UiT The Arctic University of Norway

Faculty of Biosciences, Fisheries and Economics, The Norwegian College of Fishery Science

Species validity of five common northern/Arctic spring bloom diatoms: a combined morphological and molecular study

Martina Uradnikova

Master's thesis in Marine Ecology and Resource Biology BIO-3950 February 2020

Acknowledgements

I would like to thank Professor Hans Christian Eilertsen for all his knowledge, scientific and human inspiration, patience and encouragement he leaded me with throughout the process of writing of this master thesis. I am thankful also to researcher Galina Gusarova for her kind and prompt help with phylogenetic part of my survey. Tom-Ivar Eilertsen and Augusta Sundbø I am thankful for help with preparation of EM samples and for time spent with me by electron microscope. Finally, my gratitude belong to my parents, who taught me not to give up.

Abstract

Relevant taxonomical morphological characters of five common spring bloom centric diatoms, preliminary identified as *Attheya longicornis* Crawford & Gardner, *Chaetoceros furcellatus* Yendo, *Porosira glacialis* Jørgensen, *Coscinodiscus sp.* and *Skeletonema marinoi* Sarno & Zingone were observed and described using light and scanning electron microscope. Further, samples of cultured diatoms were sent to the Canadian centre of DNA Barcoding (CCDB) for DNA extraction, PCR and sequencing. Out of five submitted replicates for each species, CCDB gained five sequences of *Attheya* strain nr. 20.2, two sequences of *Chaetoceros* strain nr. 61 and one sequence of *Porosira* strain nr. 201 based on marker of *rbc*L gene from chloroplast. One sequence of Skeletonema strain nr. 204 resulted from sequencing of 16S rRNA gene. Sequencing was not successful for *Coscinodiscus* strain nr. 203.

According to results from the morphologic examination, with some uncertainty, it is possible to conclude that strain nr. 20.2 is *Attheya longicornis*. *Chaetoceros* strain nr. 61 is estimated to be *Ch. furcellatus*, but it is not possible to identify certainly due to missing SEM images of spores and missing appropriate phylogenetic data. *Porosira* strain nr. 201 is *Porosira glacialis*, mainly based on morphologic description, with indication from molecular survey. *Coscinodiscus strain* nr. 203 is *Coscinodiscus concinnus* and *Skeletonema* strain nr. 204 is *Skeletonema marinoi*. Both last named species were identified only according to morphologic features.

Using reference DNA sequences available in the GenBank, it was possible to demonstrate that rbcL gene from chloroplast is a promising marker for DNA barcoding of diatoms. Many species were found in strongly supported monophylectic groups on the rbcL phylogeny, necessarily to remind that the periodic reconciliation of up to date scientific publications on diatom taxonomy is inevitable in order to perform proper taxonomic classification.

Keywords: *Attheya longicornis*, *Chaetoceros furcellatus*, *Porosira glacialis*, *Coscinodiscus concinnus*, *Skeletonema marinoi,* diatoms, taxonomy, DNA Barcoding

Table of Contents

1. Introduction

Diatoms (phylum Bacillariophyta Karsten, 1928) are unicellular eukaryotic mainly photosynthetic algae with two characteristic silicified valves (Graham et al. 2016). Diatoms are unicellular: they occur either as solitary cells or in colonies, which can take the shape of ribbons, fans, zigzags, or stars. Individual cells range in size from 2 to 200 micrometers. Diatoms have two distinct shapes: a few (centric diatoms) are radially symmetric, while most (pennate diatoms) are broadly bilaterally symmetric.

Diatoms reproduce vegetatively by binary fission, and two new individuals are formed within the parent cell frustule. Each daughter cell receives one parent cell theca as epitheca, and the cell division is terminated by the formation of a new hypotheca for each of the daughter cells. The decrease in the average cell size of a diatom population during vegetative growth implies a need for a means of restoring the cell size. This is made possible by auxospore formation, in which a cell sheds its siliceous theca, thereafter forming a large sphere surrounded by an organic membrane. Within this sphere, a new diatom frustule of maximal size is formed, and the cycle starts anew. Diatom resting spores are normally formed as a response to unfavorable environmental conditions, and germination occurs when the conditions improve. Resting spore formation is common in centric, but rare in pennate marine planktonic diatoms. (Rytter Hasle in Tomas1997) Knowledge of algal physiology and life cycles is essential for taxonomists to avoid incorrect descriptions of “new”, morphologically delineated species that are merely separate life cycle stages of species already known to science (Degerlund 2011).

Due to large number of estimated species: somewhere between 10.000 – 1.000.000 (Kaczmarska 2007), diatoms as a group of organisms is not homogenous, neither are they thoroughly described. Algaebase.org as an example of contributed and updated databases counts ≥14.000 species of diatoms, Alverson (2008) reports ≥25,000 species in 2008 (according to the computerized database of verified diatom names at the California Academy of Sciences).

Different biologists have different ideas about how to define a species and other taxonomic categories, but **species** remains to be main taxon, thus historically diatom species were defined phoenetically, there were no fundamental difference between their definition and the definitions of genera, families, or whatever. Thus, for example, a genus was merely a cluster of species between which, in the opinion of taxonomists, where the differences were not large enough to allow further subdivision (Tomas et al. 1997).

According to Haider (2018), there are over 30 concepts of species in use today. Most of these concepts can be suitable under appropriate circumstances, depending on the organism under study, and the question that needs to be addressed. Among them three are considered to be the main ones and hence most used: biological, morphological and phylogenetic species concept.

The biological species concept is drawing a line of species separation according to their inability to interbreed. Application of this approach is possible only on populations that occurs together in time and space. Reproductive barriers are often not complete and little knowledge of their mechanisms complicates the use of the biological species concept. (Casteleyn et al. 2009).

The morphological approach to diatom species delimitation includes thousands observations both from light and electron microscope resulting in thorough descriptions of diatom structures and categorization according to shape similarities. Relying only on this approach, however, can lead to misidentifications as there are many facts that may confuse beginning taxonomist: phenotypic plasticity, existence of polymorphism within one clone or even within one cell; existence of cryptic species; morphological and physiological changes throughout the life cycle; mutation tendency both in the nature and in isolated monocultures. But still, 237 years old history of diatom screening since Otto Friedrich Müller described the first diatom in 1783, has served us with a phletora of practical information about diatom morphology, ecology, physiology and genetics, this at least for species that are abundant and most visible.

Diatom species were diagnosed and classified based on morphological features of the siliceous cell wall, whereby names were assigned to more-or-less discrete,

morphologically similar phenotypes. Employing SEM technology and multivariate statistical methods, regarding spatial and life cycle aspects, considering reproductive isolation to be the hallmark of diatom speciation, taxonomists were supported with a complex sets of continuous morphometric characters for the first time. (Alverson 2008)

Particularly fine frustule structures now attracts both taxonomists (and artists) and is the main study object for traditional taxonomists. The frustule is a silica cell wall that consists of two valves, enclosing the protoplasm, joined together by siliceous girdle bands (De Stefano 2005). Simplified schematic gross morphology of the frustule consisting of valves and cingula according to Rytter Hasle in (Tomas 1997) is shown in Figure 1, Appendix 1.

The wide variety of geometric shapes and features connected with diatom morphology potentially results in numerous characteristic patterns. When a motif of the valve or frustule is repeated systematically, the result is a periodic pattern. (McLaughlin 2012) From the symmetry two main groups of diatoms arose: radially symmetric centric diatoms and elongated pennate diatoms. Basic description of axes and planes of diatom frustule is shown at Figure 2, Appendix 1. (Rytter Hasle in Tomas 1997)

Light microscopy is commonly used in routine work, but mostly only gives a rough 2D insight into cell morphology and reveals basic ecological features of a species, e.g. colony forming capability. While diatom cells in separable colonies are connected by organic substances, inseparable diatom colonies, like chains or ribbons are formed due to inseparable interlockings made of silica.

Since 1965, scanning electron microscope (SEM) has revealed new taxonomically important diatom microstructures, and SEM data are now foundational to nearly all levels of the diatom classification system (Round 1990). The siliceous layer of the valve and cingula is characterized by a wide variety of perforations (as pores, striae or processes), surrounding forms and covering forms. (Rytter Hasle, in Tomas 1997; von Quillfeldt, 2001) Example of multilayer structure of the valve is depicted at Figure 3, Appendix 1.

Collecting information on diatom morphology helps explaining functions of the cell wall (Medlin et al. 1986) and inspires biotechnological research in many ways. In order to understand the processes involved in biomineralisation of the frustule that may eventually allow mimicking its structures and producing new materials with advanced mechanical, magnetic, optical, electrical, piezoelectrical, or adhesive properties (Losic 2007); material engineers and chemists employed also atomic force microscopy to reveal diatom topography and nanostructures of a great detail.

Again, biotechnological research drives biochemical studies of diatoms, their cell properties, shifting the focus into the inside structures down to molecular level searching for the genes involved in biological processes. Biologically active compounds extracted from diatom cells have been proposed for a range of biotechnological applications (Lopez 2005) perceiving diatoms as actively present ecological components.

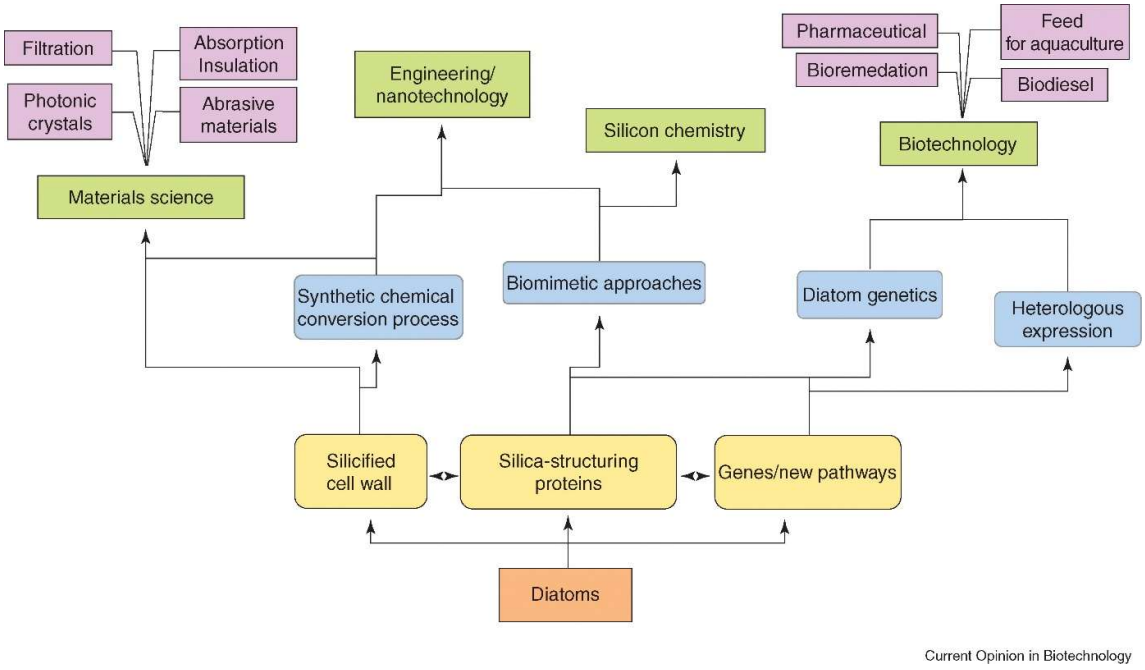


Figure 4 Various fields and application domains for diatom research. Taken from (Lopez 2005)

After relevant molecular tools were developed and the first complete nuclear, plastid and mitochondrial genome sequences of the centric diatom *Thalassiosira pseudonana* (Armbrust et al. 2004) and pennate diatom *Phaeodactylum tricornutum* (Bowler et al. 2008) were presented, the lack of model species was overcome.

The completed *P. tricornutum genome* is approximately 27.4 megabases (Mb) in size, which is slightly smaller than *T. pseudonana* (32.4 Mb distributed on 24 chromosomes). *P. tricornutum* is predicted to contain fewer genes (10,402 as opposed to 11,776). For *P. tricornutum* gene identification and functional analysis was facilitated by the availability of more than 130,000 expressed sequence tags (ESTs) generated from cells grown under 16 different conditions.

Evidence that *P. tricornutum* shares only 57% of its genes with *T. pseudonana* points up to a great molecular divergence while being in the same group of organisms (Bowler 2008). An overwhelming portion of rapidly increasing amounts of sequence data on diatom molecular properties led to questions and answers about evolutionary history of diatom genomes, leading to a new lineage-based species concept for diatoms.

The lineage - based (phylogenetic) species concept for diatoms assumes the existence of a single line of direct ancestor–descendant relationship, or a single branch on a phylogenetic tree (Alverson 2008). Phylogenetic species are defined as monophyletic clusters of individuals that are diagnosably distinct from other such clusters and should display a parental pattern of ancestry and descent (Cracraft 1989; Vanderlaan et al. 2013) (Medlin 2018) Delimitation factors are e.g., reproductive isolation, autapomorphic features, reciprocal monophyly, etc. (de Queiroz 2007) from (Alverson 2008).

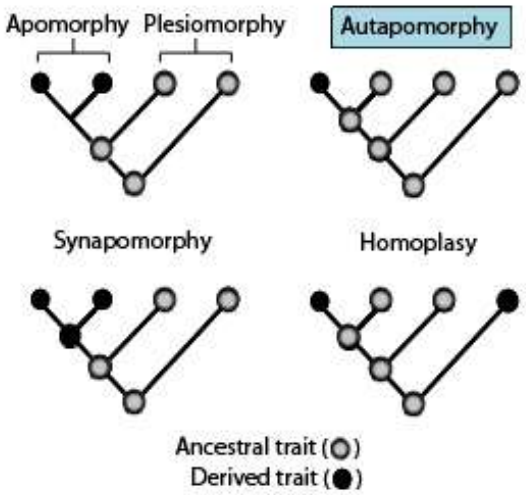


Figure 5 Trees showing the terminology used to describe different patterns of ancestral (o) and derived (•) character states. Taken from (Page, Holmes 2009)

The phylogenetic species in the diatoms can be assessed through cladistic analysis of their siliceous wall components (Methods in Kitching et al. 1998 and examples in Kooistra et al. 2010; Edgar et al. 2015; Pennesi et al. 2016), through geometric morphometric analyses (Beszteri et al. 2005; Edgar et al. 2015; Urbánková et al. 2016) or through sequence analysis of one or more genes (Medlin 2016b; Theriot et al. 2015), or combinations of these methods. (Medlin 2018) Recent descriptions of new diatom taxa are supported both with morphological and molecular data of examined strains.

According to Medlin (2018) molecular data can (1) identify multi-species complexes (cryptic species) and help better define a species' limits, (2) provide an objective framework upon which to interpret the taxonomic level to which physiological and morphological differences can be applied, (3) interpret gene flow and dispersal mechanisms and (4) depict the phylogenetic history of a group and interpret its biogeographic distribution.

Of crucial importance is that employment molecular data for taxonomic purposes presumes existence of reliable and globally accessible references. A major genomic repository is National Center for Biotechnology Information (NCBI) with databases of the International Nucleotide Sequence Database Collaboration which includes DNA

Data Bank of Japan (DDBJ), European Molecular Biology Laboratory (EMBL) and GenBank. Taxonomic information are available also in other biodiversity organizations e.g. Global Biodiversity Informaion Facility (GBIF) or Barcode of Life Data System (BOLD).

The most important task of molecular taxonomy is to find suitable genetic marker – a gene or DNA sequence that is specific for species. Portions of the nuclear rDNA cistron remain the most widely sequenced markers for many organisms, including diatoms. Small subunit (SSU or 18S) rDNA is useful for reconstructing higher level relationships across the entire phylogeny of diatoms (e.g., Alverson et al. 2006; Sorhannus 2004), whereas the large subunit (LSU or 28S) D1–D3 and internal transcribed spacer (ITS) regions can resolve species and sometimes population-level relationships (e.g., Behnke et al. 2004; Beszteri et al. 2005b; Godhe et al. 2006; Vanormelingen et al. 2007; Vanormelingen et al. 2008). (in Alverson 2008)

DNA Barcoding represents a new approach and perhaps the most reliable framework for effective identification employing sequence diversity in short, standardized gene regions to aid species identification and discovery in large assemblages of life (Sujeevan 2007). Goldstein and DeSalle (2019), analyzed 3756 papers from over 15 years to detect the extent to which 'its purposes, premises, rationale and application have evolved'. However, finding suitable markers for the identification of unicellular organisms has been difficult (Kuksa et al. 2009), and several taxonomic groups still require the use of several different DNA markers. This technique helped to recognize cryptic species or different developmental life stages of a single species, which was impossible by using morphological characters alone (Pečnikar 2014).

