Salla Kalaniemi

Phylogenetic Analysis of the Baltic *Skeletonema marinoi*

Helsinki Metropolia University of Applied Sciences

Bachelor of Sciences

Laboratory Sciences

Thesis

3.2.2013

Helsinki Metropolia University of Applied Sciences

Author Title Number of Pages Date	Salla Kalaniemi Phylogenetic Analysis of the Baltic *Skeletonema marinoi* 30 pages + 3 appendices 3 February 2013
Degree	Bachelor of Sciences
Degree Programme	Laboratory Sciences
Instructors	Anke Kremp, PhD Conny Sjöqvist, MSc Jarmo Palm, MSc

This thesis project was carried out in the Marine Research Centre of Finnish Environment Institute (SYKE) in the Modelling and Innovations Unit where researchers PhD. Anke Kremp and PhD student Conny Sjöqvist are investigating phytoplankton ecology, biodiversity and population genetics, and one of the organisms under study is *Skeletonema marinoi*. The aim of this graduate study was to investigate the possible differentiation of the *S. marinoi* strains isolated in the Baltic Sea by phylogenetic analysis.

S. marinoi is a diatom species with worldwide distribution. As a photosynthezing organism it has a significant role in the carbon dioxide assimilation. Moreover, the typical diatom cell wall called frustule makes *S. marinoi* an important organism for the silica cycle on Earth. *S. marinoi* is highly abundant also in the Baltic Sea especially during the spring bloom. The Baltic Sea is an interesting brackish water basin with steep salinity gradient. Due to these facts *S. marinoi* is fundamentally studied in many aspects.

By phylogenetic analysis it is possible to gain huge amounts of information to understand the relationships between organism and even within species, moreover to investigate the evolutionary hypothesis. In this graduate study non-coding ITS region of rDNA was analyzed to observe the possible variation within *S. marinoi* strains. The isolated strains were maintained in the laboratory and DNA was extracted. The ITS region was copied by PCR reaction and sequenced. The sequences were combined with sequences from the GenBank to construct multiple alignments by ClustalW method in the MEGA program. Finally, phylogenetic trees were constructed based on the multiple alignments in the MEGA by Neighbor-Joining method. The trees were rooted by *Thalassiosira rotula* as an outgroup. In the end the reliability of the trees was estimated by Bootstrap analysis.

Two phylogenetic trees are represented in this graduate study. The first tree is relatively reliable according the Bootstrap values, and supports the hypothesis that differentiation has occurred within the *S. marinoi* strains in the Baltic Sea. The northern and southern strains have separated into their own clades which indicates that the salinity gradient in the Baltic Sea has caused adaption. The second tree is partly contradictory and cannot be explained well in this thesis project.

Keywords	*Skeletonema marinoi*, ITS, Phylogenetic analysis, Phylogenetic tree

Tekijä Otsikko Sivumäärä Aika	Salla Kalaniemi Itämerellä esiintyvän *Skeletonema marinoi*n fylogeneettinen analyysi 30 sivua + 3 liitettä 3.2.2013
Tutkinto	Laboratorioanalyytikko (AMK)
Koulutusohjelma	Laboratorioalan koulutusohjelma
Ohjaajat	Erikoistutkija Anke Kremp Tutkija Conny Sjöqvist Lehtori Jarmo Palm

Opinnäytetyö toteutettiin Suomen ympäristökeskuksen Merikeskuksessa Innovaatiot ja mallit yksikössä, jossa erikoistutkija Anke Kremp ja tutkija Conny Sjöqvist tutkivat kasviplanktonien ekologiaa, monimuotoisuutta sekä populaatiogenetiikkaa. Piilevä *Skeletonema marinoi* on yksi heidän tutkimistaan organismeista, ja tässä opinnäytetyössä oli tavoitteena fylogeneettisen analyysin avulla tutkia Itämerellä eristettyjen *S. marinoi* -kantojen erilaistumista. Itämeri on murtovesiallas, jossa on voimakas suolagradientti, minkä seurauksena se on mielenkiintoinen tutkimuskohde. *S. marinoi* on maailmanlaajuisesti levinnyt piilevä, joka kukkii keväisin voimakkaasti myös Itämerellä. Yhteyttävänä organismina sillä on merkittävä rooli hiilensidonnassa. Lisäksi piileville tyypillinen piidioksidisoluseinärakenne tekee *S. marinoi*sta keskeisen organismin myös piinkierrossa.

Fylogeneettisen analyysin avulla on mahdollista saada paljon tietoa sekä lajien välisistä että sisäisistä suhteista. Analyysin avulla pystytään tutkimaan lajien evolutiivista historiaa. Tässä opinnäytetyössä analysoitiin ribosomaalisen DNA-alueen ITS-jaksoa, joka ei koodaa mitään geenejä eikä rRNA:ta, mutta on osoittautunut hyväksi fylogeneettiseksi markkeriksi lajinsisäisessä tutkimuksessa.

Työssä eristettiin DNA:ta laboratoriokasvatuksessa olleista *S. marinoi* -kannoista. DNA-näytteistä kopioitiin PCR-reaktiolla ITS-jaksoja, jotka lähetettiin sekvensoitaviksi. Sekvenssit linjattiin yhdessä GenBankista kerättyjen sekvenssien kanssa MEGA-ohjelmassa ClustalW-metodilla. Sekvenssirinnastuksista rakennettiin fylogeneettisiä puita MEGA-ohjelmassa Neighbor-Joining -metodilla. Puut juurrutettiin piilevä *Thalassiosira rotula*n avulla, ja puiden luotettavuus arvioitiin Bootstrap-analyysillä.

Tässä työssä esitellään kaksi fylogeneettistä puuta, joista ensimmäinen on suhteellisen luotettava Bootstrap-arvojen perusteella. Puussa Itämeren kannat ovat erottuneet Kanadan kannoista sekä Itämeren pohjoiset ja eteläiset kannat ovat eriytyneet omiin haaroihinsa. Havaittu järjestäytyminen tukee esitettyä hypoteesiä suolagradientin vaikutuksesta *S. marinoi*n erilaistumiseen Itämerellä. Toinen puu ei ole kovin luotettava, eikä kantojen osin ristiriitaista järjestäytymistä pystytä selittämään tämän opinnäytetyön puitteissa.

Avainsanat	*Skeletonema marinoi*, ITS, fylogeneettinen analyysi, fylogeneettinen puu

List of Abbreviations

AGE	Agarose Gel Electrophoresis
CTAB	Cetyl-Trimetyl-Ammonium-Bromide
DNA	Deoxyribonucleic Acid
ITS	Internal Transcriber Spacer
LSU	Large Sub Unit
mtDNA	Mitochondrial Deoxyribonucleic Acid
Myr	Million years
NJ	Neighbor-Joining
PCR	Polymerase Chain Reaction
psu	Practical salinity unit
rDNA	Deoxyribonucleic Acid encoding Ribosomal Ribonucleic Acid
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
SSU	Small Sub Unit
SYKE	Suomen ympärisökeskus
TBE	Tris-Borate-EDTA

Contents

1 Introduction

Skeletonema marinoi is a marine diatom species with worldwide distribution, and it is highly abundant in the Baltic Sea during spring bloom. As a photosynthesizing organism *S. marinoi* has a significant role in the carbon dioxide assimilation. Moreover, as a diatom *S. marinoi* produces a silica frustule which makes it an important organism for the silica cycle on Earth. Due to these facts, *S. marinoi* is fundamentally studied in many aspects. In addition, *S. marinoi* is relatively simple to maintain in a laboratory, hence it is a suitable model organism for different phytoplankton studies.

The Baltic Sea is an interesting habitat since it is a brackish water basin. There are several rivers which bring fresh water into the sea, and narrow sounds regulate water exchange between the Baltic Sea and the North Sea. In the north and close to St. Petersburg the surface water salinity is lowest while it is highest in the Danish Straits. The special environment has created different methods of adaption among the organism in the Baltic Sea and differentiation also occurs within species.

Phylogenetic analysis is a tool which can be used to investigate the evolutionary relationships of species. Usually, a phylogenetic tree is drawn to illustrate the calculated phylogenetic model. For the phylogenetic analysis DNA sequences of species of interest must be collected and analysed in a statistical program. The phylogenetic analysis is relatively easy to conduct and plenty of information can be gained but it is particularly important to understand that the constructed tree is always only a hypothesis of the actual genetic relationships.

Wide marine research is conducted in the Marine Research Centre of Finnish Environment Institute (SYKE). In the Modelling and Innovations Unit researchers PhD Anke Kremp and PhD student Conny Sjöqvist are investigating phytoplankton ecology, biodiversity and population genetics, and one of the organisms under study is *S. marinoi*. This graduate study was carried out as a side project of the PhD study of Conny Sjöqvist. The aim of this graduate study was to investigate the possible differentiation of the *S. marinoi* strains isolated in the Baltic Sea by phylogenetic analysis. In addition potential parallels between the salinity gradient in the Baltic Sea and the genetic relationships of populations from a geographically widespread area were observed.

2 Diatoms and *Skeletonema marinoi* in the Baltic Sea

This chapter discusses what kind of organisms diatoms and *S. marinoi* are. Consequently, fundamental functions and characteristics are represented. In addition, information about the suggested evolutionary history and taxonomy of diatoms is given.

