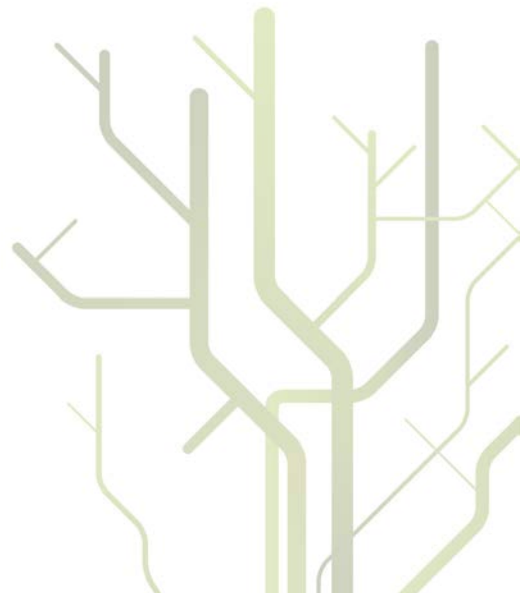
UNIVERSITY OF TROMSØ UIT

FACULTY OF BIOSCIENCES, FISHERIES AND ECONOMICS
DEPARTMENT OF ARCTIC AND MARINE BIOLOGY

Metabolic fingerprinting applied in diatom taxonomy

Siv Huseby

A dissertation for the degree of Philosophiae Doctor
June 2012



Metabolic fingerprinting applied in diatom taxonomy

Siv Huseby

A dissertation for the degree of Philosophiae Doctor

University of Tromsø
Faculty of Biosciences, Fisheries and Economics
MabCent and Department of Arctic and Marin Biology

Preface

My university studies started with an introduction to philosophy, which is customary in Norwegian university studies. One of the things I recall from those lessons was the allegory of Plato's cave. I have often got the notion, when looking at my data trying to understand how it all is connected, that we really are inside the cave and that the results we are able to see are just mere shadows of the "true" nature. But what is needed in order to see the "truth"? In modern taxonomy the goal is to describe a classification that is somehow true or natural, reflecting the evolution of the taxa one is working with. This is a demanding task, which unfortunately also, to a certain extent, has been neglected by the scientific society for several years. Hopefully this trend is about to change. New methods introduced into science often bring with them high hopes for the future; thinking that we finally have the methods needed to see the whole picture. At last we can turn our heads and see what is outside of the cave instead of making our conclusions of it based on the form of the shadows! In the case of diatom taxonomy the introduction of electron microscopy showed us a whole new view of the cell walls. The hope connected to these new methods was that if it was possible to see even the smallest of details in the morphology of the cell walls then, finally, it would also be possible to describe a taxonomy of the diatoms that was somehow true. This was also the case with the introduction of electron microscopy for diatom taxonomy. However, after some time, one came to realize that these new methods alone could not reveal the true taxonomy. When the molecular tools were introduced in diatom taxonomy, high hopes again appeared. But what do we do when the results from the old taxonomy based on morphology contradicts the findings of molecular tools? There is also a need for increased knowledge on phenotype information and function of a species. Could the biochemistry of species be a helpful tool applied in diatom taxonomy? The present thesis is an attempt to bring this into the science of diatom research. The question is if this will help us get a view of the outside of the cave.

Table of Contents

Preface i
Abstract iv
List of papers vi
List of abbreviations vii
Background for this thesis 1
Diatoms 1
Diatom taxonomy 2
What is a species? 3
How to delimit between species? 5
Chemotaxonomy 6
Objectives 10
Methods 11
Monocultures and experimental design 11
Morphological studies 12
Physiological characteristics 13
Molecular analysis 13
Metabolic fingerprinting and choice of method 13
Data handling and statistics 15
Main results and Discussion 15
Method related experiences 15
Intraspecies variability 18
Interspecies variability 20
Effect of environmental conditions on metabolic fingerprints 23
Effect of physiological status on metabolic fingerprint 26
Chemical diversity in diatoms and consequences for ecological studies and bioprospecting 28
Concluding remarks 29
Acknowledgements 32
References 33
Original papers

Abstract

The main aim of this thesis was to investigate if and how metabolic fingerprinting can be applied in diatom taxonomy. Traditionally, taxonomy of diatoms has, to a large extent, been based on the morphology of their cell walls. During the last 20 years molecular tools have also been included in taxonomical work. Even though both morphology and gene sequences have been shown to be appropriate tools in diatom taxonomy there are cases where these give contradicting results, like in the case of cryptic species. Cryptic species have similar morphology but are genetically different. Another issue with these two tools is that they do not offer much information about the function of a species, information that is interesting in light of for example ecology and management. Metabolomics investigates the metabolites synthesized by an organism. The metabolites synthesized at a certain moment in time will be a reflection of what genes are expressed at that time and will be a product of the organisms response to the environmental and biological conditions prevailing.

Direct injection mass spectrometry was used to investigate the metabolic fingerprints of different, commonly occurring, northern diatom species. The method produces mass-to-charge ratios (markers) normally with a mass precision of four decimals. Reproducibility of the method was 80% with the direct injection method applied, using a decimal precision of 0.1. The results of the analysis showed that the different species shared between 26-67% of the total markers. Even species of the same genera showed a high diversity. The two species *Chaetoceros furcellatus* and *Chaetoceros socialis* only shared 30% of the total markers. For four out of six species the difference between species increased with decreasing temperature. The expected phylogeny of these six species could not be reflected by the metabolite data. This latter result was also verified in another experiment performed during this thesis work, with six centric diatoms, but where the extracts from the species were analyzed with ultra performance liquid chromatography mass spectrometry.

Species that are genetically distinct but morphologically similar are called cryptic species. Cryptic diversity has been documented within the so-called cosmopolitan species *Chaetoceros socialis.* We investigated this diversity between strains collected from two geographic areas; the northeast Atlantic and Arctic and from Mediterranean waters. Monoclonal cultures were cultivated at three different temperatures; 2.5, 8 and 13° and analyzed with the aid of

morphology, LSU rRNA sequencing, growth rates, photosynthetic maximum quantum yield and metabolic fingerprinting. Comparison of gene sequences of the two groups showed an unequivocal difference, while only small morphological differences in spore morphology (not in the morphology of the vegetative cells) could be found between the two groups. At all three temperatures there were clear differences in growth and maximum quantum yield. The results from the metabolic fingerprinting also supported these findings. The clear genetic as well as functional differences does not support the cosmopolitan distribution of *C. socialis* and we therefore conclude that this species should be revised.

The results, both from the comparison of metabolic fingerprinting between diatom species as well as within a pseudo cryptic diatom species, in my opinion, is in support of the use of metabolomics in diatom taxonomy. Our results underline the need of metadata, e.g. growth rates, in metabolomics studies. I also think that increased knowledge of functional traits of species, like metabolomics, could be implemented in ecological modeling, building a bridge between taxonomy and ecology. The results of this thesis are also relevant to bioprospecting. The higher chemical diversity between species found at the lower temperatures, would indicate that it could be beneficial to cultivate diatoms at low temperatures, close to zero, in search for bioactivities.

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals:

Paper I

Eilertsen H.C., Huseby S., Degerlund, M., Eriksen, G.K., Ingebrigtsen, R.A., Hansen, E. (manuscript) Sample storage effects on reproducibility of direct infusion high-resolution mass spectrometry (HR-MS) microalgae extract analysis *Submitted to Limnology and Oceanography Methods*

Paper II

Huseby, S., Hansen, E., Degerlund, M., Eriksen, G. K., Ingebrigtsen, R.A, Eilertsen, H.C. (manuscript) Chemical diversity of six northern diatoms *Submitted to Journal of Applied Phycology*

Paper III

Degerlund, M., Huseby, S., Zingone, A., Sarno, D., Landfald, B. (2012) Functional diversity in cryptic species of *Chaetoceros socialis* Lauder (Bacillariophyceae) *Journal of Plankton Research*, **34**, 416-431, doi: 10.1093/plankt/fbs004

Paper IV

Huseby S., Degerlund M., Zingone, A., Hansen. E (in revision) Metabolic fingerprinting reveals differences between northern and southern strains of the cryptic diatom *Chaetoceros socialis. European Journal of Phycology*

List of abbreviations

COI	Cytochrome *c* oxidase gene region
ESI	Electrospray ionization
GC	Gas chromatography
HPLC	High performance liquid chromatography
HR	High resolution
ITS	Internal transcribe spacer gene region
LSU	Large subunit of the ribosomal gene region
m/z	mass-to-charge ratio
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
SSU	Small subunit of the ribosomal gene region
TOF	Time-of-flight, a high resolution filter used in MS
UPLC	Ultra performance liquid chromatography

Background for this thesis

Diatoms

Diatoms is a group of algae consisting of single cells or cells in colonies, characterized by their highly ornamented siliceous cell walls. The cell wall is multipartite and always composed of two units called valves, one somewhat larger than the other, together forming what resembles a petri dish. The siliceous parts of the whole cell wall are collectively called the frustule (Round *et al.*, 1990). The name *diatom* is derived from the Greek word diatomos meaning 'cut in half', which refers to the two valves of the diatoms.

Diatoms is a very species rich group. As there is no official register for all currently described diatom species a precise number of hitherto described species is hard to give. In the Californian Academy of Sciences' Catalogue for diatom names 60 000 diatoms were listed in 2008 (Kociolek, 2011) which gives an indication of the number of species described. The estimated number of existing diatom species is, however, about 200 000 (Armbrust, 2009). This high diversity in a relatively young (in a evolutionary timescale) group is explained by high evolutionary rates (Bowler *et al.*, 2008). The group comprises both planktonic and benthic species, some species can actually be both (Round *et al.*, 1990), and they are found almost in every aquatic habitat. Diatoms are important primary producers estimated to account for about 20-25% of the earths total primary production (Mann, 1999). Diatoms are argued to be particularly important in sustaining fisheries as they are the main producers of 'new' phytoplankton biomass, sustained by fluxes of nutrients from deeper waters as opposed to nutrients recycled via the 'microbial loop' (Falkowski *et al.*, 1998) and they play an important part in the biogeochemical cycle of silicon in the sea (Tréguer *et al.*, 1995). Diatoms have a complex evolutionary history and genes from both red and green algal origin have been found in diatom genomes (Moustafa *et al.*, 2009) as well as from a heterotrophic host (Armbrust *et al.*, 2004). Data also indicates that genes from a member of the Chlamydiae (Becker *et al.*, 2008) and also other bacterial genes (Bowler *et al.*, 2008) can be found in diatoms. This mixture of genes originating from different organismal groups has given the diatoms a range of attributes. One example of an unexpected attribute is that diatoms have a complete urea cycle (Armbrust *et al.*, 2004) earlier thought only to occur in multicellular organisms. This diversity of both genes and

species thus therefore remind us that small and unicellular does not equal “uncomplicated”, neither chemically nor functionally. There is still much to learn about the different intricate life cycles and survival strategies of diatoms. Spores of diatoms have been found to vital even after 100 years in the sediment (Härnström *et al.*, 2011). Even though diatoms are mostly regarded as autotrophic organisms there are also example of parasitic cases (Bavestrello *et al.*, 2000).

The cell wall, being so characteristic for diatoms, has attracted much attention leaving the biology and ecology somewhat behind. Since the first discoveries of diatom species more than 300 years ago (Mann, 2010) the taxonomy of diatoms and development of microscopes have been tightly linked. The increased quality of the microscopes gave researchers better opportunities to investigate the morphology of the frustules in more detail. When the electron microscopes were developed in the 1930ties diatoms were used as test objects and researchers in the field of diatoms quickly started using electron microscopes in their own research (Hendey, 1959). This helped them see even smaller morphological differences and has been helpful in the taxonomy of diatoms. During the last decades development of molecular methods has, as in other fields of biological science, revolutionized both diatom taxonomy and systematics.