The most commonly used markers have been parts of the genes coding for ribosomal RNAs, in particular 18S rDNA. The advantages of 18S rDNA are that they are found in all eukaryotes, occurs in many copies per genome, allowing genetic work at the individual (single-cell) level; it is highly expressed, permitting molecular ecological investigation at the RNA level; and it includes a mosaic of highly conserved and variable nucleotide sequences allowing combined phylogenetic reconstruction and biota recognition at various taxonomic levels. (Pawlowski 2012)

18S rDNA barcodes are, however, not sufficient to delimit all diatom taxa, and various alternative protistan DNA barcodes have been proposed: the D1–D2 and/or D2–D3 regions at the 59 end of 28S rDNA, ribosomal internal transcribed spacers (ITS1 and/or ITS2 rDNA, the mitochondrial gene coding for cytochrome oxidase 1 (COI), the large subunit of the ribulose1,5-biphosphate carboxylase–oxygenase gene (rbcL) and the chloroplastic 23S rRNA gene for photosynthetic protists. (Pawlowski 2012)

A two-step barcoding approach suggest preliminary identification using a universal eukaryotic barcode, called the pre-barcode, followed by a species-level assignment using a groupspecific barcode (Figure 6). In this nested strategy, the 500 bp variable V4 region of 18S rDNA is proposed as the universal eukaryotic pre-barcode and subsequently group-specific barcodes shall be used. (Pawlovski 2012)

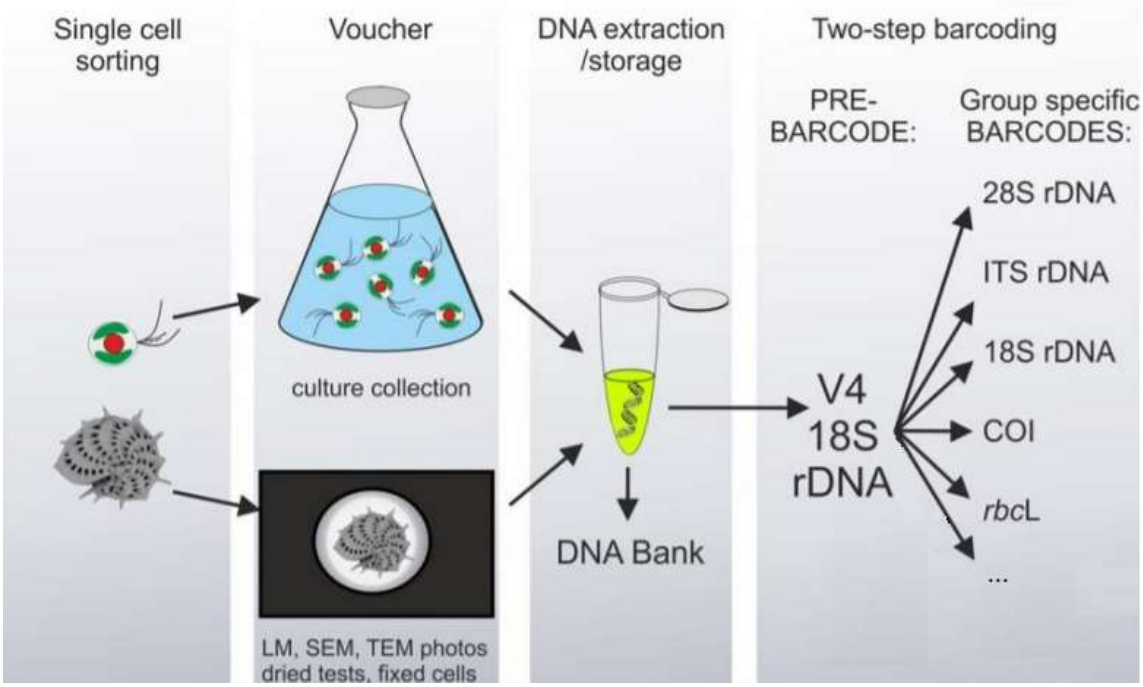


Figure 6. Two-step protist barcoding pipeline. Protistan species, spanning four orders of cell-size magnitude (from ,1 µm to .10,000 µm), are individually sorted from the environment, phenotyped either directly or after culture growth, DNA extracted, and barcoded using a twostep, nested strategy. (Pawlowski 2012)

However, not all biologists fully share barcoding optimism. Manoylov (2014) sums up several reasons why it is not solely possible to adopt molecular bioassessment of algal community composition: The identity of most taxa in the reference libraries of genetic sequences has not been rigorously evaluated, so the reference library taxonomy may not be accurate. Because the importance of correct identification for bioassessment has been established, it is necessary to mention that the complexity of algal shapes, adaptations, and survival cannot be simply translated into a DNA code because the expression of proteins depends on the unique set of environmental conditions in which algae live (Will and Rubinoff 2004, Sluys 2013). A large initial investment is required for equipment, or the price of analyses must be lowered. A low number of reads (<100) and reads with non-compatible tag combinations (Carlsen et al. 2012) are common and depend on the sequencing platform used. Manual inspection of BLAST searches (Altschul et al. 1997) and GenBank (Benson et al. 2012) to identify “contaminant sequences” is also required. Interpretations of sequences that have switched tags at both ends (Carlsen et al. 2012) could produce problems. Descriptive statistics for molecular variations have not been developed and tested to include the number of reads, alignment sites, and haplotypes; the haplotype diversity; the nucleotide diversity; or the average number of nucleotide differences.

The Barcode of Life Data System (BOLD) is an informatics workbench aiding the acquisition, storage, analysis and publication of DNA barcode records. By assembling molecular, morphological and distributional data, it bridges a traditional bioinformatics chasm. BOLD is freely available to any researcher with interests in DNA barcoding. By providing specialized services, it aids the assembly of records that meet the standards needed to gain BARCODE designation in the global sequence databases. The Barcode of Life Data System (bold) — www. barcodinglife.org — provides an integrated bioinformatics platform that supports all phases of the analytical pathway from specimen collection to tightly validated barcode library. Although bold aids the assembly of barcode data and maintains these records, a copy of all sequence and key specimen data also migrate to NCBI or its sister genomic repositories [DNA Data Bank of Japan (DDBJ), European Molecular Biology Laboratory (EMBL)] as soon as results are ready for public release. Key features include the requirement for a persistent linkage between a barcode sequence and its source specimen and a secure

environment that stores, organizes and queries these records, accessible to the entire biodiversity community. (Sujeevan 2007)

The Norwegian Barcode of Life (NorBOL) is a national network of research institutions for collaboration on DNA barcoding in Norway and a regional node in the International Barcode of Life Project (iBOL). The goal was to barcode 20 000 species within 2018 and make all data available in the Barcode of Life Data Systems (BOLD). However not specified on protists, NorBOL partner organization Tromsø Museum accepted samples of target diatom specimens and via the Canadian centre of DNA Barcoding (CCDB) supported author with sequences of the diatom samples for this master thesis.

Because the samples were accepted as an experimental material for testing of barcoding methods, it was not possible to decide for the genetic marker and originally suggested 18S rDNA was changed. Cooperation with CCDB had also other limiting terms that influenced the quality of phylogenetic part of this thesis.

Goals of this master thesis:

1. To support the reader with a brief overview of taxonomic issues connected to diatom species identification.
2. To describe methods used for recognition of five given species from North Atlantic waters.
3. To analyze morphologic and genetic data for confirming preliminary species estimations.
4. To discuss the reliability of results on a background of chosen methods.
5. To evaluate barcoding method and propriety of using this method in diatom taxonomy.

The present study was performed on species that were sampled in high-latitude areas with specific characteristics, mainly seasonal variations in the geophysical viariables: light, temperature and salinity as summed up in Degerlund (2011).

2. Material and methods

2.1 Origin, isolation and cultivation of strains

Five strains of different centric diatom species were chosen for taxonomic screening as shown in Table 1. The strains were kept as part of a multi-species diatom stock collection at UiT The Arctic university of Norway.

Ingebrigtsen (2010) previously described strains of *Attheya longicornis* Crawford & Gardner and *Chaetoceros furcellatus* Yendo. Since they have been kept for a long period of time in culture, new taxonomic descriptions can reveal differences between previous and recent characteristics caused by the effect of small scale cultivation (hypotheses supported also by Lakeman et al. 2009). Strains of *Porosira glacialis* Jørgensen and *Coscinodiscus sp*. were isolated from samples taken during research cruises in the Arctic in 2013 and 2014, and the preliminary identification need confirmation.

A strain of *Skeletonema marinoi* Sarno & Zingone was isolated in January 2015 as a contaminant in another culture in the collection. Taxonomic screening can proof if it originates from neighbouring, previously cultivated *S. marinoi* strain nr 86.

All cultures in the collection were grown from single cell isolates. Strains 61, 201 and 203 were isolated directly on board of cruise ship; strain 20.2 was isolated after germination of sea sediments cultivated in Guillard's f/10 medium (Ingebrigtsen 2010).

Table 1 Target strains

Strain number	Preliminary identification	Origin	Date
20.2	*Attheya longicornis*	sediment from N.Norway	November 2006
61	*Chaetoceros furcellatus*	78 41.92 N ; 10 14.95 E	18.05.2007
201	*Porosira glacialis*	76 27.54 N; 33 03.54 E	05.05.2014
203	*Coscinodiscus sp.*	Porsangerfjord	October 2013
204	*Skeletonema marinoi*	Håkøybotn, Tromsø	01.12.2008

The strains in collection were kept in a Termaks incubator with integrated LED light source, preset to 14:10 day:night light regime at intensity of ca 20 *µmol* m^{-2} s^{-1}. Temperature was 4 ˚C. The cells were kept in glass Duran™ GL18 culture tubes (15 ml) and were checked for wellness, density and contamination at 7-10 day intervals. Maintenance of dense cultures included removing ca 2/3 – 3/4 of vial content and supplying Guillard's f/10 medium and Silicate solution, omitting HCl as described by Ingebrigtsen (2010).

2.2 Sample preparation for morphological screening and scanning electron microscope inspection: live samples and cleaned frustules

Live samples for light microscopy were placed into 2 ml Nunc™ 4-well polystyrene chambers and let for sedimentation for at least 2 hours. Subsequently the samples were observed with Zeiss Primo Vert and Leica inverted microscopes at magnification from 100x to 400x.

In order to obtain frustules free from organic material a modified von Stosch's method (Tomas 1997) was used as described in Appendix 2.

Results from preliminary scanning done by Zeiss Sigma field emission scanning electron microscope revealed need for efficient drying for the species with setae. Additional drying was performed by Leica EM CPD300 Automated Critical Point Dryer. Protocol for drying and fixation of the samples for SEM is attached as Appendix 3.

All images made by LM was taken by me, except for Figure nr. 16 and 17, that Nerea Aalto is author of. SEM images were taken under the supervision of Tom-Ivar Eilertsen.

2.3 Sample preparation for molecular screening

Prior to sample take out all strains were cultivated in a climate-controlled room at 5±0.5°C and 50 µmol m^{-2} s^{-1} scalar irradiance and photoperiod of 14:10 h (light:dark).

An aliquot of ca 5 ml of *Attheya* stock culture strain nr. 20.2 was transferred into 40 ml Nunclon polystyrene topped up with Guillard's f/10 medium enriched with silicate

solution, and cultivated for ca 3-5 days. When visibly dense, the volume of the monoculture was increased to 160 ml, 500 ml and 2 l – the whole process performed three times. Biomass from each 2 l bottle was then filtrated onto Whatman® Cyclopore® polycarbonate membrane filters (Sigma-Aldrich), pore size 0,2 µm and then freezed at - 20°C. The same process was applied to the *Chaetoceros* strain nr. 61.

Coscinodiscus strain nr. 203 was grown at the same conditions up to 500 ml, but cultivation up to 10 l in glass beakers in 3 replicates. Samples were harvested by phytoplankton net mesh size 20 µm and residual water was centrifuged at 3500 rpm for 5 minutes. The pellet was transferred into 2 ml Eppendorf tube and stored at -80°C.

Cultures of *Porosira* (nr. 201) and *Skeletonema* (nr.204) strain were grown up to 100 l in Plexiglas cylinders. The culture medium was prepared from filtered (0.22 µm), pasteurized, local seawater (Tromsø Sound, 25 m depth) by adding silicate (final concentration 12.3 µM) and a commercial nutrient mixture (SubstralTM, 0.25 ml l^{-1}; The Scotts Company [Nordics] A/S, Denmark). All cultures were aerated with compressed air to avoid sedimentation and CO_2-limitation. Culture samples were collected by first concentrating cells onto a plankton net (mesh size 5 µm for *Skeletonema* and 20 µm for *Porosira*) before centrifuging at 3500 rpm for 5 min. The obtained wet pellets were transferred into 50 ml Falcon tubes and stored at -80°C.

2.4 Molecular processing

Canadian Centre for DNA Barcoding (CCDB) decided on barcoding methods using two molecular markers: rbcL gene of chloroplast and 16S rRNA, however 18S rDNA was suggested by submission of the samples. Protocols of DNA extraction, amplification and sequencing are attached to the present text as separate files.

DNA was extracted using the manual protocol for plants as described in (Ivanova et al. 2008) with minor modifications: residual liquid (leftovers of ethanol) was evaporated at 56°C. Samples were homogenized in TissueLyzer in the presence of sterile sand

(Acros Organics, Cat.No. 370942500) and stainless steel beads. Insect lysis Buffer (ILB without PVP) was used for the lysis stage.

Pre-made PCR mixes with Platinum Taq (Ivanova & Grainger 2007; Hebert et al. 2013) were used to amplify rbcL and 16S rDNA with the following primers: rbcL gene of chloroplast– cfD (Hamsher et al. 2011) /DPrbcL7 (Levialdi Ghiron, 2006). 16S rRNA was amplified with combination of primers CYA106F-GC/CYA781R (49 bands) (Nubel et al. 1997) CYA781R is a primer cocktail of CYA781R-a and of CYA781R-b.

CYA106F-GC/CYA781R and rbcL amplification product were used for sequencing. PCR thermocycling conditions followed (Nubel et al. 1997; Hamsher et al. 2011).

2.5 Sequence data analysis

Through the boldsystems.org CCDB provided sequences of four target species that were checked for the quality, trimmed and pre-processed so that they were ready for alignment. Sequencing was not successful for the *Coscinodiscus* strain.

A data set of 82 sequences was created from CCDB *rbc*L sequences of target species and additional sequences published in NCBI GenBank database, i.e. it included relevant ingroup and outgroup species related to the target species (Appendix 4). When using Basic Logical Alignment Search Tool (BLAST) for searching similar GenBank sequences with CCDB 16S rRNA sequences, no relevant matches were obtained for the examined strains of *Porosira, Skeletonema* and *Chaetoceros* and for that reason these were not included into the phylogenetic analysis.

Phylogenetic analyses were conducted using the MEGA 7.0.26 software (Tamura et al. 2011). Alignment of sequences for *rbc*L marker was created by using ClustalW algorithm. The aligned dataset went through the testing of proper method calculation. The table of scores for each of the statistical criterion is attached as Appendix 4. General Time Reversible model with Gamma correction and Invariant sites (GTR+G+I) evolution model was estimated to be the best fitting model according to computed

Bayesian Information Criterion (BIC score), Akaike information Critierion value (AIC) and Maximum likelihood value (*lnL*). The phylogenetic tree was constructed by using Maximum likelihood (ML) statistical method with branch support of 1000 bootstrap replicates.

3. Results

Species identification based on morphology

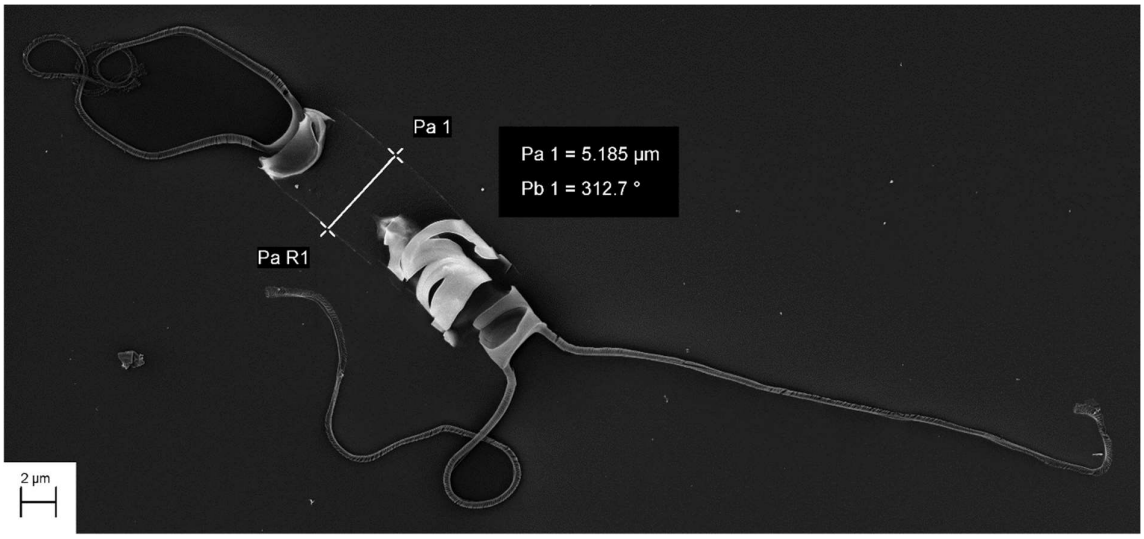
3.1 Morphological analysis of *Attheya* strain nr. 20.2

EM figures shows presence of two long horns extending out from each valve (Fig. 7, 8). Horns consists of numerous siliceous bands (Fig. 9a) that outcome from supporting rods – longitudinal strips (Fig.9b) finished at distal end by spines (Fig.9c). Horns of examined strain are rather straight, than wavy.

The quality of the picture does not allow counting more than 3 longitudinal strips, which are running rather straight, not spiraling throughout the length of the horns. Several terminal spines are visible on their bifurcated ends. It is not visible any process or spine disturbing the homogenous structure of the valve (Fig. 10), unfortunately shown only from the inner side. Very fine-structured girdle bands (Fig. 11) construct mid, elongated part of the cell.

Light microscopy demonstrates single-cell species; presence of one or two chloroplasts, formation of clumps, variation of the cell size and the different ratio between length (according to apical axis) and width (according to transapical axis) of the cells. (Fig. 13-15)

Throughout two years of observation, horns were not significantly long reaching the ration 8-10 x of the cell size; however, in newly diluted, richly fertilized cultures longer and wavy horns appeared.



Graphic visualization of the *Attheya* strain nr. 20.2 Fig 7. SEM Whole cell with visible elongated horns. Scale bar, 2 µm.

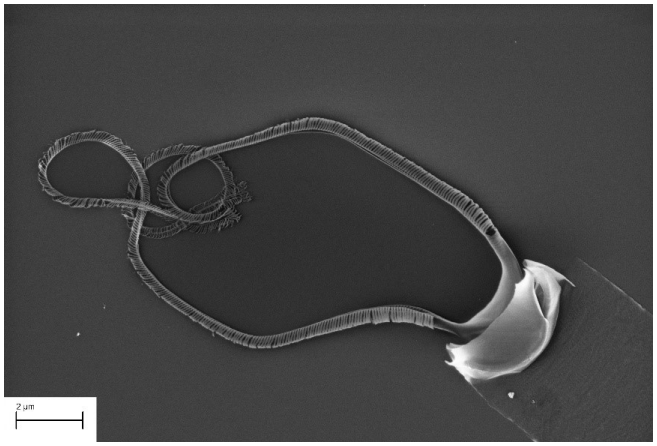


Fig. 8 SEM detail of the horn with siliceous bands. Scale bar, 2 µm.