2.1 Diatoms in the marine environment

Diatoms are a huge group of microscopic unicellular algae growing in the aquatic environments worldwide, and with an estimated 200,000 different species it is the most diverse phytoplankton group dominating most of the phytoplankton communities. Diatom cells can form chains of connected cells or grow as single cells. (1)

Diatoms are photoautotrophic and marine diatoms have a major role in the carbon assimilation. By photosynthesis marine diatoms are able to fix atmospheric carbon dioxide as much as terrestrial rainforests. By using sunlight, carbon dioxide, and water to produce biologically available energy diatoms form the base for marine food webs together with other phytoplankton groups. Glaciers and permafrost makes photosynthesis limited in polar environments but the diatoms are able to support the food web there both in the marine and terrestrial ecosystems. (1) Diatoms store the energy in a form of a glucose polymer called laminarin or in oil molecules (2 p. 585). The ability to generate chemical energy and metabolic intermediates from the breakdown of fatty acids is thought to be one of the main reasons why diatoms are able to survive long periods in the dark at the poles. In addition some large and some open-sea species can control their buoyancy, and so absorb nutrients from deeper while photosynthesizing in the surface waters. (1)

A diatom population can increase rapidly in favorable conditions. Usually this takes place in well-mixed coastal upwelling regions but also on the sea-ice edges. In addition to carbon dioxide and light diatoms require inorganic nitrogen, phosphorus, iron, silicon and trace elements to grow. The need and ability to restore different nutrients such as iron depends on the diatom species. However, heavy metals and different organic xenobiotics easily inhibit the diatom growth and cyanobacteria start to dominate the habitat. Hence, water quality can be determined by measuring the phytoplankton diversity. (3)

2.2 Cell wall structure, reproduction and silicon cycle

Diatoms have a special glass-like cell wall called frustule. The frustule is constructed of silica which endures extremely high pressure and gives protection against predators. The frustule has its species-specific form which can be highly ornate and morphologically symmetric including different nano- and micro-scale pores, ridges and tubules as shown in Figure 1. (2 p. 585) Diatoms have been divided into two major groups according their symmetry. The symmetry can be pinnate when similar parts are arranged bilateral on opposite sides of an axis and the diatom is called pennate. In comparison the diatom whose symmetry is radial is called centric. (4) Moreover, proteins and polysaccharides are added on the cell wall to prevent the silica form dissolving in the sea water (1).

The cell wall of a diatom is made up of two separated parts called valves and they fit together like a lid and a box providing the name from Greek where *diatomos* means “cut in half”. Reproduction of the diatoms is mainly asexual by mitotic division where the parental cell divides into two valves and both daughter cells generate a new half fitting inside the old one. In consequence, the new valve is always smaller than the parental one hence the new diatoms become gradually smaller over the time. Nevertheless, sexual reproduction restores the original cell size. During sexual reproduction diatoms forms a special auxospore cell. The significance of sexual production is not well-known but it is assumed to occur relatively rarely. (1, 4) In addition diatoms are able to form cysts as resisting stages (2 p. 585).

Figure 1. A dark-field photomicrograph of different diatom species illustrates the complex frustules (5).

Due to the silica frustule diatoms have a major role in the silicon (Si) cycle in the oceans together with silicoflagellates and radiolarians. Diatom cells take up silicon as dissolved in silicic acid (H_2SiO_4) to form silica (SiO_2) as it is illustrated in Figure 2. Part of the silicon is released as silica and dissolved into silicic acid right after the diatom cell dies. On the other hand, sinking diatom cells widely bring silicon, and also carbon, to the deeper ocean layers and finally to the ocean floor. Part of the silicon is re-incorporated in the cycle as silica by upwelling. Because of the relatively large mass of a diatom cell, it sinks quite rapidly and the strong frustule protects it on the ocean floor from decomposers such as bacteria. This stresses the significant role of diatoms in the biological pump in the marine environment. Silicon and organic material can be buried in sediments for millions of years over the geological timescales. This ability has even brought up an idea of fertilizing the ocean to promote the diatom blooms. This could potentially protect against global warming and elevate CO_2 levels in the ocean. (1, 2 p. 585, 5 p. 738)

Figure 2. The cycle of silica as different compounds in the marine environment is described in this graph (5).

2.3 Evolutionary history of diatoms

The persistent of diatom frustules has created a high quality fossil record which indicates that diatoms have had a dominant role in the marine environments for at least 90 million years (Myr). Even though, the molecular-clock analysis demonstrates that diatoms arose already 250 Myr ago on the Triassic period, some fossils are from the early Jurassic period being 190 Myr old. On the Cretaceous time 100 Myr ago, when diatoms reached a major role in the carbon cycle, atmospheric CO_2 level was three times higher than nowadays and the O_2 level was increasing. At that time the divergence of a second major lineage of diatoms occurred, bipolar and multipolar centrics. Diatoms survived well over the mass extinction at the end of Cretaceous period about 65 Myr ago. Moreover, the centric species were able to colonize a new habitat, the open ocean, and the impact of diatoms on the carbon cycle even increased. From the same time there is fossil evidence of the emergence of a third diatom group araphid pennates. The fourth diatom group, the raphid pennates arose 30 Myr ago, and they have a special slit called raphe in the frustule which allows them to glide along the surfaces. (1, 3) The separation events of diatoms are timed with Earth's historical periods in Figure 3. Although the strong fossil evidence, nowadays the classification and hypothesis of evolutionary lines both between and within centrics and pennates diatoms based on an outward appearance are highly questioned, instead the molecular methods are used and new studies are in progress. (6)

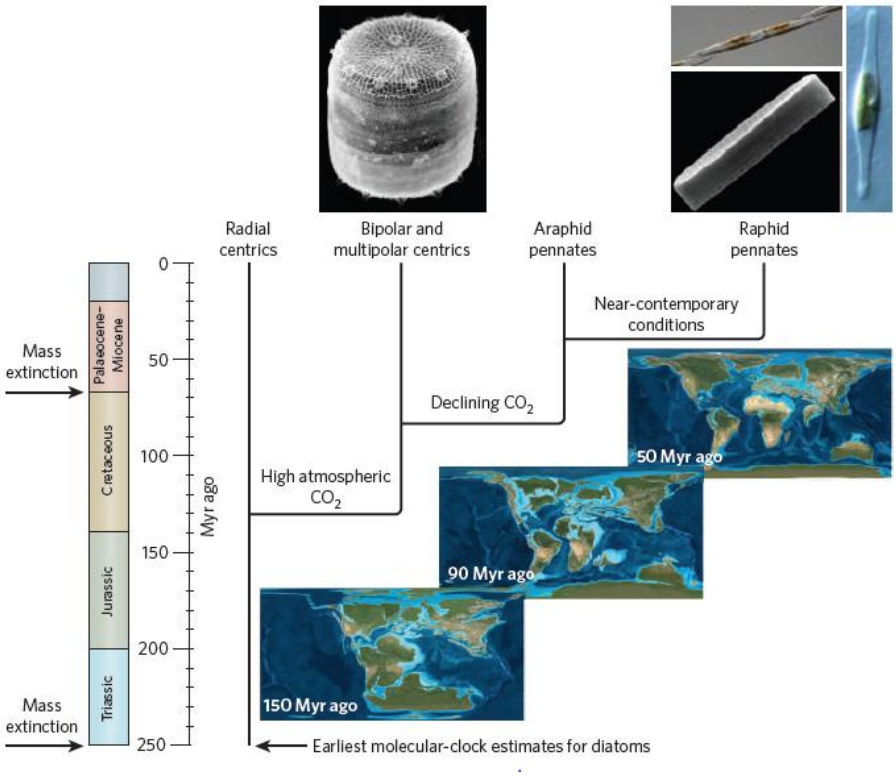

Figure 3. The chart describes the separation of three different diatom lineages on estimated timescale (1).

2.4 Classification of diatoms

As stated before the taxonomic classification of organisms has been based mainly on the morphology, which necessarily has nothing common with the real evolutionary relationships. The new molecular methods based on the analysis of DNA have created the phylogenic classification where the evolutionary history is included. Still, also the phylogenetic analysis is only a hypothesis and many lineages are unclear. Although, the taxonomic classification have been revised widely as a result of new genetic information. (2 p. 537, 3, 6)

As described before diatoms are eukaryote cells, and they are classified into the kingdom called Chromista or Chromalveolata. This kingdom is an extremely diverse clade of protists, where protist is an old term for eukaryotes which are not plants, animals or fungi and are mostly unicellular. In the Chromalveolata kingdom diatoms belong to the phylum called Herokonta or Stramenopiles. Current DNA data suggest that the Chromalveolates have a common ancestor, photosynthetic red alga, which as a result of the secondary endosymbiosis derivated the Chromalveolates. In comparison, the red

alga originates from the primary endosymbiosis. Although this hypothesis of the origin of the kingdom Chromalveolata and its two phyla is partly contradictory because of discrepancy in the appearance of plastid genes, it is thought to be the best current scientific hypothesis. (2 p. 575-588)

The second phylum of Chromalveolata is Alveolates and the species are recognized by their special membrane-bounded sacs called alveoli. Consequently, the group has strong molecular evidence behind. Three subgroups of the Alveolates are plankton cells called dinoflagellates, animal parasites called apicomlexans and ciliates which use their cilia to move and feed. In addition to the diatoms the Stramenopiles phylum include also other marine algae such as mostly unicellular golden algae, complex multicellular brown algae and as water molds called oomycetes. Structurally stramenopiles have a special flagellum with numerous fine hairs; moreover the flagellum can be paired with a short nonhairy flagella. This character provides the name of the group since *stramen* means straw and *pilos* hair in Latin. (2 p. 578-588)

2.5 *S. marinoi* and the Baltic Sea

The diatom class under phylum Heterokonta is also called as Bacillariophyceae. Under this class *S. marinoi* belongs to the order of Thassiosiranae and furthermore, to the family of Thassiosirales and finally to the genus of Skeletonemacea. (7) *S. marinoi* is a centric cosmopolitan marine species which blooms mainly in the temperate sea waters including the Baltic Sea. *S. marinoi* appears as long chains of single cells due to a special connective structure fultoportule at the base of each cell. (8) In Figure 4 is a scanning electron microscopy picture of *S. marinoi* cells which are connected by the fultoportule. The morphology of *S. marinoi* can be also seen in this picture. Furthermore, a micrograph of a chain of *S. marinoi* is shown in Figure 5.