Diatom taxonomy

Biological systematics is the study of the diversification of living organisms and the kinship between them. Taxonomy is the identification, description and naming of organisms and is as such more concentrated around species. These two fields are naturally closely linked and provide a system of organisms that in many cases is a prerequisite in other fields of biology. How research on diatom taxonomy has been or is performed is a much discussed and criticized subject. Mann (1999) stated in a review on species concepts in diatoms that “*History suggests that in many cases, diatomists might just as well have been classifying scraps of wallpapers as diatoms. Diatom taxonomy has developed largely without a conceptual basis, using a restricted range of characters drawn from just one part of the phenotype (the valve).”* This is maybe pushing things to the extremes but let us review some of the discussion. The discussion is divided in two parts, one more philosophical and the other more pragmatic. Firstly, what is a species? This question deals with the definition of species and the choice of species concept. Secondly, how do we

delimitate between species? Here the discussion is about which characters to use when deciding where the break between two species is.

What is a species?

"Nor shall I here discuss the various definitions which have been given of the term species. No one definition has as yet satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species." Darwin (1859)

Since Aristotle, *species* has the basic unit in taxonomy and is an important notion in biology. As of today a clear definition of the term *species*, which covers all groups of organisms, is still lacking. Darwin (Darwin, 1859) stated himself that a correct definition of species was difficult (see above), a statement that is still valid more than 150 years later. The reason for this is to a large extent the enormous variation of life, in its expression and function. Before the theory of evolution was launched, varieties described within a species were considered as "mistakes" or imperfections and were not necessary to explain. With Darwin and the new thoughts this view was altered and the varieties could be explained in the light of evolution. The theory of evolution however, did not have an immediate effect on diatom taxonomy. By the 1880ties discussions on species vs. varieties started and scientists argued more than earlier about the status of taxa (Mann, 2010). Modern age of species concepts started in 1942 (Hey, 2006) when Ernst Mayr published his book Systematics and the Origin of Species (Mayr, 1942). Earlier discussions had only dealt with having one definition while Mayr described different approaches to species identification and named these different species concepts accordingly. In the wake of Mayr's definitions of different species concepts followed several new concepts and there are actually some 20 concepts in use (De Queiroz, 2007). According to the morphological species concept, a species is defined as a species if it has sufficient morphological traits to differentiate it from all other species (Pedros-Alio, 2006). The number of morphological traits that is sufficient is not stated. Taxonomy of diatoms has traditionally been based on the morphology of the frustules and the morphological species concept has thus been the prevailing one. The perhaps most commonly used species concept today, if we look at biology in general, is the biological species concept defended by Mayr and which is defined as follows: Species are groups of interbreeding natural populations that are reproductively isolated from other such groups (Mayr, 1996). Their

harmonious gene pool is protected by the fact that these are not reproducing with organisms outside of its species population. As a consequence of this, species of organisms that are not reproducing sexually, for example bacteria, cannot be delimited by this species concept. According to the phylogenetic species concept a species is defined as the smallest set of organisms that share a common ancestor. With the entry of genomics and cheaper and easier accessible DNA technology, as well as development of bioinformatics, this species concept has increasingly been taken into use, as one now has better ways of measuring phylogenetic distance. In species delimitation in bacteria it is generally accepted that for two strains to belong to the same species they must show a similarity of >97% in their small subunit rDNA sequences, but this is argued to give an underestimation of number of species (Pedros-Alio, 2006). The use of several different species concepts in biology and the confusion and discussions to which this contributes to is problematic. A reconciled species concept would probably serve biology better. De Queiroz (2007) offers an explanation and a solution to the problem of conflicting species concepts. He defines a species as a segment of a separately evolving metapopulation lineage. A metapopulation is defined as an inclusive population made up of several connected subpopulations. Earlier species concepts all have in common that species are separately evolving metapopulations but they differ in their secondary defining properties. For example the secondary defining property of the biological species concept is that the lineage also has to be reproductively separated from other such lineages, for the morphological species concept the lineage need also be morphologically different and so on. De Queiroz states that the reason for different characteristics, like the morphology, molecular sequences, mating preference or ecological differentiation to mention some, give contradicting results in species delimitation is because these different characteristics evolve at different paces in evolution. Evolution is, naturally, a process in which we only look at snapshots in time. Therefore at the time of our investigations of two different taxa there might well be differences in the molecular sequences while the differences in morphology might not be very distinct. Perhaps in due time also morphological differences become more pronounced as evolution and speciation progresses. Applying de Quiroz species concept this should hold for all organismal groups even non-sexual reproducing ones like bacteria. Another important result of this would be that more than one characteristic would be important in species delimitation.

How to delimit between species?

Diatom species has traditionally been separated based on morphological traits of their frustules. With the entry of the DNA technology, and the decreasing costs for these types of analyzes, the results from investigations by molecular tools is increasingly more often found to conflict with the taxonomy built on the morphological characteristics. These are most often situations were the morphology do not show differences while molecular tools do this in such extent that it reveals two or more different species within what was earlier thought to be one. Species found to differ in genetic sequences (to such an extent that splitting into two species is advisable) while not showing any morphological differences is termed cryptic species. Another term is pseudo cryptic that are the species found to be genetically different and often when investigating further, for example in electron microscopy, researchers also reveal small morphological differences. One example of a pseudo cryptic species is found in the genera of *Pseudo-nitzschia*. This genus is important in coastal areas and some of its species produce domoic acid which causes amnesic shellfish poisoning in humans (Hasle & Syvertsen, 1997;Mos, 2001). Species of *Pseudo-nitzschia* are difficult to separate, and generally electron microscope is necessary for identification (Hasle & Syvertsen, 1997). Investigations with molecular methods have shown that the species *Pseudo-nitzschia delicatissima* assumed to be a single species, actually consists of several species (Orsini *et al.*, 2004). They followed a bloom of *P. delicatissima* during spring 2001 in the Gulf on Naples, where it regularly blooms in spring and sometimes also during autumn. Investigations on especially the ITS 1 and ITS2, but also other sequences, showed 5 distinct lineages in what was thought to be one species. Their investigations also indicated reproductive isolation between these lineages, even if they bloomed at the same time in the same area. Reproductive isolation in *Pseudo-nitzschia delicatissima* and *Pseudo-nitzschia pseudodelicatissima* was further extensively investigated and confirmed in the same area (Amato *et al.*, 2007). They could also find small but consistent differences in the morphology between the different species, and therefore concluded that the morphological differences, although subtle, would have ecological relevance as well. The authors suggested that during bloom conditions the encounter rates of the gametangia would be maximal. Therefore, if the different species or lineages bloomed at different times during the year, this would prevent or reduce interbreeding.

Morphological and molecular data are both strong tools in species delimitation, but they do not manage to give complete information of a species. A criticism put forward towards these two tools in taxonomy is that they do not give proper information on the functional traits of a species, such as its' physiology or ecology (Fenchel & Finlay, 2006;Rosselló-Móra, 2012). Another reason for including functional traits into taxonomy, to a larger degree than what is being done now, is to invite to a better communication between taxonomists and ecologist which probably would be beneficial for both scientific fields (Kociolek & Stoermer, 2001).

Chemotaxonomy

One such functional trait would be a species' biochemistry. In chemotaxonomy, also called chemosystematics, it is presumed that different species produce different chemical compounds. Typically one or a few compound classes e.g. pigments (Zapata *et al.*, 2011) or fatty acids (Dunstan *et al.*, 1994) have been used in classification of diatoms or between diatoms and other groups of algae. Investigations of fatty acid profiles of 2076 strains of microalgae (from different algae groups) were found suitable to discriminate between taxa of higher rank, however on species level the variability was too high to draw any conclusions (Lang *et al.*, 2011). Investigating the biochemistry of diatoms is also performed within the science of bioprospecting. Bioprospecting is the search for and characterization of bioactive compounds found in different organisms. Diatoms have been shown as interesting and potent organisms in the search for new bioactive compounds (Prestegard *et al.*, 2009). With bioprospecting research on organisms, new knowledge on the biochemistry of these organisms is also gained. The understanding on how organisms can change chemically with altered conditions is also very interesting related to bioprospecting research.

Up until now researchers have had to settle with one or a few different compounds as support in chemotaxonomy studies, but recent developments in mass spectrometry and NMR methods open new vistas. During the recent years, in the wake of the developments in molecular techniques or genomics, we have seen a rapid development of the so-called omics' scientific fields. This includes among others transcriptomics, proteomics and metabolomics (Fig.1). Metabolomics is defined as the study of an organism's full suit of metabolites synthesized (Fiehn, 2001). It is

complementary to the other "omics" fields like genomics, transcriptomics and proteomics, but is closer linked to the function of the organism than for example genomics as it analyzes what metabolites are expressed, not what it potentially can express. Metabolomics has therefore also been termed "the link between genotype and phenotype" (Fiehn, 2002).

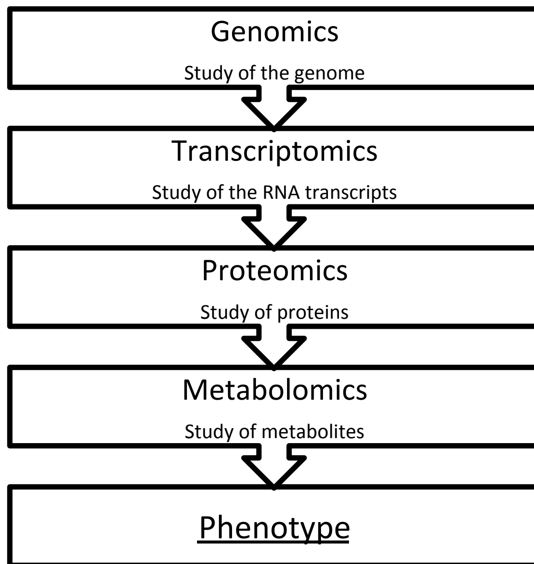


Fig.1 Schematic diagram of the different "omics" fields and what level of the biological system each field concerns

An organism's metabolome will change with environmental conditions, and also reflects the physiological status, life history and developmental stage of the actual organism. As such metabolomics is an interesting approach in the search for good methods to investigate organism's functional traits and how it responds to changes in its environment. As of now there are no methods that can analyze all metabolites within an organism's metabolome (Macel *et al.*, 2010). It is estimated that the number of genes in an organism equals the number of compounds it can contain (Verpoorte *et al.*, 2008). The two, so far, whole genome sequenced diatom species *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* are predicted to contain 10 402 and 11 776 genes respectively (Bowler *et al.*, 2008). There are several methods available when doing metabolomics research, and they all have their advantages and disadvantages that influences the outcome of the analyses. Most are based on either nuclear magnetic resonance (NMR)

spectroscopy or mass spectrometry (MS) in combination with chromatography. NMR spectroscopy is a physical measurement of the resonance of magnetic nuclei, such as ^{1}H, ^{13}C, or ^{15}N in a strong magnetic field where each compound has a highly specific spectrum (Verpoorte *et al.*, 2007). The advantages with NMR are the high reproducibility of the method, it is nondestructive and it can quantify the metabolites, but on the other hand it has a low sensitivity (lower than other techniques used within metabolomics) so metabolites at low quantities will not be detected (Verpoorte *et al.*, 2007;Macel *et al.*, 2010). In MS, metabolites are ionized (charged) and their mass-to-charge ratios (*m/z*) are measured using electric or magnetic fields in a mass analyser (Macel *et al.*, 2010). There are two main chromatography techniques that can be coupled to MS used in metabolomics; liquid chromatography (LC) and gas chromatography (GC). Further there are several different types of GC and LC instrumentations developed. The reproducibility of these two methods is lower than with NMR and quantification is more difficult. On the other hand they both have higher sensitivity and because of the lower cost of the instrumentation they are more available. The advantage of LC MS compared to GC MS is that it analyses a higher amount of metabolites without prior derivatisation, on the other hand metabolite identification is made more easy with GC MS as international databases can be used (Viant & Sommer, 2012).

As well as there are different techniques available for metabolomics studies there are also different approaches when analyzing. Metabolic fingerprinting is defined as a global, high-throughput and rapid analysis to provide sample classification (Ellis *et al.*, 2007) without necessarily investigating what metabolites the profiles consist of (Fiehn, 2001). Other approaches used within metabolomics include metabolic profiling, metabolite target analysis and metabolic footprinting (see Table I for terms and definitions).

Table I: Definitions of terms and different approaches used within metabolomics based on definitions given in (Fiehn, 2001;Fiehn, 2002;Ellis *et al.*, 2007).

