATAGGCCAGCGTGAGTTTGTACCGGGGGATAACA--CTCTTCAGAAGGCAG--
CTAGATTTCCTAGCGATTGCTGTTGAGCTGTGCTCTGGTACAGACTGAGG--CTAGTCACTT---GTGCTCGTGACGCTGGCGAAATGGTTTTCTTCACCCCGTCTTGAA

K-AeN706-35

TAGCATATAATTAAGCGGAGGAAAAGAAACTAACCAGGATTCCCTCAGTAAGGGCGACTGAAGCGGGAATAGCTCGAACCG----------
AAGAATTGTGGTGT----------GTTTGCGTTAC-TTGCTCGGCCAAAGTCCCTTGGAACAGGGCAGCGTAGAGGGTGAGACTCCCGT---TTGGCTGTGCTTTTGAGCTTACA--------
TGATGGAGTCGAGTTGTTTGGGATTGCAGCTCAAAGTGGGTGGTAAATTCCATCGAAAGCTAAATATTGGTGGGAGACCGATAGCGTACAAGTACCGTGAGGGAAAGAT
GCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCGA-AGGAAACCAGTGTC--
GTGCTTGTCATATTTCTGTGACCGCTTGCGGTTACGGCGCTGTGGCT--GCACTGGTCAGGGTTGGTTCTGGTACGGTACAATTC----------
GTGCTGTCTGGGACCGAG----------TTCTGTCGAAATGGTTTTCTTTACCCCGTCTTGA

K-AeN706-36

CTAACTAGGATTCCCTCAGTAAGGGCGACTGAAGCGGGAAGAGCTCACCGTGTGAATCTGTGCGGGTGTTTATTGTC-TCATCTTTGATGATTTGGTAGACCCAGCA--
CCGAATTGTGGTCT-GTAGAAGCATTGTCGGCCGC-GGTCCCGGGCCAAGTCCCTTGGAAAAGGGCAGCATAGAGGGTGAGACTCCCGT-
TCTTGCCTGGGACGTGTTGGCCATGGCACATGTATTCGACGAGTCGAGTTGCTTGGGATTGCAGCTCCAAGTGGGTGGTAAATTCCATCTAAAGCTAAATACCGGTGGGA
GACCGATAGCGAACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAAGGGAAGCGA-AGGAAACCAGTGTC-
GTGGCTGTGCATATTTCCCGTGCCGCTTGCGGTGTGGGCGCTGTGCACGGTCTTGGGCCAGCATGGGTTTGTATTTCATGGCAAATGGTGGTTTGGGAGGTAGGTTCCACT
TGTGGTTCCGATACCAAGTCATTTTGTGGTGGATGAGACCGAGGGAGTCGTCTATTTTGACGCTCGTGATGCTGGCGAAATGGTTTTCTTTACCCCGTCTTGAA

K-AeN706-37

AACTAACTAGGATTCCCTCAGTAAGGGCGACTGAAGCGGGAAGAGCTCACCGTGTGAATCTGTGCGGGTGTTTATTGTC-TCATCTTTGATGATTTGGTAGACCCAGCA--
CCGAATTGTGGTCT-GTAGAAGCATTGTCGGCCGC-GGTCCCGGGCCAAGTCCCTTGGAAAAGGGCAGCATAGAGGGTGAGACTCCCGT-
TCTTGCCTGGGACGTGTTGGCCATGGCACATGTATTCGACGAGTCGAGTTGCTTGGGATTGCAGCTCCAAGTGGGTGGTAAATTCCATCTAAAGCTAAATACCGGTGGGA
GACCGATAGCGAACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAAGGGAAGCGA-AGGAAACCAGTGTC-
GTGGCTGTGCATATTTCCCGTGCCGCTTGCGGTGTGGGCGCTGTGCACGGTCTTGGGCCAGCATGGGTTTGTATTTCATGGCAAATGGTGGTTTGGGAGGTAGGTTCCACT
TGTGGTTCCGATACCAAGTCATTTTGTGGTGGATGAGACCGAGGGAGTCGTCTATTTTGACGCTCGTGATGCTGGCGAAATGGTTTTCTTTACCCCGTCTTGAA