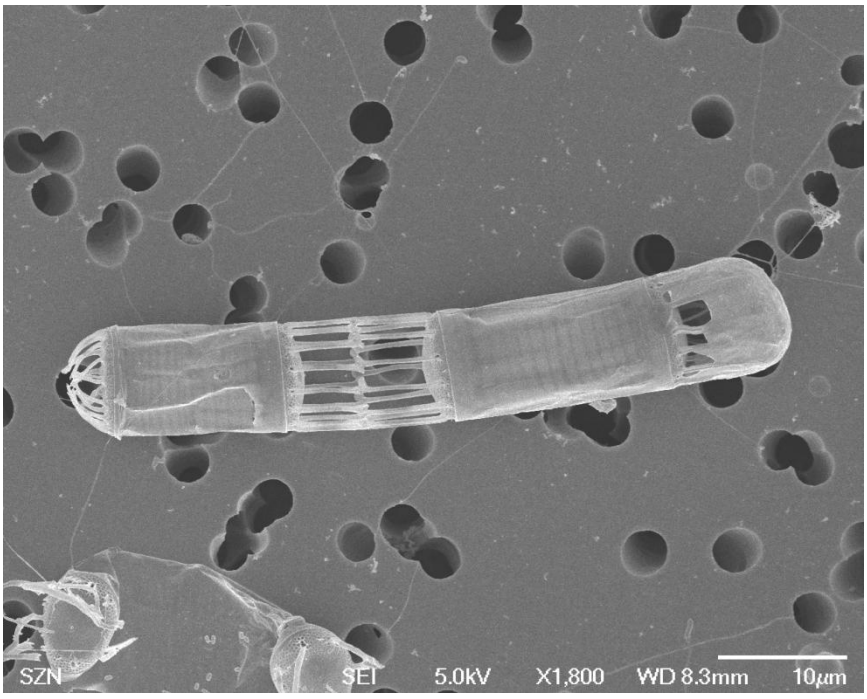

Figure 4. Scanning electron microscopy picture of *S. marinoi*. (Courtesy of Anke Kremp)

S. marinoi cells produce aldehydes which disturb other marine organisms such as copepods and sea urchins (9). *S. marinoi* is present in the Baltic Sea throughout the year but is most abundant during the spring bloom. Moreover, high abundances of benthic resting stages are found in the Baltic Sea sediments. Some studies have estimated the *S. marinoi* concentration to be up to 3.5 millions of propagules per gram of sediment. These propagules are able to survive even decades in the sediments. (10) The strong presence of *S. marinoi* in the Baltic Sea indicates its important role as primary producer there. It also serves as a valuable source of energy for higher trophic levels. (9) Consequently, studying this organism has a high ecological relevance. The long clonal chains create a great base for examinations of populations. In addition, it has been observed that the species is relatively easy to isolate and maintain in the laboratory. As a result, *S. marinoi* has been taken into account as a good phytoplankton model species within molecular research. (10) So far, wide genetic heterogeneity and physiologic differentiation among local and regional populations have been discovered. Hence, the large genetic variation suggests that sexual reproduction must be important. (9)

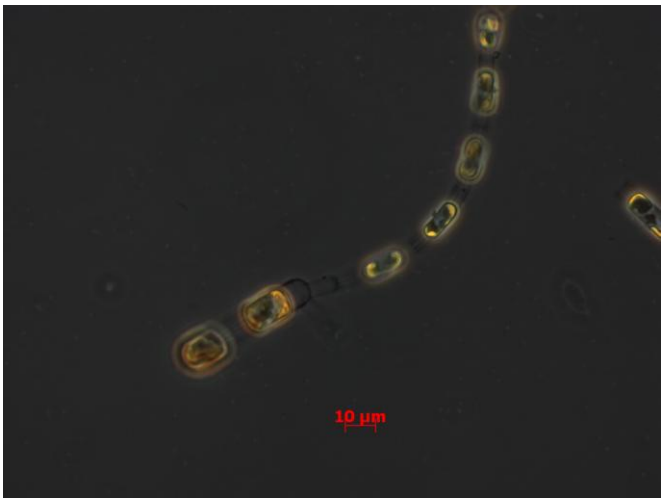

Figure 5. A photograph of a *S. marinoi* chain. (Courtesy of Anke Kremp)

The Baltic Sea is a special environment being a brackish water basin with a relatively steep salinity gradient. There are several rivers which drain fresh water in to the sea. Moreover, the narrow sounds between Denmark and Sweden regulate the water exchange. New salty sea water arrives to the shallow Baltic Sea as pulls. The salinity is lowest in the North and close to St. Petersburg, 2-4 psu, and highest in the Danish Straits, 7-15 psu. In Figure 6 is shown a rough illustration of the surface water salinity concentration in the Baltic Sea. These special properties require that the organisms living in the Baltic Sea are able to adapt. Furthermore, these properties result in great physiological variation even within species. (8, 11)

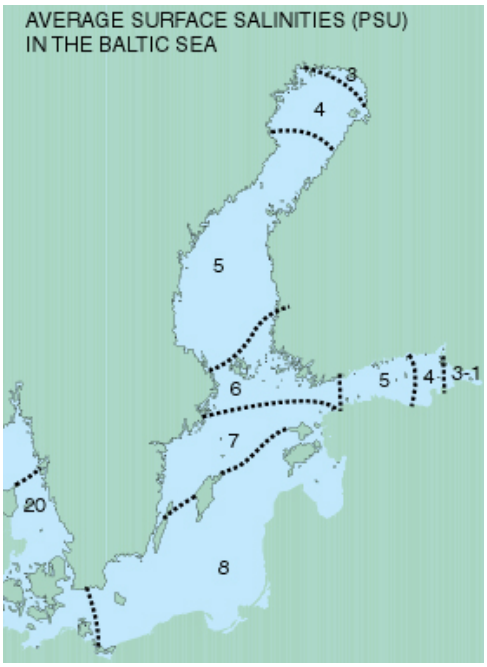

Figure 6. The salinity gradient present in the Baltic Sea (12).

3 ITS and phylogenetic analysis

In this section it is discussed what phylogenetic analysis means, what kind of information can be gained by it, and why it is applied so often nowadays. Primarily, the molecular tools used in this graduate study are presented. This gives a base to understand the results of this study presented later.

3.1 Taxonomy versus phylogeny

People seem to have a natural need to classify the life on Earth and to figure out the evolutionary lineages. Furthermore, classifying all the species is a way for us to understand the world. Carl von Linné created the binomial naming system in the 18th century where the first part of a name stands for the genus and the second is specific for each species within the genus. Taxonomists have developed the Linnean system further, and nowadays it goes hierarchically from domain into kingdom, from kingdom into phylum, from phylum into class, from class into order, from order into family, from family into genus and finally there is the unique species name. Every level is called a taxon. The taxonomy has been based on morphological characters and do not necessarily reflect the real relationships and evolutionary history. In addition to the homologues structures which have been under observation, similarity among organisms at the molecular level became a tool. (2 p. 463, 536-539)

Instead, phylogeny means the evolutionary history of species, and a phylogenetic tree is a branched diagram which illustrates the evolutionary relationships of a certain group of species. Particularly phylogeny is an important tool in evolutionary research and on examination of new species. Furthermore, it is used to understand biological processes. The concepts of phylogeny and taxonomy are different but naming of species is made to correspond to the phylogeny which has lead to reclassification. Nowadays there is a discussion if naming should be based only on the phylogeny by a system called PhyloCode. However, the phylogeny is more than only reclassification of species since the analysis gives enormous amounts of information. For example discovering new species related to cultivated plants can provide a great source of new beneficial genes for plant breeding. The phylogenetic trees have also been used to indentify species in different applications; in Japan it is used as a method to recognize real and false

whale meat and in USA bacteria species and their origins in bioterrorism cases. (2 p. 536-539)

3.2 DNA regions used in phylogeny

DNA sequences are under observation when the phylogenetic analysis is conducted. Differences between aligned nucleic acid sequences are examined; the more distant species are the more base pairs differ and different lengths appear due to the insertions and deletions occurred over history. Different methods are applied depending on which organism is studied and what is studied. Also information about morphology is connected to the molecular data to determine the best and the most reliable result of analysis. Moreover, different types of DNA sequences are used. The DNA sequence can be a specific protein coding gene region, or so called rDNA which encodes ribosomal RNA, or a gene region coding for mitochondrial DNA called mtDNA. The non-protein coding genes have been detected to be good for the phylogenetic analysis since they are stable compared to protein coding regions. (2 p. 539-540, 548) In particuar, the rRNA gene region has indicated to be a good phylogenetic marker among eukaryotes (8).

The rDNA region consists of three coding parts: nuclear-encoded small subunit SSU, 5.8S and nuclear encoded large subunit LSU. Furthermore, between the coding sections are non-coding areas called internal transcriber spacer 1 and 2 (ITS1 and ITS2) as illustrated in Figure 7. These different parts of rDNA have been examined widely in phytoplankton studies recently, and intraspecies variation has also been found among the diatoms. Studies show that the SSR and 5.8S rRNA genes are so preserved that they are not suitable when examined intra-species variation among *Skeletonema* species. Instead, the LSU and ITS regions have given good results, especially the ITS2 area. (8) Therefore, the ITS region was analyzed in this graduate study.