Terms	Definitions
Metabolome	All metabolites synthesized by an organism or a biological system
Metabolomics	The identification and quantification of the metabolome of an organism or a biological system
Metabolite profiling	The identification and quantification of a set of predefined metabolites
Metabolic fingerprinting	Global, high-throughput rapid analysis of metabolites to provide sample classification
Metabolic footprinting	Analysis of the (exo)metabolites that is secreted/excreted by an organism
Metabolite target analysis	Qualitative and quantitative analysis of one, or several, metabolites related to a specific metabolic reaction

MS in combination with effective tools for data analysis has been shown to be a promising method in the classification and taxonomy of filamentous fungi and yeast (Smedsgaard & Nielsen, 2005;Frisvad *et al.*, 2008), and MS methods are now introduced in clinical routine microbiological diagnostics (Welker & Moore, 2011). So far metabolomic studies performed on diatoms are scarce. Chemotaxonomic studies using metabolomic data on microalgae has been shown to be promising in species discrimination using NMR spectroscopy (Chauton *et al.*, 2003). A few metabolomics studies have been performed on microalgae related to their environment (Bölling & Fiehn, 2005) and further some on growth stages and exudates of diatoms (Barofsky *et al.*, 2009;Barofsky *et al.*, 2010;Vidoudez & Pohnert, 2011). Combining both the need for joining together more than two species concepts as well as the need for more information on functional traits in diatom taxonomy, I believe it could be worth while to investigate if metabolomics can be applied in taxonomy and biogeography studies of diatoms.

Objectives

In the intersection between bioprospecting of northern marine diatoms and traditional taxonomy I started out this thesis work in order to investigate:

- if metabolic fingerprinting can be applied and aid in classifying samples of diatoms, both between different species as well as within species
- how environmental conditions, such as culturing temperature, and status of the samples, such as growth rate or physiological state, influences the metabolic fingerprint of different species of diatoms

Metabolomic studies are relatively new of age and so far there are few studies to compare with in this field of science. Breaking new ground demand us to keep an open mind. In order to investigate these objectives several experiments were carried out at the Department of Arctic and Marine Biology, University of Tromsø. The uncertainty and variation of the direct injection MS method both between samples and between sample runs are discussed in paper I. In paper II metabolic fingerprints from direct injection MS of extracts from different common cold-water diatom species were compared to investigate if it was possible to distinguish the species based on their metabolic fingerprints. The effects of temperature on the metabolic fingerprints are also discussed. In papers III and IV cryptic diversity within the taxa *Chaetoceros socialis* was investigated and discussed with emphasis on phenotypic traits like growth and metabolic fingerprinting. In addition to the studies presented in the paper I-IV metabolic fingerprinting data from an experiment with cold-water diatom species analyzed from UPLC MS are discussed.

Methods

This thesis work is essentially based on experimental work. Several methods have been applied during the different experiments and the methods are explained in more detail in the papers. Here follows a short introduction and explanation to some of the methods of choice:

Monocultures and experimental design

In all experiments monocultures of different diatom species were used. These were isolated from the north east Atlantic and Arctic waters (all papers and experiment series III) or from the Gulf of Naples, Tyrrhenian Sea (papers III&IV). Cells of diatoms, either as single cells or colonies, were manually isolated from field samples or from germinated spores from sediment samples using micropipettes. These were later held in culture in f/10 medium (Guillard & Ryther, 1962) and additional silicate at ambient light and temperature. The cultures isolated from the Tyrrhenian Sea were in the beginning held at another temperature, light and nutrient media (see paper III). All cultures were stepwise adapted to the experimental conditions chosen and kept there for at least two weeks before the onset of the experiments.

The problem of experiments using monocultures is that we of course do not measure how they would have acted in nature. Despite all efforts we cannot truly mimic the natural environment. Also, as diatoms are known to have a rapid genetic drift there is a chance that they through some years of culturing in the lab will change genetically (Lakeman *et al.*, 2009). To reduce this effect we used relatively recent isolates. As significant clonal variability is common in diatoms (Wood *et al.*, 1992), we tried to reduce this effect by comparing several strains of the same form of *Chaetoceros socialis* in papers III and IV.

This thesis work is essentially based on experimental work, which includes three different experimental series:

<u>Experiment series I: Papers I and II</u>

Short outline: Six diatom species were simultaneously cultivated at two different temperatures:0.5 and 8.5°C (paper II). In order to get a measure of the uncertainty of the method one of the

monocultures from the experiment (*Porosira glacialis,* AMB49.2D) was cultivated in several replicates (paper I). Samples were analyzed with direct injection MS. For further information see the papers.

Experiment series II: papers III and IV

Short outline: Strains of *Chaetoceros socialis* originating from two geographic areas; the north east Atlantic and Arctic and the Gulf of Naples were cultivated at the three temperatures 2.5, 8 and 13°C. Morphology, gene sequencing, growth rate, maximum quantum yield and metabolic fingerprints were analyzed in order to investigate the earlier reported cryptic diversity found within this species.

Experiment series III

Data from this experiment is not part of any of the papers but some of the results are discussed in the synopsis of this thesis. In short monocultures of 6 diatom species originally isolated from samples collected in the north east Atlantic and Arctic were cultured simultaneously at two different temperatures 2 and 6°C. The monoclonal cultures were cultivated in pasteurized f/10 growth medium with additional 24.6μmol $Si(OH)_4$ L^{-1} added. Scalar irradiance was measured to 30μmol m^{-2} $s^{-1}$and photoperiod was 14:10 (light:dark). The experiment started when all cultures were in exponential growth and ended after a minimum of 10-fold increase in Chl*a* content. Growth rates were calculated as in paper II. Samples were then filtered onto burnt GF/C filters (450 °C, 8 hours), flash freezed in fluid nitrogen and kept at -80°C until extraction could be made. The UPLC MS analysis followed the same procedure as in paper IV). Data was processed with MarkerLynx (same method as presented in paper IV) giving a dataset of markers (with mass and retention time) and signal strengths. Subsequent statistical and numerical analysis was conducted in Statistica and Matlab.

Morphological studies

For species identification, samples from all monoclonal cultures (all papers and experiment series III) were examined in light microscopy and identification and nomenclature generally followed Hasle and Syvertsen (1997). In the study on strains of *C. socialis,* samples were also examined in

transmission and/or scanning electron microscopy (for further details see paper III), this was also the case for the monoclonal culture of *S. marinoi* used in paper II.

Physiological characteristics

Metadata is very important in metabolomics studies (Fiehn *et al.*, 2007) and information on physiological characteristics is interesting in light of taxonomy, biogeography or ecology of a taxa or sample. Physiological characteristics were therefore included in all experimental series. Growth rates, as doublings day^{-1}, was calculated based on Chl *a* measurements, cell counts and/or calculated biovolume. As another measure of physiological status we also measured maximum quantum yield in Photo System II (Φ_{PSII}) with a Water- Pulse Amplitude Modulated (PAM) fluorometer (Water-ED/B, Heinz Waltz GmbH). This method estimates the photochemical efficiency of photosystem II (Schreiber *et al.*, 1986) where the maximum quantum yield is a measure of the fraction of open and closed reaction centers in photosystem II.

Molecular analysis

We chose to analyze for partial sequences of the large subunit (LSU) of the ribosomal gene region for the molecular analyzes in paper III. This gene region has earlier been shown to detect cryptic diversity within morphologically delineated species (Sarno *et al.*, 2005;Sarno *et al.*, 2007;Kooistra *et al.*, 2010). Further information on molecular analysis is found in paper III. Some of the monoclonal cultures used in paper II and in experiment series III have, in addition to morphological studies, also been identified by either 18s rDNA, (SSU) and/or the large subunit, 28s rDNA (LSU).

Metabolic fingerprinting and choice of method

The prime objective in this thesis was to evaluate if it could be possible to separate between diatom species or even between different forms within species based on their produced metabolites. Therefore metabolic fingerprinting, which is used for classification of samples, is the chosen approach. As we wanted to be able to rapidly analyze for as many metabolites as possible the choice of analyzing method fell on LC MS. Two different variants of LC MS have been applied in the different experimental series included in this thesis. In a way they reflect the recent rapid

developments in mass spectrometry methods. In paper I&II direct injection High Performance (HP) LC MS was the chosen approach for analyzing samples from different diatom species. Direct injection is a fast method where no separation of the metabolites occurs before analyzed by mass spectrometry (no chromatographic step, the HPLC is only used to introduce the samples to the MS), and has earlier been shown to be a good method in classifying between samples (Larsen *et al.*, 2005;Beckmann *et al.*, 2008). In paper IV, as well as in experiment series III, Ultra Performance (UP) LC High Resolution (HR) MS with a Waters Acquity BEH C18 column was chosen. With this methods metabolites are separated in a chromatographic step before analyzed in the MS. The problem with potential influence of metabolites close to each other in the MS analysis is therefore reduced. In all experiments samples were ionized with positive electrospray (ESI+). Compared with direct injection MS, UPLC MS takes somewhat longer time for each sample, but the high resolution of the chromatographic separation in UPLC means that even complex extracts can be resolved in runs lasting no longer than a few minutes. Another advantage of the UPLC MS compared with direct injection is that the chromatographic step results in one additional variable; retention time. This means that the data sets that are generated are more complex, but with well-developed data handling programs this is not presently a problem and the result is more robust data. Using a high-resolution mass filter, such as time-of-flight (TOF), will effectively separate compounds with similar masses that has not been chromatographically separated. It is also possible to calculate elemental compositions of the compounds based on their accurate mass and isotopic distribution. However, to determine the 2D-structure of markers based on electrospray ionization (ESI) HR-MS data is difficult, partially because no large database of fragmentation data in LC-MS exist (as opposed to the large amount of data available for electron impact ionisation used in gas chromatography (GC)-MS).

In metabolomics studies it is found to be especially important to use replicates, to run the samples several times in the MS and also if possible to run the samples in a random order so that samples with the same treatment do not follow each other. This has been a learning process and these are things that we also ourselves have noticed the importance of during the different experiments. Other things to bear in mind is that the metabolome of an organism is rapidly changing with changing conditions so when ending the experiments we handled the samples as fast as possible from taking the samples from the experimental bottles until the filters were flash frozen. We also made sure to

handle the samples in the same manner both during the experiments but also when terminating the experiments.

Data handling and statistics

The new omics' sciences including genomics is facing us with enormous possibilities but the huge quantities of data that is being generated also give us a great challenge when it comes to data handling. In this thesis several approaches has been tried out as the different papers proves. From the manual picking of peaks, visual inspection of spectra and comparing between samples to decide on what belongs to the same compound (paper II), use of an in-house developed program that handles the data purely numerically (papers I, II and IV) to using a software developed for metabolomics data by MarkerLynx combined with statistical software (paper IV, experiment series III). For the interpretation of data we have looked for number of common markers, usually as percent of total obtainable markers when comparing between samples (either with the in-house developed program or with a Matlab script when interpreting data already processed in MarkerLynx XS). Principal component analysis, which is an exploratory data analysis and a variable reduction procedure that is commonly used in metabolomics studies.

Main results and Discussion

Method related experiences

When approaching a new method there are new things to be learned all the time. Some of the main results learned about the method itself are discussed below:

<u>Reproducibility</u>

Every method has an uncertainty, which is important to have in mind when analyzing the results. This uncertainty was, for us, unknown for the metabolic fingerprinting method we applied working with diatom extracts. We therefore performed a reproducibility test on one of our cultures; *Porosira glacialis,* running the extract several times in direct injection HPLC MS,

without freezing the extract or with one or several thaw-freeze cycles. The results from these investigations are found in paper I, which I will briefly discuss here.

The HPLC MS, and also UPLC MS, give out *m/z* ratios with a mass of four decimals. When conducting metabolic fingerprinting analysis we were interested in as many *m/z* ratios as possible in order to cover as much of the metabolome possible with the chosen method. Analysis on the *P. glacialis* extracts showed a marked decrease in reproducibility when using a higher decimal precision level. Another factor that showed to be important during this study was thaw-freeze cycles. When terminating the cultivation experiments samples were flash-frozen in liquid nitrogen and thereafter kept at -80°C, which is the preferred method as flash-freezing will halt enzymatic activity (Macel *et al.*, 2010). After extraction the extracts were kept at -20°C until they were run in the HPLC/UPLC MS. The highest reproducibility (89.6%) was achieved when using a 0.1 decimal precision level and not freezing the samples at all. The second highest (80%) was achieved when the samples were flash frozen in liquid nitrogen, frozen again at -20°C until direct injection MS was conducted. When applying several freeze-thaw cycles the reproducibility continued to decrease. So for the direct injection MS analysis our recommendations would be to keep the number of freeze-thaw cycles to the lowest possible and when comparing samples the decimal precision should be kept at 0.1 if possible. This was the method applied in paper II and when investigating the data from this study one should therefore be aware that there is a 20% uncertainty of the method. The uncertainty of the UPLC MS method (paper IV and experiment series III) was not tested in the same manner, but the importance of keeping the freeze-thaw cycles to a minimum should be just as valid for this method.