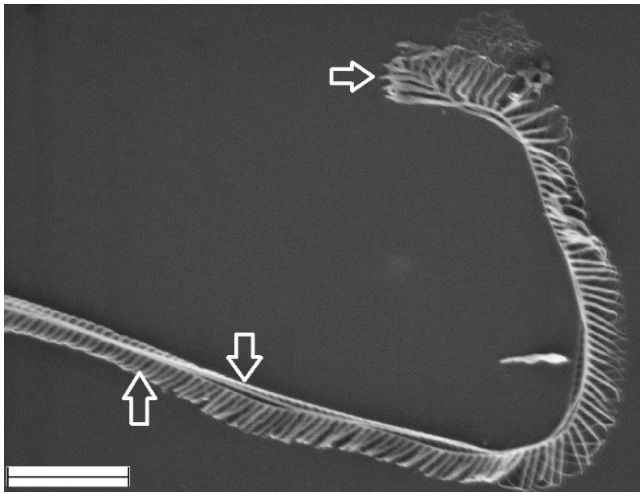


Fig.9 Detail of siliceous bans (9a) supporting rods (9b) and terminal spines (9C).Scale bar, 1 µm.

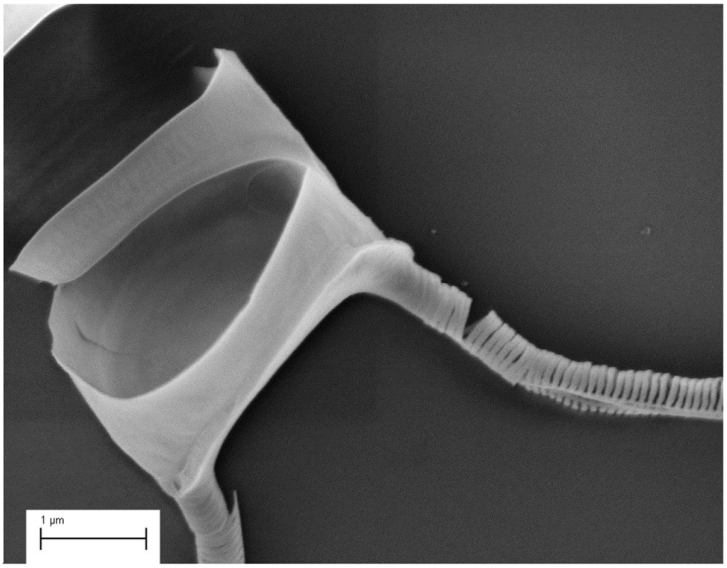


Fig. 10 SEM valve detail. Scale bar, 1 µm

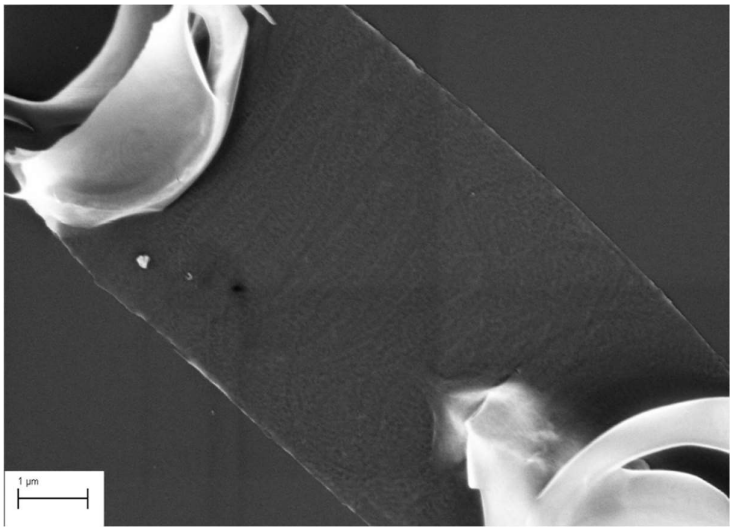


Fig. 11 SEM detail of fine-structured girdle bands. Scale bar, 1 µm.

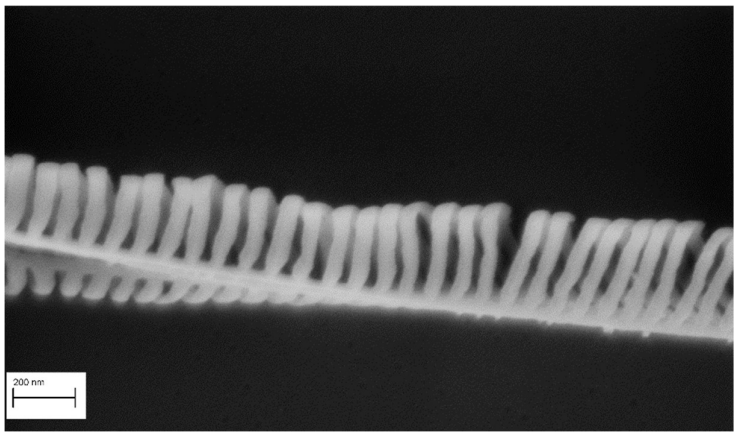


Fig. 12 SEM detail of siliceous horn bands. Scale bar, 200 nm.

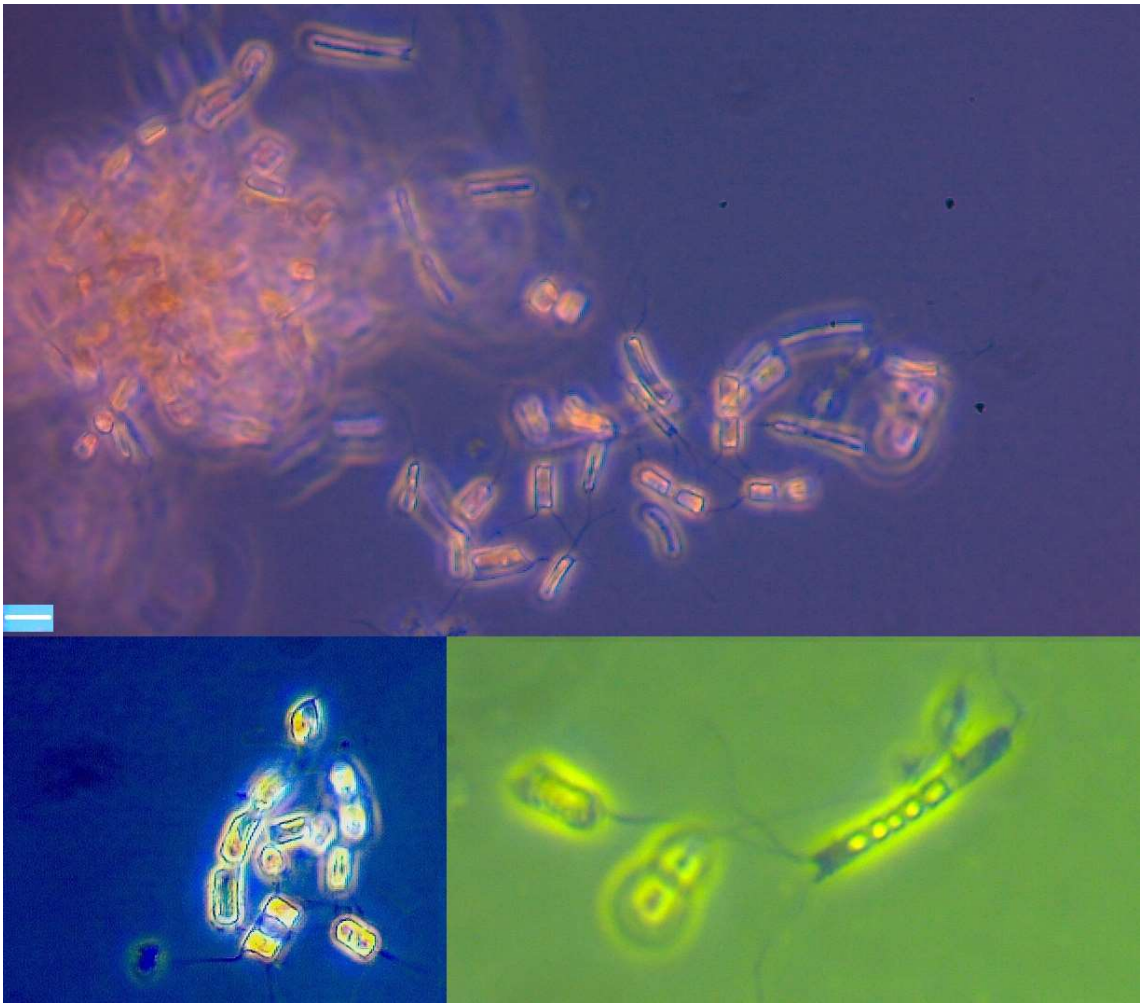


Fig. 13-15. LM images of *Attheya sp*. nr. 20.2 representing the variability in size and shapes Fig**. 13**. Tightly clumped colony of mostly single cells.**Fig. 14.** Nearly square cells at the bottom of the image with horns resembling more *Attheya septentionalis*. **Fig. 15** Elongated cell with unfinished division process. Scale bar, 10 µm.

3.2 Morphological analysis of *Chaetoceros* strain nr. 61

Vegetative cells were mainly organized in straight chains (Fig. 16), containing one large chloroplast (Fig. 17). Resting cells exists with paired spores (Fig. 16, 17).

Valve face were oval to round (with decreasing cell size), terminal valve (Fig. 18) had eccentric flattened tube, costae were radially branched (Fig. 18). It was possible to recognize central annulus. Girdle bands were straight, not arched, with longitudinal rib pattern (Fig.18).

Terminal setae of vegetative cells were thin (Fig.19), without visible branching, irregularly oriented toward chain axis. Micrograph of setae ultrastructure depicts spirally arranged capilli and spines without any larger solitary pores. (Fig. 20)

Due to short basal part of setae rising from slightly elevated valve margin, apertures among cells were narrow, slightly compressed in center or invisible at LM images.(Fig.17 and 19).

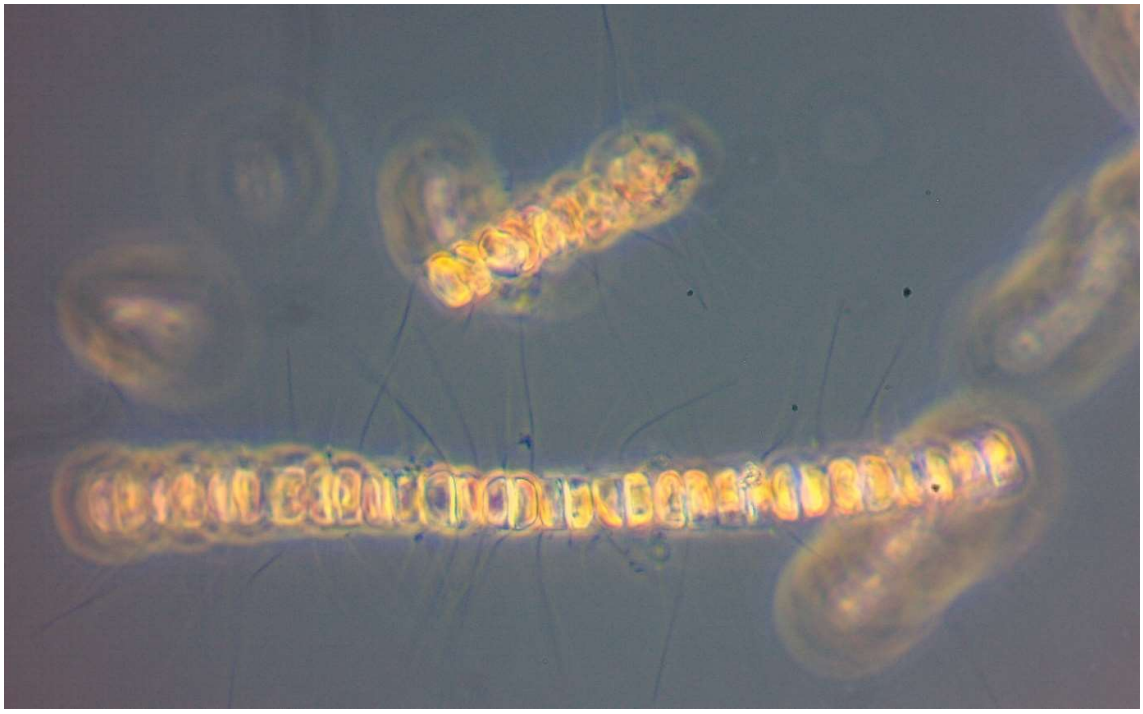
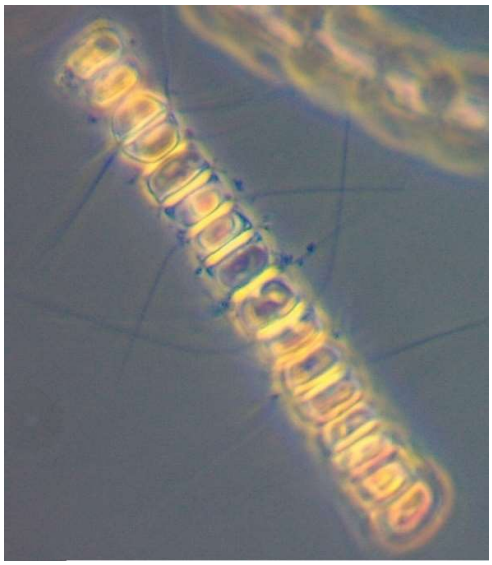


Fig. 16 Chaetoceros strain nr. 61. Long straight chain with initiate process of resting spores formation.



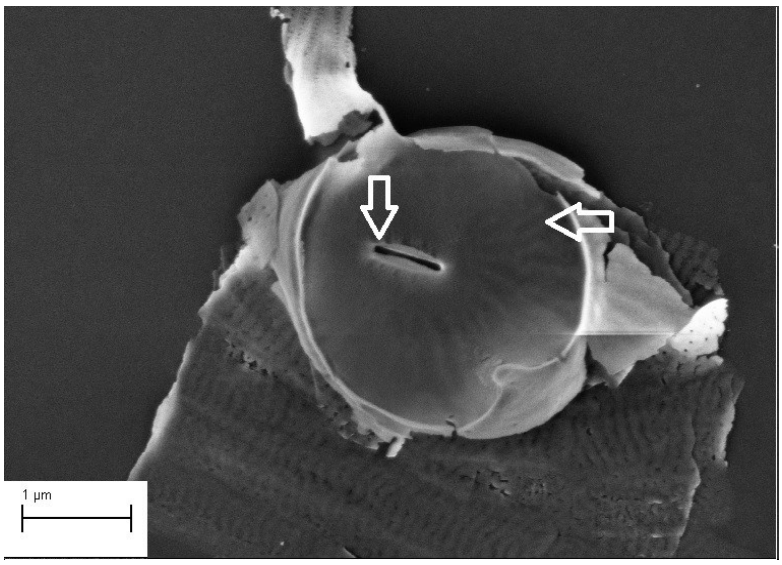


Fig.17 Chaetoceros strain nr. 61, LM image. Cells in the middle with one chloroplast. **Fig. 18** EM of the valve face Chaetoceros strain nr.61. Left arrow - flattened tube, right arrow - radially oriented costae. Scale bar, 1 µm.

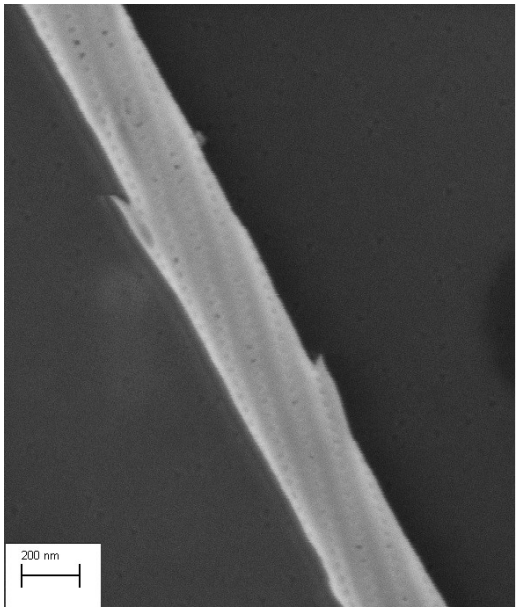


Fig. 19 EM micrograph of Chaetoceros strain nr. 61. Basal part of the setae short, setae long and thin without branching. Micrograph influenced by unsufficient drying. Scale bar, 10 µm. **Fig.20** EM micrograph. Detail of setae with spirally arranged capilli and spines. Scale bar, 200 nm.

3.3 Morphological analysis of *Porosira* strain nr. 201

The strain had single discoid cells (EM micrograph figure 21), valve with central annulus and radial areolation that was not organized in straight, but rather wavy lines. Labyrinth pattern of costae arranged around areola is visible at figure 22.

One large labiate process was in the marginal zone and numerous strutted processes were distributed radially all over the valve face. Around labiate process (figure 23) were strutted processes distributed rather unequally.

At LM images, strutted processes were visible mainly on newly divided cells (figure 24). Numerous chloroplast were distributed all over the cell (figure 24, 25). Cell size of cultivated strain measured with EM was 29,22 µm (by 6.4.2018). Figure 26 shows both of the valves and demonstrates different pattern of inner and outer side of the valve.