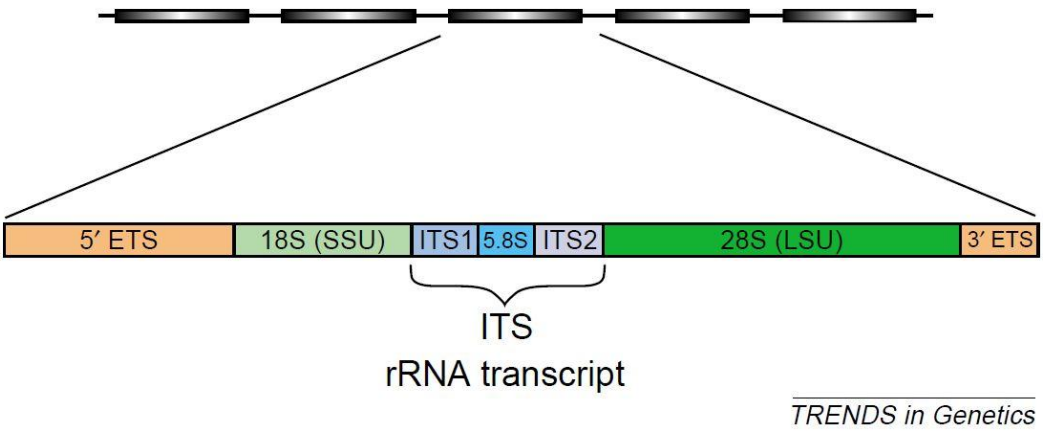

Figure 7. The rDNA section which includes the genes encoding ribosomal RNA and the non-coding Internal Transcriber Spacers (13).

3.3 Phylogenetic analysis

3.3.1 Aligning

Processing of the sequence data is actually a statistical analysis, and there are several programs which run the process based on different methods. After choosing the best suitable DNA sequences the phylogenetic analysis continues with aligning. The aligning of the chosen homologuous sequences is described to be the most critical phase on phylogenetic analysis and its quality has a high influence on the quality of the phylogenetic tree. In the multiple alignment all the sequences are collected together and compared when gaps are introduced into the sequences to move the nucleotides to their corresponding homologuous positions. (14 p. 22, 49)

In constructing of a multiple alignment mostly the ClustalW method is applied. The ClustalW constitutes three stages in which the algorithm progressively builds the multiple alignment. Firstly, the ClustalW does pairwise comparison among all the sequences creating pairwise alignments out of the pairs give the best score for introduced gaps. To control this scoring the gap opening and extension penalties can be adjusted for each purpose. Secondly, the ClustalW builds a guide tree based on the scored pairwise alignments. Finally, the guide tree is used to finish the multiple alignment. The analysis programs provide different tools to evaluate the quality and reliability of the multiple alignment before further applications. Especially it is advised to correct obvious mistakes and remove some gaps if needed. (14 p. 22, 49)

3.3.2 Construction of a phylogenetic tree

As there are several programs and methods for alignment available on the internet, there are even more programs to construct a phylogenetic tree. The most used analysis methods are called Bayesian, Maximun Likehood, Parsimony, and Neighbor-Joining (NJ). The Neighbor-Joining method is based on an algorithmic approach and the others are called tree-searching methods. All the different methods have their advantages and disadvantages and one is not proved to be better than another, even though, professionals working with them have strong opinions. Always before a phylogenetic tree is constructed, the applied method must be considered well. It is important to remember that in every case and with every method the created tree is still only a hypothesis of evolutionary history. It is impossible to achieve accurate information of the events in the past. (14 p. 57, 60)

A phylogenetic tree constructs of branches and nodes. Internal nodes represent previous taxa, the common ancestor of present taxa. In comparison, external nodes are the present species under study. Branches connect these nodes together. (14 p. 67) If there is no specific information about branch lengths on a phylogenetic tree, then they normally do not respond to the amount of occurred genetic change, or the time when a particular species evolved in each evolutionary lineage in reality. Neither can be assumed that a taxon on a phylogenetic tree has evolved from the taxon next to it. Based on the branching it can only be assumed that the branch points are common ancestors of the present taxa. (2 p. 544) Depending on the type of a tree the branch lengths can indicate time, or changes per site in the multiple alignment (14 p. 67).

The Neighbor-Joining method used in this study is a distance method which means that the distances correspond to the amount of sites differing between two sequences in a multiple alignment; i.e. the longer time has passed the more differentiation has occurred. This assumption makes sense excluding the case of multiple substitution. Based on all the included taxa in the multiple alignment the Neighbor-Joining creates step by step a series of distance matrices to construct a phylogenetic tree. The NJ method calculates different distances between the taxa and uses the net divergences to calculate corrected distance matrices. Comparing the lengths of the matrices, the taxa are connected to each other by nodes and the tree is constructed. (14 p. 65-66)

3.3.3 Evaluation of a phylogenetic tree

The tree constructed by the used program is still quite raw and needs to be modified. Moreover, the reliability needs to be estimated. The form and outline of a tree must be chosen to be suitable to serve the purpose of that specific tree and to illustrate all the important information. Basically, all the NJ trees are unrooted which means that the last common ancestor is unknown even if the node farest to the left in rectangular trees seems to be that. The tree graph can be changed to a radial cladogram, but that does not always serve the purpose. If the direction of evolution among the taxa in the tree is known, the root can be set manually. Furthermore, the tree can be rooted by adding an outgroup. The outgroup includes one or more sequences of species which are more distantly related to the ingroup sequences than the ingroup sequences are to each other, where the ingroup means the species under study. (14 p. 89, 99)

The uncertainty of phylogenetic trees cannot be emphasized too much. There are some tools to estimate the reliability of a phylogenetic tree, and a Bootstrap analysis is one of the most used. The Bootstrap method estimates reproducibility but not the accuracy of the constructed tree. The Bootstrap analysis creates so called pseudoalignments, out of the original multiple alignment used in the phylogenetic tree, where the same data is reorganised randomly. Further, from every pseudoalignment a phylogenetic tree is built in the same way as the original tree. Every pseudotree is compared with the real tree and each branch of the original tree is scored according to its presence in a pseudotree; no appearance gives zero points. Normally, in an analysis there are 500 to 1000 Bootstrap replicates used in calculation. After comparing every replicate, a percentage value from 0 to 100 is calculated to estimate the reproducibility and expressed next to every branch. The closer the Bootstrap value is to 100 the higher is the reliability. There are different quality ranges of the Bootstrap values depending on the study. (14 p. 81-82)

4 Materials and methods

This chapter describes the materials and methods applied in this graduate study during the laboratory work and data analysis. Also the analysed *S. marinoi* strains are represented.

4.1 The work plan

The work was carried out as described in the flowchart in Figure 8.

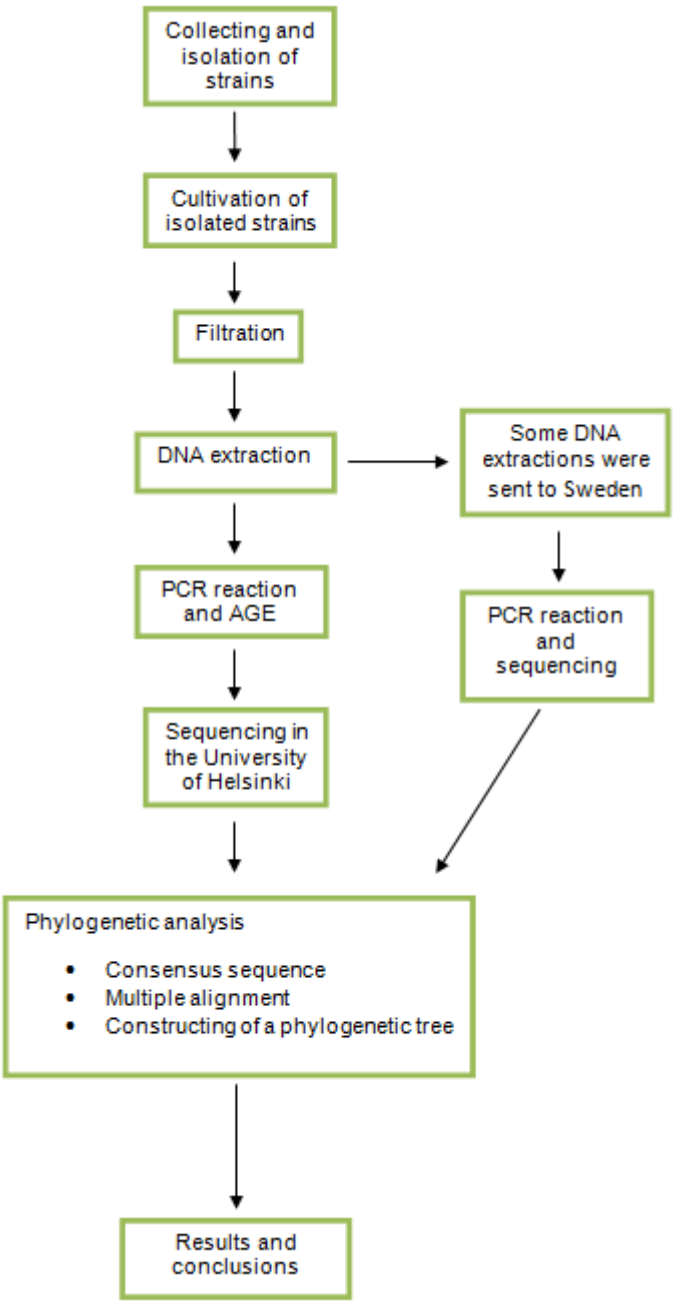

Figure 8. A schema which shows the different steps of this study.

4.2 Examined strains and cultivation procedure

S. marinoi cells for this study were collected from seven different locations in the Baltic Sea and isolated to grow in the laboratory in optimised conditions. Harvesting sites are shown in Figure 9, and they were the Bothnian Sea (C14), the Gulf of Finland (LL7), Eastern Gotland (GD), the Gulf of Gdansk (GK), Ystad (YS), Sound, Bjärred (BJ), and two from Kattegat, Arild (AR) and Båstad (BA). The isolated *S. marinoi* strains were growing in 50 ml of standard f/2 Guillard and Ryther 1962, Guillard 1975 media including $NaH_2PO_4 \cdot H_2O$, $Na_2SiO_3 \cdot 9H_2O$, trace metal solution, and vitamin solution. The contents of the solutions are described in detail in Appendix 1. The media was prepared in filtrated sea water and the salinity was optimised to correspond to the natural habitat. Incubation temperature was +4 °C, and light dark cycle 12 h : 12 h. The cultures were diluted periodically to keep the strains growing and in good condition.