Earlier studies from proteomic mass spectrometry have also reported that proteomic profiles are not very reproducible over time and that the intensities should only be regarded as semiquantitative (Hu *et al.*, 2005). We also observed a difference between mass spectrometry runs (paper II) although this was not equally clear for all species. As all species were included in both runs, something that also was recommended by Hu et al (2005), this should not have affected the outcome of the comparison of the different species. In paper IV all samples were included in each run, in total tree times, but the order of the samples were randomized so that

drift or changes in the UPLC should not affect the results and further the delimitation based on the results. We still observed a variation between the different runs of the same sample.

Importance of numbers of samples compared

In paper II we also observed the importance of number of samples compared. When we examined the distribution of hit rates we found a skewness (0.407) in the data toward lower values, implicating that mean hit rates calculated from few samples would have a tendency to be lower than if more samples were included. To look further into this we programmed a random generator to pick data from our complete dataset (n of hits=628) in sets of 3 to 21 samples. The mean of these datasets varied between 200 and 455, and there was an increase up to n= 9 - 10 (Fig. 2). We therefore decided only to apply datasets with n > 9 (paper II). This practice is in accordance with normal statistical considerations, i.e. to avoid low n values in order to gain more precise estimates of true population values (Cumming *et al.*, 2007).

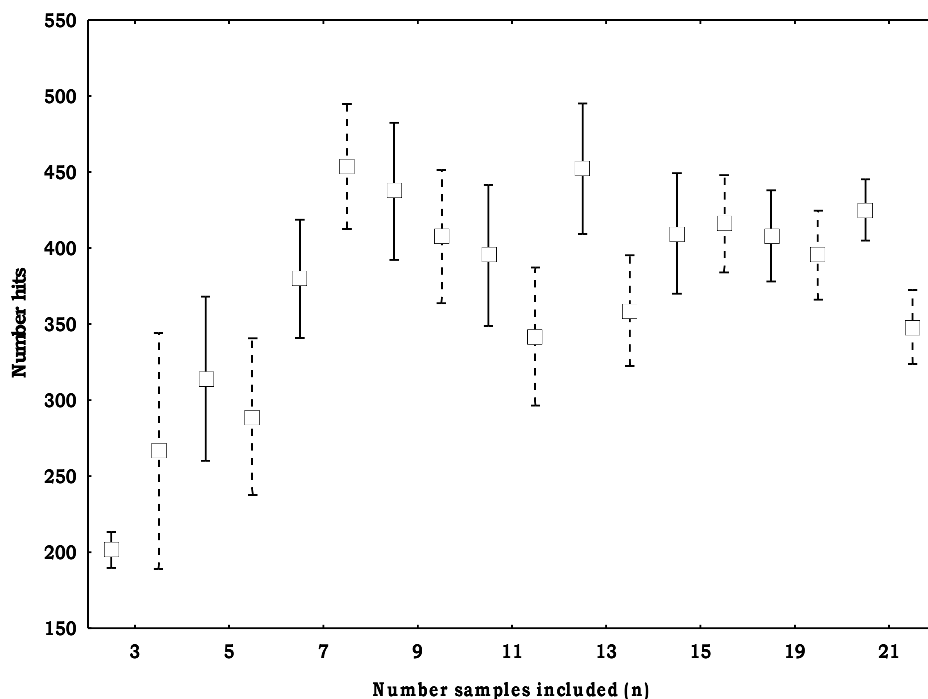


Fig.2. Mean number of hits (same marker in separate samples) in samples collected randomly. Vertical bars represents +/- 1 SE (standard error) and broken lines are 0.5 °C and whole lines 8.0 °C.

Comparison of instrumentation and data analysis approaches

Data handling by the in-house developed program resulted in between 1000-1200 markers from the direct injection analysis of extracts from six different diatom species (paper II). In contrast the manual inspection method extracted 211 markers from the same data set. Both data handling approaches did in this case show high diversity between species. The strength of the numerical approach is the high number of metabolites included in the analysis and that the element of subjectivity in extracting proper signals is removed. The manual inspection method is also much more time consuming, something that is an important aspect when there are many samples to inspect. In the analysis of the UPLC data (paper IV and in experiment series III) we also applied a MarkerLynx program to process the data. With the UPLC metabolites are separated in the chromatographic step and each marker has both a mass and a retention time. This results in higher demands on the data processing program but renders more robust data. To take advantage of the higher complexity the MarkerLynx program was used.

Chemical diversity (paper II and IV + data from experiments series III)

Intraspecies variability

From the different experiments of this thesis high diversities in the markers were observed both at the level of replicates, of the same strain of one species, between different strains of the same species and between species. Samples of the same monoclonal culture of species, both temperatures included, had ca. 76% of the markers in common (paper II). Earlier studies also report on high phenotypic diversity in phytoplankton cultures. Clonal variability and rapid genetic drift (Wood *et al.*, 1992;Lakeman *et al.*, 2009) could of course not be ruled out, although the cultures had only been separated for a couple of weeks, which was the time the experiment lasted.

Cryptic species or pseudo-cryptic species are defined as species with similar morphology but genetically distinct. The introduction of molecular methods has resulted in the detection of considerable hidden or cryptic diversity within species of diatoms (Sarno *et al.*, 2005;Lundholm *et al.*, 2006;Sarno *et al.*, 2007;Alverson, 2008;Kooistra *et al.*, 2008;Kooistra *et al.*, 2010). In papers III-IV we investigated the cryptic diversity within *C. socialis* and we found that the two

groups of strains, one originating in the north east Atlantic and Arctic and the other from the Gulf of Naples showed clear distinction in their partial LSU rRNA gene sequence. The vegetative cells could not be distinguished based on their morphology, but the morphology of the spores revealed that the strains were belonging to two forms of *C. socialis*; *C. socialis* forma *socialis* (found in the group of strains originating from the north east Atlantic and Arctic) and *C. socialis* forma *radians* (found in the group originating from the Gulf of Naples). The choice of gene sequences is important when the goal is to delimit between species. In our study on *C. socialis* there were clear differences on several levels, both in terms of gene sequences and in phenotypic characteristics. But in other cases the selected gene sequences in a study do not distinguish clearly between strains, although physiologically or chemically they do act differently (Loret *et al.*, 2002;Rossello-Mora *et al.*, 2008). Interestingly, these two groups of strains could also be distinguished based on their metabolic fingerprints (paper IV).

When we compared different strains of *C. socialis* cultured at the same conditions, originating from two different geographical areas we could show that the two groups of strains, termed northern and southern, shared about half the markers or applying the numerical method at best shared 63.6% of the markers (paper III). The variation in number of common markers, within the northern group of strains, ranged between 60 and 98% of total markers. High diversity between strains belonging to the same species have been reported from many studies, one useful review on these matters is Wood *et al*. (1992). Gallagher reported on a high diversity between growth rates of clones within the species *Skeletonema costatum* (Gallagher, 1982) and later also on differences in photoadaptation between different clones of the same species. (Gallagher *et al.*, 1984). Populations of diatom species have also been found to consist of several clonal strains exhibiting both genetic as well as physiological variability (Rynearson & Armbrust, 2000). There are also examples were clones of the same species (e.g. the dinoflagellate *Karenia brevis*) have not shown variability in ribosomal DNA sequences while exhibiting differences in physiology and toxin production (Loret *et al.*, 2002). The internal transcribed spacer (ITS) which was chosen in the study of Loret *et al*. is proposed as a good marker for barcoding diatoms (Moniz & Kaczmarska, 2009) however did not manage to differentiate these clones of *Karenia brevis*. In the study of Loret *et al.* (2002) there was also a high diversity within the replicates of especially one of the clones in terms of growth rates. Chemically it is reported several times that different clones of the same toxic species produces different amounts of toxins (Ogata *et al.*, 1987;Thessen

et al., 2009) or that some clones of a species produce toxins at a certain environmental condition while others do not (Lelong *et al.*, 2012). Even though the awareness of strains diversity within algae species has increased with the introduction of molecular markers the intraspecific variability (strains variability) is still commonly overlooked (Burkholder & Glibert, 2006).

Interspecies variability

When we compared the metabolic fingerprints between different diatom species the similarity is between 26-67%, (paper II, both temperatures included) in terms of common markers. The two most different species, in terms of common markers, were *Thalassiosira gravida* and *Chaetoceros furcellatus* (Fig 2. Paper II). The two species with most common markers were *Skeletonema marinoi* and *Attheya* cf. *longicornis.* This difference between species did not seem to be coupled to the presumed phylogenetic relationships between the species. One could assume that near related species would be the ones having the most markers in common, but *Chaetoceros socialis* and *Chaetoceros furcellatus* only shared 30 % of total markers (or if we add the uncertainty of the method these two species shared at most 50% of their markers). Our data from the direct injection MS analysis (paper II) could not reflect the presumed phylogenetic relationships between species. Kooistra *et al.* (2007) proposed a phylogenetic tree where the six species included in paper II were grouped as follows: the genera *Thalassiosira*, *Skeletonema* and *Porosira* in one group, the genera *Chaetoceros* in another while *Attheya* is a separate group close to the pennates. These groupings are in general also supported by other studies e.g. (Rampen *et al.*, 2009;Sorhannus & Fox, 2012). For our metabolite data to reflect phylogeny we would therefore assume that for example *S. marinoi* and *T. gravida* would be more similar in terms of metabolites than *S. marinoi* and *C. socialis*, something that was not the case. The species that showed the most resemblance with each other were the smaller species *S. marinoi*, *A.*cf. *longicornis* and also the two *Chaetoceros* species, although surprisingly, not when compared to each other. In our dataset the two species least similar to other species were *T. gravida* and *P. glacialis*. We also compared metabolic fingerprints of six centric diatom species in the experiments series III, but in this study we chose chromatographic separation by UPLC. In this study however, contrary to the study described in paper II, the two species *T. gravida* and *T. hyalina* belonging to the same genera actually were the ones most difficult to distinguish from each other based on their metabolic fingerprints (Fig. 3 and 4).