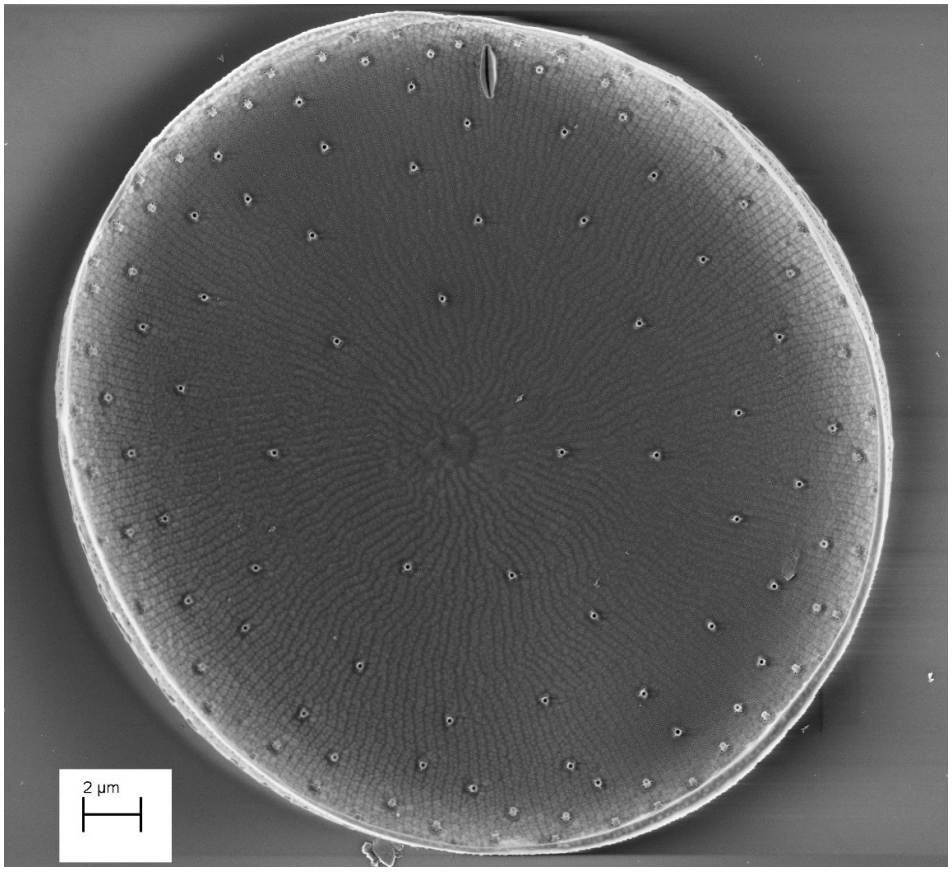
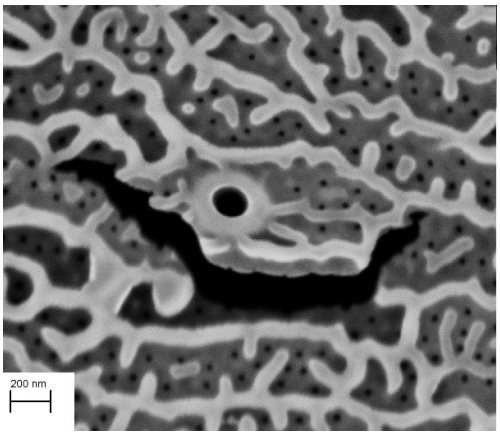


Fig. 21 *Porosira* strain nr. 201 EM micrograph. Valve face with central annulus, strutted processes, and radially wavy pattern of areolation. Scale bar, 2 µm.



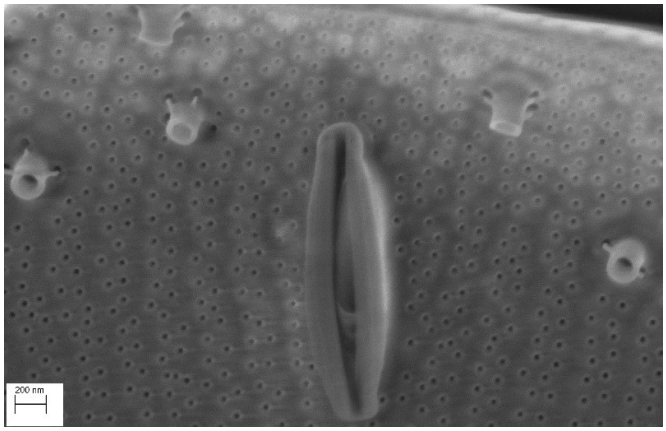
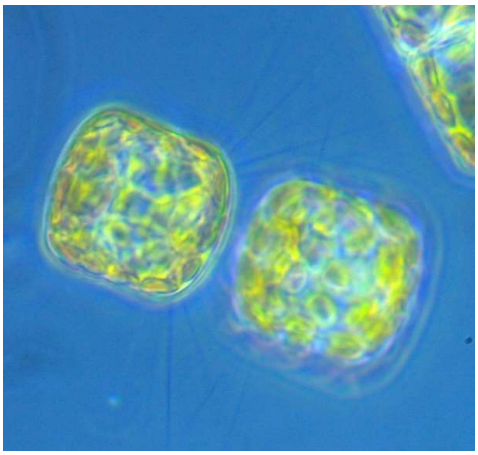


Fig. 22 EM micrograph, ultrastructure of the inner part of the valve. Section with detail of porous structures in the center and areolation surrounded by costae. **Fig. 23** EM detail of labiate process. Strutted processes arranged irregularly. Scale bar, 200 nm.



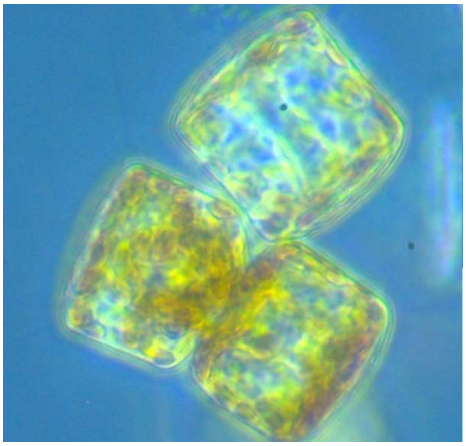


Fig. 24, 25 LM images of *Porosira* strain nr. 201, vegetative cells in girdle view with numerous chloroplasts.

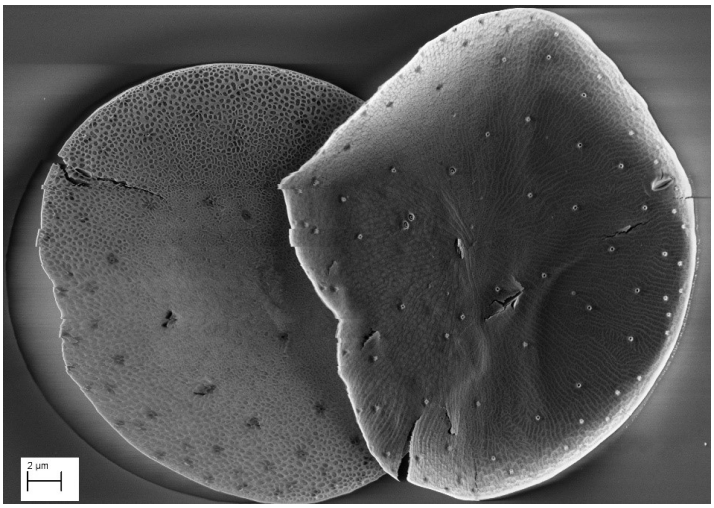
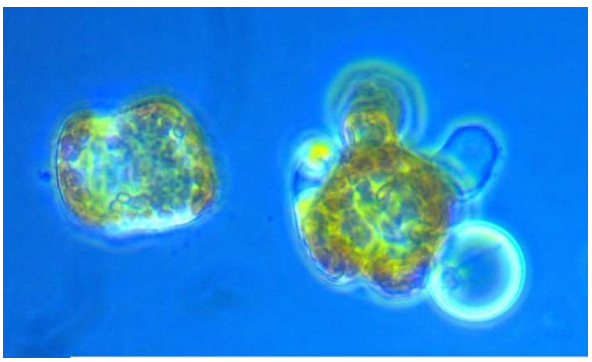


Fig. 26 Face (smoother, right) and inner (left) side of the valve. Scale bar, 2 µm.

The morphology of observed *Porosira* cells went through a radical change during sexual reproduction (auxospore formation). Cell content concentrated even in cells that went through the process of dividing (figure 27). Auxospore formation continued with creation of side bubble form (figure 27 right, 28) where the sexual cells matured (figure 29, 30) and afterwards were expanded into the environment (figure 31).



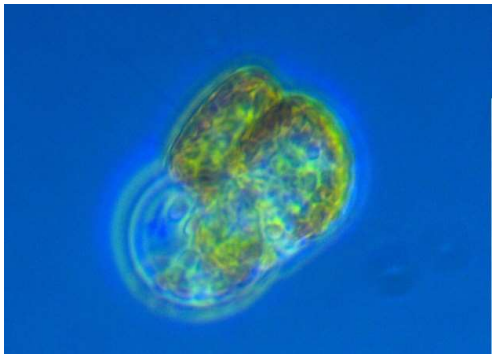
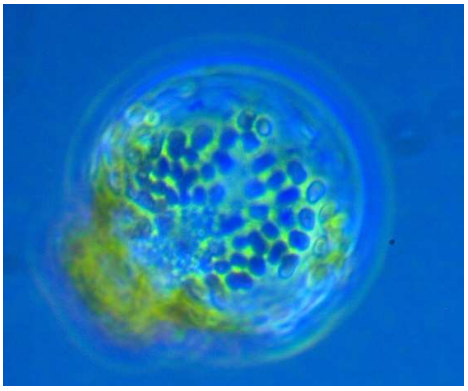


Fig. 27 Auxospore formation in *Porosira* strain nr. 201, LM. Left cell on the bottom visible change in the structure of cell wall. Right cell has concentrated cell content, fist appearance of side spherical form surrounded by membrane. **Fig. 28** New sexual cells fills the space of spherical form beside the cell that has just finished process of dividing.



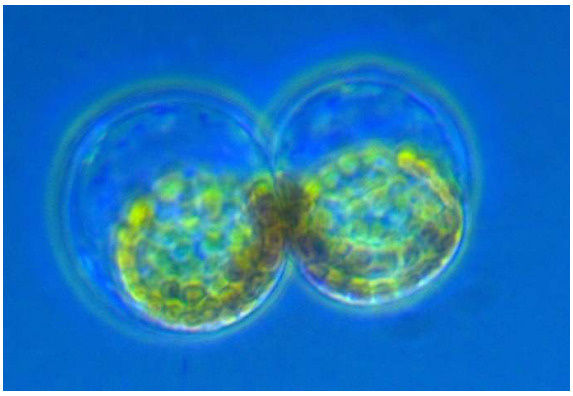


Fig. 29 Maturing cells of sexual reproduction. **Fig. 30** The same process later in progress. Original mother cell dettached.

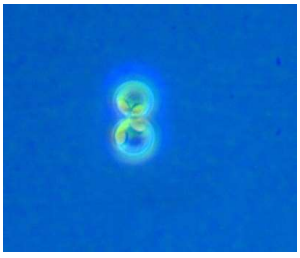


Fig. 31 Cells of sexual reproduction released into the water column.

3.4 Morphological analysis of *Coscinodiscus* strain nr. 203

Cells of the strain was large, single and discoid (figure 32), valve face was covered with radial rows of areolae (figure 34), with nonaerolate center (hyaline, figure 33). Rimoportula was not present. Smaller labiate processes (figure 35) created marginal ring round the cell together with two large labiate processes as visible at figure 32. Ribbon girdle bands at figure 36 with narrow tooth-like projection exceeded into the upper girdle band.

Light microscope images (figure 37 - 41) present different life stages of the cell and different cell structures; cell wall, chloroplasts, cytoplasm, cytoskeleton and lipid droplets.

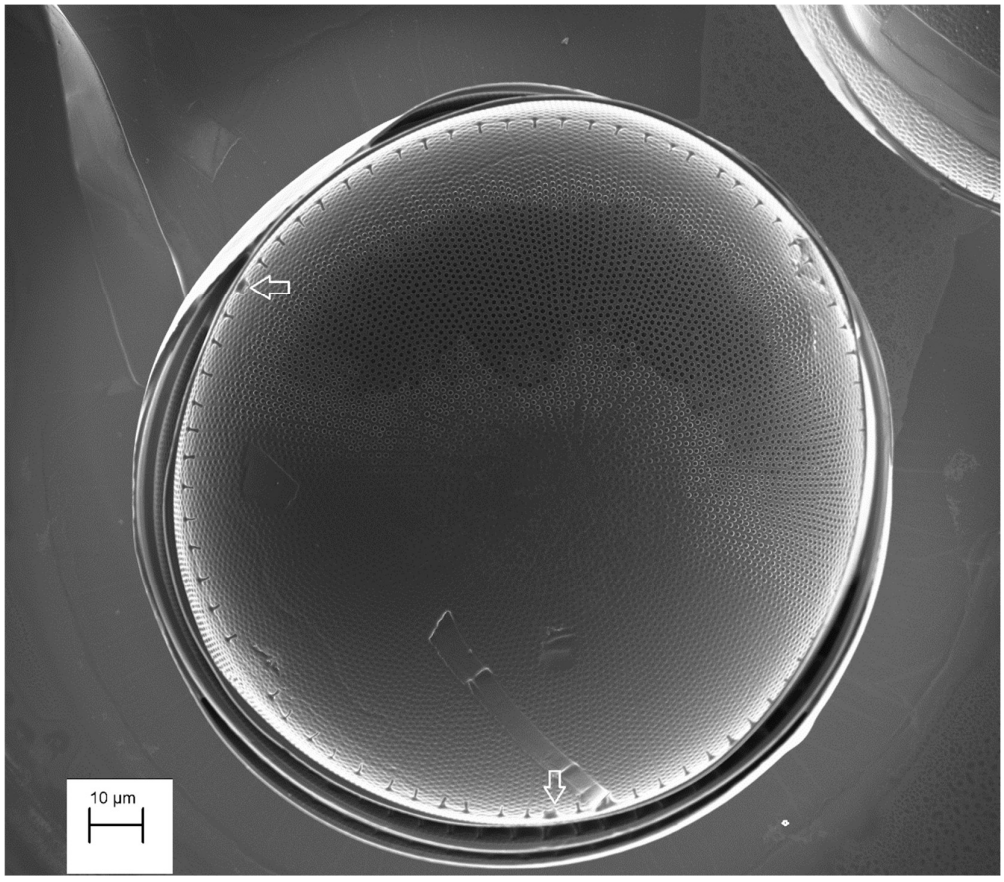


Fig. 32 *Coscinodiscus sp*. strain nr. 203, EM. Large discoid cell with one marginal ring of small labiate processes interrupted by 2 larger labiate processes (arrows). Scale bar, 10 µm.

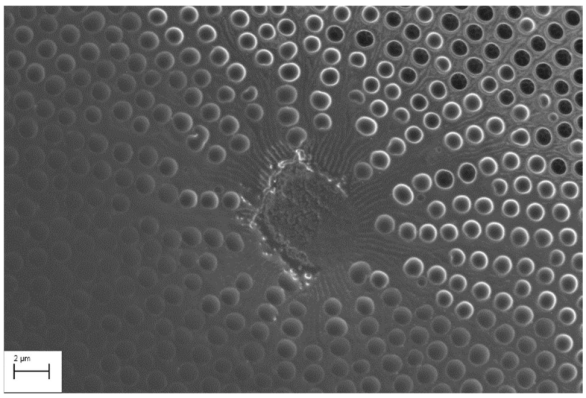
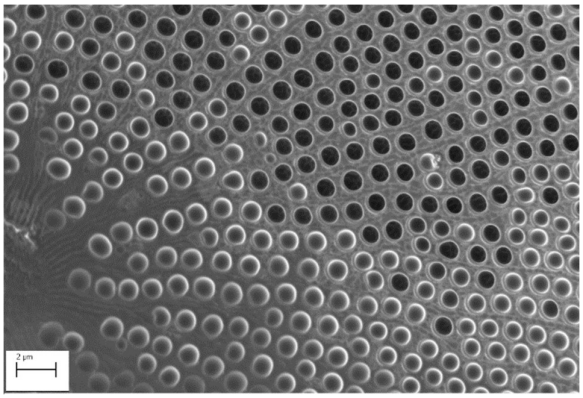



Fig. 33 Detail of hyaline central annulus.Scale bar, 2 µm. **Fig. 34** Radial areolation. Scale bar, 2 µm.

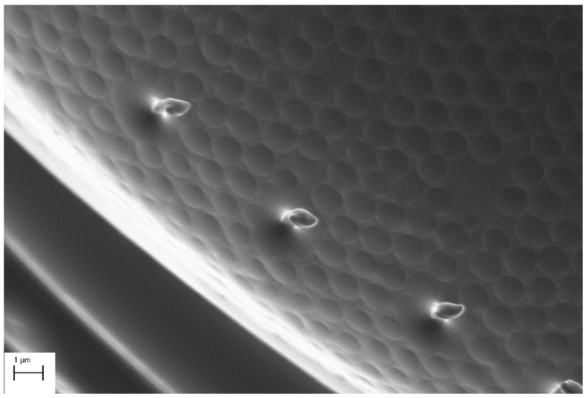
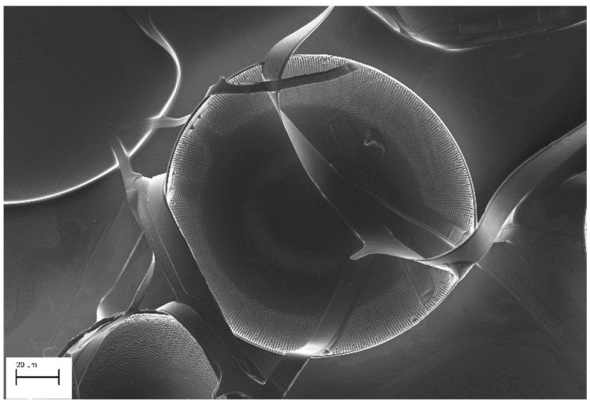
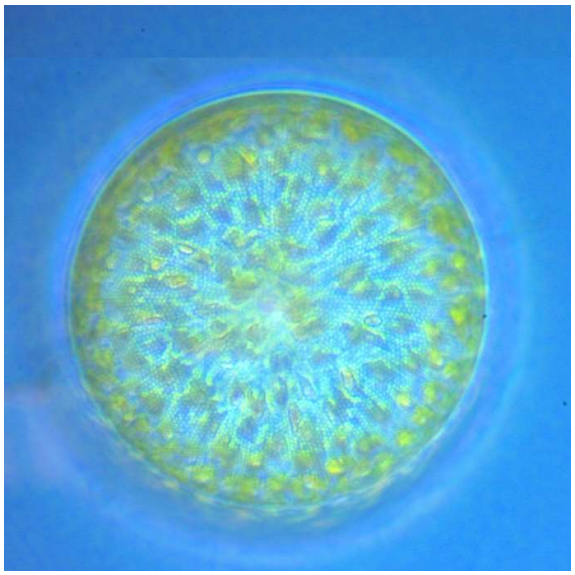
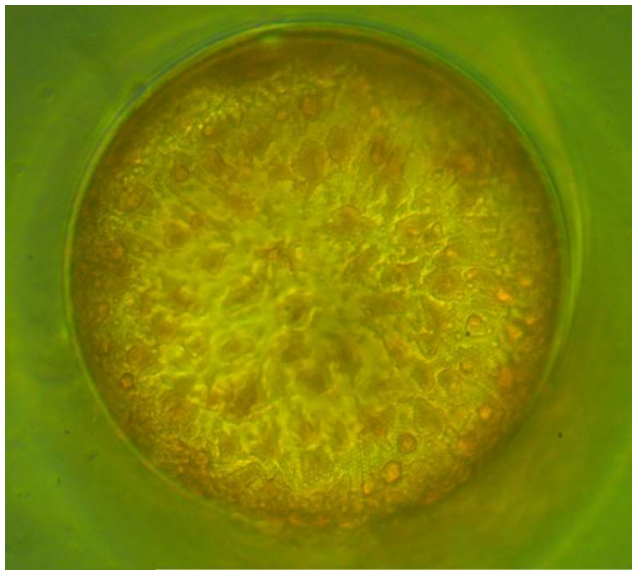



Fig. 35 Detail of small labiate processes at inner side of the valve. Scale bar, 1 µm. **Fig. 36** Girdle band in front of the cell valve. Scale bar, 20 µm.