Figure 9. A map showing the harvesting sites of the examined *S. marinoi* strains. No samples from station A5 were included in this in this experiment. (Courtesy of Conny Sjöqvist)

4.3 Filtration

The *S. marinoi* cultures were filtrated for DNA extraction. Suitable cell density was estimated by eye to gain enough DNA. The cell quality was controlled under a light microscope. One sample flask containing 50 ml growth media was filtered on two filters to

get duplicates, excluding some special cases of weak growth when 50 ml was filtrated onto one filter only. For the filtration Pall Corporation Versapor-3000T w/wa 3.0 µm filters were used and vacuum condition was optimized with a pump. The filters were stored in DNAase-free Eppendorf tubes at -20 °C freezer from a few days to a couple of weeks.

4.4 Extraction of DNA

DNA was extracted from the cells on the filter by following a modified CTAB protocol in Appendix 2. The reagents were prepared in the laboratory using high quality laboratory chemicals suitable for molecular applications. The CTAB (cetyl-trimethyl-ammonium-bromide) method is suitable for plant-like samples which contains great amounts of carbohydrates, polyfenols and other large metabolites able to bind nucleic acids, since the CTAB in the sample solution make a complex with DNA and the disturbing molecules can be removed (15 p. 105).

In this study 253 DNA extractions were carried out. All the extractions were analyzed with a NanoDrop Spectrophotometer (Thermo Fisher Scientific) to determine the DNA concentration and purity of each sample. DNA concentration below 25 ng/µl was not accepted. Neither samples whose 260/230 ratio were below 1.50 or 260/280 ratio were 2.50-3.00 were accepted. According the results, samples were chosen for further experiments, or new cell samples were taken from the cultures. All the extracted samples were stored at -20 °C in a freezer.

4.5 PCR reaction and AGE

Most of the DNA extractions were sent to collaborators in Sweden for further analysis outside this graduate study. For this reason ITS region PCR reactions and sequencing were applied for some of the samples of this graduate study in Sweden. Nevertheless, the PCR reaction for the ITS (ITS1, 5.8S, ITS2) region was optimized also in Helsinki as a part of this study using Bio-Rad C1000 Thermal Cycler. A touchdown PCR protocol was applied to ensure specific annealing of the used primers. In touchdown PCR the annealing temperature is set just above the melting temperature of the primers and then continuously dropped on every round. The protocol was developed specially for this study. In the protocol the annealing temperature was 65 °C in the beginning, and was lowered 0.5 °C per cycle. The whole protocol is described in Table 1.

Table 1. The applied touchdown PCR reaction.

Phase	Temperature (°C)	Time (min)	
Denaturation	94	5	
Denaturation	94	0.5	**Repeated 20 times, every cycle -0.5 °C**
Annealing	65	1	
Extension	72	1	
Denaturation	94	0.5	**Repeated 25 times**
Annealing	55	0.5	
Extension	72	0.5	
Final extension	72	5	

Table 2. The applied touchdown PCR reaction mixture for samples.

Regent	Volume (µl)	Stock concentration
dH2O	22	
Primer pair ITS5/DIR-R	1+1	10 µM
Primer pair SSU-R-F/DIR-R	1+1	10 µM
Template DNA	1	25 (ng/µl)
PCR beads		
Total	27	

The reaction mixture of the applied PCR reaction is described in Table **2** in detail. The mixture was based on PCR beads (GE Healthcare illustra PuRe-Taq Ready-To-Go PCR Beads 27-9557-01) which include all the needed enzymes and chemicals. The reaction mixture was adjusted to 27 µl including 22 µl dH_2O, 4 µl of diluted primers, and 1 µl template DNA. Two different pairs of primes were chosen to cover the area of interest as well as possible; i.e. ITS5 + DIR-R and SSU-R-F + DIR-R. The binding sites of the primers are illustrated in Figure 10. In the same figure also the sequencing primers are shown. All the primers were ordered from Oligomer Oy. The primer sequences are shown in Appendix 3. The template DNA was diluted to the concentration of 25 ng/µl. Also a control reaction without template DNA was run. Altogether nine LL7, three AR, three GK and three BJ DNA extractions were run in Helsinki.

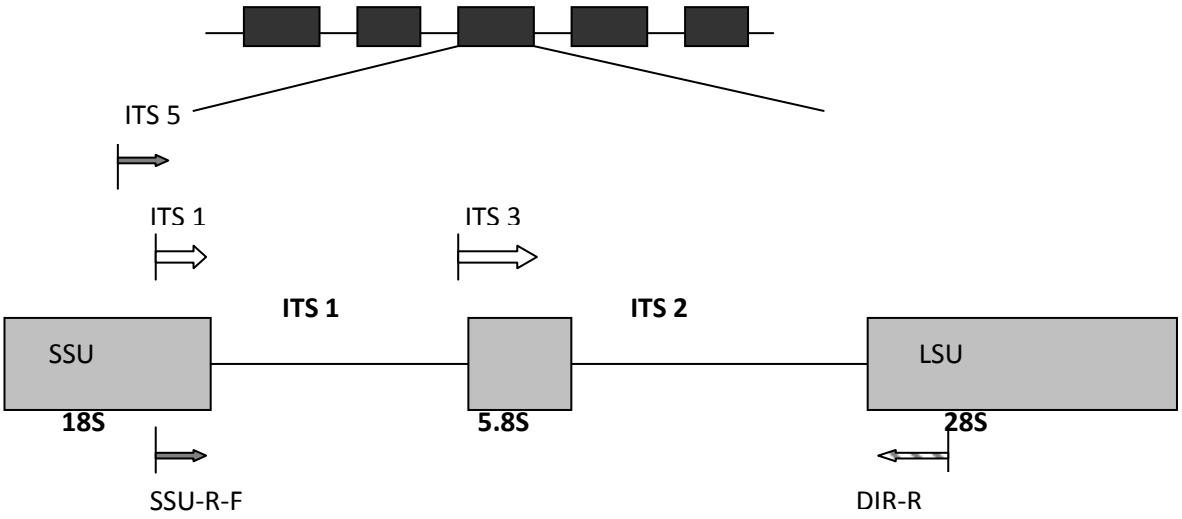

Figure 10. The binding sites of the chosen primers for PCR reaction and sequencing. ITS5 and SSU-R-F were used in PCR reaction, ITS1 and ITS3 in sequencing and DIR-R in both reactions. (Courtesy of Conny Sjöqvist)

All the PCR products were run in an agarose gel electrophoresis (AGE) to determine if the wanted rDNA region was copied well enough. The gels were 2 % agarose in 1 % TBE buffer stained with ethidium bromide. 2 µl of sample together with 4 µl of loading buffer (Fermentas) was loaded into the gel. Also two different, 1 kb and 100 bp, ladders (Fermentas) were applied into the gels. The gels were run for 45 min, 110 V, 500 mA, and photographed under UV-light.

4.6 Sequencing

To complete the data set received from Swedish collaborators samples which indicated good results on the PCR according the AGE results were chosen for sequencing. These four PCR products (BJ 42(23.7.), GK 34(23.7.), LL7 01(9.5.B) and AR 24(25.4.)) were purified using a commercial kit (GE Healthcare illustra GFX PCR DNA and Gel Band Purification Kit) following the instructions (16). Because the samples were planned to sequence, according to the manual (16) buffer type 6 was chosen for elution and for final volume 25 µl. The purified samples were measured with the NanoDrop Spectrophotometer. Due to weak results only purified products AR, GK and BJ were sent for sequencing to the Sequencing and Genomics laboratory of the University of Helsinki. Dilutions of primers ITS1, ITS3 and DIR-R were also sent to the sequencing laboratory. The binding sites of these primers are shown in Figure 10.

4.7 Data analysis of sequences

Raw sequences of samples YS02, BA06, C1402 and GD01 were received from Swedish collaborators and AR24, GK34 and BJ42 from the University of Helsinki. As shown in Figure 10 there were two primers starting from the different ends of the region of interest, and so there were also two sequences of that region. Moreover, the primer ITS3 which binds in the middle was used. In consequence, there was a third sequence of half of the region. A consensus sequence was prepared out of the three raw sequences of every sample to represent each strain. Obvious mistakes were corrected and unreliable areas were deleted by using BioEdit and Chromas programs. The programs are available for free of charge on the internet. More ITS region sequences of *S. marinoi* were collected from the GeneBank. These sequences were from different locations around the world; from Portugal, two different locations from China, from India, from Canada, and few more from different locations in the Baltic Sea. The reason to collect more sequences was to increase the perspective of the constructed trees. Furthermore, a marine diatom *Thalassiosira rotula* was chosen for rooting as an outgroup. *T. rotula* was selected based on the publication where *S. marinoi* phylogeny was studied (8). Also the ITS region sequence of *T. rotula* was acquired from the GeneBank.

The work was continued using 5.05 version of a program called MEGA. Also this program was loaded for free of charge from the internet. Guidelines provided in the book (14) were followed as needed appropriate in multiple aligning and tree-constructing. The ClustalW method on the MEGA was applied to construct several multiple alignments. Different combinations of the collected sequences were run using default settings. On both, Pairwise and the final Multiple Align, Gap Opening Penalty was 15 and Gap Extension Penalty 6.66. The multiple alignments were controlled by eye, and purposeless gap areas especially in the beginnings and in the ends, were removed. The multiple alignments were saved for further studies. Even though the sequences in each multiple alignment showed high similarity, pairwise distances and p-distances were computed on MEGA to evaluate the alignments.

The created multiple alignments were used to construct phylogenetic trees. Again the MEGA program was used, and various settings were applied following to some extend the instructions in the guide book (14). The analyses were based on the Neighbor-Joining method and Bootstrap was set to measure the reliability. The applied settings are introduced in Table 3 in detail. The trees were rooted with *T. rotula* as an outgroup

using a rooting tool. Also the topology was changed to rectangular with the scale of substitutions per site. The final trees were saved for explication.