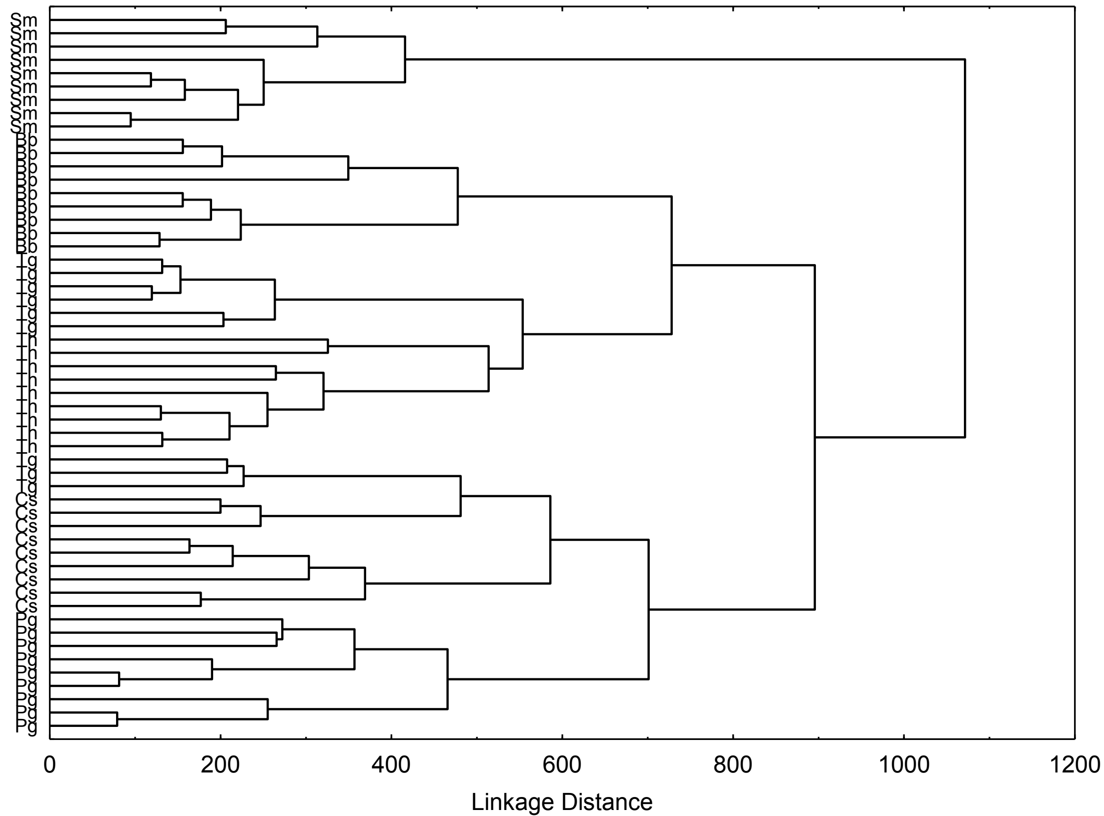


Fig.3 Tree diagram for the six diatom species in the experiment series III from 2°C, Complete linkage used, distances are Euclidean. Abbrevations for species names: *Porosira glacialis* (Pg), *Chaetoceros socialis* (Cs), *Thalassiosira gravida* (Tg), *Thalassiosira hyalina* (Th), *Bacterosira bathyomphala* (Bb), *Skeletonema marinoi* (Sm)

Classification of some species of sponges from the Mediterranean from HPLC MS data did show to be congruent with phylogenetic trees obtained by mitochondrial COI in a study by Ivanišević *et al.* (2011). In neither of the two studies on different centric diatoms, that are included in this thesis, did the metabolic fingerprints reflect the phylogeny described in the literature. The two different cultivation temperatures also produced different relations between the species, indicating that temperature influences the resemblance between species.

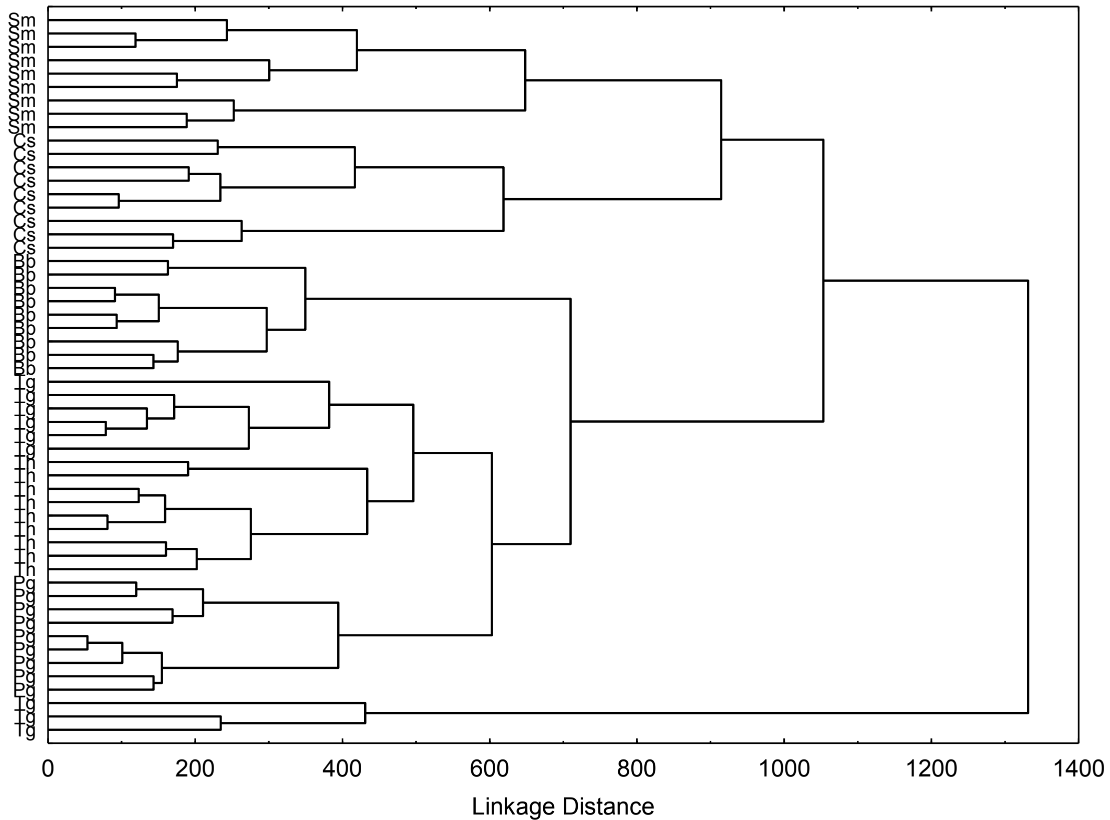


Fig.4: Tree diagram for the six diatom species in the experiment series III from 6°C, Complete linkage used, distances are Euclidean. Abbrevations for species names: *Porosira glacialis* (Pg), *Chaetoceros socialis* (Cs), *Thalassiosira gravida* (Tg), *Thalassiosira hyalina* (Th), *Bacterosira bathyomphala* (Bb), *Skeletonema marinoi* (Sm)

In conclusion; generally there is a higher similarity in terms of markers detected with lower taxa levels (Table II), within species being the lower levels. But the data is not always conclusive, for example we see an effect of temperature and physical status, which also influences the metabolic fingerprint of a sample.

Table II: Common markers at different taxonomic levels as % of total number of markers detected. Data are from either experiment series a)I, b) II or c) III and are processed either by the in-house developed program[1] (described in paper I), by the MarkerLynx program in combination with MatLab[2], or reflects the overlapping range of results from both a combination of MarkerLynx and Matlab as well as form the in-house developed program[3]. Method uncertainties are not included.

Taxa level	Common markers as % of total number of markers
Replicates of same strain of species, same conditions[c2]	85-90%
Only northern strains of *C. socialis*[b3]	60-98%
Northern compared with southern strains of *C. socialis*[b3]	47-64%
Different species of centric diatoms[a1]	26-67%

Effect of environmental conditions on metabolic fingerprints

Water temperature is asserted to be one of the main factors determining species distribution (Faurby & Funch, 2011), and it is suggested to be more important than for example salinity in the distribution of phytoplankton (Smayda, 1958). The question is if, and possibly how, does temperature influence the metabolic fingerprint?

Cultivating temperature affected the metabolic fingerprints of the cultures in the studies included in this thesis in different degree. When we compared different species of diatoms in paper II we observed that most species were more similar to each other in terms of markers at the higher temperature than at the lower one (Fig 3, paper II), which indicates that low temperature could increase chemical diversity. This change in similarity was not the case for all species however; *C. socialis* and *C. furcellatus* did not change much in terms of common markers with other species. This difference in change in similarity with other species did not seem to have any consistency with change in growth rate. *C. socialis* showed the highest difference in mean growth rate (difference=0.27), while no large change in similarity to other species caused by temperature change. On the other hand *C. furcellatus*, had a low difference in growth rate between the two

temperatures (difference 0.06). As metabolomics studies are relatively new there are not much references to be found concerning temperature and metabolic fingerprints of algae. Most studies on algal chemistry, and possible effects of temperature change, deal therefore with one or a few metabolites or ratios of main elements C:N:P. Interest of algae as food in mariculture is increasing and their nutrition value is heavily investigated, but mostly in terms of lipid and carbohydrate content. The results of these different studies do not give a clear picture of how the biochemistry of algae is affected by temperature. Different species are affected differently and so are different compounds. For example a study show that for the diatom *Chaetoceros* cf. *wighamii* lipid and carbohydrate were higher at the lower temperatures tested, while protein was not affected (de Castro Araújo & Garcia, 2005), on the other side protein was markedly affected by temperature for most species in a study on six Antarctic microalgae (Teoh *et al.*, 2004). The cellular content of N in *Skeletonema costatum* was fund to be independent of temperature (Yoder, 1979). In a study on *Leptocylindrus danicus* however, N and cell carbon increased with increasing temperature, although the C:N ratio was independent of temperature (Verity, 1981). Montagnes and Franklin (2001) report that for diatoms there is no relationship between C and N per unit volume and temperature. The low consistency in these types of studies is probably due to the large variation among species, possibly also between strains of species as discussed earlier, in how they behave physiologically and therefore also chemically (Kudo, 2003).

Investigations of metabolic fingerprints of the land plant *Arabidopsis* during cold acclimation revealed a global reorganization of the metabolome (Gray & Heath, 2005). In this study 593 out of 1187 markers, or about 50% of the markers, were observed to change when leaves of *Arabidopsis* were shifted from 23 to 4°C. Other studies of Arabidopsis report on extensive reconfiguration of the metabolome with low temperature, where 75% of the metabolites of a strain were increased with lowered temperature (Cook *et al.*, 2004). It is also reported that the metabolism of plants are far more influenced by low temperature-stress than high temperature-stress (Guy *et al.*, 2008). Now, the monocultures of different species that were used in experiment series I and III were all originating from the north east Atlantic and Arctic. Sea surface temperatures ranges between -2 and 9 °C in this area during spring months when these species normally bloom (Degerlund & Eilertsen, 2010). In the lab they had also been held within this

temperature range since collection. Differences in their natural habitats' temperatures can therefore not be an explanation to why there is difference in how much their metabolic fingerprints change with changing temperature. They could though have different temperature optima which could influence their response to the different temperatures (Suzuki & Takahashi, 1995).

In the study of strains of *C. socialis* (paper III-IV) the situation was different. The northern strains all originated from the northeast Atlantic and Arctic. In this areas *C. socialis* is an important species during the spring bloom when sea temperature is reported to be between -2°C and 9°C (Degerlund & Eilertsen, 2010). The seawater temperature reported from the Gulf of Naples, from where the southern strains originated, ranges between 14-26°C (Ribera D'Alcala *et al.*, 2004), and *C. socialis* generally blooms in spring and autumn when water temperature is around 20°C. The three temperatures used in the study were 2.5, 8 and 13 °C; all below what the southern strains could be expected to be adapted to and closer to what the northern strains have been experienced in their natural habitat. The reason why we did not use a higher temperature as maximum temperature was that we were unable to make any of the northern strains grow at temperatures above 13°C. It is reported that cold adapted species; so-called physcrophiles has an upper temperature limit of 15°C. At all three temperatures we observed difference in growth rates, maximum quantum yield and metabolic fingerprints between the two groups of strains. The southern strains had an increase in growth with increased temperature (Spearman R test $p < 0.05$), while the growth of the northern strains did not show a clear correlation with temperature. For maximum quantum yield there were no clear correlation with temperature for either of the groups. One interesting observation was that at 8 °C the maximum quantum yield was higher for the southern strains group than for the northern, while the northern group still performed higher in terms of growth. This we have no good explanation for, it might be that at 8°C the southern strains could have a higher maximum quantum yield but because of different loss rates the growth in doublings per day was still lower. The differences in growth and metabolic fingerprints, between the two groups of strains, we believed could be a result of different adaptive strategies to change in environmental conditions. The two groups of strains shared about 50% of the markers at all three temperatures resulting from the MarkerLynx data processing program.

When applying the in-house developed program (described in paper II) however, there was a somewhat higher similarity between the two groups of strains at 2.5°C than at the two higher temperatures. The latter method yields a higher number of markers and does not take into account retention time when comparing markers, which could explain this discrepancy between results of the two data handling approaches. In the experiment series I (paper II) *C. socialis* was represented by the strain AMB80, which also was included in the experiments series II (paper III&IV). The difference in temperature condition in experiment series I did not have a large effect on how different this strain was to the other species in this particular study (Fig.3, paper II). In the *C. socialis* study, at the highest temperature 13°C, it was actually the strain AMB80 that had a very different metabolic fingerprint than the other northern strains of *C. socialis*, which might reflect its different physiological status.