LM Coscinodiscus strain nr. 203 Fig. 37 Valve view, zoom to the surface structure of the valve. Numerous chloroplasts spread around the whole cells. **Fig. 38** Valve view, young cell, zoom set to the inner cell structures.

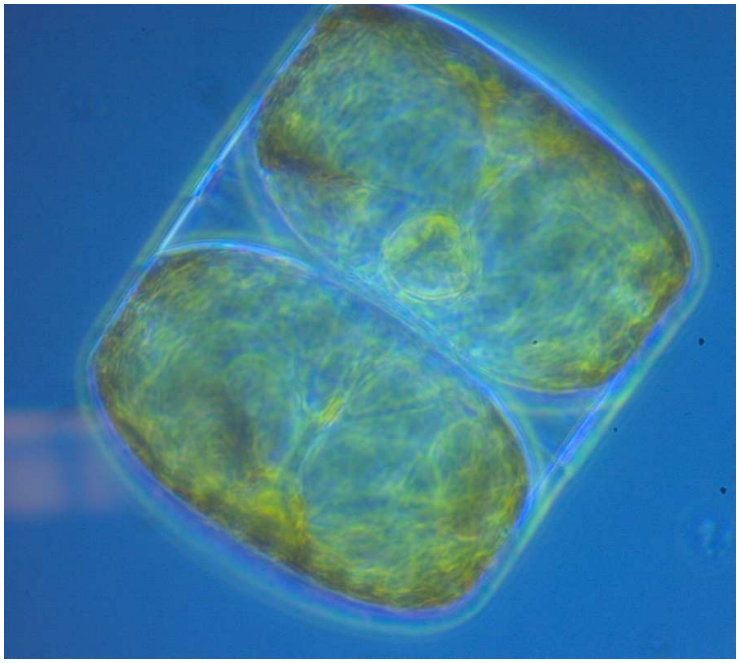
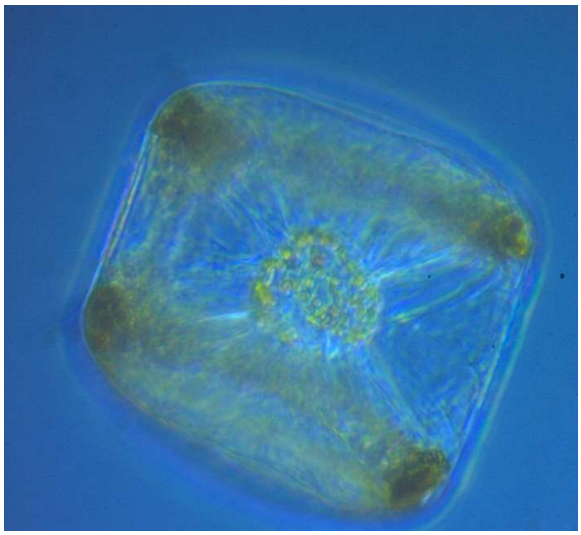
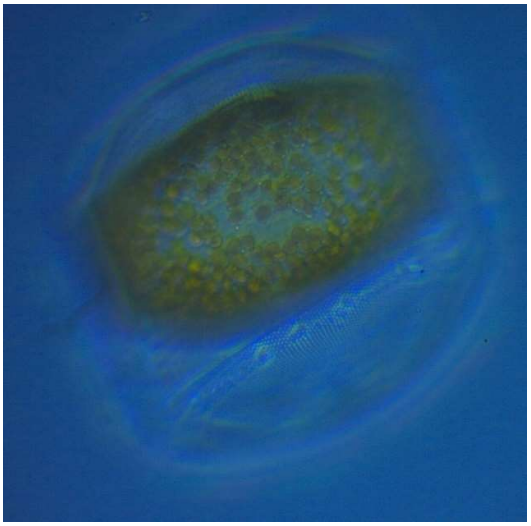


Fig. 39 Newly divided cell in girdle view. Borders of the cell covered by chloroplasts, cytoplasm and nucleus visible in the central part of the cell, this way possible to observe transport of translation substances.





Senescent cells in girdle view Fig. 40 Shape of the valve in the middle slightly convex. Emptying cytoskeleton in the middle of the cell. **Fig. 41** Concentration of chloroplasts and ergastic substances; valve space empty.

3.5 Morphological analysis of *Skeletonema* strain nr. 204

EM visualization of the strain nr. 204 (figure nr. 42) reveals *Skeletonema* species with its characteristic external processes (fultoportulae), which enable creation of specific locking connection with neighboring cell. Observed strain had 9 of such fultoportulae on the valve face (figure nr. 43) with flattened and flared distal ends (figure nr. 44).

Terminal rimoportula was placed close to the central annulus and had elongated process flared at the end. Whole valve was richly radially ridged. Girdle bands (figure nr. 45) ribbed with regularly repetitive lines of pores.

LM observation presents cylindrical cells with 1 or 2 chloroplasts that form long, rather straight or curved chain colonies of cells with different size and variable ratio between length and width.

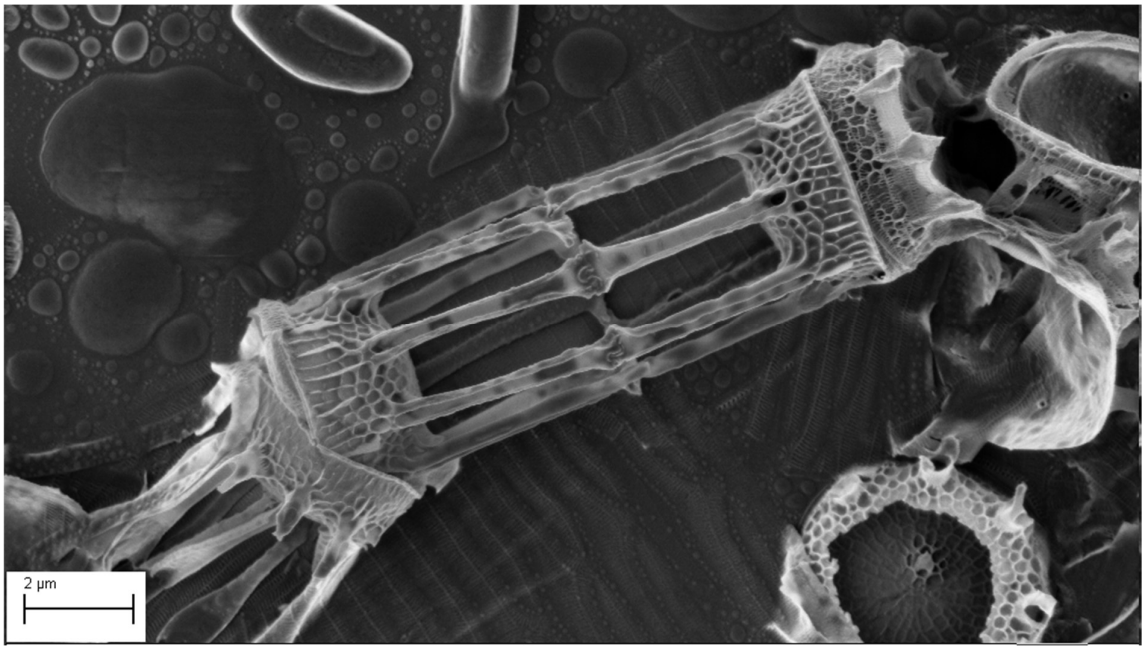
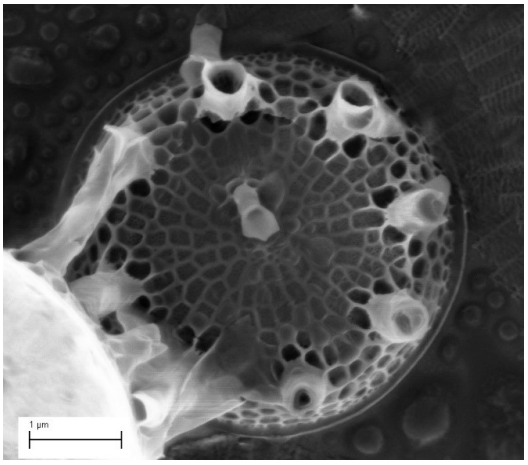


Figure nr. 42 Skeletonema strain nr. 204 Connection of two cells via rimoportulae exceeding from ridged valves. Scale bar, 2 µm.



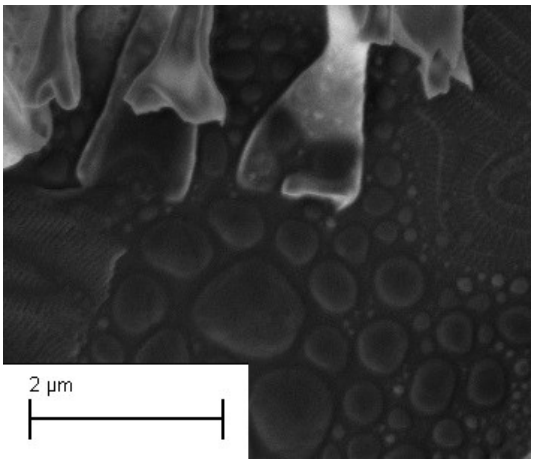


Fig. 42 Valve view. 9 rimoportulae in marginal zone, one excentric furtoportula with flared end and radial ridge all over the valve face visible at EM micrograph. Scale bar, 1 µm **Fig. 43** Detail of flared and flattened distal end of rimoportulae. Scale bar, 2 µm.

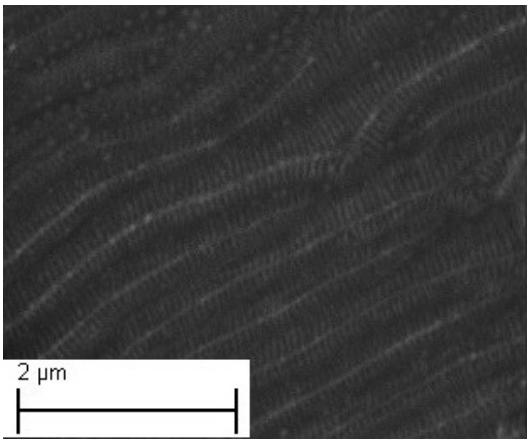


Fig. 44 Girdle bands repetitively ribbed with series of pores. Scale bar, 2 µm.

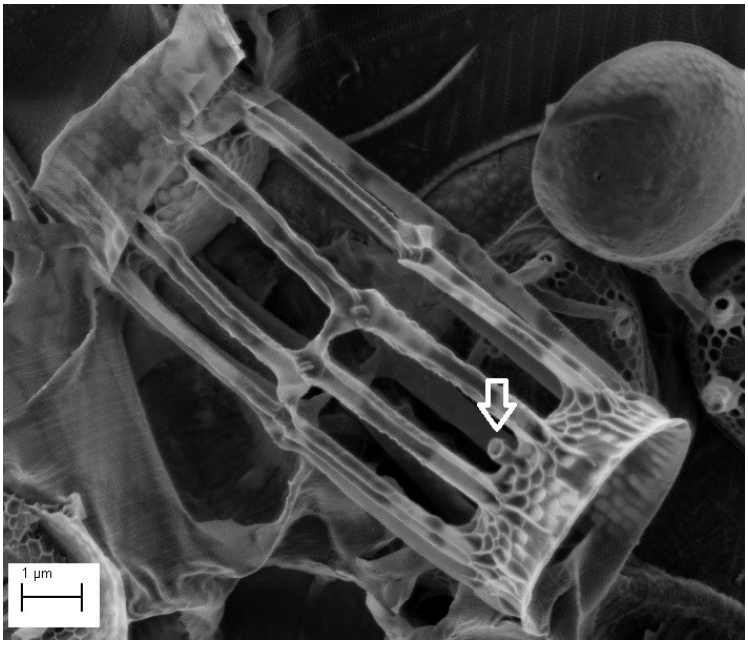


Fig. 45 EM Fultoportula with cup-shape end (arrow). Rimoportulae opened. Inner side of the valve at the right side. Scale bar, 1 µm.

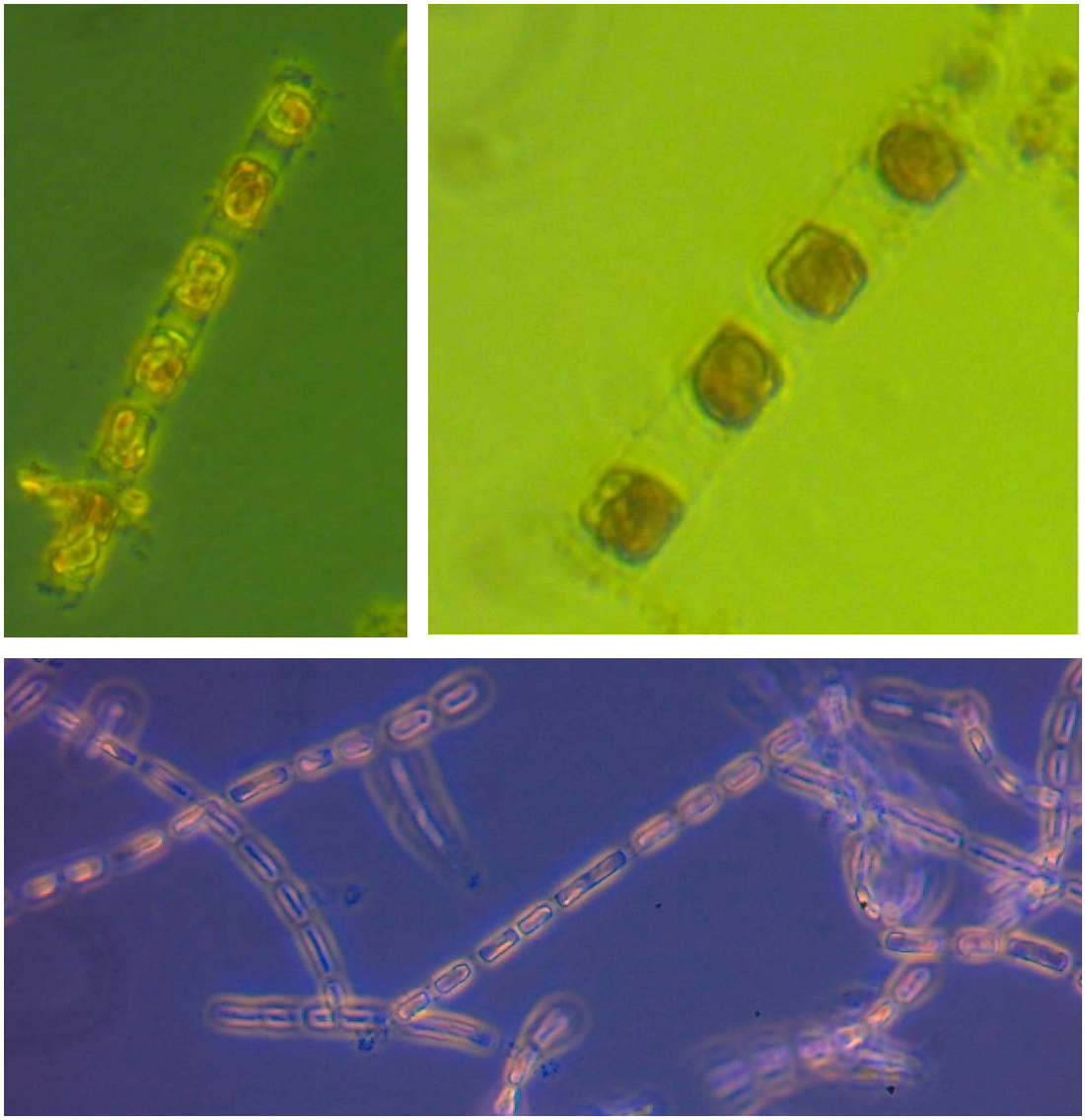


Fig. 46-48 LM Skeletonema strain nr. 204. Chain colonies of cells.