Table 3. The applied NJ settings on MEGA.

Application	Setting
Analysis	
Analysis	Phylogenetic Reconstruction
Scope	All selected taxa
Statical method	Neighbor-joining
Phylogeny Test	
Test of Phylogeny	Bootstrap method
No. Of Bootstrap Replications	1000
Substitution model	
Substitution Type	Nucleotide
Model/Method	Maximum Composite Likehood
Substitutions to Include	d: Transitions + Transversions
Rates and Patterns	
Rates among Sites	Uniform rates
Pattern among Lineages	Same (Homogenous)
Data Subset to Use	
Gaps/Missing Data Treatment	Complete Deletion

5 Results and conclusions

This chapter evaluates both the laboratory work and phylogenetic analysis applied in this thesis project. Moreover, the final results, the phylogenetic trees, are represented and explained.

5.1 Cultivation of *S. marinoi* and filtration of the samples

The isolated *S. marinoi* strains were maintained successfully in the laboratory. Most of the samples grew well enough to be filtered for further studies during the experiment, and survived vital even longer. Even though, some strains were growing quite slowly and caused re-planning.

The filtration was a straightforward step to gain the sample cells for DNA extraction. It was critical to work accurately to avoid mixing or contaminating of the samples. In addition, it was important to have stable and low enough pressure that the *S. marinoi* cells stayed intact and the DNA was not lost because of cell rupture.

5.2 Extraction of the DNA

In total DNA was extracted from 253 filters. The DNA extraction protocol was relatively simple (Appendix 1) but there were several problems with the results. All the extractions were measured on the NanoDrop Spectrophotometer to estimate the DNA concentration, and purity by 260/280 and 260/230 ratios. For some unknown reason samples of the LL7 strain failed repeatedly despite the used reagents or results of the other samples at the same set.

There are few possible explanations for the failed extraction results. Firstly, DNA is very sensitive for different contaminations and breaks down easily. Especially, an enzyme contamination from the laboratory staff is a high risk and a likely explanation in this case also. Secondly, the quality control of the reagents used in the experiment was not accurate enough. The storage times of the mixed reagents may not have been controlled well enough. Moreover, during the experiment the used reagents were prepared by different persons and somehow the solutions also looked different. These uncertainties could explain the problems in extraction results. Nevertheless, there were so many strains and samples under cultivation in the laboratory and several extractions were applied that also enough DNA was gained to continue the experiments of this graduate study.

5.3 PCR reaction and sequencing

Sequences representing strains YS, BA, C14, and GD were received from Swedish collaborators. To prepare a complete data set for phylogenetic analysis more sequences were needed. The PCR reaction to copy the ITS region was applied for a few AR, GK, BJ and LL7 DNA extractions. Because the PCR protocol was new, it was first tried only with six samples and a blank control. Figure 11 shows a photo of the AGE gel which was run for the first PCR products. In the Figure 11 it can be observed that there is plenty of DNA in wells 6-8 which size is around 750 bp i.e. the size of the wanted ITS region. In the well 9 is a blank control which contained no DNA, and it is empty as ex-

pected. In the wells 3-5 are those few LL7 samples which gave satisfactory results on NanoDrop. The gel clearly indicates how there is hardly any DNA and the PCR reaction has failed. Nevertheless, the bands in the lines 6-9 were so promising that the protocol was accepted, especially because of the previous problems with LL7 strain. Some LL7 samples were anyway later taken to the PCR reactions and gave promising results. From all of the PCR products a DNA sample was taken to run it in an AGE gels to evaluate the PCR result. All the gels are not represented here, but the results appeared sufficiently.

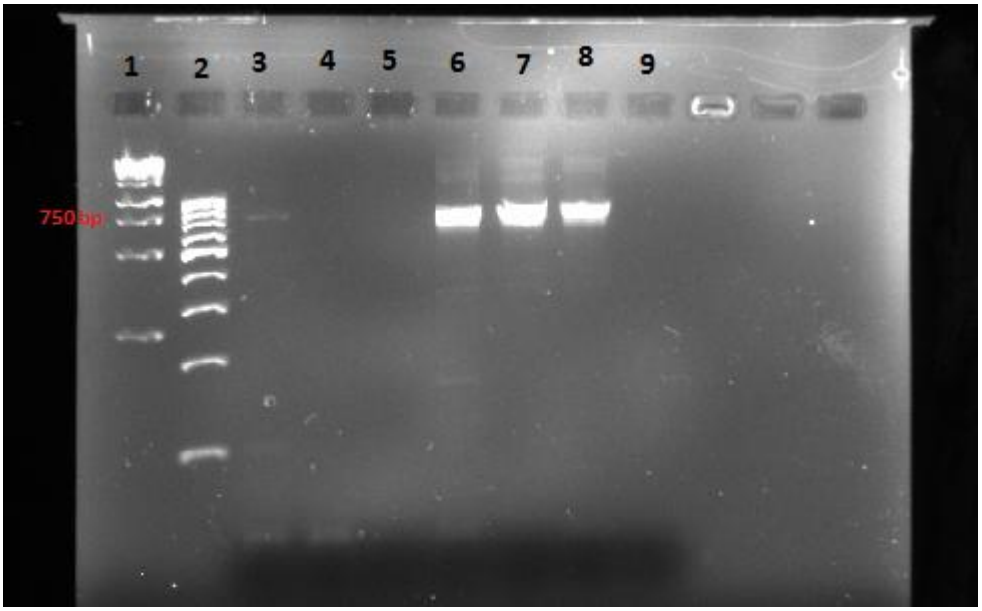

Figure 11. A picture of the AGE of the first PCR products. Wells include: 1. 1kb ladder 2. 100 bp ladder 3. LL719 4. LL722 5. LL723 6. AR17 7. AR19 8. AR24 9. Blank.

One DNA sample of each strain (AR24, LL701, GK34, BJ42) was purified for sequencing. These purified samples gave very low concentrations when measured on the NanoDrop Spectrophotometer. Excluding the LL7 sample, the others were sent for sequencing. Although the quantity and quality of the DNA appeared to be weak the sequences appeared good. Possible reasons for low DNA concentration may be a contamination which had broken down the DNA. On the other hand, the quality of PCR products may have been so weak that the loss in purification became significant.

5.4 Data analysis

As a result of the own laboratory work sequences of eight different strains of the Baltic *S. marinoi* were gained: C14, GD, GK, YS, AR, BJ, and BA. Only the LL7 strain was

lost due to weak results. The rest of the sequences were searched from the GenBank. The phylogenetic analysis was carried out with these sequences. The first tree of the two is more accurate, and presents all the own Baltic strains and six Canadian strains of *S. marinoi*, one *T. rotula* strain as an outgroup. In addition to the strains in the first tree, in the second tree seven Baltic strains are present, five Portuguese strains, nine Chinese strains, and one Indian strain. The second tree is quite raw and cannot be explained comprehensively.

The phylogenetic analysis was probably the most challenging phase of this thesis project. There are numerous different ways to analyze the sequences for trees on different programs with adjusted settings. There was no complete protocol how a phylogenetic tree is constructed, so it was done as a sum of advice other were able to give, methods applied in the publications, guide books, and the instructions the programs include. Various alignments and trees were tried. Finally, two different phylogenetic trees appeared to be good enough to show as a result of this study.

When the data analysis was started the knowledge was almost at zero level, thus, the achieved trees were a success. Of course, the setting which may not have been optimal may have had impact on the obtained trees. To extend the understanding of phylogeny, and the possible relationships, it would have been reasonable to apply also other tree constructing methods. Within the framework of this study there was no time to investigate the other options.

5.5 Final results

Figure 12 represents the constructed phylogenetic tree illustrating the evolutionary relationships within the Baltic *S. marinoi* strains and between *S. marinoi* strains isolated in Canada based on the Neighbor-Joining method. Firstly, it can be seen that the *T. rotula* as an outgroup has evolved to its own lineage. Secondly, the Canadian strains and Baltic strains are separated into two clades. Due to this observation one may hypothesize that the different circumstances of these habitats have created a situation where *S. marinoi* populations can be distinguished from each other on the molecular level.

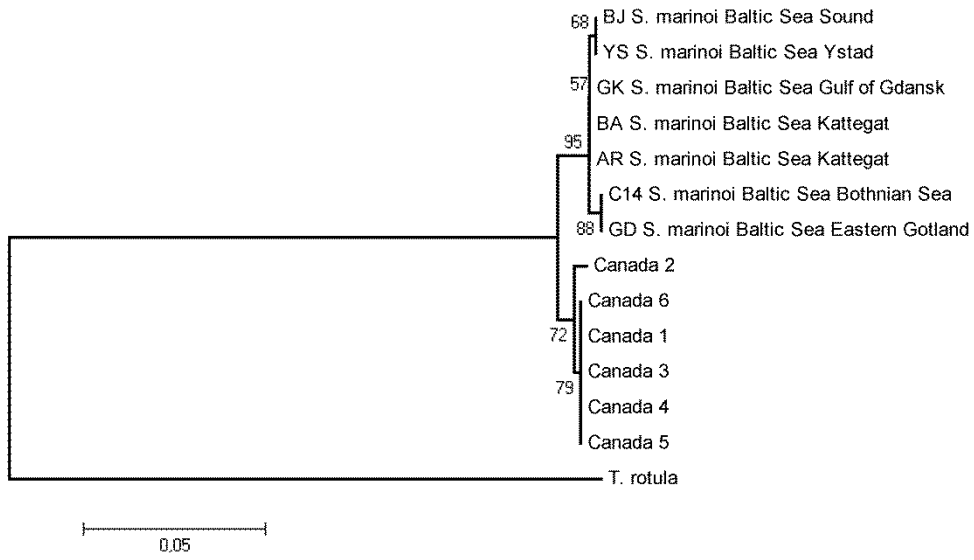

Figure 12. A phylogenetic tree representing the evolutionary relationships between Baltic *S. marinoi* strains and Canadian *S. marinoi* strains as Neighbor-Joining tree, *T. rotula* as an out-group.