Effect of physiological status on metabolic fingerprint

Growth rate is a commonly used measurement of physiological status. In general, growth rates are expected to increase, linearly or exponentially, with increasing temperature (Eppley, 1972;Bissinger *et al.*, 2008), but also bell shaped responses are reported (Fiala & Oriol, 1990). Further, growth rates are reported not to be species specific (Gilstad & Sakshaug, 1990), but has been claimed to be dependent on the size of the diatom cells, with smaller cells in general having higher growth rates (Sarthou *et al.*, 2005). In the different experiment series reported in this thesis there were no indications of species-specific growth rates, not either a correlation between growth rates and cell volume. In this regard should also be mentioned that the datasets are relatively small in order to test this. Conclusions from an extensive series of growth experiments (n=1056, including replicates, different species, different temperatures and light and replicate experiments) on diatom species concluded that maximum growth yields were unpredictable in at least 55% of the cases and that growth conditions and possibly internal rhythms were the determining variables instead of species (personal communication Hans Christian Eilertsen, University of Tromsø). In the study on the cryptic species of *C. socialis* the somewhat larger cells from the northern strains had higher growth rates than the southern strains and we concluded that the adaptive strategy of the species in its response to changed environmental conditions was more important for physiological performance than the morphometric features of the cell, a finding

also supported by Kagami & Urabe (2001). The results from the study on *C. socialis* strains also reveal another problem with growth rates; the variability between strains within the same species. Especially at 13°C the northern strains differed in growth rates. Difference in growth rates between strains has also been reported for strains of *Skeletonema costatum*, where the difference in growth rates between strains ranged between as much as 0.1 – 5.0 divisions per day at the same conditions (Gallagher, 1982).

Since the metabolome is highly responsive to the biological and environmental conditions the history of a sample and how it has been treated is very important in metabolomics (Fiehn *et al.*, 2007), something that we also observed in our studies. At 13°C in experiment series II on different strains of *C. socialis*, the northern strain AMB80 diverged both from the southern as well as the other northern strains. This strain had significantly lower growth rate ($p = 0.013$, Mann-Whitney U test) than the other northern strains and it also diverged from the other strains in terms of its metabolic fingerprint. AMB80 was also included in the 2.5 °C experiment, where it did not diverge from the other northern strains. Neither did strain AMB66 at 8°C, which has the same origin as AMB80. Strains AMB80 and AMB66 have a more northern origin than the other northern strains included in the study, but the results from the phylogenetic analysis did not reveal any clear differences among the northern strains in the LSU rRNA gene (paper III). Earlier studies on exudates of the diatom *Skeletonema costatum* reported that the three different growth stages; stationary, exponential and declining could be separated by UPLC coupled with TOF MS, or GC-MS (Barofsky *et al.*, 2009;Vidoudez & Pohnert, 2011) and diurnal variation during the exponential growth stage was also observed (Vidoudez & Pohnert, 2011). In the study on strains of *C. socialis* all sampling for UPLC MS was performed around midday and all cultures were in exponential growth phase at the termination of the experiments. However, the low growth rate during the time of the experiment is most probably the reason for the divergence between the northern strains. This shows again that chemically the difference between two taxa is not fixed, but is dependent on the conditions and state at the time of, or prior to, the sampling.

Since the distance between the metabolic fingerprints in terms of markers produced is affected by environmental and physiological conditions theoretically we can of course not rule out the possibility that there exist some combination of conditions were two species actually produce the same markers. The high diversity in chemistry reported from our studies between and within species of diatoms has some consequences to other areas of biology.

Chemical diversity in diatoms and consequences for ecological studies and bioprospecting

It has been assumed that marine organisms with high dispersal capacities would be genetically homogenous with no geographically isolated populations (Medlin, 2007). The use of molecular tools has revealed extensive genetic diversity, for example with the increased reports on cryptic species (Sarno *et al.*, 2005;Lundholm *et al.*, 2006;Sarno *et al.*, 2007;Alverson, 2008;Kooistra *et al.*, 2008;Kooistra *et al.*, 2010), but also high diversity of different clones within the same populations blooming at the same time (Rynearson & Armbrust, 2000). It has also been argued that variations in genetic sequences do not necessarily reflect functional variability (Fenchel & Finlay, 2006). Our results from the study on *C. socialis* do, however, reflect variability both in LSU sequences and in functional traits like growth rates, maximum quantum yields and biochemical composition within, what for now at least, is considered one species. Other studies report on biochemical differences between strains of species that cannot be separated based on gene sequences commonly used for species delimitation (Rossello-Mora *et al.*, 2008). A more trait based microbial biogeography has been called for (Green *et al.*, 2008) arguing that patters of traits shed light on issues like why organisms live where they do and how they will respond to environmental change. The unit of species is an important unit also in ecology, but in order to understand and predict changes in an ecosystem one approach is to group species into functional groups (Lavorel *et al.*, 1997). Attempts have been made towards a functional classification of freshwater phytoplankton (Reynolds *et al.*, 2002). To succeed in this it will be essential to have robust knowledge of the functional traits of the different species. In ecological modeling one needs to generalize in order to see the larger picture. It is therefore normal to gather algae in one bulk or at best in different groups of algae, for example diatoms and flagellates (Wassmann *et al.*, 2006) represented as carbon or Chl *a*. It has been suggested to include the physiology of the component organisms, their behavioral traits and the interactions between the different organism that are included, have been suggested (Allen & Polimene, 2011). We know that much of the

compounds synthesized by algae are essential for higher trophic levels, for example fatty acids. In the case of fatty acids, studies report that different groups of algae produce different types of fatty acids profiles (Zhukova & Aizdaicher, 1995;Dijkman & Kromkamp, 2006) and one can conclude from this that it will not be indifferent for other trophic levels what type of algae is dominating in an area or at a certain time. When we now, with the aid of metabolomics, see the enormous chemical diversity within these algae that are important in the production of the sea is it then advisable not to take this into account in ecosystems modelling? Barofsky *et al.* (2010) have provided us with an example of how this chemical diversity influences the feeding on diatoms by copepods. They investigated the different growth phases of *Skeletonema marinoi* with UPLC – MS and if this influenced feeding of the copepod *Calanus* sp. They reported that *Calanus* sp. showed a greater preference for *S. marinoi* in late stationary phase in laboratory settings. *Calanus* sp. was shown to be able to discriminate between different growth phases of the diatom, probably because of different (info)chemicals in or in the surroundings of the diatoms cells.

The large chemical diversity both between strains of the same species and between different species is good news for bioprospecting on diatoms. The effect of temperature also indicates that low temperatures might be especially interesting as the number of markers increased and we also saw larger differences between species at low temperatures. However, it should be underlined that we have not looked for bioactive substances in the samples and there need not be a correlation between number of markers/metabolites produced in an organism and the bioactivity of an organism.

Concluding remarks

"But one thing is certain: to understand the whole you must look at the whole"

(Kacser, 1986)

We are in the beginning of the post-genomic era and the new techniques and data handling will need some time to evolve. What is demonstrated in this thesis is a remarkable chemical diversity, both within and between species. That species, and also strains within the same species, are different in terms of their functional traits is not news. However, the large variation within part of the metabolome of diatom species demonstrated in this thesis brings new information. It is not only a few metabolites that differs between relatively closely related taxa of species but the different species investigated in this thesis share as little as 26% of the total number of markers compared. Between clones belonging to the pseudo cryptic species *C. socialis* we show a diversity of 50% of the markers produced under the same conditions. This is a diversity that we should take into consideration when working with diatom species. To some extent as a biologist one can feel overwhelmed by this huge diversity on all levels, and it is hard too see how to handle all this information and still being able to draw conclusions on the biology of species or how ecosystems work. In looking for biomarkers on pollutant exposure seasonal variability has been a limiting factor in the application of seemingly well-established single biomarkers like metallothionein or heat shock proteins (Lacorn *et al.*, 2001). But it has been argued that metabolomics, which measures several hundreds of metabolites, can, together with statistical analysis, offer another possibility where the integrated profile of a subset of the metabolites measured can discriminate between healthy and stressed organisms (Viant, 2007). This thesis shows that both different species as well as forms of species can be distinguished from each other based on their metabolites. With such high diversity, comparison is made easier when taxa are compared in pairs. It also shows that temperature influences the metabolic fingerprints and also how different diatom species are compared with each other.

It is not common to use metabolic fingerprinting in taxonomy. I have compared species by using MS analysis producing markers without identifying the actual chemical compounds of the metabolic fingerprint, but only referring to markers. This may seem controversial to researchers that are used to work with chemical identities most often attached to special physiological traits. My aim was, however, to get an overall picture of the metabolome (or as large a section of it as possible). This is reflected in that there are few works to compare with, but I am convinced that as science progresses methods like this will come into use more frequently, also supplied by specific information of actual metabolite identities. I believe that the considerable chemical differences demonstrated in the study of *C. socialis* should be a good argument to include also

metabolomics data into taxonomy, as it is then possible to have variables that reflect what genes are actually expressed. For metabolomics to be useful in taxonomy, as well as in other scientific fields, methods have to be improved. Sharing data, like what has been done in genomics with open databases like Genbank and European Nucleotide Archive, could enhance the usefulness of metabolomics data. The BinBase initiative of the Fiehn lab (http://fiehnlab.ucdavis.edu/) is still relatively new, but now holds GC/MS data from over 24000 samples from over 60 species. It would be beneficial also to have similar databases for LC MS data. Methods, including data handling, also need to be standardized for possible comparison of data (Fiehn *et al.*, 2007;Verpoorte *et al.*, 2008). Increased possibilities in improved quantification and identification of metabolites would also give larger areas of application. Based on the results from the studies included in this thesis we were able to distinguish between samples belonging to different species, at least to a certain extent, but could not reflect the phylogeny of the centric species as reported from genomic studies (Kooistra *et al.*, 2007;Rampen *et al.*, 2009). More targeted approaches of metabolomics coupled to genomics, or other omics scientific fields like transcriptomics, may be more suited perhaps. The rapid development of the omics' sciences gives some promises for future research on functioning, for example the Marine Microbial Eukaryote Transcriptome Project (http://marinemicroeukaryotes.org). Metabolomics seem to be well suited as an additional trait for delimiting between taxa and in biogeography studies. Now the next natural step, in my opinion, would be to go further to investigate what these large differences exists in, for example by using GC MS to get a better picture of what compounds differ between species and other taxa levels as well as what compounds are effected by altered environmental conditions or physiological changes.

Acknowledgements

These four years as a Phd student have been a great experience and there are a lot of people who have contributed to that. Some of them I wish to thank in particular. First, I would like to thank MabCent, Center for marine bioactives and drug discovery, for financing this PhD project. I am very happy that you chose to fund a PhD position in chemotaxonomy and support taxonomic research! I would also like to thank my supervisors, Hans Chr. Eilertsen , Espen Hansen and Maria Perander for supporting me. Hans Chr, I really appreciate that your door is always open and your enthusiasm and curiosity makes it fun to part of your phytoplankton team. Thank you Espen for all your help and patience in introducing me to the world of metabolomics instrumentation and methods, and also for performing all the mass spectrometry analysis. Many hours was spent in the lab and in the cold rooms. These hours would not have been so fun, and they would definitively have been more, were it not for the trio Gunilla Eriksen, Richard Ingebrigtsen and Maria Degerlund, and also others that have contributed through the different experiments. Gunilla and Richard, you are both such gifted microalgae cultivators from whom I have learned a lot and I also would like to thank you for always being so cheerful, interested and helpful. I would also have been lost without you Maria. It has been a real privilege to work with you on the different experiments and manuscripts, sharing frustrations and successes. Thank you for being such a good co-author, colleague and friend. I hope we can continue our cooperation also in the future! As for the writing process I would also like to thank my other co-authors Adriana, Diana and Bjarne for fruitful cooperation. Thank you to my roommates through these four years, to Sünnje, Anaïs and Rahman, for all your support and good company. Rahman, I have really appreciated your help on Matlab and R programming. For making these years such a nice experience I also want to thank all other friends and colleagues at the Department of Arctic and Marine biology and at MabCent.

Last, but not least I would like to thank my family for all your support through the years. To Andreas and our wonderful girls; thank you for coming with me to Tromsø and supporting me all through these PhD adventures. I love you.

Siv Huseby

References

Allen, J. I. and Polimene, L. (2011) Linking physiology to ecology: towards a new generation of plankton models. *J. Plankton Res.,* **33,** 989-997.