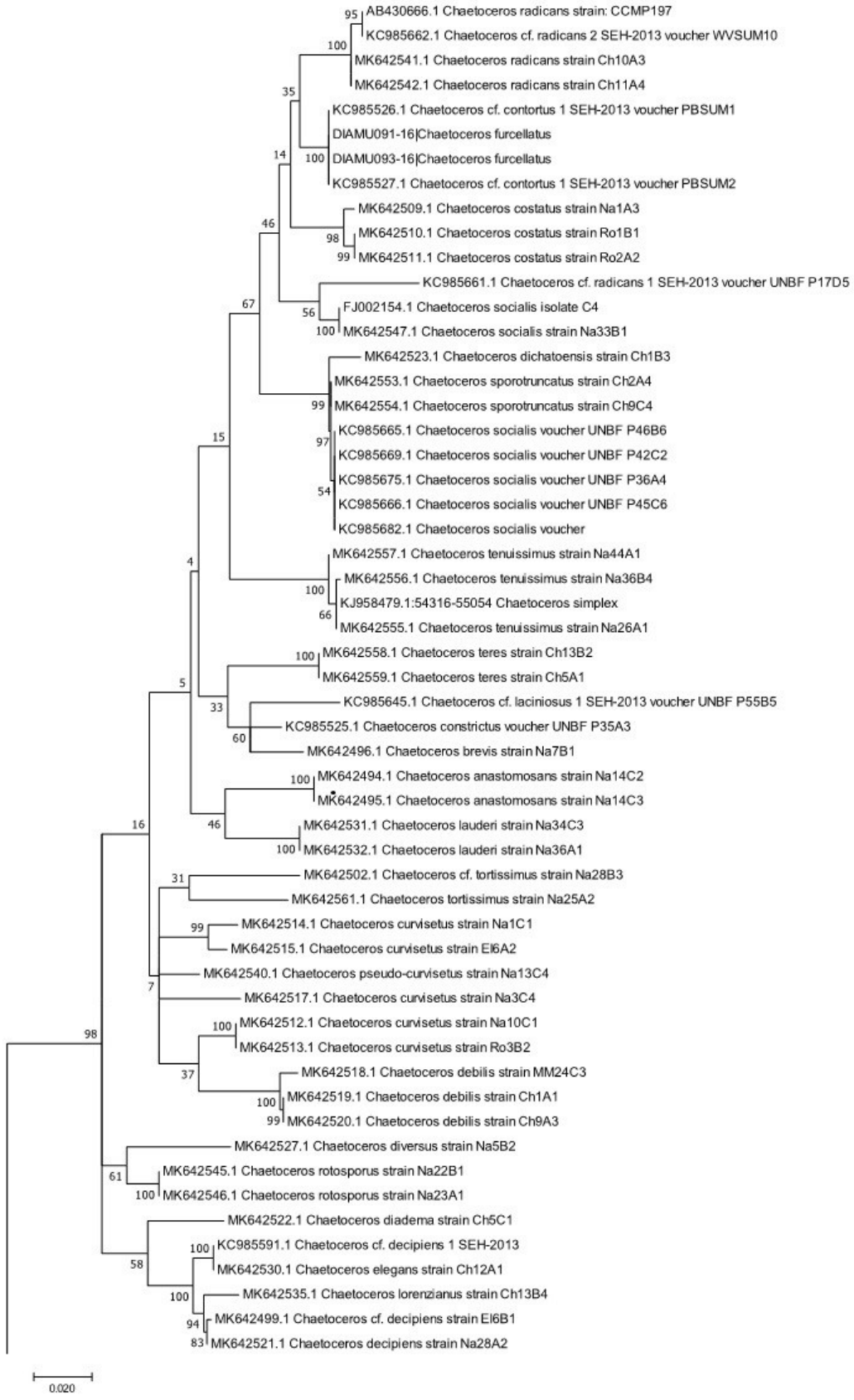
3.6 Phylogenetic analysis based on molecular data

Out of five replicates submitted to CCDB for each examined strain it was possible to receive five sequences of *Attheya* strain nr. 20.2, two sequences of *Chaetoceros* strain nr. 61 and one sequence of *Porosira* strain nr. 201 based on *rbc*L molecular marker (with DIAMU code in phylogenetic tree; Figure nr. 49).

Based on screening of *rbc*L gene of chloroplast genome, only two relevant results for strain nr. 20.2 were obtained. They were *Attheya longicornis* and *Attheya septentrionalis.* The examined strain is placed with 99 % of bootstrap support in a clade with three accessions of *A. septentrionalis* and three accessions of *A. longicornis*; also including the strain that was used for analysis of complete chloroplast genome. Within this clade there is also a subclude with 100% support including sequences of three other *A. septentrionalis* accessions from GenBank.

Phylogenetic analysis of *Chaetoceros* sp. nr. 61 included 2 sequences of target species and 52 sequences downloaded from GenBank database in order to create a broader outgroup and place correctly target species into its phylogenetic position within the *Chaetoceros* species complex in accordance with published phylogenetic literature (De Luca et al. 2019; Gaonkar et al. 2018).

With 100% support of the ML statistical metod, newly sequenced *Chaetoceros* stain nr. 61 (reported as *Ch. furcellatus* on phylogenetic tree) was identical with sequences of *Chaetoceros cf. contortus* and distinct from the closest *Chaetoceros radicans* (100% support based on 4 sequences). However, relationships among the groups of species within the genus *Chaetoceros* are poorly resolved. Monophyly of the *Chaetoceros* is supported by 98% boostrap value. Many *Chaetoceros* sequences, based on this phylogenetic analysis, form well supported clades 97 – 100 %, corresponded to morphological species. These results based on molecular data were set into the context with morphological analysis for final discussion and conclusion on interpretation of the species status.


AB430666.1 Chaetoceros radicans strain: CCMP197
KC985662.1 Chaetoceros cf. radicans 2 SEH-2013 voucher WVSUM10
MK642541.1 Chaetoceros radicans strain Ch10A3
MK642542.1 Chaetoceros radicans strain Ch11A4
KC985526.1 Chaetoceros cf. contortus 1 SEH-2013 voucher PBSUM1
DIAMU091-16|Chaetoceros furcellatus
DIAMU093-16|Chaetoceros furcellatus
KC985527.1 Chaetoceros cf. contortus 1 SEH-2013 voucher PBSUM2
MK642509.1 Chaetoceros costatus strain Na1A3
MK642510.1 Chaetoceros costatus strain Ro1B1
MK642511.1 Chaetoceros costatus strain Ro2A2
KC985661.1 Chaetoceros cf. radicans 1 SEH-2013 voucher UNBF P17D5
FJ002154.1 Chaetoceros socialis isolate C4
MK642547.1 Chaetoceros socialis strain Na33B1
MK642523.1 Chaetoceros dichatoensis strain Ch1B3
MK642553.1 Chaetoceros sporotruncatus strain Ch2A4
MK642554.1 Chaetoceros sporotruncatus strain Ch9C4
KC985665.1 Chaetoceros socialis voucher UNBF P46B6
KC985669.1 Chaetoceros socialis voucher UNBF P42C2
KC985675.1 Chaetoceros socialis voucher UNBF P36A4
KC985666.1 Chaetoceros socialis voucher UNBF P45C6
KC985682.1 Chaetoceros socialis voucher
MK642557.1 Chaetoceros tenuissimus strain Na44A1
MK642556.1 Chaetoceros tenuissimus strain Na36B4
KJ958479.1:54316-55054 Chaetoceros simplex
MK642555.1 Chaetoceros tenuissimus strain Na26A1
MK642558.1 Chaetoceros teres strain Ch13B2
MK642559.1 Chaetoceros teres strain Ch5A1
KC985645.1 Chaetoceros cf. laciniosus 1 SEH-2013 voucher UNBF P55B5
KC985525.1 Chaetoceros constrictus voucher UNBF P35A3
MK642496.1 Chaetoceros brevis strain Na7B1
MK642494.1 Chaetoceros anastomosans strain Na14C2
MK642495.1 Chaetoceros anastomosans strain Na14C3
MK642531.1 Chaetoceros lauderi strain Na34C3
MK642532.1 Chaetoceros lauderi strain Na36A1
MK642502.1 Chaetoceros cf. tortissimus strain Na28B3
MK642561.1 Chaetoceros tortissimus strain Na25A2
MK642514.1 Chaetoceros curvisetus strain Na1C1
MK642515.1 Chaetoceros curvisetus strain El6A2
MK642540.1 Chaetoceros pseudo-curvisetus strain Na13C4
MK642517.1 Chaetoceros curvisetus strain Na3C4
MK642512.1 Chaetoceros curvisetus strain Na10C1
MK642513.1 Chaetoceros curvisetus strain Ro3B2
MK642518.1 Chaetoceros debilis strain MM24C3
MK642519.1 Chaetoceros debilis strain Ch1A1
MK642520.1 Chaetoceros debilis strain Ch9A3
MK642527.1 Chaetoceros diversus strain Na5B2
MK642545.1 Chaetoceros rotosporus strain Na22B1
MK642546.1 Chaetoceros rotosporus strain Na23A1
MK642522.1 Chaetoceros diadema strain Ch5C1
KC985591.1 Chaetoceros cf. decipiens 1 SEH-2013
MK642530.1 Chaetoceros elegans strain Ch12A1
MK642535.1 Chaetoceros lorenzianus strain Ch13B4
MK642499.1 Chaetoceros cf. decipiens strain El6B1
MK642521.1 Chaetoceros decipiens strain Na28A2
0.020

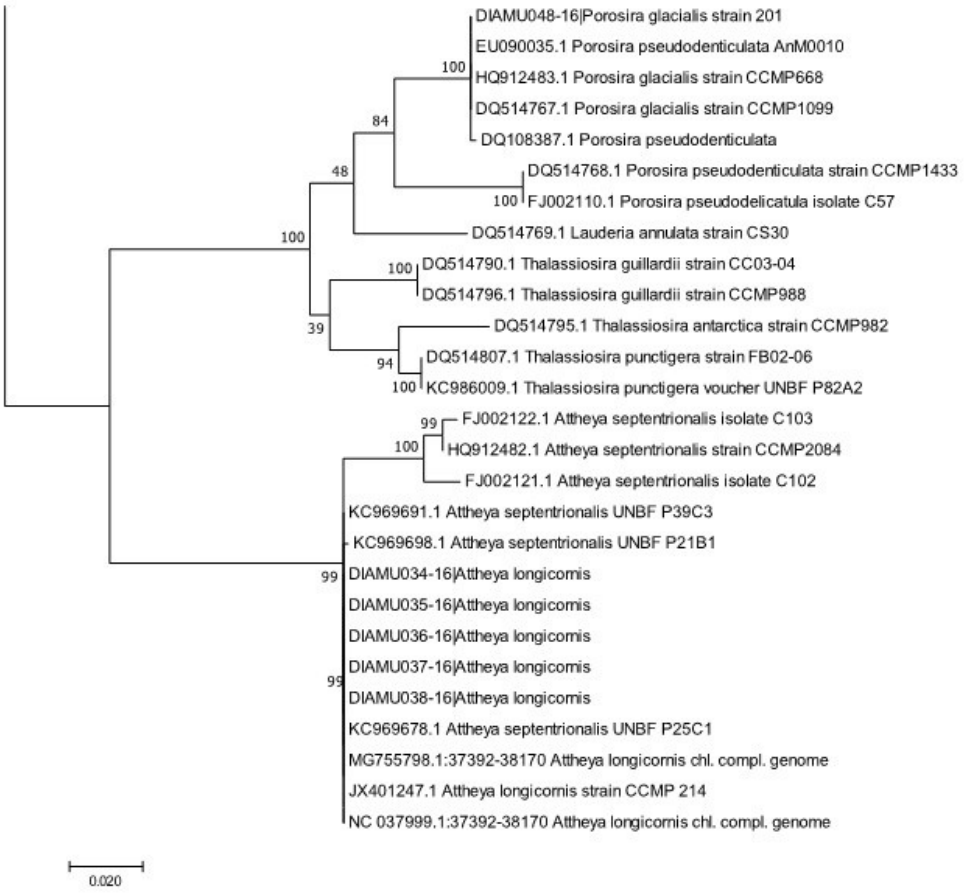


Figure 49. Molecular Phylogenetic analysis by Maximum Likelihood method representing relationships between 82 species of Bacillariophyta. Numbers at nodes are bootstrap support values obtained from the 50% majority rule bootstrap consensus tree. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model (Nei,Kumar 2000). The tree with the highest log likelihood (-4854.58) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.2311)). The rate variation model allowed for some sites to be evolutionarily invariable ([+*I*], 36.50% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted in MEGA7 (Kumar et al. 2016).

Phylogenetic position of *Porosira* strain nr. 201, shows that it forms a highly supported (100%) group with two sequences of *Porosira pseudodenticulata* and two sequences of *Porosira glacialis* from GenBank. All included *Porosira* sequences are monophyletic with 84% bootstrap support. Together with three *Thalassiosira* species and *Lauderia annulata Porosira* forms 100% supported clade.

CCDB also sequenced *Skeletonema* strain nr. 204 and supplied me with one sequence of 16S ribosomal RNA gene of examined strain nr. 204, but via BLAST search it was not possible to get any relevant similar sequence from GenBank database.

Sequencing was not successful for *Coscinodiscus strain* nr. 203, *Porosira glacialis* strain nr. 49.2D, for *Skeletonema marinoi* strain nr. Ma 39.2 and 120. The last three mentioned strains originated from the strain collection of phytoplankton scientific group at UiT; were previously sequenced and identified and were supposed to be used as reference material for identifying target strains.

4. Discussion

The advantage of DNA barcoding aims to be an effective tool for species identification based on a short genetic marker to identify it as belonging to a particular species. The method is expected to be fast and needs small amounts of biological material. Due to inherent limitations in the barcoding method testing as performed by CCDB, the results gained via this master thesis at times has the form of qualified estimations. More comprehensive and accountable results though would also have to comply to the following demands:

- The period between sampling and taxonomic screening shall be shortest possible, in terms of days and weeks, not months or years. Possible shape modifications and mutations caused by artificial limited environment that cells have to cope with can significantly influence the quality of morphological research. In this study shorter time would be helpful mostly for *Attheya* and *Chaetoceros* – species that are small and are prone to mutate. Long cultivation periods can sometimes cause changes in size of examined parts of cells and for that reason morphometric data are not comparable (Lakeman et al. 2009).

- The method chosen shall be appropriate for each species. As pointed at in the Introduction part, universal markers for diatoms have not been found. If sequences shall be compared with for example GenBank ones, then also sequenced regions must be equal (Pearson 2013; Guo et al. 2015).

- The computation of phylogenetic relationships is a basal tool to get an overview over the position of a species in the phylogenetic tree. The MEGA software is one of the most convenient ones for “early” taxonomists, it is though today not accepted to be the most scientific, mostly due to using substitutional models rather than time-consuming and equipment demanding analyses. But MEGA software is developing and its authors reports comparable results with other used softwares. (Tamura et al. 2011).

- A comprehensive phylogenetic survey is not possible based on only one or a few sequences. High quality computational results also requires existence of replications. If the Canadian Centre for DNA Barcoding (CCDB) should accept samples it was necessary with at least 19 different species, none of them with more than 5 replicates to fill up the plate capacity of 96 wells. Although the service of sequencing was free of charge, this requirement influenced negatively the quality of the survey as out of five submitted samples only a few were sequenced. For this reason it was not possible to compare *Porosira* strains nr. 49.2D and 201; *Skeletonema* strains MA39.2, 86 and 204 and confirm preliminary hypothesis of similarity with previously identified strain of UiT stock monocultures. The preliminary required 18 rDNA sequencing method was also changed by CCDB without prior notice as the plate was accepted as experimental for testing of the methods.

4.1 Discussion of the diatom strains

Attheya sp. strain nr. 20.2

Both morphologic and molecular examination of strain nr. 20.2 brought up close results of two similar *Attheya* species: *longicornis* and *septentionalis*.

Comparing to results in Crawford et al. (1994), i.e. descriptions of *Attheya longicornis* and *Attheya septentrionalis*, horns of the examined strain are rather straight than wavy. There are present 3 not 4 of longitudinal strips that are running rather straight, not spiraling throughout the length of the horns, and the valve does not have any additional structures.

Ingebrightsen (2010) previously observed the same strain and reffered, that horn size were 1-9 times cell length and the cell size varied. Length from 10.1 to 28,4 µm, and width from 2,5 to 10,8 µm. These measurements exceed 4 - 6 µm average length stated for *Attheya septentrionalis* according to Hasle in Tomas (1997). Crawford et al. (1994) also supported these morphometric data and Rampen et al. (2009) repeated

again morphologic features of both discussed species in their study of the phylogenetic positions of *Attheya longicornis* and *septentionalis,* referring that *Attheya lognicornis* can possibly have both 3 and 4 supporting rods of horns. Morphologic features of both *Attheya* species are discussed also at Bolzano et al. (2017)

According to results from the morphologic examination, with some uncertainty, it is possible to conclude that strain nr. 20.2 belongs to *Attheya longicornis,* however it is not possible to proof this conclusion by phylogenetic data based on rbcL gene from chloroplast. It is only possible to give the priority to the complex study of whole chloroplast genome (Yu et al., 2018) before partial sequences resulting from other studies. Based on this speculation, it would be possible to conclude that the strain nr. 20.2 belongs to *Attheya longicornis*.

***Chaetoceros sp.* strain nr. 61**

Identifying strain nr. 61 led to several challenges and questions where several remained not answered. The morphological analysis has to be critic and take into account that not everything what is visible at LM or EM images is necessarily part of the target organism. As example serves the bulb form attached to inner seta of *Chaetoceros sp.* strain nr. 61 at inner side of the valve (Figure 50). According to specialist (personal communication with Nina Lundholm, associate professor), this form was not observed before at *Chaetoceros sp.* and possibly, it could be some parasite attached to the cell. This form, however, seem to be connected organically with the setae and they seem to belong to *Chaetoceros sp.* strain nr. 61.

Another example is the branching of setae (Figure 19). From the EM micrograph it is not clear if seta is really branched, or if ‘branching’ is artificial and is made up by fragments of several setae clumped together.