Already on the multiple alignment it was observed that the Baltic sequences highly resemble each other, and now also in the phylogenetic tree it is shown that the strains are closely related. However, it is possible to detect that some separation within the Baltic *S. marinoi* strains have occurred over the time supporting the hypothesis set for this study. The strain C14 is isolated in Bothnian Sea and the GD near Gotland, as can be seen in Figure 9, so they are the most northern samples among the Baltic Sea samples in this tree. These two strains have clearly settled in their own clade separating from the strains of southern parts of the Baltic Sea, i.e. these two strains are from areas where the salinity is lower when compared to others. Consequently, this may indicate that there has occurred adaptation to different conditions of the Baltic Sea habitat within the *S. marinoi.*

The bootstrap values define the reliability of the tree as explained before. The limits of the values are evaluated separately in every case. For this graduate study the limits were not set that high to analyze the trees. In the tree in Figure 12 most of the bootstrap values are acceptable. The highest values are in the steps were C14 and GD strains evolve from the others. Hence, the bootstrap analysis supports the hypothesis

of the differentiation of the C14 and GD strains. The arrangement of rest of the *S. marinoi* strains are not strongly supported by the bootstrap values in this tree.

Figure 13 shows the second phylogenetic tree of *S. marinoi* strains which are from various locations around the world. This tree is not as reliable as the first one since the bootstrap values are remarkably low. This tree is shown as a raw hypothesis of the relationships between a wide amount of *S. marinoi* sequences. All the lineages in the tree cannot be understood and explained. The European strains cannot be distinguished from the Chinese ones and neither can the Baltic strains be distinguished from the Portuguese ones among the European strains. At the moment it is impossible to know how accurate the tree is. However, phytoplankton cells may be transported around the world by ballast waters. On this way new genes are introduced into populations, and the appearance can be seen in a phylogenetic tree. Another possible explanation for the arrangement of the strains in the second tree may be the secondary structure of ITS rRNA which is not discussed in this graduate study. The secondary structure has been detected to be an important factor in the phylogeny, and may be taken into account more widely in the future. On the other hand, in the publication of Godhe et al. (8), is a Maximum Likehood phylogenetic tree which is built out of the LSU rDNA sequences of *S. marinoi*, where the different European strains cannot be distinguished from eachother either.

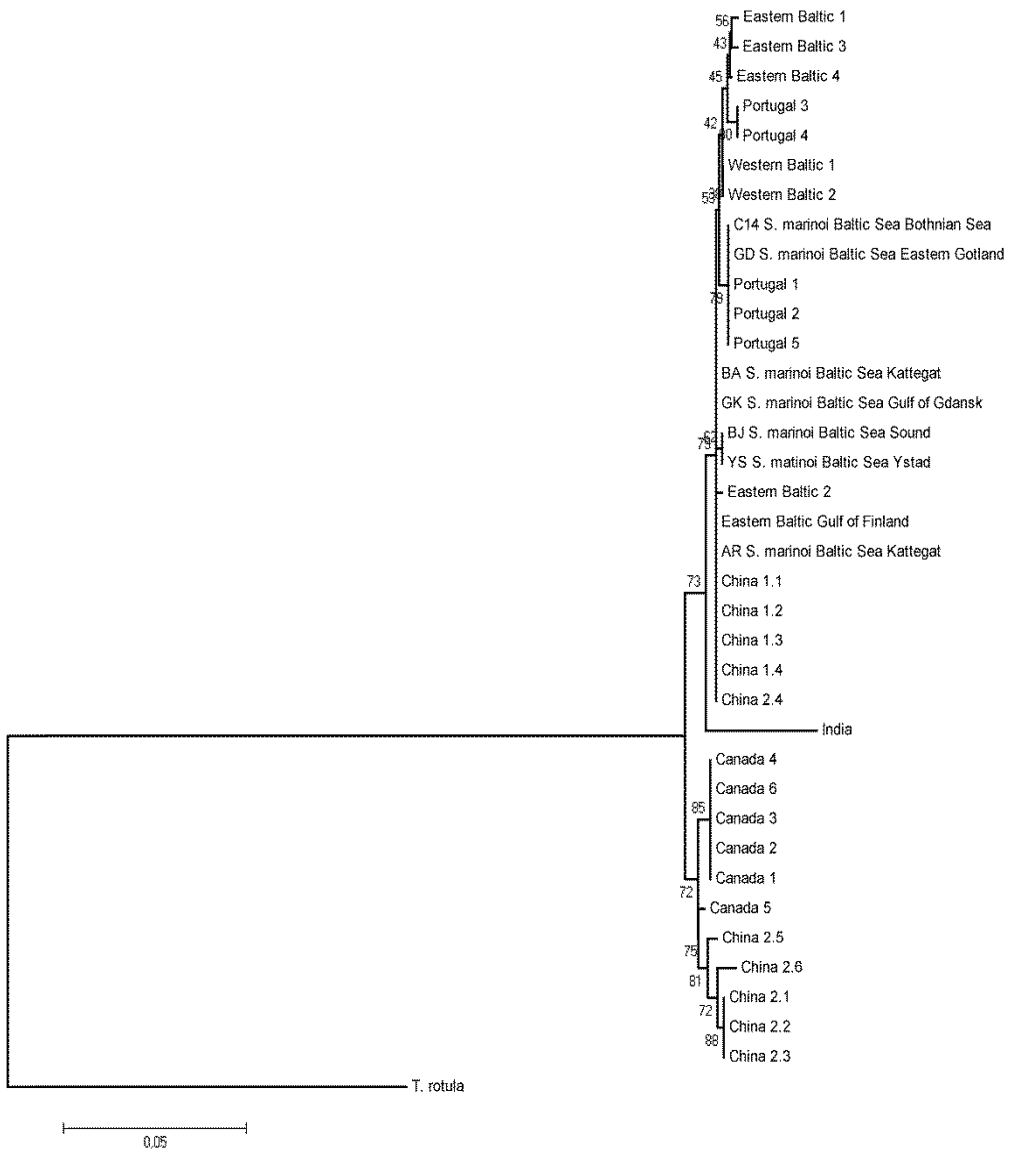

Figure 13. A phylogenetic tree describing the evolutionary relationships between several *S. marinoi* strains around the world.

6 Discussion

The purpose of this graduate study was to investigate the diversity of *S. marinoi* strains which were isolated from the Baltic Sea. The examination was planned to apply on molecular level analysing the rDNA sequences by constructing phylogenetic trees. The phytoplankton samples were maintained in the laboratory where the DNA was also extracted.

The study was relatively straightforward but notably time-consuming. Moreover, working with living organisms could not be strictly scheduled. Some problems were faced during the laboratory work such as the variable DNA extraction results and low concentrated PCR products. In the end, only one strain of the planned was discarded (LL7). The rest of the samples continued to the phylogenetic analysis which appeared to be the most challenging phase of this study. The MEGA program was used to create multiple alignments based on the ClustalW method. Other ITS sequences were also added to the alignments of *S. marinoi* from the GeneBank and the ITS sequence of *T. rotula* as an outgroup. The MEGA was also used to build the phylogenetic trees out of the multiple alignments by the Neighbor-Joining method. Moreover, the Bootstrap analysis was used to determine the reliability of the trees. Finally, two phylogenetic trees were achieved.

Especially the first tree appeared relatively reliable according the Bootstrap values. Based on the branching of this tree, it was possible to state a suggestion that among *S. marinoi* species in the Baltic Sea there has occurred differentiation. The strains which were collected from northern locations evolved into their own clade while the strains from southern parts are elsewhere. Also, the tree suggests that the Baltic strains are possible to distinguish from the Canadian strains. The results support the hypothesis of the diversity among species in the Baltic Sea habitats.

The second built phylogenetic tree does not appear as reliable as the first one and this gives contradictory information. But this tree reminds about the important fact that the phylogenetic trees are always only estimates, and that it is impossible to know the exact events occurred in the past. Moreover, this tree indicates that the phylogeny of *S. marinoi* still needs fundamental investigation. Within this graduate study it was possible to take a look at the diversity of *S. marinoi* on the molecular level and explore the phylogenetic analysis. The set goals were achieved, so the study can be considered succesful. Moreover, the studies with *S. marinoi* continue in the Marine Research Centre of the Finnish Environmental Institute.

References

1 Armbrust E. V. The life of diatoms in the world's oceans. Nature. 2009; 459: 185-192.

2 Campbell N.A., Reece J.B., Urry L.A., Cain M.L., Wasserman S.A., Minorsky P.V., Jackson R.B. Biology. 8th ed. San Francisco: Pearson Benjamin Cummings; 2008.

3 Saade A., Bowler C. Molecular Tools for Discovering the Secrets of Diatoms. BioScience. 2009; 59(9): 757-765.

4 Spaulding, S.A., Lubinski, D.J. and Potapova, M. What are diatoms? [Internet]. Diatoms of the United States. 2010. [Cited 2013 Jan 6]. Available from http://westerndiatoms.colorado.edu/about/what_are_diatoms.

5 Madigan M.T., Martinko J.M., Stahl D.A., Clark D.P. Brock Biology of Microorganisms. 13th ed. San Francisco: Pearson Benjamin Cummings; 2012.

6 Phylogeny and molecular clock [Internet]. Alfred Wegener Institute for Polar and Marine Research (AWI). [Cited 2013 Jan 6]. Available from http://www.awi.de/en/research/research_divisions/biosciences/polar_biological_oceanography/research_themes/molecular_genetics/phylogeny_and_molecular_clocks/.

7 Guiry, M.D. & Guiry, G.M. AlgaeBase World-wide electronic publication: *Skeletonema marinoi* Sarno & Zingone [Internet]. National University of Ireland: Galway. 2013. [Cited 2013 Jan 6]. Available from http://www.algaebase.org/search/species/detail/?species_id=69180.

8 Godhe A., McQuoid M. R., Karunasagar Indrani, Karunasagar Iddya, Rehnstam-Holm A-S. Comparison of three common molecular tools for distinguishing among geographically separated clones of the diatom *Skeletonema Marinoi* Sarno et Zingone (Bacillariophyceae). Phycological Society of America. 2006; 42: 280-291.

9 Godhe A. Population genetic structure of *Skeletonema marinoi* - a model species for phytoplankton bloom dynamics [Internet]. University of Gothenburg: Sweden. [Cited 2013 Jan 6]. Available from http://www.bioenv.gu.se/english/staff/AnnaGodhe,eng/#Population.

10 Saravanan V., Godhe A. Genetic heterogeneity and physiological variation among seasonally separated clones of Skeletonema marinoi (Bacillariophyceae) in the Gullmar Fjord, Sweden. European Journal of Phycology. 2010; 45(2): 177-190.

11 The Baltic Sea Portal [Internet]. Ministry of Environment, SYKE, Finnish Meteorological Institute. [Cited 2013 Jan 15]. Available from http://www.itameriportaali.fi/en/tietoa/yleiskuvaus/en_GB/yleiskuvaus/.

12 The Baltic Sea Portal [Internet]. Ministry of Environment, SYKE, Finnish Meteorological Institute. [Cited 2013 Jan 15]. Available from http://www.itameriportaali.fi/en/tietoa/veden_liikkeet/en_GB/hydrografia/.

13 Coleman A.,W., ITS2 is a double-edged tool for eukaryote evolutionary comparisons. TRENDS in Genetics. 2003; 19(7): 370-375.

14 Hall B.G. Phylogenetic Trees Made Easy: A How-To Manual. 3rd ed. Sunderland, MA: Sinauer Associates; 2008.

15 Suominen I., Pärssinen R., Haajanen K., Pelkonen J. Geenitekniikka. Turku: Turun ammattikorkeakoulu; 2010.

16 illustra GFX PCR DNA and Gel Band Purification Kit Product booklet. GE Healthcare UK Limited; 10/2008.

17 f/2 Medium [Internet]. NCMA Bigelow Laboratory for Ocean Sciences. [Cited 2013 Jan 23]. Available from https://ncma.bigelow.org/node/79.

f/2 Medium (17)

f/2 Medium (Guillard and Ryther 1962, Guillard 1975)

Component	Molar Concentration in Final Medium
$NaNO_3$	8.82×10^{-4} M
NaH_2PO_4 H_2O	3.62×10^{-5} M
Na_2SiO_3 $9H_2O$	1.06×10^{-4} M
trace metal solution	---
vitamin solution	---

f/2 Trace Metal Solution

Component	Molar Concentration in Final Medium
$FeCl_3$ $6H_2O$	1.17×10^{-5} M
Na_2EDTA $2H_2O$	1.17×10^{-5} M
$CuSO_4$ $5H_2O$	3.93×10^{-8} M
Na_2MoO_4 $2H_2O$	2.60×10^{-8} M
ZnSO4 $7H_2O$	7.65×10^{-8} M
$CoCl_2$ $6H_2O$	4.20×10^{-8} M
$MnCl_2$ $4H_2O$	9.10×10^{-7} M

f/2 Vitamin Solution

Component	Molar Concentration in Final Medium
thiamine HCl (vit. B_1)	2.96×10^{-7} M
biotin (vit. H)	2.05×10^{-9} M
cyanocobalamin (vit. B_{12})	3.69×10^{-10} M

Guillard, R.R.L. 1975. Culture of phytoplankton for feeding marine invertebrates. pp 26-60. *In* Smith W.L. and Chanley M.H (Eds.) *Culture of Marine Invertebrate Animals*. Plenum Press, New York, USA.

Guillard, R.R.L. and Ryther, J.H. 1962. Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt and *Detonula confervacea* Cleve. *Can. J. Microbiol.* **8:** 229-239.

Short CTAB DNA Extraction, Wiebe Kooistra 2007

DNA Extraction

- Add 500 µl of CTAB extraction buffer* and a drop of β-mercaptoethanol in the Eppendorf tube that contains a sample filter, vortex.
- Put in water bath at 65 °C for 1 hour, vortex every 15 min.
- Remove filters using clean forceps.
- Put on ice until cold.
- Add 500 µl of SEVAG** and shake gently.
- Centrifuge at 14 000 rpm for 10 min.
- 3 layers form: top layer of water and DNA, white fluff at interphase, bottom layer of chloroform-isoamyl-alcohol.
- Pipette upper layer into a clean Eppendorf tube avoiding the fluffy white interphase.
- Repeat SEVAG stage until interphase is without any white residue – take care not to transfer any chloroform. Transfer upper (watery) layer in a clean tube.
- Add equal (200-300 µl) isopropanol (stored at -20 °C) and mix gently.
- Store tubes at -20 °C one hour.
- Remove samples from -20 °C and centrifuge at 14 000 rpm for 25 min.
- Discard the supernatant by pouring gently without disturbing the pellet.
- Wash with 400 µl 75 % ethanol.
- Centrifuge again for 15 minutes at 14 000 rpm.
- Remove alcohol by gently turning the tubes upside down without disturbing the pellet.
- Set the tubes upside down on a paper in a fume cupboard to dry completely (until you can no longer see or smell alcohol).
- Add 50 µl dH_2O.
- To dissolve the pellet, place on ice for a few hours.
- Can be stored in the fridge for a day, in the -20 °C for months and in the -80 °C indefinitely.

Reagents

*CTAB (2 %) – Extraction Buffer (50 ml)

- 1 g CTAB
- 10 ml 1 M Tris-HCl pH 8.0
- 5 ml 0.5 M EDTA
- 4.09 g NaCl
- Make up 50 ml with dH_2O
- After autoclaving add a dash of PVP (Poly Vinyl Polypyrrolidone)

**SEVAG

- Mix 96 ml Chloroform and 4 ml isoamyl alcohol

Analysis certificates of Primers

Oligomer Oy
Latokartanonkaari 11
00790 Helsinki
Finland

Analyysitodistus

Oligomer

Asiakas:
Sjöqvist Conny, Tutkija, Suomen ympäristökeskus

Asiakasnumero	463-3
Tilausnumero:	Viite/vastuualue:
Tilaustunnus:	20507-001
Tilauspäivä:	7.5.2012
Kommentit:	

Tilauslaskuri: 0001

Oligon nimi: DIR R **Oligon numero:** 20509F4A09 3/4

5'-TAT GCT TAA ATT CAG CGG GT-3'

Toimitettu:	10.1 OD	**Emästen lkm.:**	20	**Synteesiskaala:**	0,04 µmol
	294.2 µg	**GC %:**	40.0	**Puhdistus:**	HPLC
	45.5 nmol			**Olomuoto:**	Kuivattu/Solid
Konsentraatio	100.0 µM			**Laadunvarmistus:**	DMT-monitoring
MP:	6147	**A**	**C**	**G**	**T**	**Liuotetaan (µl):**	455
Tm:	53.2	5	3	5	7		

DNA

Sivu 1 Päivämäärä 11.5.2012

Oligomer Oy
Latokartanonkaari 11
00790 Helsinki
Finland

Analyysitodistus

Oligon nimi: ITS3 **Oligon numero:** 20507M1H02 4/4

5'-GCA TCG ATG AAG AAC GCA GC-3'

Toimitettu:	3.3	OD	**Emästen lkm.:**		20	**Synteesiskaala:**	0,04 µmol
	92.7	µg	**GC %:**		55.0	**Puhdistus:**	HPLC
	14.3	nmol				**Olomuoto:**	Kuivattu/Solid
Konsentraatio	100.0	µM	**A**	**C**	**G**	**T**	**Laadunvarmistus:**	DMT-monitoring
MP:	6160		7	5	6	2	**Liuotetaan (µl):**	143
Tm:	59.4							

DNA

Oligomer Oy
Latokartanonkaari 11
00790 Helsinki
Finland

Analyysitodistus

Oligomer

Asiakas:
Sjöqvist Conny, Tutkija, Suomen ympäristökeskus

Asiakasnumero	463-3
Tilausnumero:	Viite/vastuualue:
Tilaustunnus:	20507-001
Tilauspäivä:	7.5.2012
Kommentit:	

Tilauslaskuri: 0001

Oligon nimi: ITS5 **Oligon numero:** 20507M1B04 1/4

5'-GGA AGT AAA AGT CGT AAC AAG G-3'

Toimitettu:	3.2 OD	**Emästen lkm.:**	22	**Synteesiskaala:**	0,04 µmol
	84.3 µg	**GC %:**	40.9	**Puhdistus:**	HPLC
	11.7 nmol			**Olomuoto:**	Kuivattu/Solid
Konsentraatio	100.0 µM			**Laadunvarmistus:**	DMT-monitoring
MP:	6865	**A C G T**		**Liuotetaan (µl):**	117
Tm:	56.5	10 2 7 3			

DNA

Oligon nimi: SSU R F **Oligon numero:** 20507M1C01 2/4

5'-GTA GGT GAA CCT GCA GAA GGA TCA-3'

Toimitettu:	3.5 OD	**Emästen lkm.:**	24	**Synteesiskaala:**	0,04 µmol
	97.9 µg	**GC %:**	50.0	**Puhdistus:**	HPLC
	12.5 nmol			**Olomuoto:**	Kuivattu/Solid
Konsentraatio	100.0 µM			**Laadunvarmistus:**	DMT-monitoring
MP:	7451	**A C G T**		**Liuotetaan (µl):**	125
Tm:	62.7	8 4 8 4			

DNA