Alverson, A. J. (2008) Molecular systematics and the diatom species. *Protist,* **159,** 339-353.

Amato, A., Kooistra, W. H. C. F., Ghiron, J. H. L., Mann, D. G., Proschold, T. and Montresor, M. (2007) Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist,* **158,** 193-207.

Armbrust, E. V. (2009) The life of diatoms in the world's oceans. *Nature,* **459,** 185-192.

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S. G., Allen, A. E., Apt, K. E., Bechner, M., Brzezinski, M. A., Chaal, B. K., Chiovitti, A., Davis, A. K., Demarest, M. S., Detter, J. C., Glavina, T., Goodstein, D., Hadi, M. Z., Hellsten, U., Hildebrand, M., Jenkins, B. D., Jurka, J., Kapitonov, V. V., Kroger, N., Lau, W. W. Y., Lane, T. W., Larimer, F. W., Lippmeier, J. C., Lucas, S., Medina, M., Montsant, A., Obornik, M., Parker, M. S., Palenik, B., Pazour, G. J., Richardson, P. M., Rynearson, T. A., Saito, M. A., Schwartz, D. C., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F. P. and Rokhsar, D. S. (2004) The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. *Science,* **306,** 79-86.

Barofsky, A., Simonelli, P., Vidoudez, C., Troedsson, C., Nejstgaard, J. C., Jakobsen, H. H. and Pohnert, G. (2010) Growth phase of the diatom *Skeletonema marinoi* influences the metabolic profile of the cells and the selective feeding of the copepod *Calanus* spp. *J. Plankton Res.,* **32,** 263-272.

Barofsky, A., Vidoudez, C. and Pohnert, G. (2009) Metabolic profiling reveals growth stage variability in diatom exudates. *Limnology and Oceanography-Methods,* **7,** 382-390.

Bavestrello, G., Arillo, A., Calcinai, B., Cattaneo-Vietti, R., Cerrano, C., Gaino, E., Penna, A. and Sara, M. (2000) Parasitic diatoms inside antarctic sponges. *The Biological Bulletin,* **198,** 29-33.

Becker, B., Hoef-Emden, K. and Melkonian, M. (2008) Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. *BMC Evol. Biol.,* **8**.

Beckmann, M., Parker, D., Enot, D. P., Duval, E. and Draper, J. (2008) High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry. *Nat Protoc,* **3,** 486-504.

Bissinger, J. E., Montagnes, D. J. S., Sharples, J. and Atkinson, D. (2008) Predicting marine phytoplankton maximum growth rates from temperature: Improving on the Eppley curve using quantile regression. *Limnol. Oceanogr.,* **53,** 487-493.

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J. A., Brownlee, C., Cadoret, J. P., Chiovitti, A., Choi, C. J., Coesel, S., De Martino, A., Detter, J. C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M. J. J., Jenkins, B. D., Jiroutova, K., Jorgensen, R. E., Joubert, Y., Kaplan, A., Kroger, N., Kroth, P. G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jezequel, V., Lopez, P. J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L. K., Montsant, A., Oudot-Le Secq, M. P., Napoli, C., Obornik, M., Parker, M. S., Petit, J. L., Porcel, B. M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T. A., Schmutz, J., Shapiro, H., Siaut, M., Stanley, M., Sussman, M. R., Taylor, A. R., Vardi, A., von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L. S., Rokhsar, D. S., Weissenbach, J., Armbrust, E. V., Green, B. R., Van De Peer, Y. and Grigoriev, I. V. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature,* **456,** 239-244.

Burkholder, J. M. and Glibert, P. M. (2006) Intraspecific variability: an important consideration in forming generalisations about toxigenic algal species. *Afr. J. Mar. Sci.,* **28,** 177-180.

Bölling, C. and Fiehn, O. (2005) Metabolite profiling of *Chlamydomonas reinhardtii* under nutrient deprivation. *Plant Physiol.,* **139,** 1995-2005.

Chauton, M. S., Optun, O. I., Bathen, T. F., Volent, Z., Gribbestad, I. S. and Johnsen, G. (2003) HR MAS ^{1}H NMR spectroscopy analysis of marine microalgal whole cells. *Mar Ecol-Prog Ser,* **256,** 57-62.

Cook, D., Fowler, S., Fiehn, O. and Thomashow, M. F. (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of *Arabidopsis*. *P Natl Acad Sci USA,* **101,** 15243-15248.

Cumming, G., Fidler, F. and Vaux, D. L. (2007) Error bars in experimental biology. *J. Cell Biol.,* **177,** 7-11.

Darwin, C. R. (1859) *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.* Vol., Murray, J., London.

de Castro Araújo, S. and Garcia, V. M. T. (2005) Growth and biochemical composition of the diatom *Chaetoceros* cf. *wighamii* brightwell under different temperature, salinity and carbon dioxide levels. I. Protein, carbohydrates and lipids. *Aquaculture,* **246,** 405-412.

De Queiroz, K. (2007) Species Concepts and Species Delimitation. *Syst. Biol.,* **56,** 879-886.

Degerlund, M. and Eilertsen, H. C. (2010) Main species characteristics of phytoplankton spring blooms in NE Atlantic and Arctic waters (68-80° N). *Estuaries and Coasts,* **33,** 242-269.

Dijkman, N. A. and Kromkamp, J. C. (2006) Phospholipid-derived fatty acids as chemotaxonomic markers for phytoplankton: application for inferring phytoplankton composition. *Mar Ecol-Prog Ser,* **324,** 113-125.

Dunstan, G. A., Volkman, J. K., Barrett, S. M., Leroi, J. M. and Jeffrey, S. W. (1994) Essential polyunsaturated fatty-acids from 14 species of diatom (Bacillariophyceae). *Phytochemistry,* **35,** 155-161.

Ellis, D. I., Dunn, W. B., Griffin, J. L., Allwood, J. W. and Goodacre, R. (2007) Metabolic fingerprinting as a diagnostic tool. *Pharmacogenomics,* **8,** 1243-1266.

Eppley, R. W. (1972) Temperature and phytoplankton growth in sea. *Fish. Bull.,* **70,** 1063-1085.

Falkowski, P. G., Barber, R. T. and Smetacek, V. (1998) Biogeochemical controls and feedbacks on ocean primary production. *Science,* **281,** 200-206.

Faurby, S. and Funch, P. (2011) Size is not everything: a meta-analysis of geographic variation in microscopic eukaryotes. *Global Ecol. Biogeogr.,* **20,** 475-485.

Fenchel, T. and Finlay, B. J. (2006) The diversity of microbes: resurgence of the phenotype. *Philosophical Transactions of the Royal Society B-Biological Sciences,* **361,** 1965-1973.

Fiala, M. and Oriol, L. (1990) Light-temperature interactions on the growth of Antarctic diatoms. *Polar Biol.,* **10,** 629-636.

Fiehn, O. (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Funct. Genomics,* **2,** 155-168.

Fiehn, O. (2002) Metabolomics - the link between genotypes and phenotypes. *Plant Mol. Biol.,* **48,** 155-171.

Fiehn, O., Robertson, D., Griffin, J., van der Werf, M., Nikolau, B., Morrison, N., Sumner, L. W., Goodacre, R., Hardy, N. W., Taylor, C., Fostel, J., Kristal, B., Kaddurah-Daouk, R., Mendes, P., van Ommen, B., Lindon, J. C. and Sansone, S. A. (2007) The metabolomics standards initiative (MSI). *Metabolomics,* **3,** 175-178.

Frisvad, J. C., Andersen, B. and Thrane, U. (2008) The use of secondary metabolite profiling in chemotaxonomy of filamentous fungi. *Mycol. Res.,* **112,** 231-240.

Gallagher, J. C. (1982) Physiological variation and electrophoretic banding patterns of genetically different seasonal populations of *Skeletonema costatum* (Bacillariophyceae). *J. Phycol.,* **18,** 148-162.

Gallagher, J. C., Wood, A. M. and Alberte, R. S. (1984) Ecotypic differentiation in the marine diatom *Skeletonema costatum* - Influence of light-intensity on the photosynthetic apparatus. *Mar. Biol.,* **82,** 121-134.
Gilstad, M. and Sakshaug, E. (1990) Growth-rates of 10 diatom species from the Barents Sea at different irradiances and day lengths. *Mar Ecol-Prog Ser,* **64,** 169-173.
Gray, G. R. and Heath, D. (2005) A global reorganization of the metabolome in *Arabidopsis* during cold acclimation is revealed by metabolic fingerprinting. *Physiol. Plant.,* **124,** 236-248.
Green, J. L., Bohannan, B. J. M. and Whitaker, R. J. (2008) Microbial biogeography: From taxonomy to traits. *Science,* **320,** 1039-1043.
Guillard, R. R. and Ryther, J. H. (1962) Studies of marine planktonic diatoms .1. *Cyclotella nana* Hustedt, and *Detonula confervacea* (Cleve) Gran. *Canadian Journal of Microbiology,* **8,** 229-239.
Guy, C., Kaplan, F., Kopka, J., Selbig, J. and Hincha, D. K. (2008) Metabolomics of temperature stress. *Physiol. Plant.,* **132,** 220-235.
Hasle, G. R. and Syvertsen, E. E. (1997) Marine Diatoms. In: C. R. Tomas (ed) *Identifying Marine Phytoplankton.* Academic Press, San Diego, pp. 5-385.
Hendey, I. N. (1959) The structure of the diatom cell wall as revealed by the electron microscope. *Journal of the Quekett Microscopical Club,* **5,** 147-175.
Hey, J. (2006) On the failure of modern species concepts. *Trends Ecol. Evol.,* **21,** 447-450.
Hu, J., Coombes, K. R., Morris, J. S. and Baggerly, K. A. (2005) The importance of experimental design in proteomic mass spectrometry experiments: Some cautionary tales. *Briefings in Functional Genomics & Proteomics,* **3,** 322-331.
Härnström, K., Ellegaard, M., Andersen, T. J. and Godhe, A. (2011) Hundred years of genetic structure in a sediment revived diatom population. *Proceedings of the National Academy of Sciences*.
Ivanisevic, J., Thomas, O. P., Lejeusne, C., Chevaldonne, P. and Perez, T. (2011) Metabolic fingerprinting as an indicator of biodiversity: towards understanding inter-specific relationships among Homoscleromorpha sponges. *Metabolomics,* **7,** 289-304.
Kacser, H. (1986) On parts and wholes in metabolism. In: G. R. Welch and J. S. Clegg (eds) *The organization of Cell Metabolism.* Plenum Press, New York.
Kagami, M. and Urabe, J. (2001) Phytoplankton growth rate as a function of cell size: an experimental test in Lake Biwa. *Limnology,* **2,** 111-117.
Kociolek, J. P. (2011) Microscopic in size: Macroscopic in impact. Diatom–human interactions. In: Z. Dubinsky and J. Seckbach (eds) *All Flesh Is Grass: Plant-Animal Interrelationships.* Vol. 16. Springer Netherlands, pp. 257-283.
Kociolek, J. P. and Stoermer, E. F. (2001) Taxonomy and ecology: A marriage of necessity. *Diatom Research,* **16,** 433-442.
Kooistra, W. H. C. F., Gersonde, R., K. Medlin, L. and G. Mann, D. (2007) Chapter 11 - The origin and evolution of the diatoms: Their adaptation to a planktonic existence. In: G. F. Paul and H. K. Andrew (eds) *Evolution of Primary Producers in the Sea.* Academic Press, Burlington, pp. 207-249.
Kooistra, W. H. C. F., Sarno, D., Balzano, S., Gu, H., Andersen, R. A. and Zingone, A. (2008) Global diversity and biogeography of *Skeletonema* species (Bacillariophyta). *Protist,* **159,** 177-193.
Kooistra, W. H. C. F., Sarno, D., Hernandez-Becerril, D. U., Assmy, P., Di Prisco, C. and Montresor, M. (2010) Comparative molecular and morphological phylogenetic analyses of taxa in the Chaetocerotaceae (Bacillariophyta). *Phycologia,* **49,** 471-500.
Kudo, I. (2003) Change in the uptake and cellular Si:N ratio in diatoms responding to the ambient Si:N ratio and growth phase. *Mar. Biol.,* **143,** 39-46.
Lacorn, M., Piechotta, G., Simat, T. J., Kammann, U., Wosniok, W., Lang, T., Muller, W. E. G., Schroder, H. C., Jenke, H. S. and Steinhart, H. (2001) Annual cycles of apoptosis, DNA strand breaks, heat

shock proteins, and metallothionen isoforms in dab (*Limanda limanda*): influences of natural factors and consequences for biological effect monitoring. *Biomarkers,* **6,** 108-126.