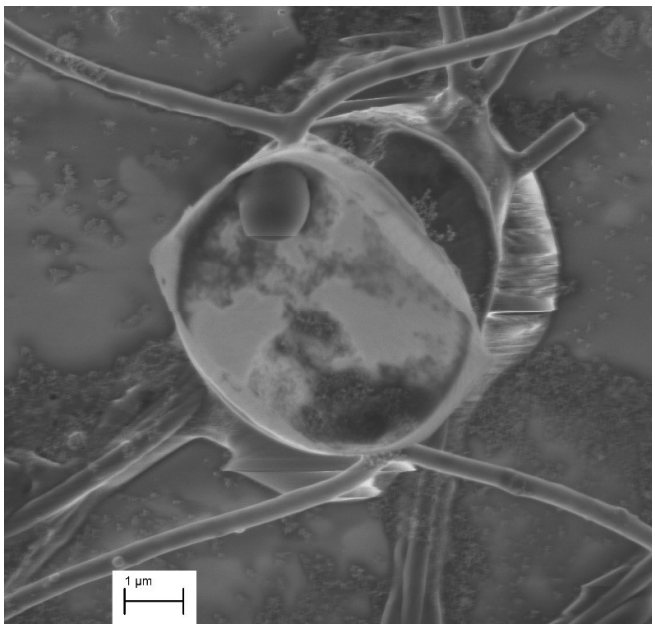


Fig. 50 Bulb form continually elongates the seta at inner part of the cell. Scale bar, 1 µm.

Misunderstanding of reference literature can be also a source of possible confusion. Two distinct and still valid Chaetoceros species are described very similarly according to historical sources. It is *Chaetoceros furcillatus* Bailey 1856 and *Chaetoceros furcellatus* Yendo 1911. As noted in Algaebase.org these two descriptions shall not be confused. Also Peterson et al. (1999) points to the fact, that 'the epithet furcillatus has usually, but unjustifiably, been changed to the variant furcellatus in all more recent reports.' EM micrographs of *Chaetoceros furcillatus* vegetative cells in last mentioned publication, but also description at Schevchenko et al. (2006) resembles examined *Chaetoceros* strain nr. 61 that was identified according to morphological features compared with Hasle (in Tomas, 1999) as *Chaetoceros furcellatus*. LM images did not reveal fork branched setae at spore forming cells. To conclude on species for Chaetoceros strain nr. 61 it would be important to obtain EM micrographs of spores, since they are often the only delimitation feature, and examine thoroughly original description and preserved reference slides for Chaetoceros furcellatus Yendo 1911 (the same way as it was done for *Ch. furcillatus* by Peterson).

Neither sequences of *Chaetoceros furcillatus* nor *furcellatus* were present in the GenBank database. For this reason, it was not possible to confirm a preliminary hypothesis based on molecular data.

The closest matching sequences for *Chaetoceros* strain nr. 61, with 100 % of bootstrap support, were two accessions of uncertainly identified *Chaetoceros cf. contortus* in a study that aimed estimation of the species richness of Chaetoceros and Thalassiosira (Bacillariophyta) in the Bay of Fundy (Hamsher et al., 2013). When searching in a phylogenetic tree for results with 'SEH' behind species name, that belong to the above mentioned study, I could see several uncertain results, or results that did not match the group of its species. Moreover, *Chaetoceros contortus* described in literature had morphologically distinct features as possible to see for example in Schevchenko, Orlova (2010) or recently in Xu et al. (2019).

Relationships between the groups of species within the genus *Chaetoceros* were poorly resolved. Despite of this fact, based on the placement of examined strain closely to *Chaetoceros radicans*, a species placed together with *Chaetoceros cinctus* in section Furcellata in a multigene phylogenetic study on the evolutionary history of Chaetocerotaceae (De Luca et al. 2019 and Gaonkar et al. 2018), it is possible to speculate, that the examined Chaetoceros sp. is placed correctly in the phylogeny in the present thesis and resembles morphologically similar species as stated by Rytter Hasle (in Tomas, 1997). Main morphological difference of *Chaetoceros radicans* are branched intercalary setae, larger pores on spiral utrastructure of setae and arched girdle bands. (Gaonkar et al., 2017) and (Lee et al., 2014)

Chaetoceros costatus, socialis, dichatoensis, sporotruncatus, tenuissimus and *simplex* were other species, that the strain nr. 61 was morphologically compared with. Species belonging to *Ch. socialis* species complex (meaning *Ch. socialis, Ch. dichatoensis and Ch. sporotrucatus*) create globular chain complexes that was not observed in examined strain nr. 61. Other features is possible to compare with Gaonkar et al. (2017).

Chaetoceros costatus is different from identified strain nr. 61 by thickened terminal setae and noticeably excentric slit-shape process (flattened tube) at the face of valve. For thorough comparison, a comparison with Kooistra et al. (2010) was used.

Chaetoceros tenuissimus is discussed at Sar et al. (2002). Its description show mismatch with the strain nr. 61: 'The cells are very small and solitary, square to rectangular in girdle view.' Rather small and solitary is also *Chaetoceros simplex*.

As can be concluded from above mentioned publications, the genus *Chaetoceros* is being, complex by complex, revised and new species are defined. This means that the scientific debate is not over and it is possible that *Chaetoceros furcellatus* will be revised, supplemented by EM micrographs and sequenced so that the strain nr. 61 would be possible to identify better in the future.

***Porosira sp.* strain nr. 201**

Based on morphologic features as stated in the result chapter, referring to Rytter Hasle (in Tomas, 1997), Villareal & Fryxell (1990, at http://symbiont.ansp.org/dntf/index.php) and Bolzano et al. (2017) it is possible to identify the strain nr. 201 to be *Porosira glacialis*, and to distinguish *P. glacialis* species from the close "relative" *Porosira pseudodenticulata, mainly because P. glacialis* possess central annulus that is not present at *P. pseudodenticulata*. The organization of the radial aeration is rather wavy than straight, and organization of strutted processes around labiate process is rather unequal.

With 100 % of bootstrap support, the phylogenetic analysis turned four similar sequences from GenBank in the same branch as the examined *Porosira* strain nr. 201. Two of these were identified as *P. glacialis* and two as *Porosira pseudodenticulata*. When analyzing the mismatching result with accession nr. EU90035 of Choi et al. (2008), related to the research of some polar diatoms, the molecular similarity represented by the phylogenetic relationship was not confirmed by morphologic identity. This means that it is not possible to proof the result of morphologic analysis

only by analysis of rbcL gene of chloroplast. The other mismatching record of *Porosira pseudodenticulata* from GenBank was not possible to link to available scientific publications.

GenBank includes more results for both *Porosira glacialis* and *Porosira pseudodenticulata*, mainly sequences from examination of 18S small and 28S large subunit ribosomal RNA gene. If study of Porosira strain nr. 201 shall be completed correctly, additional molecular examination would be necessary.

***Coscinodiscus sp.* strain nr. 203**

According to features described in the result chaptert of this thesis and to comparison with Rytter Hasle, Lange (1992) it is possible to suggest that strain nr. 203 is *Coscinodiscus concinnus*. Holmes, Reimann (1966) proposed that *Coscinodiscus concinnus* and *Coscinodiscus granii* were different stages of the same species, but this hypothesis was later rejected by several authors as summed up in Rytter Hasle (in Tomas, 1997) lately by Ferrario et al. (2008). The examined Coscinodiscus strain did not have wedge-shaped frustules nor asymmetric valvocopula.

The study of Theriot et al. (2010) includes both *Coscinodiscus granii* and *concinnus* as experimental species and distinguish between them, as seen from GenBank sequence records, but this fact is not specifically mentioned in this scientific article.

Taxonomic delimitation finished halfway due to missing molecular analysis since CCDB did not gain any sequence of the examined strain nr. 203. GenBank includes several records for *Coscinodiscus concinnus* with sequences from chloroplast and ribosomal genes and this way it would be possible to estimate suitable markers also for the examined strain.

***Skeletonema sp.* strain nr. 204**

Preliminary morphological examination of Skeletonema sp. strain nr. 204 began by comparing morphological features with Hasle (in Tomas, 1997), leading to the conclusion that this strain belonges to the group of species as is *Skeletonema costatum*. These species are characterised by cylindrical cells, which create long chains by locking connection of the ring of long processes emerging from the edge of each valve face.

Since this strain was a hypothetically re-isolated 'emigrant' of the strain nr. 86 that contaminated other cultured strains, and was previously delimitated as *Skeletonema marinoi*, morphologic analyses was focused mainly on comparison with Sarno et al. (2005) that originally described this species. Further morphological examination was completed with two possible results: *Skeletonema marinoi* and *Skeletonema dohrnii.*

Examined strain nr. 204 had 9 rimoportulae in the marginal zone of the valve, with flattened ends and ridged throughout the entire length; one excentric furtoportula ended with cup shape flaring at the end. Girdle bands were repetitively ribbed with series of pores. Girdle bands of *Skeletonema dohrnii* are defined as 'with irregular series of pores' interrupted by interspaces without pores. Mainly because of this densely and periodical pattern of pores on overlapping girdle bands I conclude that *Skeletonema* strain nr. 204 is *Skeletonema marinoi*.

Among morphometric data Sarno et al. (2005) states that number of rimoportulae for *Skeletonema dohrnii* is 8-13 and for *Skeletonema marinoi* 9-11. In Degerlund (2011), at SEM micrograph, figure 3e, identified *Skeletonema marinoi* has five rimoportulae. This can point to a morphological heterogeneity that can be accepted for this species. In case of identified strain nr. 204, flared ends of rimoportulae are a bit less serrated comparing to the literature.

Degerlund (2011) also summes up, that considerable genetic heterogeneity is reported for *S. marinoi* (Ellegaard et al., 2008; Godhe and Härnström, 2010; Saravanan and Godhe, 2010). While phylogenetically distinct from its sibling species *S. dohrnii* Sarno et Kooistra (Sarno et al., 2005; Ellegaard et al., 2008), overlapping of the ultrastructural girdle band characters separating the two species.

CCDB supported me with sequence of 16S ribosomal RNA gene of examined strain nr. 204, but none of strain nr. 86 nor Ma 39.2 (Degerlund 2011). For this reason it was not possible to compare the two examined strains from one collection. Neither was it possible to align with results of GenBank since that contained record of the complete mitochondrial genome of *Skeletonema marinoi*, and several sequences of 5,8S, 28S and ITS of ribosomal DNA.

Summing up all stated above, *Attheya* strain nr. 20.2 is *Attheya longicornis*, based on morphological features and partly also supported by phylogenetic analysis. *Chaetoceros* strain nr. 61 was estimated to be *Ch. furcellatus*, but not possible to identify certainly due to missing EM images of spores and missing appropriate phylogenetic data. *Porosira* strain nr. 201 is *Porosira glacialis*, mainly based on morphologic description and partly supported also by molecular survey. *Coscinodiscus strain* nr. 203 is *Coscinodiscus concinnus* and *Skeletonema* strain nr. 204 is *Skeletonema marinoi*. Both last named species were identified only according to morphologic features.

5. Conclusion

It was possible to identify four of five examined species according to morphological analysis, and was also possible to find closely similar species for identification of the strain of *Chaetoceros*.

Based on DNA barcoding method, using rbcL gene of chloroplast as a marker, it was possible to find similar species to *Attheya* and *Porosira* strains. Identifying of *Coscinodiscus* failed due to missing sequences. *Skeletonema* strain was not possible to compare with sequences of reference database and *Chaetoceros* identification was difficult, because of poorly resolved phylogenetic relationships based on chosen genetic markers. Generally, it was possible to demonstrate that rbcL gene from chloroplast is a promising marker for DNA barcoding of diatoms, as many species were found in strongly supported monophylectic groups on the rbcL phylogeny.

The DNA barcoding represents the most reliable framework for effective identification (Sujeevan 2007). As confirmed by this thesis, in services of taxonomy, DNA barcoding is crucially dependent on choosing of the right genetic markers, the right protocols for sequencing, on the correct and reliable matching data in gene libraries and also on solid base of scientific literature. DNA barcoding represents multistep process requiring qualified personal throughout the whole analysis.

This thesis also confirmed, that molecular method was not sufficient to be used solely and results of morphological screening were important for all of the examined strains.

6. References

Algaebase [online]. Available from http://www.algaebase.org/ [viewed 9.10.2019]

ALVERSON, Andrew J. Molecular systematics and the diatom species. Protist, 2008, 159.3: 339.

ARMBRUST, E. Virginia, et al. The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science, 2004, 306.5693: 79-86.

ASHWORTH, Matt Peter. *Rock snot in the age of transcriptomes: using a phylogenetic framework to identify genes involved in diatom extracellular polymeric substance-secretion pathways*. 2013. PhD Thesis.

BALZANO, Sergio, et al. Morphological and genetic diversity of Beaufort Sea diatoms with high contributions from the Chaetoceros neogracilis species complex. *Journal of phycology*, 2017, 53.1: 161-187.

BOWLER, Chris, et al. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature, 2008, 456.7219: 239.

CASTELEYN, Griet, et al. Natural hybrids in the marine diatom Pseudo-nitzschia pungens (Bacillariophyceae): genetic and morphological evidence. *Protist*, 2009, 160.2: 343-354.

CHOI, Han-Gu, et al. Morphology and phylogenetic relationships of some psychrophilic polar diatoms (Bacillariophyta). *Nova Hedwigia*, 2008, 133: 7-30.

CRAWFORD, Richard M.; GARDNER, Claire; MEDLIN, Linda K. The genus Attheya. I. A description of four new taxa, and the transfer of Gonioceros septentrionalis and G. armatas. *Diatom Research*, 1994, 9.1: 27-51.

DEGERLUND, Maria. Species concepts and functional aspects of cold-water diatoms (Bacillariophyceae). 2011.

DE LUCA, Daniele, et al. A multigene phylogeny to infer the evolutionary history of Chaetocerotaceae (Bacillariophyta). *Molecular phylogenetics and evolution*, 2019, 140: 106575.

DE STEFANO, Mario; DE STEFANO, Luca. Nanostructures in diatom frustules: functional morphology of valvocopulae in Cocconeidacean monoraphid taxa. *Journal of Nanoscience and Nanotechnology*, 2005, 5.1: 15-24.

Diatom new taxon file [online]. Available from http://symbiont.ansp.org/dntf/index.php, [viewed 04.02.2020]

GOLDSTEIN, Paul Z.; DESALLE, Rob. Review and Interpretation of Trends in DNA Barcoding. *Frontiers in Ecology and Evolution*, 2019, 7: 302.

GRAHAM, Linda E.; GRAHAM, James M.; WILCOX, Lee W.; COOK, Martha E. Algae. 3rd ed. LJLM Press, 2016.

FERRARIO, Martha E., et al. Species of Coscinodiscus (Bacillariophyta) from the Gulf of Mexico, Argentina and Antarctic waters: morphology and distribution.(With 68 figures and 1 table). *Nova Hedwigia Beihefte*, 2008, 133: 187.

GAONKAR, Chetan C., et al. 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*, 2017, 53.4: 889-907.

GAONKAR, Chetan C., et al. Annotated 18S and 28S rDNA reference sequences of taxa in the planktonic diatom family Chaetocerotaceae. *PloS one*, 2018, 13.12.

GUO, Liliang, et al. Comparison of potential diatom ‘barcode’genes (the 18S rRNA gene and ITS, COI, rbcL) and their effectiveness in discriminating and determining species taxonomy in the Bacillariophyta. *International journal of systematic and evolutionary microbiology*, 2015, 65.4: 1369-1380.

HAIDER, Nadia. A Brief Review on Species Concepts with Emphasis on Plants. *LS: International Journal of Life Sciences*, 2018, 7.3: 115-125.

HAMSHER, Sarah E., et al. Barcoding diatoms: exploring alternatives to COI-5P. *Protist*, 2011, 162.3: 405-422.

HAMSHER, Sarah E., et al. A comparison of morphological and molecular-based surveys to estimate the species richness of Chaetoceros and Thalassiosira (Bacillariophyta), in the Bay of Fundy. *PloS one*, 2013, 8.10.

HASLE, Grethe Rytter; LANGE, Carina B. Morphology and distribution of Coscinodiscus species from the Oslofjord, Norway, and the Skagerrak, North Atlantic. *Diatom Research*, 1992, 7.1: 37-68.

HEBERT, Paul DN, et al. A DNA 'Barcode Blitz': Rapid digitization and sequencing of a natural history collection. PloS one, 2013, 8.7: e68535.

HOLMES, Robert W.; REIMANN, Bernhard EF. Variation in valve morphology during the life cycle of the marine diatom Coscinodiscus concinnus. *Phycologia*, 1966, 5.4: 233-244.

INGEBRIGTSEN Richard A.; New application of a lettuce Lactuca sativa cv. Grand Rapids (Linnaeus) seed bioassay: identifying bioactivity in mass cultured diatoms. Master thesis, University of Tromsø, 2010.

IVANOVA, Natalia V.; FAZEKAS, Aron J.; HEBERT, Paul DN. Semi-automated, membrane-based protocol for DNA isolation from plants. Plant Molecular Biology Reporter, 2008, 26.3: 186.

IVANOVA, Natalia V., GRAINGER .M CCDB Protocols, COI amplification. 2007

KACZMARSKA, Irena; REID, Charlotte; MONIZ, Mónica. Diatom taxonomy: morphology, molecules and barcodes. In: Proceedings of the 1st Central-European Diatom meeting. FU-Berlin: Botanic Garden and Botanical Museum Berlin-Dahlem, 2007. p. 69-7.

KOOISTRA, Wiebe HCF, et al. Comparative molecular and morphological phylogenetic analyses of taxa in the Chaetocerotaceae (Bacillariophyta). *Phycologia*, 2010, 49.5: 471-500.

KUMAR, Sudhir; STECHER, Glen; TAMURA, Koichiro. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, 2016, 33.7: 1870-1874.

LAKEMAN, M. B., Von Dassow, P. and Cattolico, R. A. (2009) The strain concept in phytoplankton ecology. *Harmful Algae,* 8, 746-758.

LEE, Sang Deuk; JOO, Hyoung Min; LEE, Jin Hwan. Critical criteria for identification of the genus Chaetoceros (Bacillariophyta) based on setae ultrastructure. II. Subgenus Hyalochaete. *Phycologia*, 2014, 53.6: 614-638.