K-AeN706-38

ACTATGATTCCCTCAGTAAGGGCGACTGAAGCGGGAACAGCTCATTCTGTGAATCTCTGCTTCG----------GCA--GCGAGTTGTGG-------
AAAGGCATGTCAGCTAA-GCTACTGGTCTAAGTCTCTTGGAAAAGAGCGCTTGAAAGGGTGACAGCCCCGT-
CTTTGACTGGTTAGCTGTGGCTTTAGCACATGTTTTATACGAGTCGAGTTGCTTGGGATTGCAGCTCTAAGTGGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAC
ACCGATAGCGTACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAATGGAAGCGA-AGGAAACCAGTGCC-
ATTCTTCACTATTTTTCTGTTCCTACTTGTGGGAATAGCGCTGTGGTGTTGAATAGGCCAGCGTGAGTTTGTACCGGGGGATAACA--CTCTTCAGAAGGCAG--
CTAGATTTCCTAGCGATTGCTGTTGAGCTGTGCTCTGGTACAGACTGAGG--CTAGTCACTT---GTGCTCGTGACGCTGGCGAAATGGTTTTCTTCACCCCGTCTTGA

K-AeN706-45

AACTAACTAGGATTCCCTCAGTAACGGCGAGTGAAGCGGGACTAGCTCAGGATGTGAATCTGCGCTTTTAAG----------GCG--CCGAATTGTGGTCT-
GTAG----ACTTTGATGTTA-TCTGCCGGGCCAAGTTCCTTGGAAAAGGACAGCTGAGAGGGTGAGACTCCCGT---CCGCCTGGTAGAATGAATCATGTG---------
TCAACGAGTCGAGTTGTTTGGGATTGCAGCTCTAATTTGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAGACCGATAGCGTACAAGTACCGTGAGGGAAAGATG
CAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGCTGAAACGGAAGCGA-AGGAAACCAGTGTT--
TGTTGGTTCATATTTCCCTGACCACTTGTGGTTTGGGCGCTGTGAGCCTGCGTGGGTTTGTTTTGATTGATCCCTTTGGAAGAGCGCAGACAGAGTTGATGT--CTGTTGCTA-
----------GCACTGGGTTTGATTGAT----------GCAGACGAAATGGTTTTCTTTACCCCGTCTTGAAACACGGACCAAGGA

K-AeN706-52

AGCATATAATTAAGCGGAGGAAAAGAAACTAACAAGGATTCCCCTAGTAACGGCGAGTGAAGCGGGAAGAGCTCACCATGTGAATCTGTGTAACCCGCAAGGGT----------
----------GCA--CCGAATTGTGGTCT-GGAGAAGTACTGTCGGCCAT-GTTTCCGGGCCAAGTCTCTTGGAAAAGAGCAGCTGAGAGGGTGAAACTCCCGT-
TCTTGCCTGGAAATGTTGCGCCATGGCATATGCTTTCTATGAGTCGAGTTGCTTGGGATTGCAGCTCAAATTTGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGAT
ACCGATAGTGCACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGTTGAAAGGGAAGCGA-
AGGAAACCAGTGCCGAAGCTTAGTCATATTTCTCTTGCTACTTGTGGCAAGGGCGCTGTGGCTTTGCATGGGTCAGCATCGGCTCTTTGCCTGGGGTAAATCTTCGGTTGG
TAGACGA--CCCTCTTATGGGGTGAGTGCCTTCCGTTGCTATCCTGGGTTGGGCCGAGG--TCAGTCACTCT--GTGCTCGTGATGCTGGCAAA

K-AeN706-56

AACTAACAAGGATTCCCCTAGTAACGGCGAGTGAAGCGGGAAGAGCTCACCATGTGAATCTGTGTAACCTTC--GGGT----------GCA--CCGAATTGTGGTCT-
GGAGAAGTATTGTCGGCCGA-GTTTCCGGGCCAAGTCTCTTGGAAAAGGGCAGCTGAGAGGGTGAAACTCCCGT-
TCTTGCCTGGAAACTTTGCGCTTTGGCACATGCTTTCTACGAGTCGAGTTGCTTGGGATTGCAGCTCAAATTTGGTGGTAAATTCCATCTAAAGCTAAATATTGGTGGGATA
CCGATAGTGCACAAGTACCGTGAGGGAAAGATGCAAAGAACTTTGAAAAGAGAGTTAAAGAGTACCTGAAATTGTTGAAAGGGAAGCGA-
AGGAAACCAGTGCCGAAGCCTTGTCATACTTCTCTTGCCACTTGTGGTAAGGGCGCTGTGGCTTGGCGTGGGTCAGCATCGGCTCTTCGCCTGGGGTAAATCTTCGGTTGG
TAGACGA--CCACATTATGTGGTGAGTGCCTTCTGTTGCTATCCTGGGTCGGGCTGAGG--TCAGTCACTCT--
GTGCTCGTGATGCTGGCAAAATGGTTTTCTTTACCCCGTCTTGAAACACGGACCAAGGA