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

Lang, I. K., Hodac, L., Friedl, T. and Feussner, I. (2011) Fatty acid profiles and their distribution patterns in microalgae: a comprehensive analysis of more than 2000 strains from the SAG culture collection. *BMC Plant Biol.,* **11**.

Larsen, T. O., Smedsgaard, J., Nielsen, K. F., Hansen, M. E. and Frisvad, J. C. (2005) Phenotypic taxonomy and metabolite profiling in microbial drug discovery. *Natural Product Reports,* **22,** 672-695.

Lavorel, S., McIntyre, S., Landsberg, J. and Forbes, T. D. A. (1997) Plant functional classifications: from general groups to specific groups based on response to disturbance. *Trends in Ecology & Evolution,* **12,** 474-478.

Lelong, A., Hégaret, H., Soudant, P. and Bates, S. S. (2012) *Pseudo-nitzschia* (Bacillariophyceae) species, domoic acid and amnesic shellfish poisoning: revisiting previous paradigms. *Phycologia,* **51,** 168-216.

Loret, P., Tengs, T., Villareal, T. A., Singler, H., Richardson, B., McGuire, P., Morton, S., Busman, M. and Campbell, L. (2002) No difference found in ribosomal DNA sequences from physiologically diverse clones of *Karenia brevis* (Dinophyceae) from the Gulf of Mexico. *J. Plankton Res.,* **24,** 735-739.

Lundholm, N., Moestrup, O., Kotaki, Y., Hoef-Emden, K., Scholin, C. and Miller, P. (2006) Inter- and intraspecific variation of the *Pseudo-nitzschia delicatissima* complex (Bacillariophyceae) illustrated by rRNA probes, morphological data and phylogenetic analyses. *J. Phycol.,* **42,** 464-481.

Macel, M., van Dam, N. M. and Keurentjes, J. J. B. (2010) Metabolomics: the chemistry between ecology and genetics. *Mol Ecol Resour,* **10,** 583-593.

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

Mann, D. G. (2010) Discovering diatom species: is a long history of disagreements about species-level taxonomy now at an end? *Plant Ecology and Evolution,* **143,** 251-264.

Mayr, E. (1942) *Systematics and the origin of species: from the viewpoint of a zoologist.* Vol., Columbia University Press, New York.

Mayr, E. (1996) What is a species, and what is not? *Philosophy of Science,* **63,** 262-277.

Medlin, L. K. (2007) If everything is everywhere, do they share a common gene pool? *Gene,* **406,** 180-183.

Moniz, M. B. J. and Kaczmarska, I. (2009) Barcoding diatoms: Is there a good marker? *Mol Ecol Resour,* **9,** 65-74.

Montagnes, D. J. S. and Franklin, D. J. (2001) Effect of temperature on diatom volume, growth rate, and carbon and nitrogen content: Reconsidering some paradigms. *Limnol. Oceanogr.,* **46,** 2008-2018.

Mos, L. (2001) Domoic acid: a fascinating marine toxin. *Environ. Toxicol. Pharmacol.,* **9,** 79-85.

Moustafa, A., Beszteri, B., Maier, U. G., Bowler, C., Valentin, K. and Bhattacharya, D. (2009) Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science,* **324,** 1724-1726.

Ogata, T., Kodama, M. and Ishimaru, T. (1987) Toxin production in the dinoflagellate *Protogonyaulax tamarensis*. *Toxicon,* **25,** 923-928.

Orsini, L., Procaccini, G., Sarno, D. and Montresor, M. (2004) Multiple rDNA ITS-types within the diatom *Pseudo-nitzschia delicatissima* (Bacillariophyceae) and their relative abundances across a spring bloom in the Gulf of Naples. *Mar. Ecol. Prog. Ser.,* **271,** 87-98.

Pedros-Alio, C. (2006) Marine microbial diversity: can it be determined? *Trends Microbiol.,* **14,** 257-263.

Prestegard, S. K., Oftedal, L., Coyne, R. T., Nygaard, G., Skjærven, K. H., Knutsen, G., Døskeland, S. O. and Herfindal, L. (2009) Marine benthic diatoms contain compounds able to induce leukemia cell death and modulate blood platelet activity. *Mar. Drugs,* **7,** 605-623.

Rampen, S. W., Schouten, S., Panoto, F. E., Brink, M., Andersen, R. A., Muyzer, G., Abbas, B. and Damste, J. S. S. (2009) Phylogenetic position of *Attheya longicornis* and *Attheya septentrionalis* (Bacillariophyta). *J. Phycol.,* **45,** 444-453.

Reynolds, C. S., Huszar, V., Kruk, C., Naselli-Flores, L. and Melo, S. (2002) Towards a functional classification of the freshwater phytoplankton. *J. Plankton Res.,* **24,** 417-428.

Ribera D'Alcala, M., Conversano, F., Corato, F., Licandro, P., Mangoni, O., Marino, D., Mazzocchi, M. G., Modigh, M., Montresor, M., Nardella, M., Saggiomo, V., Sarno, D. and Zingone, A. (2004) Seasonal patterns in plankton communities in a pluriannual time series at a coastal Mediterranean site (Gulf of Naples): an attempt to discern recurrences and trends. *Scientia Marina,* **68,** 65-83.

Rosselló-Móra, R. (2012) Towards a taxonomy of Bacteria and Archaea based on interactive and cumulative data repositories. *Environ. Microbiol.,* **14,** 318-334.

Rossello-Mora, R., Lucio, M., Pena, A., Brito-Echeverria, J., Lopez-Lopez, A., Valens-Vadell, M., Frommberger, M., Anton, J. and Schmitt-Kopplin, P. (2008) Metabolic evidence for biogeographic isolation of the extremophilic bacterium *Salinibacter ruber*. *Isme Journal,* **2,** 242-253.

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

Rynearson, T. A. and Armbrust, E. V. (2000) DNA fingerprinting reveals extensive genetic diversity in a field population of the centric diatom *Ditylum brightwellii*. *Limnol. Oceanogr.,* **45,** 1329-1340.

Sarno, D., Kooistra, W. H. C. F., Balzano, S., Hargraves, P. E. and Zingone, A. (2007) Diversity in the genus *Skeletonema* (Bacillariophyceae): III. Phylogenetic position and morphological variability of *Skeletonema costatum* and *Skeletonema grevillei*, with the description of *Skeletonema ardens* sp. nov. *J. Phycol.,* **43,** 156-170.

Sarno, D., Kooistra, W. H. C. F., Medlin, L. K., Percopo, I. and Zingone, A. (2005) Diversity in the genus *Skeletonema* (Bacillariophyceae). II. An assessment of the taxonomy of *S-costatum*-like species with the description of four new species. *J. Phycol.,* **41,** 151-176.

Sarthou, G., Timmermans, K. R., Blain, S. and Treguer, P. (2005) Growth physiology and fate of diatoms in the ocean: a review. *J. Sea Res.,* **53,** 25-42.

Schreiber, U., Schliwa, U. and Bilger, W. (1986) Continuous recording of photochemical and nonphotochemical chlorophyll fluorescence quenching with a new type of modulation fluorometer. *Photosynthesis Res.,* **10,** 51-62.

Smayda, T. J. (1958) Biogeographical studies of marine phytoplankton. *Oikos,* **9,** 158-191.

Smedsgaard, J. and Nielsen, J. (2005) Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics. *J. Exp. Bot.,* **56,** 273-286.

Sorhannus, U. and Fox, M. G. (2012) Phylogenetic analyses of a combined data set suggest that the *Attheya* lineage is the closest living relative of the pennate diatoms (Bacillariophyceae). *Protist,* **163,** 252-262.

Suzuki, Y. and Takahashi, M. (1995) Growth responses of several diatom species isolated from various environments to temperature. *J. Phycol.,* **31,** 880-888.

Teoh, M.-L., Chu, W.-L., Marchant, H. and Phang, S.-M. (2004) Influence of culture temperature on the growth, biochemical composition and fatty acid profiles of six Antarctic microalgae. *J. Appl. Phycol.,* **16,** 421-430.

Thessen, A. E., Bowers, H. A. and Stoecker, D. K. (2009) Intra- and interspecies differences in growth and toxicity of *Pseudo-nitzschia* while using different nitrogen sources. *Harmful Algae,* **8,** 792-810.

Tréguer, P., Nelson, D. M., Bennekom, A. J. V., DeMaster, D. J., Leynaert, A. and Quéguiner, B. (1995) The silica balance in the world ocean: A reestimate. *Science,* **268,** 375-379.

Verity, P. G. (1981) Effects of temperature, irradiance, and daylength on the marine diatom *Leptocylindrus danicus* Cleve. I. Photosynthesis and cellular composition. *J. Exp. Mar. Biol. Ecol.,* **55,** 79-91.

Verpoorte, R., Choi, Y. and Kim, H. (2007) NMR-based metabolomics at work in phytochemistry. *Phytochem. Rev.,* **6,** 3-14.

Verpoorte, R., Choi, Y., Mustafa, N. and Kim, H. (2008) Metabolomics: back to basics. *Phytochem. Rev.,* **7,** 525-537.

Viant, M. and Sommer, U. (2012) Mass spectrometry based environmental metabolomics: a primer and review. *Metabolomics,* 1-15.

Viant, M. R. (2007) Metabolomics of aquatic organisms: the new 'omics' on the block. *Mar Ecol-Prog Ser,* **332,** 301-306.

Vidoudez, C. and Pohnert, G. (2011) Comparative metabolomics of the diatom *Skeletonema marinoi* in different growth phases. *Metabolomics,* 1-16.

Wassmann, P., Slagstad, D., Riser, C. W. and Reigstad, M. (2006) Modelling the ecosystem dynamics of the Barents Sea including the marginal ice zone: II. Carbon flux and interannual variability. *J. Mar. Syst.,* **59,** 1-24.

Welker, M. and Moore, E. R. B. (2011) Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. *Syst. Appl. Microbiol.,* **34,** 2-11.

Wood, A. M., Leatham, T., Manhart, J. R. and Mccourt, R. M. (1992) The species concept in phytoplankton ecology. *J. Phycol.,* **28,** 723-729.

Yoder, J. A. (1979) Effect of temperature on light-limited growth and chemical composition of *Skeletonema costatum* (Bacillariophyceae). *J. Phycol.,* **15,** 362-370.

Zapata, M., Rodriguez, F., Fraga, S., Barra, L. and Ruggiero, M. V. (2011) Chlorophyll C pigment patterns in 18 species (51 Strains) of the genus *Pseudo-Nitzschia* (Bacillariophyceae). *J. Phycol.,* **47,** 1274-1280.

Zhukova, N. V. and Aizdaicher, N. A. (1995) Fatty-acid composition of 15 species of marine microalgae. *Phytochemistry,* **39,** 351-356.

PAPER I

Eilertsen H.C., Huseby S., Degerlund, M., Eriksen, G.K., Ingebrigtsen, R.A., Hansen, E. (submitted manuscript)

Sample storage effects on reproducibility of direct infusion high-resolution mass spectrometry (HR-MS) microalgae extract analysis

PAPER II

Huseby, S., Hansen, E., Degerlund, M., Eriksen, G. K., Ingebrigtsen, R.A, Eilertsen, H.C. (submitted manuscript)
Chemical diversity of six northern diatoms

PAPER III

Degerlund, M., Huseby, S., Zingone, A., Sarno, D., Landfald, B. (2012)
Functional diversity in cryptic species of *Chaetoceros socialis* Lauder (Bacillariophyceae)
Journal of Plankton Research, **34**, 416-431

PAPER IV

Huseby S., Degerlund M., Zingone, A., Hansen. E (in revision)
Metabolic fingerprinting reveals differences between northern and southern strains of the cryptic diatom *Chaetoceros socialis.*
European Journal of Phycology

ISBN xxx-xx-xxxx-xxx-x