LEVIALDI GHIRON J.H. Plastid Phylogeny and Chloroplast Inheritance in the Planktonic Pennate Diatom Pseudo-nitzschia (Bacillariophyceae). Doctoral thesis, Universita Degli Studi Di Messina. 2006.

LOPEZ, Pascal J., et al. Prospects in diatom research. Current opinion in Biotechnology, 2005, 16.2: 180-186.

LOSIC, Dusan, et al. Atomic force microscopy (AFM) characterisation of the porous silica nanostructure of two centric diatoms. Journal of Porous Materials, 2007, 14.1: 61-69.

MANOYLOV, Kalina M. Taxonomic identification of algae (morphological and molecular): species concepts, methodologies, and their implications for ecological bioassessment. *Journal of phycology*, 2014, 50.3: 409-424.

MCLAUGHLIN, Robert B. An Introduction to the Microscopical Study of Diatoms. 2012.

MEDLIN, Linda K. Mini review: Diatom species as seen through a molecular window. Brazilian Journal of Botany, 2018, 1-13.

MEDLIN, L. K.; CRAWFORD, R. M.; ANDERSEN, R. A. Histochemical and ultrastructural evidence for the function of the labiate process in the movement of centric diatoms. British Phycological Journal, 1986, 21.3: 297-301.

NEI, Masatoshi; KUMAR, Sudhir. Molecular evolution and phylogenetics. Oxford university press, 2000.

NÜBEL, Ulrich; GARCIA-PICHEL, Ferran; MUYZER, Gerard. PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl. Environ. Microbiol.*, 1997, 63.8: 3327-3332.

PAGE, Roderick DM; HOLMES, Edward C. Molecular evolution: a phylogenetic approach. John Wiley & Sons, 2009.

PAWLOWSKI, Jan, et al. CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS biology*, 2012, 10.11: e1001419.

PEARSON, William R. An introduction to sequence similarity ("homology") searching. *Current protocols in bioinformatics*, 2013, 42.1: 3.1. 1-3.1. 8.

PEČNIKAR, Živa Fišer; BUZAN, Elena V. 20 years since the introduction of DNA barcoding: from theory to application. *Journal of Applied Genetics*, 2014, 55.1: 43-52.

PETERSON, Tony Douglas, et al. Chaetocerus furcillatus Bailey in the Canadian Maritimes. *Botanica marina*, 1999, 42.3: 253-264.

RAMPEN, Sebastiaan W., et al. PHYLOGENETIC POSITION OF ATTHEYA LONGICORNIS AND ATTHEYA SEPTENTRIONALIS (BACILLARIOPHYTA) 1. *Journal of Phycology*, 2009, 45.2: 444-453.

ROSS, R., et al. An amended terminology for the siliceous components of the diatom cell. 1979.

ROUND, Frank Eric; CRAWFORD, Richard M.; MANN, David G. Diatoms: biology and morphology of the genera. Cambridge University Press, 1990. /Page 117/

SUJEEVAN, Ratnasingham; HEBERT, P. A. U. L. BOLD: The Barcode of Life Data System. *Molecular Ecology Notes*, 2007, 7: 355-364.

SAR, Eugenia A.; HERNÁNDEZ-BECERRIL, David U.; SUNESEN, Ineés. A morphological study of Chaetoceros tenuissimus Meunier, a little-known planktonic diatom, with a discussion of the section Simplicia, subgenus Hyalochaete. *Diatom Research*, 2002, 17.2: 327-335.

SARNO, Diana, et al.; DIVERSITY IN THE GENUS SKELETONEMA (BACILLARIOPHYCEAE). II. AN ASSESSMENT OF THE TAXONOMY OF S. COSTATUM-LIKE SPECIES WITH THE DESCRIPTION OF FOUR NEW SPECIES 1. *Journal of phycology*, 2005, 41.1: 151-176.

SHEVCHENKO, O. G.; ORLOVA, T. Yu; HERNANDEZ-BECERRIL, David U. The genus Chaetoceros (Bacillariophyta) from Peter the Great Bay, Sea of Japan. *Botanica Marina*, 2006, 49.3: 236-258.

SHEVCHENKO, O. G.; ORLOVA, T. Yu. Morphology and ecology of the bloom-forming diatom Chaetoceros contortus from Peter the Great Bay, Sea of Japan. *Russian journal of marine biology*, 2010, 36.4: 243-251.

TAMURA, Koichiro, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*, 2011, 28.10: 2731-2739.

TOMAS, Carmelo R. (ed.). Identifying marine phytoplankton. Academic press, 1997.

VON QUILLFELDT, C. H. Identification of some easily confused common diatom species in Arctic spring blooms. *Botanica marina*, 2001, 44.4: 375-389.

XU, Xiao Jing, et al. Revisiting section Compressa of Chaetoceros (Bacillariophyceae), with descriptions of C. brevispinosus sp. nov. and C. ornatus comb. nov. *Phycologia*, 2019, 58.6: 614-627.

ZHANG, DeYuan, et al. Bio-manufacturing technology based on diatom micro-and nanostructure. *Chinese science bulletin*, 2012, 57.30: 3836-3849.

Appendix 1 Simplified schematic gross diatom morphology

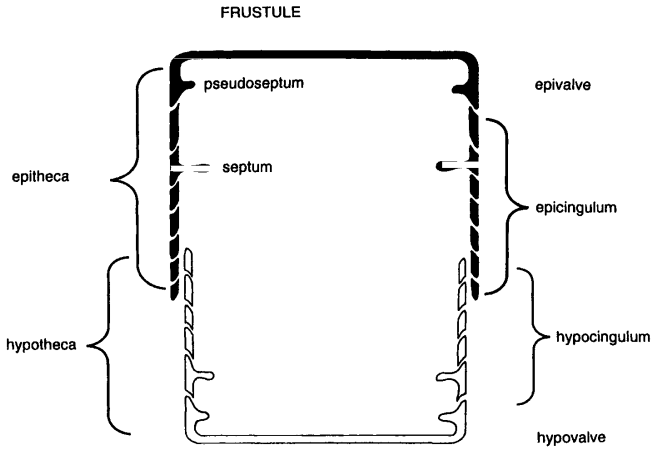
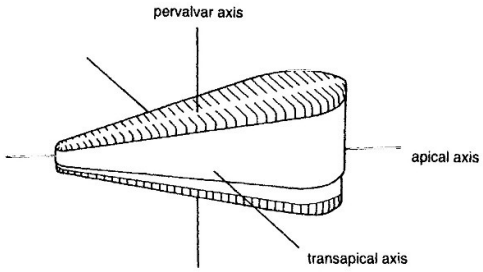
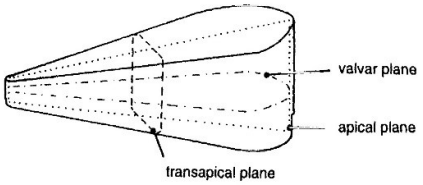


Figure 1 Simplified schematic gross morphology of diatoms according to Rytter Hasle in Tomas (1997)





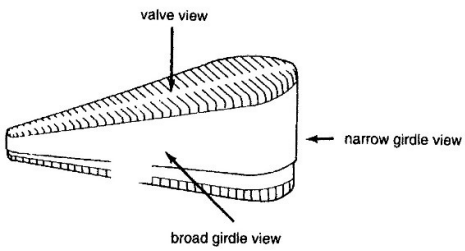


Figure 2 Schematic description of axes and planes of diatom frustule (Rytter Hasle in Tomas, 1997)

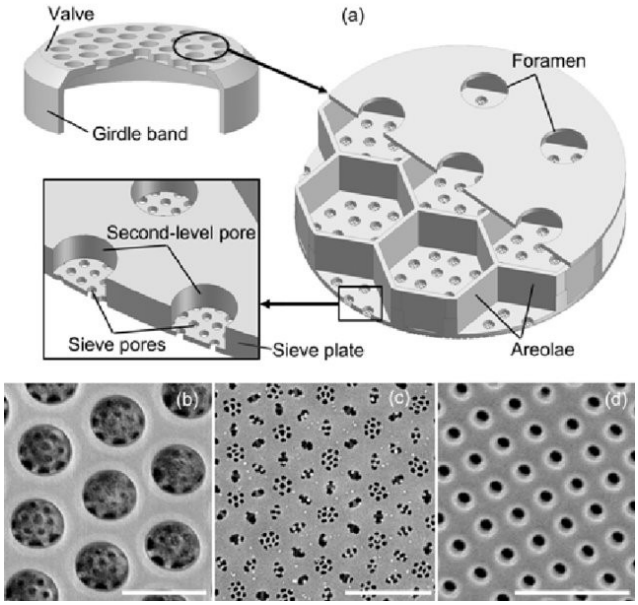


Figure 3 Example of multilayer construction of the valve (Coscinodiscus sp.) with different type of perforation taken from Zhang et al. (2012).

Appendix 2 Diatom cleaning with nitric/sulfuric acids

Tools and equipment: pipette**,** beakers, glass tubes, centrifuge tubes, Bunsen burner, metal grid on tripod, centrifuge, fume hood

Chemicals: nitric acid, sulphuric acid, MQ-water

1. 200 – 600 ml of dense diatom sample shall be concentrated. For the purpose of this master thesis 2 ml of sedimented cells were simply collected by glass pipette from the bottom of growing bottles to avoid mechanical damage. Centrifuged or filtrated samples shall be dissolved back in MQ-water. (Protocol of Stazione Zoologica)

2. Cells were transferred into a small glass beaker and placed on top of metal grid over Bunsen burner in the fume hood.

3. Diatoms concentrate was slowly mixed with 70% nitric acid and sulphuric in a ratio of 1:1:3 respectively (2 ml diatoms, 2 ml nitric acid, 6ml sulphuric acid). Diatom concentrate was supposed to be added first.

4. Mixture was heated above Bunsen burner and let be boiled 0,5 – 1 minute.

5. Cooled down mixture was transferet into glass tube and left for sedimentation ca 24 hours.

6. Acid mixture was removed down to sedimented cleaned frustules. Subsequently they were re-suspended in MQ water (5-10ml) and transferred to a tube that could be used in the centrifuge.

12. Clean frustules were centrifuged on 3000 rpm for 10 minutes and rinsed with MQ-water several times. Rinsing with 96% ethanol brought better results for drying of the samples for the purpose of SEM microscopy but still not sufficient for *Attheya* and *Chaetoceros* species.

Appendix 3 Protocol for drying and fixation of diatom samples for the purpose of SEM

Tools and equipment: Poly-L-lysine coated glass cover slips, pipette, Petri dish, Parafilm, beaker for distilled water rinsing, glass beaker with lid for ethanol washing, tweezers, Automated Critical Point Dryer with source of CO_2, EM-Tec SEM stub, stub gripper tweezers, stub holder, Carbon Conductive Tabs, exicator

Chemicals: 70%, 90% and 100% etanol, distilled water, conductive liquid silver paint

1. Petri dish was covered by Parafilm and Poly-L-lysine coated glass cover slips were transferred with the coated side up.

2. 1 drop of cleaned diatom frustules sample was transferred onto glass cover slip and let dry for 5 minutes. Results shown that drying for longer time would give better result in case of *Chaetoceros* species.

3. Cover slip with mounted sample was dipped in distilled water and transferred to mounting plate of Atomated Critical Point Dryer. Mounting plate with the sample was placed into beaker, covered with 70 % ethanol and let be there for 2 minutes.

4. 70% ethanol was replaced by 90 % ethanol and the sample was standing another 2 minutes.

5. 90 % ethanol was replaced by 100 % ethanol and sample was rinsed another 2 minutes. This procedure was done four times.

6. Mounting plate was placed into Atomated Critical Point Dryer and let run for 17 drying cycles. After ca 2 hours the sample was ready to be mounted onto metal stub.

7. Metal stub covered with carbon conductive tab was placed onto stub holder. Cover slip with mounted and dried diatom sample was placed on the top of the stub. 3-4 dots of Silver paint were placed on the edge of cover slip leading down to stub in order to create conductive zone. 8. Sample mounted on metal stub was placed into an exicator overnight for drying.

APPENDIX 4

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	*lnL*	(+*I*)	(+*G*)	*R*	*f*(A)	*f*(T)	*f*(C)	*f*(G)	*r*(AT)	*r*(AC)	*r*(AG)	*r*(TA)	*r*(TC)	*r*(TG)	*r*(CA)	*r*(CT)	*r*(CG)	*r*(GA)	*r*(GT)	*r*(GC)
GTR+G+I	171	11606.534	10067.884	-4862.451	0.41	0.30	1.19	0.283	0.325	0.177	0.216	0.136	0.022	0.095	0.119	0.127	0.022	0.035	0.232	0.030	0.124	0.033	0.025
GTR+G	170	11628.504	10098.846	-4878.938	n/a	0.15	1.18	0.283	0.325	0.177	0.216	0.133	0.023	0.096	0.116	0.125	0.022	0.037	0.229	0.034	0.125	0.033	0.028
T92+G+I	165	11673.799	10189.102	-4929.094	0.37	0.25	1.34	0.304	0.304	0.196	0.196	0.063	0.041	0.115	0.063	0.115	0.041	0.063	0.177	0.041	0.177	0.063	0.041
T92+G	164	11678.951	10203.248	-4937.172	n/a	0.14	1.34	0.304	0.304	0.196	0.196	0.063	0.041	0.114	0.063	0.114	0.041	0.063	0.177	0.041	0.177	0.063	0.041
TN93+G	167	11684.984	10182.303	-4923.683	n/a	0.15	1.28	0.283	0.325	0.177	0.216	0.069	0.038	0.090	0.060	0.130	0.046	0.060	0.238	0.046	0.118	0.069	0.038
HKY+G+I	167	11689.913	10187.231	-4926.147	0.35	0.24	1.33	0.283	0.325	0.177	0.216	0.067	0.037	0.126	0.059	0.104	0.045	0.059	0.190	0.045	0.165	0.067	0.037
HKY+G	166	11693.099	10199.410	-4933.243	n/a	0.14	1.33	0.283	0.325	0.177	0.216	0.067	0.037	0.126	0.059	0.103	0.045	0.059	0.190	0.045	0.165	0.067	0.037
TN93+G+I	168	11696.074	10184.400	-4923.726	0.51	0.43	1.68	0.283	0.325	0.177	0.216	0.058	0.032	0.080	0.051	0.162	0.039	0.051	0.297	0.039	0.105	0.058	0.032
GTR+I	170	11790.129	10260.471	-4959.750	0.72	n/a	1.25	0.283	0.325	0.177	0.216	0.134	0.024	0.092	0.117	0.135	0.027	0.038	0.247	0.014	0.121	0.041	0.012
K2+G	163	11850.166	10383.455	-5028.282	n/a	0.15	1.26	0.250	0.250	0.250	0.250	0.055	0.055	0.140	0.055	0.140	0.055	0.055	0.140	0.055	0.140	0.055	0.055
K2+G+I	164	11852.543	10376.839	-5023.968	0.39	0.27	1.61	0.250	0.250	0.250	0.250	0.048	0.048	0.154	0.048	0.154	0.048	0.048	0.154	0.048	0.154	0.048	0.048
T92+I	164	11864.109	10388.406	-5029.751	0.72	n/a	1.28	0.304	0.304	0.196	0.196	0.065	0.042	0.112	0.065	0.112	0.042	0.065	0.174	0.042	0.174	0.065	0.042
TN93+I	167	11874.644	10371.963	-5018.513	0.72	n/a	1.21	0.283	0.325	0.177	0.216	0.071	0.039	0.089	0.062	0.126	0.047	0.062	0.232	0.047	0.116	0.071	0.039
HKY+I	166	11888.041	10394.352	-5030.714	0.72	n/a	1.26	0.283	0.325	0.177	0.216	0.070	0.038	0.123	0.061	0.101	0.046	0.061	0.186	0.046	0.162	0.070	0.038
JC+G+I	163	11971.117	10504.406	-5088.757	0.40	0.28	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+G	162	11978.589	10520.871	-5097.995	n/a	0.15	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
K2+I	163	12027.274	10560.563	-5116.836	0.72	n/a	1.22	0.250	0.250	0.250	0.250	0.056	0.056	0.138	0.056	0.138	0.056	0.056	0.138	0.056	0.138	0.056	0.056
JC+I	162	12152.329	10694.611	-5184.865	0.72	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
GTR	169	13003.089	11482.423	-5571.732	n/a	n/a	0.87	0.283	0.325	0.177	0.216	0.140	0.029	0.072	0.122	0.116	0.039	0.046	0.213	0.039	0.095	0.058	0.032
TN93	166	13081.974	11588.285	-5627.680	n/a	n/a	1.18	0.283	0.325	0.177	0.216	0.072	0.039	0.070	0.063	0.139	0.048	0.063	0.256	0.048	0.092	0.072	0.039
T92	163	13107.653	11640.942	-5657.025	n/a	n/a	1.16	0.304	0.304	0.196	0.196	0.068	0.044	0.108	0.068	0.108	0.044	0.068	0.167	0.044	0.167	0.068	0.044
HKY	165	13145.080	11660.384	-5664.735	n/a	n/a	1.17	0.283	0.325	0.177	0.216	0.073	0.040	0.119	0.063	0.098	0.048	0.063	0.179	0.048	0.156	0.073	0.040
K2	162	13239.543	11781.825	-5728.472	n/a	n/a	1.16	0.250	0.250	0.250	0.250	0.058	0.058	0.134	0.058	0.134	0.058	0.058	0.134	0.058	0.134	0.058	0.058
JC	161	13358.058	11909.333	-5793.231	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083

Fig. 51 Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (lnL), and the number of parameters (including branch lengths)

are also presented [1]. Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (+G) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+I). Whenever applicable, estimates of gamma shape parameter and/or the estimated fraction of invariant sites are shown. Assumed or estimated values of transition/ transversion bias (R) are shown for each model, as well. They are followed by nucleotide frequencies (f) and rates of base substitutions (r) for each nucleotide pair. Relative values of instantaneous r should be considered when evaluating them. For simplicity, sum of r values is made equal to 1 for each model. For estimating ML values, a tree topology was automatically computed. The analysis involved 82 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 789 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.

Abbreviations: GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor.