MASTER THESIS

NGS-based genetic analysis of diatom diversity obtained from recent and ancient sediments of Siberian treeline lakes

Submitted to the

Faculty of Mathematics and Natural Sciences

of the

University of Potsdam

by

Katharina Dulias

Supervisor:

Prof. Dr. Ulrike Herzschuh (Alfred Wegener Institut, Universität Potsdam)

Second Referee:

Prof. Dr. Ralph Tiedemann (Universität Potsdam)

August 2015

ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG

Für meine Eltern

Contents

List of Abbreviations VI
List of Appendix VII
List of Figures VIII
List of measurement units IX
List of Tables IX
Abstract XI
Zusammenfassung XII
1 Introduction 1
1.1 The Arctic treeline ecotone 1
1.2 Diatoms as bioindicators 2
1.3 Environmental DNA 4
2 Objectives 7
3 Material and Methods 8
3.1 Study area 8
3.2 Sampling 11
3.2.1 Sampling procedure in the field 11
3.2.2 Subsampling of sediment cores for ancient DNA analyses 12
3.3 Dating of sediment cores 12
3.3.1 ^{210}Pb/^{137}Cs dating 12
3.3.2 Radiocarbon (^{14}C) dating 13
3.3.3 Age-Depth-Models 13
3.3.3.1 Core 11-CH-12A 13
3.3.3.2 Core 11-CH-06D 14
3.4 Genetic assessment of sediment samples 14
3.4.1 DNA Extraction 14
3.4.2 Polymerase Chain Reaction 16
3.4.3 Pre-Check: Cloning and Sanger sequencing for selected samples 17
3.4.4 Purification of PCR products 17
3.4.5 Measurement of DNA Quantity 18
3.4.6 Parallel high-throughput sequencing 18
3.4.7 Bioinformatic analysis of sequencing data 18
3.5 Morphological diatom analyses 19
3.6 Statistical Analysis 20
4 Results 23

Contents

4.1 Preliminary assessment using cloning and Sanger sequencing 23
4.2 Genetic and morphological assessment of modern lake sediments 24
4.2.1 Genetic assessment of modern lake sediments 24
4.2.1.1 Diversity assessment on assigned species and genus level of sequences 24
4.2.1.2 Data selection of environmental parameters for species sequence data 27
4.2.1.3 RDAs of sequences on species and genus level 28
4.2.2 Morphological assessment of surface data 29
4.2.2.1 Diversity assessment on species and genus level of morphological data 29
4.2.2.2 Data selection of environmental parameters of the morphological data set 33
4.2.2.3 RDAs of the morphological identified species and genera 33
4.3. Comparison of the genetic and the morphological assessment 35
4.3.1 Comparison of the retrieved diversities 35
4.3.2 Comparison of the correlating parameters and their influences on the data sets 36
4.3.3 Comparison of both data sets using *procrustes* 36
4.4 Core data 37
4.4.1 Genetic assessment of the core data 37
4.4.1.1 Diversity assessment of the tundra core 11-CH-12A 37
4.4.1.2 Diversity assessment of the light taiga core 11-CH-06D 40
4.4.2 Morphological assessment of sediment cores 42
4.4.2.1 Diversity assessment of the tundra core 11-CH-12A 42
4.4.2.2 Diversity assessment of the light taiga core 11-CH-06D 45
4.5 Comparison of the diversity assessment of sediment cores 48
4.5.1 Comparison between genetic and morphological data of core 11-CH-12A 48
4.5.2 Comparison between genetic and morphological data of core 11-CH-06D 49
5 Discussion 51
5.1 Comparison of the diatom richness and composition of the genetic and the morphological data of the surface sediments data set 51
5.2 Genetic and morphological relation with environmental parameters of the surface data set 54
5.3 Diatom composition of the temporal data sets 56
6 Conclusion 60
7 References XIII
Appendix XXI
A.1 Material XXI
A.1.1 Chemicals and buffers XXI

A.1.2 Kits and other materials XXI
A.1.3 Laboratory equipment XXII
A.2 Preliminary statistics XXIII
A.2.1 Detrended Correspondence Analysis XXIII
A.2.2 Principal Component Analysis XXIII
A.2.3 Non-metric Multidimensional Scaling XXV
A.3 Stratigraphic plots of the genera XXVI
A.4 Rarefaction curves of the genera of the surface data XXXIII
A.5 Environmental parameters XXXIII
A.6 Redundancy analysis of the genera data sets XXXV
A.7 Diatom counts of core 11-CH-06D XXXV
A.8 Rarefaction curves of the genera of the core data XXXVI
A.9 Primer and tag-combinations XXXVI
A.10 Sequencing data sets XXXVII
A.11 Morphological data sets XLV
Acknowledgements XLVII
Statutory Declaration XLVIII

List of Abbreviations

°E	Geographic longitude
°N	Geographic latitude
^{137}Cs	Cesium isotope
^{14}C	Radiocarbon
18S rRNA	ribosomal ribonucleic acid, component of the small eukaryotic ribosomal subunit
^{210}Pb	Lead isotope
bp	base pair
B.P.	Before present
BSA	Bovine Serum Albumin
cox1	Cytochrome c oxidase I, synonym COI
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxynucleotide triphosphate
DOC	Dead organic content
dsDNA	double stranded deoxyribonucleic acid
GFZ	Geoforschungszentrum
GPS	Global positioning system
HiFi	High Fidelity
ID	Identifier
ITS region	Internal transcribed spacer, situated between the small-subunit rRNA and large-subunit rRNA
NGS	Next Generation Sequencing
NTC	Negative template control
PCR	Polymerase chain reaction
Prot.K	Proteinase K
qPCR	quantitative polymerase chain reaction
rbcL	*large-chain* gene of the Ribulose-1,5-bisphosphate carboxylase/oxygenase
rDNA	ribosomal deoxyribonucleic acid
rRNA	ribosomal ribonucleic acid
SOC	Super Optimal broth with Catabolite repression
TAE	Tris-acetate-EDTA
Taq	*Thermus aquaticus*
UV	Ultraviolet

List of Appendix

Appendix 1: Length of the first axis for each data set of the DCA. XXIII
Appendix 2: PCA of the sampled lakes and the environmental parameters, as well as with the vegetation types. XXIII
Appendix 3: PCAs of the species retrieved from the surface samples of both approaches. XXIII
Appendix 4: PCAs of the genera retrieved from the surface samples of both methods. XXIV
Appendix 5: PCAs of the species data of core 11-CH-12A on both methods. XXIV
Appendix 6: PCAs of the genera data of core 11-CH-12A on both methods. XXIV
Appendix 7: PCAs of the species data of core 11-CH-06D on both methods. XXV
Appendix 8: PCAs of the genera data of core 11-CH-06D on both methods. XXV
Appendix 9: NMDS analysis of the species data of both methods. XXV
Appendix 10: NMDS analysis of the genera data of both methods. XXVI
Appendix 11: Diatom sequence types identified to genus level of the lake transect from north to south, in the four vegetation zones tundra, single tree tundra, forest tundra and light taiga. XXVII
Appendix 12: Diatoms identified to genus level by light microscopy of the lake transect from north to south, in the four vegetation zones tundra, single tree tundra, forest tundra and light taiga. XXVIII
Appendix 13: Diatom sequence types identified to genus level of core 11-CH-12A with assemblage groups identified by CONISS analysis and the estimated richness. XXIX
Appendix 14: Diatoms identified to genus level by light microscopy of core 11-CH-12A with assemblage groups identified by CONISS analysis and estimated richness. XXX
Appendix 15: Diatom sequence types identified to genus level of core 11-CH-06D, assemblage groups identified by CONISS analysis and estimated richness. XXXI
Appendix 16: Diatoms identified to genus level by light microscopy of core 11-CH-06D, assemblage groups identified by CONISS analysis and estimated richness. XXXII
Appendix 17: Rarefaction curves of the genera retrieved from the surface samples of both methods. XXXIII
Appendix 18: All measured physico-chemical parameters of the lakes used for analysis of modern sediments. XXXIII
Appendix 19: Proportions of the physico-chemical lake characteristics for the genera of the genetic data. XXXIV
Appendix 20: Proportion of the physico-chemical lake characteristics of the genera of the morphological data. XXXIV
Appendix 21: RDAs of the environmental parameters and the vegetation types for the genera data of both methods. A - RDA environmental parameter with genetic data, B - RDA of environmental parameters with morphological data, C - RDA of vegetation types with genetic data, D - RDA of vegetation types with morphological data. XXXV
Appendix 22: Counted diatom valves in three morphological categories. XXXV
Appendix 23: Rarefaction curves of the genera of both methods for both sediment cores. A - Rarefaction curve of the genetic data of core 11-CH-12A, B - Rarefaction curve of the morphological data of core 11-CH-12A, C - Rarefaction curve of the genetic data of core 11-CH-06D, D - Rarefaction curve of the morphological data of core 11-CH-06D. XXXVI

Appendix 24: Table of all used primer-tag combinations with color code for modern and old samples. XXXVI
Appendix 25: Table of all retrieved sequences with best identity, best match, count and assigned scientific name. XXXVII
Appendix 26: Sequencetypes assigned to species level of the lake transect. XLIII
Appendix 27: Genera of the sequence data of the surface transect. XLIV
Appendix 28: Species of core 11-CH-12A of the genetic data. XLIV
Appendix 29: Genera of core 11-CH-12A of the genetic data. XLIV
Appendix 30: Species of core 11-CH-06D of the genetic data. XLIV
Appendix 31: Genera of core 11-CH06D of the genetic data. XLV
Appendix 32: Morphological identified species of the lake transect. XLV
Appendix 33: Morphological identified genera of the lake transect. XLV
Appendix 34: Morphological identified species of core 11-CH-12A. XLVI
Appendix 35: Morphological identified genera of core 11-CH-12A. XLVI
Appendix 36: Morphological identified species of core 11-CH06D. XLVI
Appendix 37: Morphological identified genera of core 11-CH-06D. XLVI

List of Figures

Figure 1: Common diatom species of Arctic treeline lakes. A - *Staurosira construens*, B - *Staurosira pinnata*, C - *Cyclotella* sp., D - *Pinnularia microstauron*, E - *Aulacoseira distans*, F - *Navicula* sp. (Diatom identification was done together with Luidmila Pestryakova and pictures were taken with the SEM at GFZ Potsdam in cooperation with Ilona Schäpan). 4
Figure 2: Examples for the vegetation types around the lakes of the modern data set. A - tundra, B - single tree tundra, C - light taiga, D - forest tundra. (Pictures were taken by Stefan Kruse during the expedition 2013) 9
Figure 3: Maps of the study transect. A - camp I in the tundra, B - camp II in the single tree tundra, C - camp III in the light taiga, and D - camp IV in the forest tundra. The black stripped line indicates the current position of the treeline. 10
Figure 4: Calibrated age-depth-model of core 11-CH-12A. 13
Figure 5: Age-depth-model of core 11-CH-06D with extrapolation based on Pb/Cs results of core 11-CH-06E. 14
Figure 6: The diatom sequence types identified to species level of the lake transect from north to south, in the four vegetation zones tundra, single tree tundra, forest tundra and light taiga. The color code highlights the four vegetation types. 26
Figure 7: The diatoms identified by light microscopy of the lake transect from north to south, in the four vegetation zones tundra, single tree tundra, forest tundra and light taiga. The color code highlights the vegetation zones. 31
Figure 8: Plots of the redundancy analysis of the species assigned to sequences and the morphological identified species. A - RDA of significant environmental parameters on species assigned to sequences, B - RDA of significant environmental parameters on species identified by LM, C – RDA of vegetation types on the species assigned to sequences, D – RDA of vegetation types on the species identified by LM. 32
Figure 9: Rarefaction curves of both species data sets. A - Rarefaction curve of the genetic data, B - Rarefaction curve of the morphological data. 35

Figure 10: The diatom species of sequence types of core 11-CH-12A and the assemblage groups identified by CONISS analysis, as well as the estimated richness and the identified number of species. 39
Figure 11: The diatom species of sequence types of core 11-CH-06D and the assemblage groups identified by CONISS analysis, as well as the estimated richness and the identified number of species. 41
Figure 12: The species identified by light microscopy of core 11-CH-12A and the assemblage groups identified by CONISS analysis, as well as the estimated richness and the number of identified species. 44
Figure 13: Diatom species identfied by light microscopy of core 11-CH-06D and the assemblage groups identified by CONISS analysis, as well as the estimated richness and the number of identified species. 47
Figure 14: Rarefaction curves of the species data. A - Rarefaction curve of the genetic data of core 11-CH-12A, B - Rarefaction curve of the morphological data of core 11-CH-12A, C - Rarefaction curve of the genetic data of core 11-CH-06D, D - Rarefaction curve of the morphological data of core 11-CH-06D. 48

List of measurement units

°C	degree in Celsius
µl	microliter
cm	centimeter
g	gram
km	kilometer
mg	milligram
min	minute
ml	milliliter
ng	nanogram
pH	numeric scale to specify acidity and alkalinity

List of Tables

Table 1: Chemicals used for the PCR reactions. Chemicals marked with UV were decontaminated using UV radiation for 5 min using a crosslinker instrument. 16
Table 2: PCR program 16
Table 3: Retrieved diatom sequence types and their occurrence in each of the selected samples..... 23
Table 4: Proportions of chemical and morphological lake characteristics for the genetic data of the species, as well as the proportion of all significant variables, the significant variables under the condition of the vegetation types and the vegetation types under the condition of the significant variables. The unique proportions were only calculated fin case of significant single proportions. Forest.tundra has no p-value, because it was highly correlated with the other vegetation zones and thus redundant. 27
Table 5: Proportions of chemical and morphological lake characteristics for the morphological data of the species, as well as the proportion of all significant variables,

the significant variables under the condition of the vegetation types and the vegetation types under the condition of the significant variables. The unique proportions were only calculated fin case of significant single proportions. Forest.tundra has no p-value, because it was highly correlated with the other vegetation zones and thus redundant..... 33
Table 6: Comparison of the counts, species number and estimated richness for the species and genera data of each the genetic and the morphological assessment of the lake surface data set..... 36
Table 7: The counts, species number and estimated richness of species and genera for all samples for both genetic and morphological assessment..... 49
Table 8: The counts, species number and estimated richness for species and genera for all samples both for the genetic and the morphological assessment. 50

Abstract

The Arctic treeline ecotone is a large environmental gradient covering a vast area in Siberia. It sensitively reacts to changes in the environment, which is observable, for instance, in the responding vegetation. The vegetation across the treeline is known to influence the water chemistry of thermokarst lakes in this region. Sensitive algae, such as diatoms, respond to these changes and thus, diatom compositions correlate with the surrounding vegetation. Most studies on the diatom compositions across the treeline ecotone used the classic morphological approach, whereas little is known about the genetic diversity and composition of diatoms.

In this Master study NGS sequencing data were used to analyze the diatom composition of 32 lakes and two lake sediment cores in comparison to morphological data sets. Furthermore, the correlation of diatom assemblages with environmental factors was analyzed and compared to morphological data. All analyses were conducted on two taxonomic levels, i.e. species and genera, in order to check the taxonomic resolution of the analyses. The results showed a significant relationship between the diatom composition and specific environmental parameters, i.e. DOC and maximal depth for both methodologies. Calcium and sulfate also had a significant influence on the genetic obtained data, while hydrogen carbonate and conductivity had a significant influence on the morphological data.

It was possible to gain similar results with both approaches regarding the diatom compositions of both the recent and the ancient sediment material. Furthermore, it could be shown that there are only slight differences between the taxonomic assignment on species and genus level in the genetic approach, due to incomplete reference databases. Altogether, it could be concluded that even though the analyzed and compared methods are very different, their results are highly comparable.

Zusammenfassung

Das Baumgrenzökoton in der Arktis ist ein großer ökologischer Gradient, der eine erhebliche Fläche in Sibirien abdeckt. Dieses Ökoton reagiert sensitiv auf Veränderungen in der Umwelt, was zum Beispiel in der umgebenden Vegetation zu beobachten ist. Die Vegetation durch die Baumgrenze ist bekannt dafür die Wasserchemie von thermokarst Seen in der Region zu beeinflussen. Sensitive Algen, wie Diatomeen, reagieren auf diese Veränderungen in der Wasserchemie. Deshalb korrelieren Diatomeenvergesellschaftungen mit der umgebenden Vegetation. Die meisten Studien von Diatomeenzusammensetzungen durch das Baumgrenzökoton haben den klassischen morphologischen Ansatz genutzt, während nur wenig über die genetische Diversität und Zusammensetzung der Diatomeen bekannt ist.

In dieser Masterarbeit wurden NGS Sequenzdaten genutzt um die Diatomeen-zusammensetzung von 32 Seen und zwei Seesedimentkernen zu analysieren und mit morphologischen Daten aus diesen Seen zu vergleichen. Außerdem wurden die Artzusammensetzungen mit Umweltfaktoren korreliert und ebenfalls mit den morphologischen Daten aus diesen Seen verglichen. Alle Analysen wurden auf zwei taxonomischen Ebenen durchgeführt, i.e. Art und Gattung, um die taxonomische Auflösung der Analysen zu überprüfen. Die Ergebnisse haben einen signifikanten Zusammenhang zwischen Diatomeenvergesellschaftung und spezifischen Umwelt-parametern, wie DOC und maximale Tiefe, für beide Methoden gezeigt. Calcium und Sulfat hatten ebenfalls einen signifikanten Einfluss auf die genetischen Daten, während Hydrogenkarbonat und Leitfähigkeit einen zusätzlichen signifikanten Einfluss auf die morphologischen Daten hatten.

Beide Methoden ergaben ähnliche Ergebnisse bezüglich der Diatomeen-zusammensetzung für rezente und alte Sedimente. Außerdem konnte gezeigt werden, dass es geringfügige Unterschiede zwischen der taxonomischen Zuordnung auf Art- und Gattungsniveau im genetischen Datensatz gab, was auf unvollständige Referenzdatenbanken zurückzuführen ist. Abschließend geht aus dieser Masterarbeit hervor, dass obwohl die analysierten und verglichenen Methoden grundlegend verschieden sind, die Ergebnisse dennoch vergleichbar sind.

1 Introduction

1.1 The Arctic treeline ecotone

The Arctic circumpolar region is characterized by snow and ice that cover much of the land and sea surface during most of the year (ACIA, 2004). Going from North to South a wide extent of treeless plains over frozen ground is followed by large boreal forests. One boundary used to define the Arctic is the arcto-boreal treeline, which is the border between tundra and taiga (MacDonald et al., 2008; ACIA, 2004). This transition zone is often referred to as forest-tundra and its latitudinal extent can range from a few kilometers to over a hundred kilometers (MacDonald et al., 2008). At the northernmost limits the coniferous species mainly grow as scattered krummholz or dwarf trees (MacDonald et al., 2008). Generally, the treeline on the Eurasian continent runs for over 5500 km almost parallel to the Arctic coast, ranging between 60° and 70° N latitude. At the Taymyr peninsula (Central Siberia, Russia) the treeline extends northward up to 73° N and thus represents the world's most northern boreal forests (ACIA, 2004; MacDonald et al., 2008). Siberia in general, encompasses the largest forest-tundra ecotonal belt of the world (Frost & Epstein, 2014).

The anthropogenic rapid climate change of the last decades has a severe impact on arctic environments (ACIA, 2004; MacDonald et al., 2008). Already, in some regions a migration of trees and shrubs northwards could be observed (Sturm et al., 2001; Esper & Schweingruber, 2004; Post et al., 2009; Blok et al., 2010; Myers-Smith et al., 2011; Naito & Cairns, 2011). This results in forests encroaching tundra and tundra encroaching polar deserts (ACIA, 2004).

Coniferous trees are known to affect the limnological characteristics of lakes in the vicinity, because the higher evapotranspiration leads to salt accumulations in the soil, which are washed in to lakes through rain and snow melt (Pienitz et al., 1995; Rouse et al., 1997; Herzschuh et al., 2013). That leads to limnological differences depending on the surrounding vegetation between tundra and taiga areas (Laing & Smol, 2000). The differences in limnological characteristics of arctic lakes give rise to a variety of species compositions specific for each lake. Thus, changes in the environment and the lake-

water chemistry can affect organisms, such as sensitive algae and hereby lead to changes in the species composition in the lake.

1.2 Diatoms as bioindicators

A group of algae that reacts most sensitively to environmental factors and their changes are diatoms. Diatoms are unicellular eukaryotic algae belonging to the class of the *Bacillariophyceae*. They occur in almost all aquatic environments worldwide (Jiang et al., 2004; Flower & Battarbee, 1983; Jahn et al., 2007; Laing et al., 1999a). Diatoms are important biomass and oxygen producers in marine and lake environments (Battarbee et al., 2001). They are characterized by their siliceous cell walls (Battarbee et al., 2001). Diatoms consist of two almost identical thecae, which differ slightly in size because of cell size reduction during the reproduction cycle (Battarbee et al. 2001). Some taxa form colonies while others are solitary (Battarbee et al., 2001). Furthermore, they divide into benthic, pelagic and planktonic species (Battarbee et al., 2001). Due to the specific patterns of valve faces of each theca, a taxonomic identification at species level is possible even for most fossil samples. The silica in the cell walls leads to a preservation of diatom frustules as fossils and thus, reveals information about past environmental and climate conditions (Battarbee et al., 2001). Because of the different ecological preferences and habitats each species represents (Battarbee et al., 2001), diatom assemblages respond to environmental changes and thus are important bioindicators for water chemistry, light exposure, temperature and turbation in the water column (Battarbee et al., 2001; Rühland, 2001). To analyze influences of environmental change the optima and tolerances of species are important knowledge (Puusepp & Kangur, 2010). Diatoms are well preserved in lake sediments, and therefore bioindicators of past lake-water chemistry and state of the lake. Accordingly, they are widely used in paleolimnological studies and to develop transfer-functions. Today, diatom analysis is not just to investigate their ecology (Lauterborn, 1896; Battarbee et al., 2001), but also to understand the value of recent diatoms as indicators for water quality and surface water acidification (Charles, 1985; Flower & Battarbee, 1983; Zimmermann et al., 2011), eutrophication (Anderson & Rippey, 1994; Whitmore, 1989), and their potential as temporal indicators of

environmental and climate change from sediment records (Gasse, 1987; Fritz et al., 1991; Pienitz et al., 1995; Gasse et al., 1997; Rouse et al., 1997; Hobbs et al., 2010; Biskaborn et al., 2012; Herzschuh et al., 2013). The results are then used to reconstruct and interpret past environments (Roberts & McMinn, 1998; Sylvestre et al., 2001; Ryves et al., 2002; Rühland et al., 2003; Yang et al., 2003). Diatoms are widely used, especially for Holocene records (e.g. Korhola et al., 2000; Jiang et al., 2002; Rudaya et al., 2009), and are exceptionally interesting in Siberia, because of the rapid anthropogenic induced climate change that has strong impacts on the Arctic environments (ACIA, 2004; IPCC, 2013). Thus, climate change might induce shifts in diatom assemblages in thermokarst lakes in Siberia, depending on the differences in carbon uptake as shown for marine diatoms by Burkhardt et al. (2001).

A change in modern diatom composition within lakes is visible across the treeline ecotone (Laing et al., 1999; Rühland et al., 2003). In general, lakes are covered by ice sheets approximately three-quarters of the year. Thus, the reproduction time and algae bloom is restricted to a short period (Rühland et al., 2015). Furthermore, lakes in the forest zone have a higher DOC and conductivity then tundra lakes (Pienitz et al., 1997; Hongve, 1999; Rühland et al., 2003; Herzschuh et al., 2013). The DOC content is higher in forest areas as the catchment of coniferous leaf litter is water drained into the lakes. As Laing et al. (1999) showed, local lake-water chemistry has a stronger influence than climatic factors on the diatom composition within the lake. So far most studies using diatoms as environmental indicators only focused on the morphological data obtained from the lake sediments or the water column (Pienitz et al., 1995; Rühland et al., 2003; Schmidt et al., 2004; Pestryakova et al., 2012). Due to the arctic environment and its harsh conditions, most diatom species found in Siberia are very small and hence, very difficult to identify by their morphology (Biskaborn et al., 2012). Dominant and most diverse species in arctic lakes are small fragilarioid taxa, such as *Staurosira*, *Navicula*, *Cymbella* and *Pinnularia* (Biskaborn et al., 2012; Pestryakova et al., 2012)(see Figure 1 for examples). However, there is a high potential that species are cryptic or identical. Accordingly, morphological similar species might include distinct genetic or even reproductively isolated lineages. Such differences are not detectable with traditional microscopic methods, but can be uncovered using genetic analyses on diatoms (Evans et al., 2007). Evans et al. (2007) used isolates of *Sellaphora*

to determine relationships in this genus. Genetic analyses can also be used to assess species communities, like Kermarrec et al. (2013) tested for a mock community of freshwater diatoms. Thus, the assessment of genetic diversity and species compositions is used for cultures and mixed environmental samples.

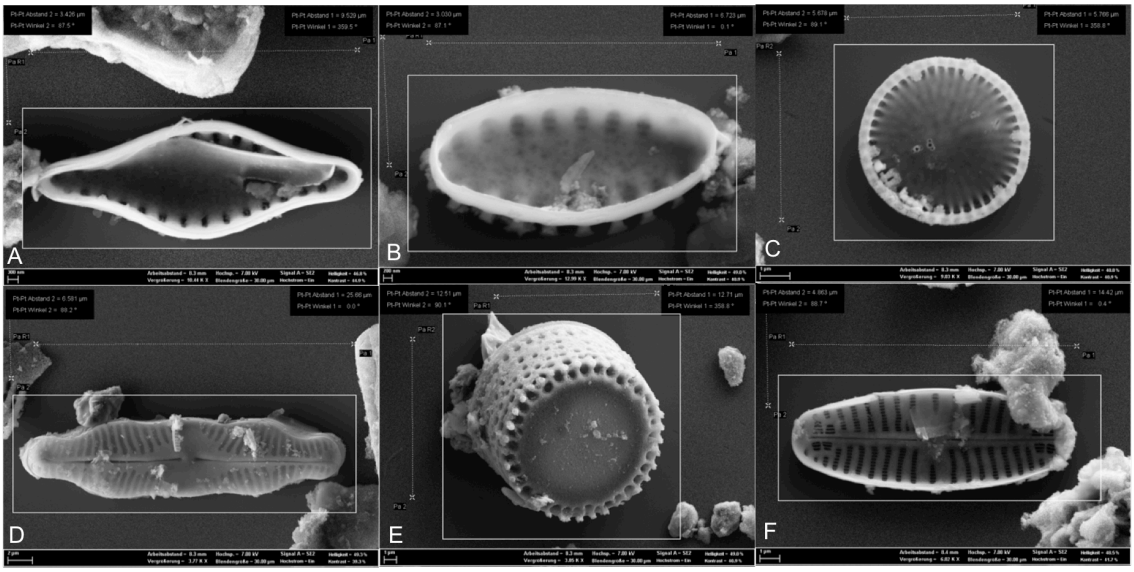


Figure 1: Common diatom species of Arctic treeline lakes. A - *Staurosira construens*, B - *Staurosira pinnata*, C - *Cyclotella* sp., D - *Pinnularia microstauron*, E - *Aulacoseira distans*, F - *Navicula* sp. (Diatom identification was done together with Luidmila Pestryakova and pictures were taken with the SEM at GFZ Potsdam in cooperation with Ilona Schäpan).

1.3 Environmental DNA

DNA can be deposited in the environment through animals and plants via e.g. decaying bodies, leaves or in some cases pollen and the secretion of plasmid and chromosomal DNA of living prokaryotes (Pietramellara et al., 2009; Pedersen et al., 2015). This so called environmental DNA (eDNA) is obtained from ice, water or sediments. DNA found in sediments is called sedimentary DNA (sedDNA), which includes recent eDNA and ancient DNA (aDNA), i.e. old, highly fragmented and degraded DNA, and always consists of a mixed sample of diverse species communities. Due to the bacterial and fungal DNases, UV radiation, heat, and chemical modifications such as oxidation, deamination, depurination and other hydrolytic processes eDNA is extremely fragmented and modified in most of the cases (Hofreiter et al., 2001; Pedersen et al., 2015). Working with old DNA requires strict protocols to prevent contamination (Hofreiter et al., 2001; Pääbo et al., 2004; Gilbert et al., 2005; Rawlence et al., 2014; Pedersen et al., 2015), and these protocols increased the amount of reliable studies in this field. Until recently, successful studies were focusing on animal populations,

because the recovery of material was easier, e.g. through museum collections, than that of plant material (Pääbo et al., 2004; Mason et al., 2011).

Generally, eDNA reveals information about present and past diversities and biogeography, as well as allowing a broad biodiversity assessment, including the detection of rare species (Pedersen et al., 2015). A suitable starting point for eDNA recovery of plants are lake sediments, which have good preservation qualities. Thus, the preserved rDNA of planktonic algae could be recovered from Holocene sediments of a lake in Antarctica (Coolen et al., 2004) and also haptophyte aDNA from the Black Sea (Coolen et al., 2006). Also permafrost is a rich archive of genetic variation, due to the good preservation in cold environments, and it contains vast numbers of preserved individuals (Hofreiter et al., 2001).

DNA metabarcoding is the use of large-scale screening of one or more markers that distinguish between species and thus allows a diversity assessment of community samples, especially useful to analyze microorganisms such as bacteria, algae and zooplankton, which are more difficult to identify solely by morphological features. DNA metabarcoding needs different markers than those used for barcoding of single individuals, in order to fit the shorter, degraded and necessarily more specific DNA fragments.

The recently developed technique of the analysis of environmental DNA offers new insights into genetic diatom diversity. A variety of markers is used and still tested to amplify diatom-specific DNA from cultures (Evans et al., 2007; Hamsher et al., 2011; MacGillivary & Kaczmarska, 2011; Zimmermann et al., 2011) and environmental samples (Jahn et al., 2007), but only a few studies tested a marker for its suitability on sedimentary environmental samples (Coolen et al., 2004; Stoof-Leichsenring et al., 2012 and 2014; Epp et al., 2015). Stoof-Leichsenring et al. (2014) proved the rbcL gene as particularly suitable for sedimentary DNA applications. The amplification probability, specifically in older samples increases (Stoof-Leichsenring et al. 2012 and 2014) due to the fact that rbcL is a part of the chloroplast genome, and therefore, occurs in multiple copies per cell. Its specificity leads to an amplification only of photosynthetic organisms, while other markers like COI are present in almost all organisms because of their mitochondrial origin. Additionally, the rbcL marker is very

variable in specific groups of diatoms, providing a relatively high taxonomic resolution up to variety level.

So far, for metabarcoding studies at the Siberian treeline cloning and Sanger sequencing was used for the analysis of the genetic diversity of the *Staurosira/Staurosirella*-like haplotypes, which showed a great diversity and spatial diversification among Siberian lakes (Stoof-Leichsenring et al., 2014). Stoof-Leichsenring et al. (2014) showed a spatial and temporal change in *Staurosira* haplotypes along the Arctic treeline and associated these changes with the surrounding vegetation. The analysis of diatom assemblages and their correlation with environmental data is for now only available based on morphological data (Pestryakova et al., 2012).

The advantage of the genetic approach is the possibility to analyze inter- and intra-specific differences of haplotypes, whereas the morphological approach cannot give this information, but is a classic and cheap method to analyze species compositions. Both methods are very different from each other, but because they both lead to a determination of species composition in the evaluated sample a comparison might show the advantages and disadvantages of each approach.

2 Objectives

The geographic distance and the environmental gradients along the Siberian treeline ecotone are characterizing the thermokarst lakes in this area. Previous studies in this ecotone could provide important information about the diversity of diatom species in thermokarst lakes (Stoof-Leichsenring et al., 2014 and 2015) and the influence of environmental factors on diatom composition in Siberian lakes (Pestryakova et al. 2012). But these studies were either focusing on the genetic diversity of one particular diatom genus (*Staurosira/Staurosirella* group) using cloning with subsequent Sanger sequencing or they used light microscopy to study diatom assemblages.

This study will analyze diatom assemblages across the treeline ecotone in Siberia with DNA metabarcoding approach after high-throughput sequencing on the Illumina platform. The aims of this thesis are:

1. to analyze the diatom diversities on different taxonomic levels (species/genera) and to compare the genetic and morphological approach
2. to analyze the correlation of diatom assemblages with environmental factors on different taxonomic levels and to compare the genetic and morphological approach
3. to analyze the temporal change of diatom communities of two sediment cores on different taxonomic levels and to compare the genetic with the morphological approach

The overall aim of this study is to examine if similar results of spatial and temporal changes in diatom assemblages can be retrieved using different taxonomic units and two different proxies, i.e. DNA and diatom frustules. Hence, the study shows if genetic analysis can give comparable results on diversities, independent of analyses done by taxonomists.

3 Material and Methods

3.1 Study area

The study area of Northern Yakutia (northeastern Siberia, Russia) is located at the southern part of the Taymyr peninsula within the district of Krasnoyarsk (Figure 3). The Taymyr peninsula is situated between the continental East Siberian and the marine-influenced West Siberia. Hence, this region is particularly sensitive to climate fluctuations (Andreev et al., 2002). Lowlands and small hills dominate the area. Perennially frozen ground at, or below 0°C soil temperature for at least two serial years, is defined as permafrost. Generally, in Russia approximately 60% of the landmass is covered by permafrost, whose key characteristic is ground ice (Soil Atlas of the Northern Circumpolar Region, 2010). The upper part of the permafrost thaws locally from 10-53 cm, during summer. This thaw-zone is referred to as active layer, because of the functioning soil processes in this season (Soil Atlas of the Northern Circumpolar Region, 2010). The landscape is characterized by deep (down to 400 m), ice-rich, continuous permafrost (Soil Atlas of the Northern Circumpolar Region, 2010). The thawing of the ice leads to the landscape forming process of surface subsidence. This process leads to the formation of depressions and is called thermokarst. Moisture and snow can accumulate in the micro-depression, slowing down freezing (winter) and thawing (summer), which allows the subsidence to become deeper. The subsequent development of small lakes supports further permafrost thawing (Soil Atlas of the Northern Circumpolar Region, 2010). These processes cause gradual changes in the landscape.

The thermokarst lakes are oligotrophic, differing in depth, size and catchment area. During the Last Glacial Maximum (LGM) the lowlands of northern Siberia were left ice-free, therefore, glacial deposits do not overprint the landscape (Matveyeva, 1994; Möller et al., 1999). The weather in the region of Khatanga is temperate, but cold. There is all year-round high precipitation, even in the driest month. The mean annual temperature is -14°C and the mean annual precipitation is 200-300 mm (Andreev & Klimanov, 2000). The warmest month is July with a mean temperature of 12.6°C in the south and only 2°C in the north of the polar desert, while the coldest month January

has a mean temperature of -33.9°C and does not vary along the north-south gradient (Matveyeva 1994; Möller et al. 1999). The studied transect is located in the transition zone between Southern tundra (Subarctic) and larch forest-tundra (Hypoarctic), as indicated by Matveyeva (1994) and Hahne and Melles (1998). The variation in summer temperatures across the region is the key factor for the variation in vegetation from the scarce plant cover in the tundra to the Larch forest south of the treeline (Matveyeva, 1994). The permafrost soil, short blooming season and climate factors like low precipitation lead to specialized vegetation communities.

The vegetation across the studied transect was categorized by the expedition team, according to the local flora. Thus, the vegetation around lakes 1-10 was described as tundra vegetation, around lakes 11-16 as single tree tundra, around lakes 17 and 25-32 as forest tundra and around lakes 18-24 as light taiga (Figure 2).

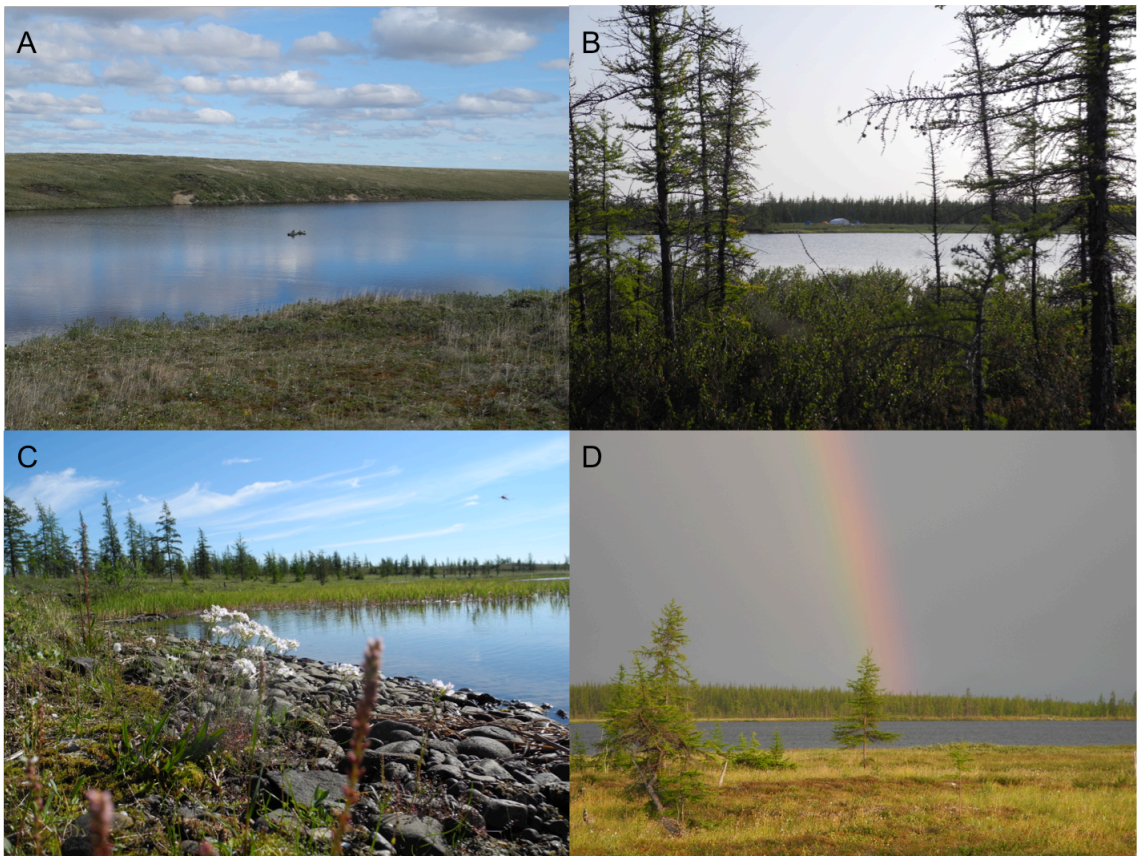


Figure 2: Examples for the vegetation types around the lakes of the modern data set. A - tundra, B - single tree tundra, C - light taiga, D - forest tundra. (Pictures were taken by Stefan Kruse during the expedition 2013)

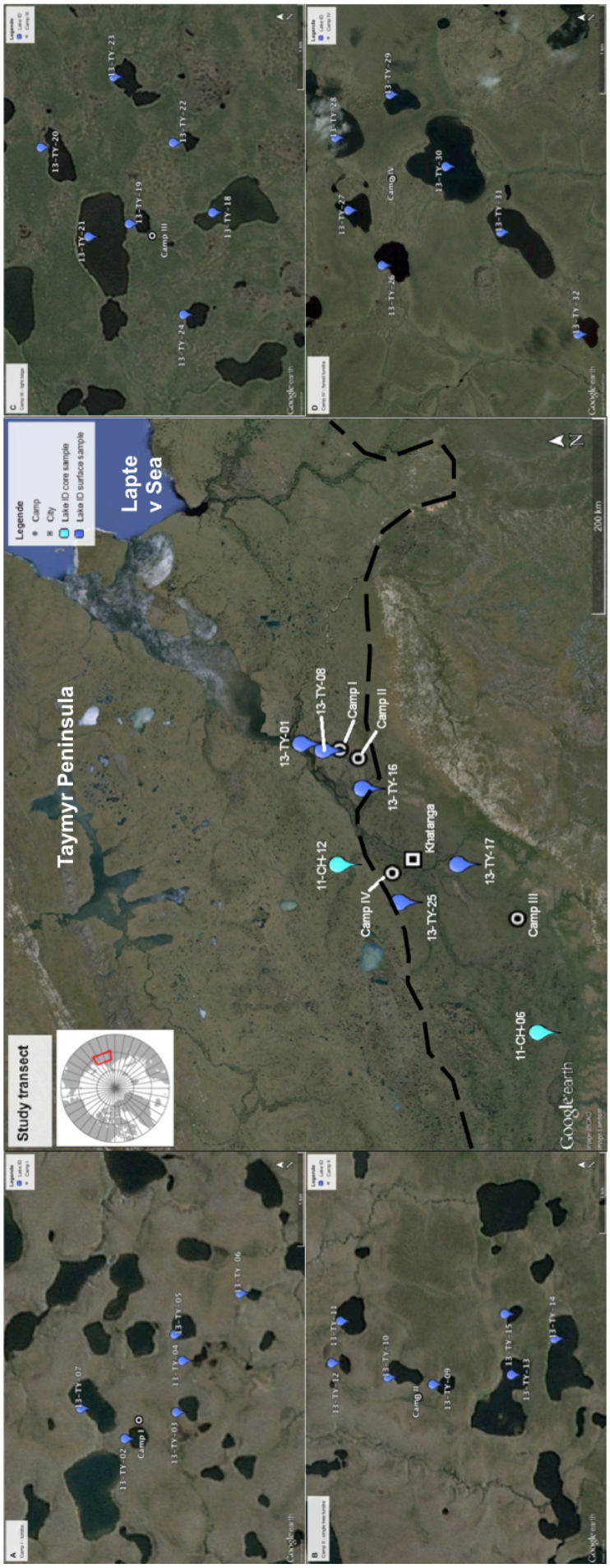


Figure 3: Maps of the study transect. A - camp I in the tundra, B - camp II in the single tree tundra, C - camp III in the light taiga, and D - camp IV in the forest tundra. The black stripped line indicates the current position of the treeline.

3.2 Sampling

3.2.1 Sampling procedure in the field

Sampling was carried out in 2011 and 2013, during field trips conducted by the Alfred Wegener Institute, Potsdam, in cooperation with the North-Eastern Federal University of Yakutsk. Lacustrine surface sediment and water samples were collected from 32 lakes along the latitudinal transect in the northern lowlands of Yakutia. The localities range from the edge of the Arctic tundra in the vicinity of the Laptev Sea, to the tundra-forest and forests south of Khatanga, crossing the current treeline ecotones (Figure 3). As all of these lakes were located in remote areas of the Taymyr Peninsula and access is only possible by helicopter, the sites could only be visited once and thus represent a “snaphot” of the limnological conditions measured approximately at the same time of the year. Geographic coordinates were taken using a handheld GPS device (Garmin etrex).

32 surface sediment samples (13-TY-01 to 13-TY-32) and two long sediment cores (11-CH-06D and 11-CH-12A) were analyzed. Sediments of the uppermost 1 cm of the surface sediment were taken with an Ekman-Birge bottom sampler. For DNA analysis the lakes were sampled on different days. Washing the sediment grabber with sterile water several times before sampling minimized the risk of cross-contamination between samples. All surface sediments were sampled in sterilized 150 ml Nalgene tubes using a new sterile plastic spoon for each sample.

Water depth was measured from rubber boats with an echo sounder and sediment cores were taken at the deepest measured depth of the lake with an UWITEC gravity corer equipped with a hammer tool. Core 11-CH-06D could be retrieved with a length of 160.5 cm and core 11-CH-12A with a length of 132 cm. Both cores have a diameter of 6 cm.

Core sediment samples were stored cool (4°C) and dark for transportation from Russia to Potsdam, where the cores were opened and subsampled in the laboratory. For each lake 28 chemical as well as 6 physical and geographical variables e.g. maximal depth, pH and DOC, were measured. The analysis of water chemistry was done by Antje Eulenburg in the laboratory of the Alfred Wegener Institute.

3.2.2 Subsampling of sediment cores for ancient DNA analyses

Cores were sampled in the climate chamber of the GeoForschungszentrum (GFZ), Potsdam at 10°C under clean conditions. First, the liner was opened with a circular saw. Then, the core was cut into two pieces by a metal wire. The two resulting core halves were separated carefully. One was wrapped in foil and stored as archive and for dating samples and the other half was used for genetic sampling. In order to prevent contamination from the outer layers the top 0.5 cm were detracted from the open core using sterile knifes. Genetic samples were taken using a 5 ml sterile one-way syringe, taking at least 2 ml of sediment without touching the liner walls and thereby, leaving the outer sediment in the core, also to prevent contamination. Taken samples were put immediately into 8 ml tubes and stored at -18°C until further use to prevent the DNA from further degradation. Four samples of core 11-CH-06D for additional analyses were taken in the pollen laboratory of the AWI. Samples were taken using modified aluminum plates to cut 0.5 cm thick slices out of the core half. In a following step the edges of the samples that touched the liner were cut off, the samples were split and one half was stored for genetic analysis, while the other one was taken for pollen analysis. This subsampling was not part of the thesis, as all subsamples were already taken after the expeditions.

3.3 Dating of sediment cores

3.3.1 ^{210}Pb/^{137}Cs dating

The cores sampled in lake 11-CH-06, i.e. core 11-CH-06E and core 11-CH-06D, were taken parallel to each other. The shorter core 11-CH-06E was subsampled in the field over a length of 46 cm resulting in 90 subsamples for ^{210}Pb/^{137}Cs dating. The cores sampled in lake 11-CH-12, i.e. core 11-CH-12D and core 11-CH-12A, were also parallel cores from the same lake. Thus again, the shorter core 11-CH-12D was subsampled in the field over a length of 32 cm resulting in 40 subsamples for ^{210}Pb/^{137}Cs dating. The samples for both cores were freeze-dried and a subset of each 40 samples was sent to the Environmental Radioactivity Research Centre of the University of Liverpool, UK (Appleby & Piliposyan 2012). The sampling and preparation of the samples was not part of this thesis. The results were kindly provided by Juliane Klemm (AWI Potsdam).

3.3.2 Radiocarbon (^{14}C) dating

The longer cores 11-CH-06D and 11-CH-12A were sampled in the laboratory at the AWI Potsdam. 11-CH-06D had a total length of 160.5 cm of which eight ^{14}C samples were taken for dating. 11-CH-12A had a total length of 131.5 cm and a total of 16 samples were taken for ^{14}C dating in the laboratories of the AWI Potsdam. The samples for both cores were freeze-dried and sent to the Poznan Radiocarbon Laboratory, Poland. The sampling and preparation of the samples for dating was not part of this thesis. The results were provided by Juliane Klemm (AWI Potsdam).

3.3.3 Age-Depth-Models

3.3.3.1 Core 11-CH-12A

The age-depth-model of core 11-CH-12A is based on ^{210}Pb/^{137}Cs and ^{14}C measurements. The ^{137}Cs concentrations of this core have a well-defined maximum value recording the 1963 atmospheric fallout maximum. The best estimate of the mean sedimentation rate lies at a of 0.0064 ± 0.0007 g cm^{-2} y^{-1}, equaling a mean of 0.037 cm y^{-1}. To verify the sedimentation progress also for this core ^{14}C measurements were included (Appleby & Piliposyan, 2012). The resulting age-depth-model is shown (Figure 4) was kindly provided by Juliane Klemm (AWI Potsdam) and published in Stoof-Leichsenring et al. (2015).

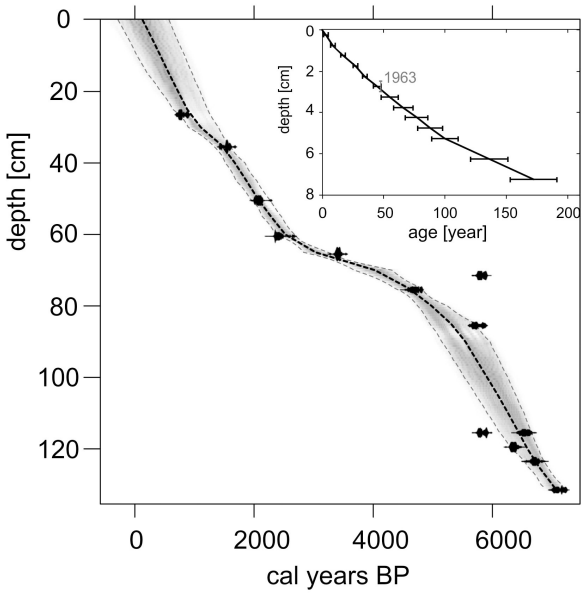


Figure 4: Calibrated age-depth-model of core 11-CH-12A.

3.3.3.2 Core 11-CH-06D

The age-depth-model of core 11-CH-06D was based on $^{210}Pb/^{137}Cs$ and ^{14}C measurements. The age-depth-model is shown in Figure 11. Core 11-CH-06D showed a very low sedimentation rate in recent times and a poor record of atmospheric fall out based on $^{210}Pb/^{137}Cs$ measurements. The best estimate of the mean sedimentation rate lays at 0.0054 ± 0.001 g cm^{-2} y^{-1}, equaling a mean of 0.021 cm y^{-1}. To verify the extrapolation of the presumed sedimentation rate ^{14}C measurements were included in the model (Appleby & Piliposyan, 2012). The model was also kindly provided by Juliane Klemm (AWI Potsdam) and kindly set at my disposal.

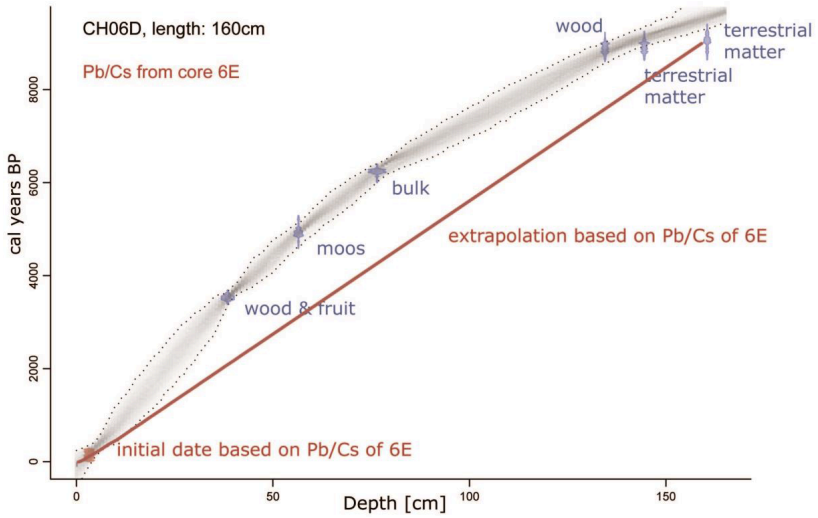


Figure 5: Age-depth-model of core 11-CH-06D with extrapolation based on Pb/Cs results of core 11-CH-06E.

3.4 Genetic assessment of sediment samples

3.4.1 DNA Extraction

The extraction of DNA from sediment samples results in a mixture of total genomic DNA from intra- and extracellular DNA of living and dead organisms in the lake.

All ancient DNA (aDNA) isolations of lake sediment cores were conducted at the ancient DNA facilities (aDNA laboratory), while modern analyses were conducted in a different laboratory building (general genetic laboratory) of the Alfred Wegener Institute Potsdam in 2014. Furthermore, four ancient samples of the core 11-CH-06 were conducted in a dedicated aDNA laboratory at the Institute of Biochemistry and Biology, University Potsdam, in 2013. All DNA extractions of ancient material were performed under special clean conditions such as the use of whole body protective

clothing, including mask, hairnet, arm guards, shoes, shoe overcoats and two pairs of gloves. This equipment was changed as often as necessary, to prevent any contamination. Techniques to minimize the contamination risk such as overnight UV light exposure in the aDNA laboratory, bleach sterilization, cleaning of all materials and surfaces using DNA Exitus. The use of PCR workstations with a UV-hood and filtered pipette tips, the division of sample preparation, extraction and PCR setup in the ancient DNA facilities are further steps to minimize contamination risks. The PCR amplification was processed in a separate facility. Within the designated preparation room, samples were stored at -20°C until use. Negative extraction controls (blanks) were done for every charge of extractions. Three blanks were processed identical to the 21 aDNA samples, except that the blanks were processed without any sample material. Sedimentary DNA was extracted using the PowerMax Soil Isolation Kit according to the manufacturer's protocol but with modifications during the initial lysis. Additionally, to the PowerBead Solution 400 µl proteinase K was used for ancient DNA extractions. The extractions of ancient material were done by Laura Epp (AWI Potsdam) and kindly placed at my disposal.

The DNA extraction of surface sediments was conducted in the pre-PCR laboratory. For seven samples (03, 09, 15, 19, 26, 27, 29) from the 2013 campaign only the first 5-10 cm of the sediment were taken, whereas for all others the sample of the upper 1 cm was used. In order to extract the DNA from surface sediments the sample bottles were cleaned with DNA-Exitus and sampled under a UV-hood dedicated for DNA isolation of environmental samples. To reduce contamination risk each bottle was handled separately and the sediment sample used for the extraction protocol was taken with a sterile spatula and transferred into the Bead tube of the PowerMax Soil Isolation Kit and with additional 400 µl of Prot.K. The weight, which was roughly 5 g, of the samples was measured and noted down after all necessary reagents were filled in the tubes. Afterwards the tubes were mixed thoroughly by vortexing.for 2 min Further extraction steps were carried out following the manufacturer's protocol of the PowerMax Soil Isolation Kit.

3.4.2 Polymerase Chain Reaction

The Polymerase Chain Reaction (PCR) increases the amount of target products in comparison to other genomic DNA pieces in the DNA extract. The setup for ancient DNA was done in the aDNA laboratory, while PCRs of the surface samples were done in the pre-PCR laboratory. To amplify the rbcL locus (76 bp) the protocol of Stoof-Leichsenring et al. (2012) was followed with minor modifications (Table 1, Table 2). Additionally, primers were modified on the 5' end with unique 8 bp tags that differed from each other in at least 5 basepairs following Binladen et al. (2007), and were preceded by primer suffix NNN to improve cluster generation on the sequencer (Coissac, 2012; De Barba et al., 2014)(Appendix 24). In case of a malfunctioning of the used tag combination another one was used. For all PCR setups HiFi Taq DNA Polymerase was used, because it has an efficient proofreading 3' to 5' exonuclease activity that improves PCR reactions. For each sample two independent PCR reactions were run.

Table 1: Chemicals used for the PCR reactions. Chemicals marked with UV were decontaminated using UV radiation for 5 min using a crosslinker instrument.

Master Mix	Amount in µL
H_2O (UV)	14.75
10X buffer (UV)	2.5
dNTPs (2.5 mM)	2.5
BSA (UV, 20 mg/ml)	1
$MgSO_4$ (UV, 50 mM)	1
HiFi (5 U/µl)	0.25
Primer Mix (10 µM)	1
DNA	2
Total volume	25

Table 2: PCR program

°C	t	
94	5 min	
94	30 sec	50 cycles
49	30 sec	
68	30 sec	
72	10 min	
15	forever	

3.4.3 Pre-Check: Cloning and Sanger sequencing for selected samples

Cloning of PCR products was done with the TOPO® TA Cloning® Kit for Sequencing according to the manufacturer's protocol. To verify amplified products and confirm the authenticity of sequence variation the protocols and recommendations of Cooper and Poinar (2000) and Krings et al. (1997) were followed. Furthermore, PCR products of extraction blanks and NTCs were cloned to check whether positive PCR products resulted from either contamination or from primer-chimeras (Krings et al., 1997; Cooper & Poinar, 2000). A subset of eight designated samples was selected for cloning and subsequent Sanger sequencing and an average of 20 clones per sample were picked. After a following PCR the selected clones were sent to LGC Genomics sequencing service (Berlin, Germany) for purification of PCR products and Sanger sequencing.

The retrieved clone sequences were imported to Geneious® 7.1.4 (http://www.geneious.com, Kearse et al., 2012). Primer sequences were uploaded and aligned to the sequences. Afterwards, ends were trimmed and sequences shorter or longer than the expected length were deleted. Then, sequences were translated into amino acids and reversed if applicable. This was done to check for stop codons that indicate errors in the sequence. Using the function "Find Duplicates" unique sequences were extracted for each sample. "Sequence Search" was used with the Geneious® function *Megablast* that matched unique sequences to the reference databases "nr". The output was analyzed for similar sequences between samples.

3.4.4 Purification of PCR products

The purification of PCR products, necessary to remove inhibiting salts, dNTPs from the PCRs and to concentrate the DNA content, which is required for sequencing, was done with the Qiagen MinElute PCR Purification Kit following the supplied protocol. For each sample the two independent PCR products were pooled and then purified. The used elution volume was 20 µl of the provided Elution buffer.

3.4.5 Measurement of DNA Quantity

The measurement of DNA quantity is needed to calculate the necessary volume of each sample to pool the samples equimolar for high-throughput amplicon multiplex-sequencing. Therefore, the DNA quantity of the purified samples was measured using 1 µl of the purified PCR products for the analysis with the Qubit dsDNA broad range assay kit and the Qubit® 2.0 Fluorometer. Each sample was measured two times, and the mean value of each sample was used for further calculations in order to minimize measurement errors. The overall DNA concentration was 72 ng per µl.

3.4.6 Parallel high-throughput sequencing

Parallel high-throughput amplicon sequencing was performed on the Illumina HiSeq 2000 (high Output) platform by the service of by Fasteris SA (Geneva, Switzerland). The process of sequencing consists of the three following steps:

- Preparation of the library (ligation of adapters onto amplicons)
- Quality check of the library using qPCR and
- Illumina Amplicon run on HiSeq High Output platform using an economy lane with 2x 100 bp, producing a total of about ten million sequences.

3.4.7 Bioinformatic analysis of sequencing data

The Illumina Amplicon Sequencing run resulted in two data files, one containing the forward and the other containing the reverse reads of the DNA sequences. These files were processed using the OBItools program (available at http://metabarcoding.org/obitools). *Illuminapairedend* was used to merge the two single strand files to one double strand file. After filtering using *ngsfilter* to assign each sequence to the corresponding tag/primer combination, unaligned sequences were removed using *obigrep*. To make groups of identical sequences and to count their number of appearance *obiuniq* was used. *obiclean* was used to remove potential PCR and sequencing errors. Taxon assignation was performed using *ecotag.* Here, each sequence was assigned to a taxon by comparing it to a taxonomic reference library prepared for diatoms by Laura Epp (AWI Potsdam) of the EMBL Release 117 (September 2013).

The resulting table contained the sequences and the matched taxonomic level as well as probability indices. All sequences with less than 96 % probability of matching the taxonomic level of the reference database were removed from the data set. All sequences found in the extraction blanks or NTCs were subtracted from the samples. Sequences shorter than 76 bp were also deleted from the samples due to an expected sequence of 76 bp. Sequences with a count smaller than one thousand were neglected. Since this study focuses on diatoms in sediment samples the non-diatom sequences were filtered out and neglected. The ecotag-file of the bioinformatics analysis was reduced to sequences identified at least to genus level. All sequences resulting in higher taxonomic levels than genus were neglected for further analyses, in order to have comparable taxonomic levels between genetic and morphological assessments. The resulting data set was split up into the surface samples, the data of core 11-CH-06D, and the data of core 11-CH-12A. For each data set the percentage of all species and sequence types were calculated. The threshold for types being kept in the data sets was set on a minimum presence of 0.5% in at least three samples, in order to use only reasonably abundant types. The resulting tables are presented in the appendix (Appendix 25 - Appendix 31).

3.5 Morphological diatom analyses

Calcareous and organic components of the samples were removed by heating with HCl (10 %) and H_2O_2 (30 %) (Pestryakova et al. 2012). The samples of both cores were prepared by Luidmila Pestryakova at the North-Eastern Federal University of Yakutsk. For sample preparation of the core material 11-CH-06D and core 11-CH-12A the mean used wet weight was 0.3 g.

To prepare mounts, about ten drops of diatom suspension were dropped carefully with a glass pipette on to a clean coverslip. Coverslips were cleaned with ethanol (96%) before use. The water has to evaporate at room temperature, before further usage. When dry, the coverslips were mounted using Naphrax. One drop of Naphrax was placed on an ethanol-cleaned glass slide and the coverslip with the dried diatoms over the drop. To drive the Toluol out of the Naphrax, the prepared slides were heated for 20 min at 130°C on a hotplate (Battarbee et al. 2001). After letting the slides cool, the

coverslips were checked to be firm. Afterwards, the slides were ready to use. Counts and identification of the diatom valves were carried out at 1000x magnification using an Axio Scope.A1 (Carl Zeiss, Oberkochen, Germany). Diatom determination followed Pestryakova et al. (2012). During this project I counted the samples of core 11-CH-06D, in the three large morphological categories centric, raphid and araphid under the supervision of Luidmila Pestryakova (Results are presented in Appendix 22). The number of counts is generally 500 valves per sample. In order to get the necessary resolution of the diatom compositions of both cores and the surface data, samples were counted in greater taxonomic detail by Luidmila Pestryakova at the North-Eastern Federal University of Yakutsk and her results were kindly at my disposal. For the preparation of diatom material for the scanning electron microscope (SEM), a small drop of the prepared suspension (see description above) can be evaporated directly on to an ethanol-cleaned coverslip. After the water evaporated at room temperature, the coverslip can be placed on the stub using a tag. Finally, the sample is coated with gold/palladium (Au/Pd) alloy. SEM stubs were prepared only for core 11-CH-06D. To get a more detailed insight of the assemblage inside each sample selected diatoms were investigated. In the four investigated samples an average of each 28 diatoms were photographed. Analysis was done at the GFZ Potsdam together with Ilona Schäpan (GFZ Potsdam) and Luidmila Pestryakova. Analysis using the SEM was only done for all samples of core 11-CH-06D.

For each morphological data set, the threshold for species being kept in the data set was set on 1.25 % presence in at least three samples, in order to use only abundant species for further analysis. The resulting tables are all presented in the appendix (Appendix 32 - Appendix 37).

3.6 Statistical Analysis

The statistical analyses were carried out using R version 3.1.2 (R Core Team, 2014). R package “vegan” (Oksanen et al. 2015) was used for Spearman’s Rank Correlation matrix, Detrended Correspondence Analysis (DCA), Principal Component Analysis (PCA), Non-metric Multidimensional Scaling (NMDS), Redundancy Analysis (RDA), and Rarefaction curves. For Cluster analysis R package “rioja” (Juggins, 2015) was used. For

t-tests R package “stats” (R Core Team, 2014) was used. The Stratigraphy was done using Tilia version (1.7.16; developed and distributed by Eric C. Grimm). CoralDRAW® Graphics Suit X6 version (16.4.0.1280) was used to modify and merge graphics from R and Tilia outputs. All taxonomic data were Hellinger transformed using the function decostand() and the environmental parameters, except pH, were transformed using log (x+1) to normalize the data. The Hellinger transformation computes the relative frequencies by site profiles (rows) and applies square root. It is useful before PCA and RDA (for details see Borcard et al., 2011). Spearman’s rank correlation is a non-parametric alternative to Pearson linear correlation, which uses rank converted measurements. This test was used to assess the association between variables. In this study it was used to identify groups of correlation variables in the environmental parameter data set. Spearman’s rank correlation was done using vegan’s function corr.test(). This analysis was done for each species and genus data of both, the sequence and the morphological data set.

A DCA was performed to check if the data distribution is linear or unimodal in order to determine the subsequent analysis tool and was applied using the decorana() command. It has the advantage of removing the arch effect from ordination and the length of the first axis refers to the homogeneity of the data. The DCA was done using vegan’s function decorana() to check if the data are linear or unimodal, in order to find the suitable further analysis. The results are displayed in the appendix (Appendix 1). PCA is a linear method of unconstrained ordination. It was used to uncover the intercorrelation of the data. This analysis was applied for a preliminary check of the data to show main compositional gradients of the data. The PCA was done using the function rda() and its outputs are shown in the appendix (Appendix 2 - Appendix 8). While the PCA is based on variance the NMDS uses distance measures among samples. Therefore, NMDS ordination plots project the relative position of the sample points. Also this analysis was applied to all data sets for a preliminary check. The function metaMDS() was used for the NMDS of the surface data of the genetic approach, and its outputs are shown in the appendix (Appendix 9 and Appendix 10). RDA is a constrained linear ordination method combining multiple regression with PCA. It was used to display the effect of the environmental variables on the species composition.

The RDA was done using the command rda() to visualize the influence of the environmental parameters on the taxonomic data.

To assess the sufficiency of the sample sizes to find the entire species diversity rarefaction were produced using rare.curves(). Rarefaction curves reaching a plateau indicate a sufficient sample size. The richness of the samples was estimated using rarefy(), which estimates the species richness in a given number of subsamples. The number of subsamples is defined as the minimum sample size in the used data set.

To check if the morphological and the genetic data set are similar procrustes() was applied to previous ordination values to check if the morphological data can be rotated to maximum similarity on the genetic data. This analysis was only done for the surface data, because the data sets of both cores were too small.

The constrained Cluster analysis using CONISS considers only stratigraphically adjacent clusters (Grimm, 1987). This method was used to identify diatom assemblage zones along the sediment cores. The Cluster analysis was done using chclust(). The reliability and significance of the identified assemblage zones was assessed by using the broken-stick model (function bstick()) (Bennett, 1996).

4 Results

4.1 Preliminary assessment using cloning and Sanger sequencing

From a total number of 78 picked clones from recent and ancient samples a total of 36 distinct sequence types were retrieved and a total of twelve genera and 15 species could be assigned (Table 3). The authenticity of sequences was checked with the translation into amino acids. Only authentic sequences were kept in the data set. No sequence types could be assigned in the NTC and blank.

Table 3: Retrieved diatom sequence types and their occurrence in each of the selected samples.

Sequence type	Best Blast Hit	Accession number	identity in %	11-CH-06D, 0-2 cm	11-CH-06D, 123-125 cm	13-TY-22, light taiga	13-TY-13, single tree tundra	13-TY-04, tundra	13-TY-08, tundra	Total number of clones per species
seq_check_1	*Achnanthidium saprophilum*	KM084941	96	0	0	0	0	**1**	0	1
seq_check_2	*Amphora sp.*	AM710511	98	0	0	0	0	**5**	0	5
seq_check_3	*Amphora sp.*	AM710511	100	0	0	0	0	**1**	0	1
seq_check_4	*Aulacoseira distans var. alpigena*	JQ003564	96	**1**	0	0	0	0	0	1
seq_check_5	*Aulacoseira distans var. alpigena*	JQ003564	94	**1**	0	0	0	0	0	1
seq_check_6	*Aulacoseira distans var. alpigena*	JQ003564	96	0	0	**1**	0	0	0	1
seq_check_7	*Aulacoseira subarctica*	AY569592	96	**3**	0	0	0	0	0	3
seq_check_8	*Aulacoseira valida*	AY569602	100	0	0	**1**	0	0	0	1
seq_check_9	*Ellerbeckia sp.*	KJ577892	100	0	**5**	0	0	0	0	5
seq_check_10	*Ellerbeckia sp.*	KJ577892	98	0	**1**	0	0	0	0	1
seq_check_11	*Entomoneis sp.*	JN162778	96	0	0	0	**1**	0	0	1
seq_check_12	*Neidium affine*	AM710508	98	0	0	**1**	0	0	0	1
seq_check_13	*Pinnularia rupestris*	AM710458	97	**1**	0	0	0	0	0	1
seq_check_14	*Planothidium frequentissimum*	KM084961	90	0	0	0	**1**	0	0	1
seq_check_15	*Sellaphora laevissima*	EF143293	100	0	0	0	**1**	0	0	1
seq_check_16	*Sellaphora laevissima*	EF143307	100	0	0	0	0	**1**	0	1
seq_check_17	*Sellaphora laevissima*	EF143263	100	0	0	0	0	**1**	0	1
seq_check_18	*Stauroneis constricta*	JN162824	94	**1**	0	0	0	0	0	1
seq_check_19	*Staurosira construens*	HQ912451	96	0	0	**1**	0	0	0	1
seq_check_20	*Staurosira construens*	HQ912451	95	0	0	**1**	0	0	0	1
seq_check_21	*Staurosira construens*	HQ912451	97	0	0	**1**	0	0	0	1
seq_check_22	*Staurosira construens*	HQ912451	97	0	0	**1**	0	0	0	1
seq_check_23	*Staurosira elliptica*	HQ828193	98	0	0	**3**	**1**	0	**1**	5
seq_check_24	*Staurosira elliptica*	HQ828193	97	0	0	**5**	0	0	0	5
seq_check_25	*Staurosira elliptica*	HQ828193	96	0	0	**2**	0	0	0	2
seq_check_26	*Staurosira elliptica*	HQ828193	96	0	0	**1**	0	**11**	0	12
seq_check_27	*Staurosira elliptica*	HQ828193	97	0	0	**1**	0	0	0	1
seq_check_28	*Staurosira elliptica*	HQ828193	98	**5**	0	**1**	0	0	0	6
seq_check_29	*Staurosira elliptica*	HQ828193	100	0	0	**1**	0	0	0	1
seq_check_30	*Staurosira elliptica*	HQ828193	96	0	0	0	**1**	0	0	1
seq_check_31	*Staurosira elliptica*	HQ828193	92	0	0	0	0	**1**	0	1
seq_check_32	*Staurosira elliptica*	HQ828193	97	0	0	0	**1**	0	**2**	3
seq_check_33	*Staurosira elliptica*	HQ828193	97	0	0	0	0	0	**1**	1
seq_check_34	*Staurosira elliptica*	HQ828193	97	**6**	0	0	0	0	0	6
seq_check_35	*Staurosira elliptica*	HQ828193	97	0	0	**1**	0	0	0	1
seq_check_36	*Staurosirella martyi*	HQ828192	91	0	0	0	0	0	**1**	1
Total number of clones per sample				18	6	22	6	21	5	78

4.2 Genetic and morphological assessment of modern lake sediments

4.2.1 Genetic assessment of modern lake sediments

4.2.1.1 Diversity assessment on assigned species and genus level of sequences

The surface data set based on the high-throughput sequencing, includes 15 species in ten genera with an overall of 2.914,148 sequence reads (=counts) for 121 unique sequence types. Despite the equimolar pooling of PCR products for multiplex sequencing the retrieval of sequences is not homogeneous across all samples and shows substantial variation. While the mean count for all samples is 91,067, a minimum count of 147 can be observed in sample 13-TY-25 and a maximum count of 480,895 in sample 13-TY-24. The genus *Staurosira* holds the maximum number of 71 diverse sequence types, splitting into 18 types assigned to *Staurosira construens* and 53 types assigned to *Staurosira elliptica*. The genus with the minimum number of sequence types is *Cymbopleura* consisting of one type that was assigned to the species *Cymbopleura naviculiformis*. The four most abundant species are *Staurosira elliptica, S. construens, Sellaphora* cf. *seminulum,* and *Pinnularia anglica*, leading into the three most abundant genera *Staurosira, Sellaphora* and *Pinnularia*.

The species sequence data shows that in tundra lakes centric diatoms, such as *Aulacoseira distans, A. subarctica* and *A. valida* are almost not present, but therefore rather abundant in single tree tundra, forest tundra and light taiga (Figure 6). *Urosolenia eriensis* was only found in a few lakes and with rather low abundances, most of which are found in lakes of the forest tundra zone. Araphid diatoms represented by *Staurosira construens* and *S. elliptica* are the most abundant species in all lakes with small differences. Only the count data of lake 13-TY-04 show no presence of *S. construens* but instead, only of *S. elliptica*. Raphid diatoms in general are slightly more abundant in taiga lakes than in the transition zone or in the tundra, while most *Sellaphora* cf. *seminulum* counts were retrieved in lakes of the forest tundra and light taiga. *Amphora pediculus* and *Lemnicola hungarica* counts were mostly retrieved in tundra and single tree tundra lakes. Also *Stauroneis anceps* and *S. phoenicenteron* have a stronger presence in tundra and treeless areas, than in the tree-covered area. *Pinnularia anglica* is rather abundant in lake 13-TY-25 and several light taiga lakes, also appearing in tundra lakes, but with a lower abundance. A similar pattern, but with

much lower counts, shows *Pinnularia substreptoraphe.* The graph shows that for most samples the sample size was insufficient to represent the entire diversity of the samples (Figure 9A). Only lakes 13-TY-09, 13-TY-16, 13-TY-24 and 13-TY-32 reached a sample size large enough. The species richness varies between five and fourteen species per sample.

Naturally, the genera gained from the sequencing represent the above-described species. Because several sequence types could only be identified to genus level, the genera plot show slight differences to the species plot (Appendix 11). The most obvious difference is the appearance of the raphid diatom genus *Planothidium*, which was not present in the species dataset. *Planothidium* is slightly more abundant in tundra lakes than in lakes of the other vegetation types. Higher abundance of *Amphora* and *Stauroneis* in tundra lakes are visible compared to lakes of the transition zones and light taiga. The rarefaction curve of the genera is displayed in Appendix 17A and shows that the richness varies between three and ten genera. The graph shows that for most genera the sample size was insufficient, only lakes 13-TY-04, 13-TY-05, 13-TY-09, 13-TY-16, 13-TY-23, 13-TY-24 and 13-TY-32 consist of a sample size representing the average number of genera possible to find in the samples.

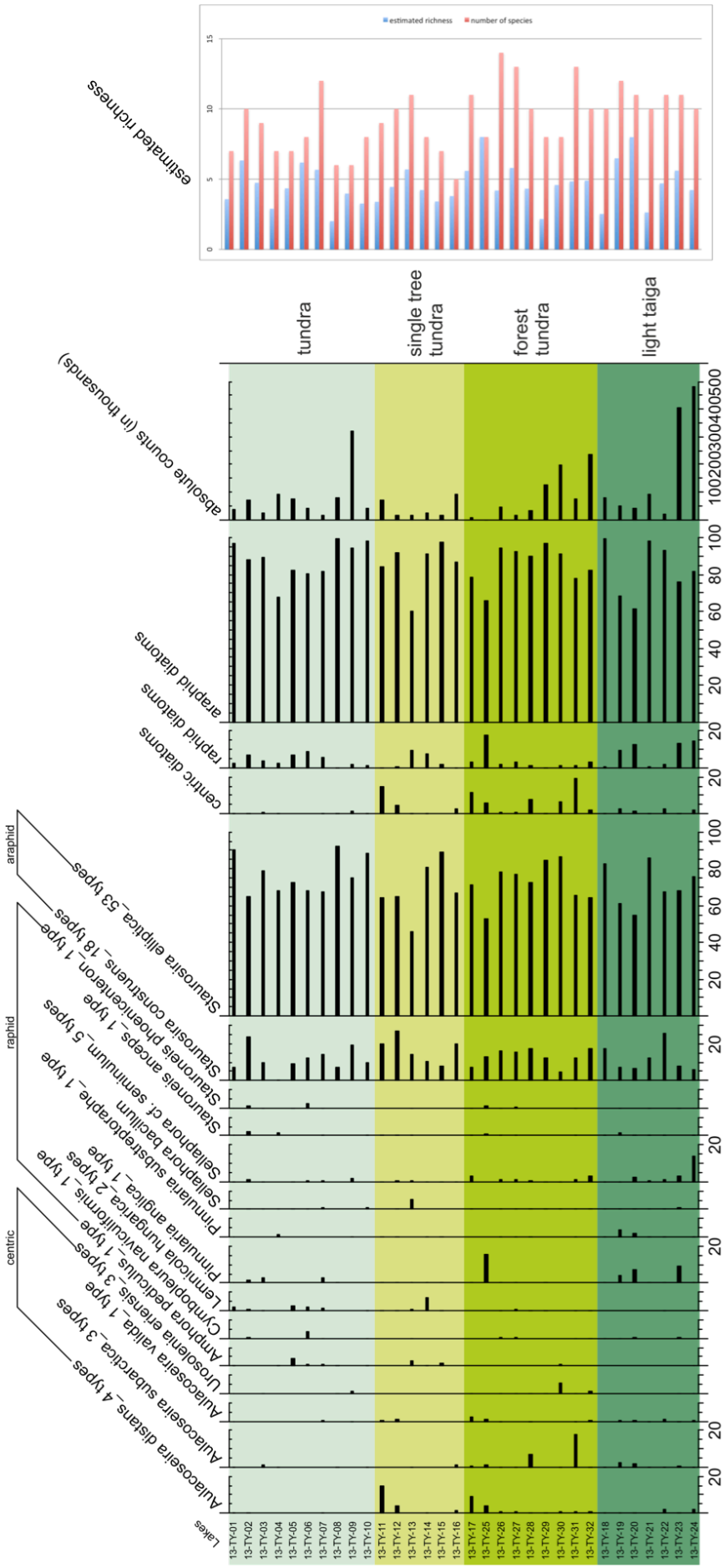


Figure 6: The diatom sequence types identified to species level of the lake transect from north to south, in the four vegetation zones tundra, single tree tundra, forest tundra and light taiga. The color code highlights the four vegetation types. For each sample the estimated richness and the number of species is given.

4.2.1.2 Data selection of environmental parameters for species sequence data

Based on the environmental parameters (Appendix 18) selected by Spearman's rank correlation, a Redundancy analysis (RDA) was run to identify the variables that explain a significant portion of variance in the diatom species dataset. The single and unique proportions were calculated and the four significant parameters were then used for further analysis on the sequence data. The four parameters that explain a significant portion of the variance are maximal depth (p=0.005), sulfate (p=0.05), calcium (p=0.02), and DOC (p=0.005). The results of the RDA are shown in Table 4. The same procedure was done for the genera sequence data as dependent variable. The parameters that significantly describe the distribution of the genera are maximal depth (p=0.005), sulfate (p=0.025), calcium (p=0.015), and DOC (p=0.005), which are the same four parameters as for the distribution of the species. The table with the overall results of the genera is shown in the appendix (Appendix 19).

Table 4: Proportions of chemical and morphological lake characteristics for the genetic data of the species, as well as the proportion of all significant variables, the significant variables under the condition of the vegetation types and the vegetation types under the condition of the significant variables. The unique proportions were only calculated fin case of significant single proportions. Forest.tundra has no p-value, because it was highly correlated with the other vegetation zones and thus redundant.

Table of proportion of chemical and morphological lake characteristics for species of sequence data

	Parameter	single proportion			unique proportion		
		percent	adjusted R^2	p-value	constrained percent	adjusted R^2	p-value
morphological and chemical parameters	max.depth	0.1027	0.0728	0.005	0.0754	0.0520551	0.01
	secchi.depth	0.1117	0.082096	0.005			
	pH	0.0426	0.01072	0.24			
	conductivity	0.0733	0.04237	0.025			
	Chlorid	0.0467	0.01497	0.17			
	Sulfat	0.0652	0.03407	0.05	0.02	-0.009238	0.15
	Nitrat	0.0068	-0.02632	0.945			
	HCO_3^-	0.0435	0.01158	0.225			
	Ca	0.0826	0.052	0.02	0.0291	0.0007794	0.42
	K	0.0443	0.01244	0.215			
	Mg	0.0303	-0.00206	0.47			
	Na	0.0515	0.0199	0.12			
	Si	0.0554	0.02392	0.085			
	DOC	0.1001	0.07015	0.005	0.0566	0.0312314	0.065

Parameter	single proportion			p-value		
	percent	adjusted R^2	tundra	single.tree	forest.tundra	light.taiga
vegetation type	0.1622	0.07246755	0.045	0.02	/	0.22
veg+cond(sign.variables)	0.0753	-0.0126603	0.53	0.545	/	0.555

Parameter	single proportion			p-value		
	percent	adjusted R^2	Sulfat	max.depth	Ca	DOC
all significant variables	0.2345	0.121079	0.21	0.005	0.315	0.07
sign.variables+cond(veg)	0.1475	0.03595064	0.485	0.045	0.77	0.11

4.2.1.3 RDAs of sequences on species and genus level

The first axis of the RDA with environmental parameters as constraining variables illustrated 12.5 % of the total variance of the species data set (Figure 8A). Axis 1 and 2 combined explained 19.8 % of the variance of the data set. *Aulacoseira* species and *Sellaphora seminulum* are represented in quadrant II. *Urosolenia eriensis* and *Staurosira construens* are represented in quadrant I. *Staurosira elliptica, Amphora pediculus, Lemnicola hungarica, Sellaphora bacillum, Cymbopleura naviculiformis* and *Stauroneis phoenicenteron* are represented in quadrant IV. *Pinnularia substreptoraphe, Stauroneis anceps* and *Pinnularia anglica* are represented in quadrant III. Axis one is assumed to reflect the calcium content of the lakes. However, most samples occur in the center of the plot.

The first axis of the RDA with the vegetation types as constraining variables illustrated 9.2 % of the total variance of the species data (Figure 8C). Axis 1 and 2 combined explained 14 % of the variance. All *Aulacoseira* species and *Urosolenia eriensis* are represented in quadrant III. *Staurosira construens* and *Sellaphora bacillum* are represented in quadrant IV. *Lemnicola hungarica, Staurosira elliptica* and *Stauroneis phoenicenteron* are represented in quadrant I. *Amphora pediculus* is positioned exactly in the middle between quadrants I and IV. The other species are represented in quadrant II. Axis one is assumed to reflect single tree tundra conditions. However, most samples occur in the center of the plot.

The first axis of the RDA with environmental parameters as the constraining variables illustrated 18.4 % of the total variance of the genera data set (Appendix 21A). Axis 1 and 2 combined explained 32.6 % of the variance. *Aulacoseira, Sellaphora* and *Pinnularia* are represented in quadrant II. *Staurosira, Planothidium* and *Lemnicola* are represented in quadrant IV. *Cymbopleura, Amphora* and *Stauroneis* are represented in quadrant III. *Urosolenia* is the only genus represented in quadrant I. Axis one is assumed to reflect the maximal depth. However, most samples occur in the center of the plot.

The first axis of the RDA with vegetation types as constraining variables illustrated 13.8 % of the variance of the genera data set (Appendix 21C). Axis 1 and 2 combined explained 18.2 % of the variance. *Aulacoseira* and *Urosolenia* are represented in quadrant III. *Lemnicola* and *Staurosira* are represented in quadrant IV. *Amphora,*

Planothidium, Cymbopleura and *Stauroneis* are represented in quadrant I. *Pinnularia* and *Sellaphora* are represented in quadrant II. Axis one is assumed to reflect tundra conditions. However, most samples occur in the center of the plot.

4.2.2 Morphological assessment of surface data

The morphological data can be described in two large groups, namely group I including tundra and single tree tundra and group II including forest tundra and light taiga. Also for this the groups for species and genera show similar patterns. The first group holds larger proportions of centric diatoms than the second group, also raphid diatoms are slightly more abundant in group I and *Staurosirella* holds a larger percentage than *Staurosira*, whereas for group II this pattern is contrarily.

4.2.2.1 Diversity assessment on species and genus level of morphological data

The surface data set based on morphological analysis includes 29 species in 19 genera with an overall of 13,276 diatom valves identified (=counts). The minimum count of 236 is found in sample 13-TY-28, the maximum count of 566 is found in sample 13-TY-12 and the mean count for all samples is 428. The genus *Aulacoseira* is the most diverse with six species. The four most dominant species are *Staurosira venter, Staurosirella pinnata, Pseudostaurosira pseudoconstruens* and *P. brevistriata*, leading to the three most abundant genera *Staurosira, Staurosirella* and *Pseudostaurosira*.

The species identified by light microscopy shows that centric diatoms, such as *Aulacoseira alpigena, A. distans, A. perglabra, Cyclotella bodanica, C. iris, C. radiosa,* and *C. tripartita*, are mostly present in tundra and single tree tundra lakes (Figure 7). Other species that are mostly present in lakes in treelees areas are *Amphora libyca, Diploneis elliptica, Eunotia praerupta, Karayevia laterostrata, Planothidium lanceolatum* and *Fragilaria capucina*. The abundance of *Staurosirella pinnata* increases in tundra lakes in comparison to lakes in forested areas, while for *Staurosira venter* the opposite is visible. *Pseudostaurosira pseudoconstruens* appears almost constant in all lakes, but with higher numbers in the vegetation types single tree tundra and forest tundra. *Eunotia faba* and *Rossithidium pusillum* are mostly found in lakes of the light taiga. The rarefaction curve in Figure 9B shows the species richness, which varies

between 11 and 22 species. The graph shows that for all samples the sample size was sufficient to represent the entire species diversity.

The genera show the same data of the species plot, because it was always possible to identify the diatoms to species level (Appendix 12). Only differences among genera are more easily visible than between species. In general, it can be seen that centric diatoms occur mostly in tundra and single tree tundra, with lake 13-TY-17 being a special case of the forest tundra vegetation and being characterized by a large abundance of *Aulacoseira*. This genus also appears in some lakes of the light taiga. For araphid diatoms the same trends as described above are visible. *Tabellaria* are present in almost all lakes, however, in lake 13-TY-04 they are the most abundant species while the usually dominant araphid species only occur in very low abundances. The richness varies between nine and seventeen genera. The plot corresponding to the genera of the morphological data shows that the sample size of all samples was sufficient to represent the entire genera diversity (Appendix 17B).

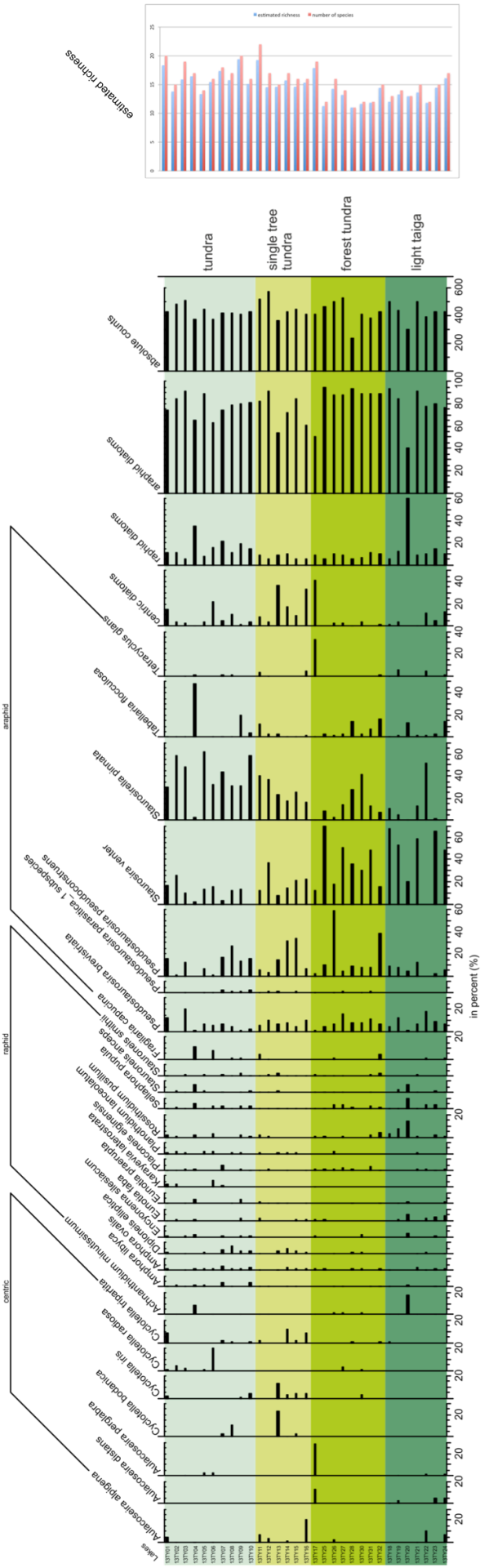


Figure 7: The diatoms identified by light microscopy of the lake transect from north to south, in the four vegetation zones tundra, single tree tundra, forest tundra and light taiga. The color code highlights the vegetation zones. For each sample the estimated richness and the number of species is given.

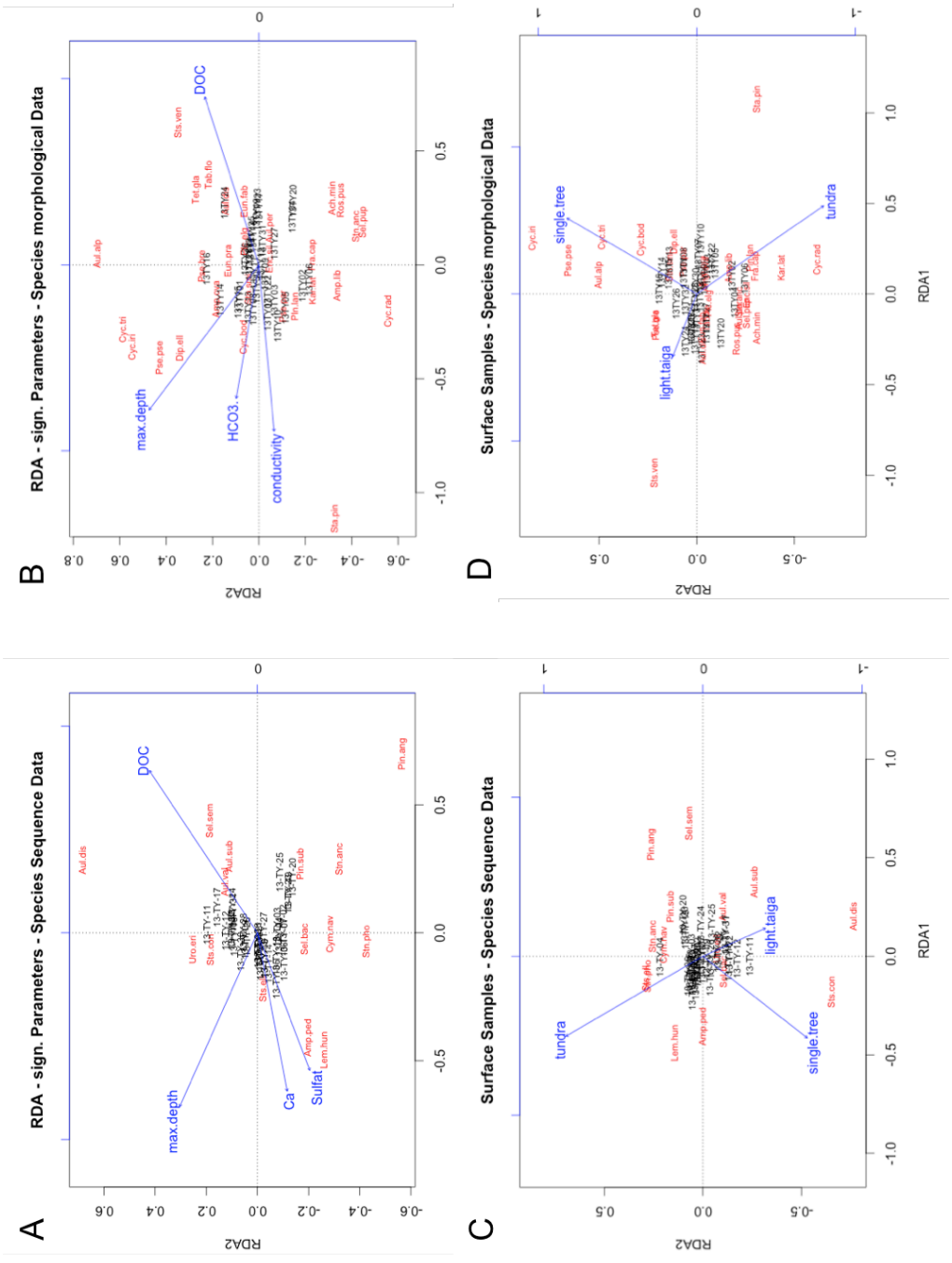


Figure 8: Plots of the redundancy analysis of the species assigned to sequences and the morphological identified species. A - RDA of significant environmental parameters on species assigned to sequences, B - RDA of significant environmental parameters on species identified by LM, C – RDA of vegetation types on the species assigned to sequences, D – RDA of vegetation types on the species identified by LM.

4.2.2.2 Data selection of environmental parameters of the morphological data set

In order to find the environmental parameters that best describe the morphological identified taxa, the same analysis as described in the genetic assessment was done. The four parameters that significantly describe the species data are maximal depth (p=0.005), conductivity (p=0.005), HCO_3^- (p=0.005), and DOC (p=0.005) (Table 5). The table with the parameters that describe a significant portion of the genera data are shown in Appendix 20. The results were similar to the results of the species data represented in Table 5, despite Si becoming very significant with a p=0.005. To keep the number of parameters that explain a significant portion of the variance equal and comparable for both taxonomic levels, Si was not included in the analysis of the genera.

Table 5: Proportions of chemical and morphological lake characteristics for the morphological data of the species, as well as the proportion of all significant variables, the significant variables under the condition of the vegetation types and the vegetation types under the condition of the significant variables. The unique proportions were only calculated fin case of significant single proportions. Forest.tundra has no p-value, because it was highly correlated with the other vegetation zones and thus redundant.

Table of proportion of chemical and morphological lake characteristics for species of morphological data

	Parameter	single proportion			unique proportion		
		percent	R2 adjusted	p-value	constrained percent	R2 adjusted	p-value
morphological and chemical parameters	max.depth	0.135073	0.105211	0.005	0.02294588	-0.002806	0.575
	secchi.depth	0.1478451	0.118461	0.005			
	pH	0.0506989	0.017964	0.12			
	conductivity	0.1554252	0.126302	0.005	0.02304514	-0.002631	0.505
	Chlorid	0.080621	0.048918	0.025			
	Sulfat	0.0443038	0.011349	0.245			
	Nitrat	0.0628029	0.030486	0.065			
	HCO_3^-	0.114173	0.083627	0.005	0.04519929	0.036465	0.06
	Ca	0.1086336	0.077897	0.005	0.01312655	-0.020134	0.87
	K	0.0489961	0.016203	0.13			
	Mg	0.0719829	0.039983	0.045			
	Na	0.0516216	0.018919	0.125			
	Si	0.0729042	0.040935	0.02			
	DOC	0.1619522	0.133054	0.005	0.0308762	0.011193	0.28

Parameter	single proportion		p-value			
	percent	R2 adjusted	tundra	single.tree	forest.tundra	light.taiga
vegetation type	0.2046131	0.1162368	0.01	0.01	/	0.53
veg+cond(sign.variables)	0.09853986	0.02154166	0.095	0.19	/	0.88

Parameter	single proportion		p-value			
	percent	R2 adjusted	max.depth	conductivity	HCO_{3-}	DOC
all significant parameters	0.2891202	0.179754	0.005	0.005	0.28	0.1
sign.variables+cond(veg)	0.183047	0.08505898	0.02	0.665	0.405	0.005

4.2.2.3 RDAs of the morphological identified species and genera

The first axis of the RDA with environmental parameters as constrained variables illustrated 19.2 % of the total variance of the species data set (Figure 8B). Axis 1 and 2

combined explained 24.5 % of the variance. All *Cyclotella* species except *Cyclotella radiosa*, are represented in quadrant I, along with *Pseudostaurosira pseudoconstruens*, *Diploneis elliptica*, and *Stauroneis smithii*. *Cyclotella radiosa*, *Staurosirella pinnata*, and *Amphora libyca* are represented in quadrant IV. DOC explains the species represented in quadrant II, i.e*. Tabellaria flocculosa* and *Staurosira venter*. *Aulacoseira perglabra, Rossithidium pusillum, Fragilaria capucina,* and *Stauroneis anceps* are represented in quadrant III. Axis one is assumed to reflect trends from lakes with a high conductivity to lakes determined by DOC. However, most samples group in the center of the plot.

The first axis of the RDA with the vegetation types as constrained variables illustrated 15 % of the variance of the species data set (Figure 8D). Axis 1 and 2 combined explained 18.4 % of the variance. *Staurosira venter, Tetracyclus glans* and *Pseudostaurosira brevistriata* are represented in quadrant I. *Pseudostaurosira pseudoconstruens, Cyclotella iris, C. tripartita, C. bodanica* and *Diploneis elliptica* are represented in quadrant II. *Staurosirella pinnata, Cyclotella radiosa, Fragilaria capucina, Amphora libyca* and *A. ovalis* are represented in quadrant III. The other species are represented in quadrant IV. Axis one is assumed to reflect light taiga conditions. However, most samples occur in the center of the plot.

The first axis of the RDA with environmental parameters as constrained variables illustrated 24.3 % of the total variance of the genera data set (Appendix 21B). Axis 1 and 2 combined explained 28.4 % of the variance. *Staurosira, Aulacoseira* and *Tetracyclus* are represented in quadrant II. *Pseudostaurosira, Diploneis* and *Cyclotella* are represented in quadrant I. While such taxa as *Staurosirella, Pinnularia, Amphora* and *Karayevia,* are represented in quadrant IV. All other genera are represented in quadrant III. Axis one is assumed to reflect trends from lakes with a high conductivity to lakes determined by DOC. However, most samples group in the center of the plot.

The first axis of the vegetation RDA illustrated 18.2 % of the variance of the genera data set (Appendix 21D). Axis 1 and 2 combined explained 20.3 % of the variance. *Staurosira, Tetracyclus, Aulacoseira, Placoneis* and *Tabellaria* are represented in quadrant I and explained by light taiga. While some taxa, such as *Diploneis, Cyclotella* and *Pseudostaurosira* are represented in quadrant II. And *Staurosirella, Amphora, Stauroneis, Fragilaria, Planothidium* and *Karayevia* are represented in quadrant III. Axis

one is assumed to reflect trends from tree tundra to light taiga conditions. However, most samples group in the center of the plot.

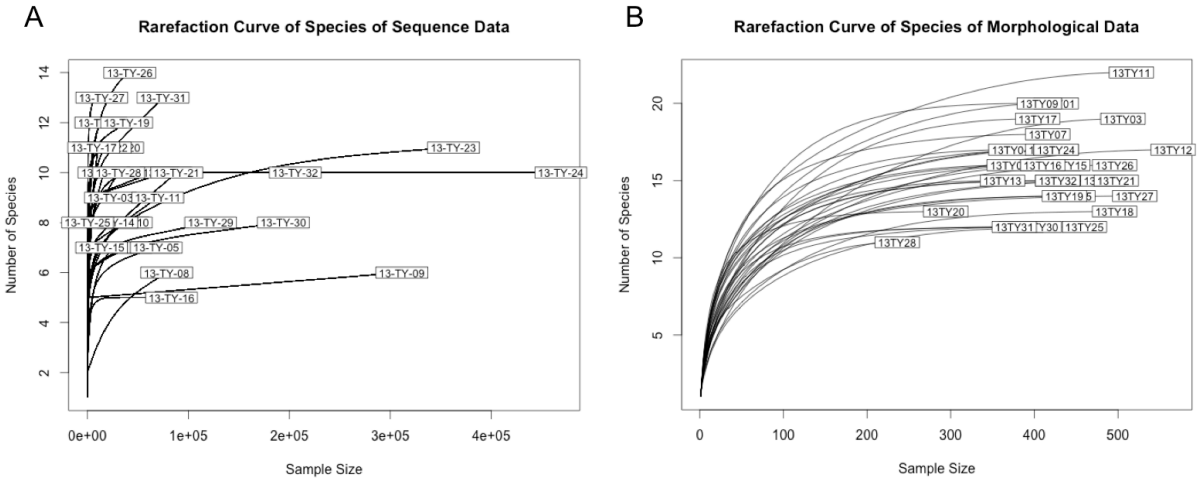


Figure 9: Rarefaction curves of both species data sets. A - Rarefaction curve of the genetic data, B - Rarefaction curve of the morphological data.

4.3. Comparison of the genetic and the morphological assessment

4.3.1 Comparison of the retrieved diversities

In general, there is a big discrepancy between the quantities of taxa signals, according to the used identification method. The genetic approach displays almost three million sequence counts, while the morphological identification reaches only a little more than 13,000 counts. Also the samples displaying the minimum and maximum counts in the datasets differ between both approaches. While 13-TY-25 displays the minimum quantity of counts in the genetic data, 13-TY-28 shows it for the morphological data. A maximum quantity of sequences can be found in 13-TY-24, while 13-TY-12 shows the maximum count in the morphological approach. In both data sets show that the most abundant genus is *Staurosira,* or in case of the morphological data the *Staurosira/Staurosirella* group. The estimated richness is based on the smallest sample size of each data set as number of randomized subsamples. Thus, the estimated richness and the observed number of species of the smallest sample are equal. Whereas, the estimated richness of all other samples is smaller than the observed number of species, because the estimated sample size is much smaller than the actual

sample size. The differences between the estimated richness of both methods are shown in Table 6.

4.3.2 Comparison of the correlating parameters and their influences on the data sets

Significant parameters contributing to the distribution of species in the sequence dataset are maximal depth, sulfate, calcium and DOC, whereas for the morphological data set the parameters explaining a significant portion of the variance were sulfate and calcium, conductivity and HCO_3^- (for details please refer to Table 4 and Table 5).

4.3.3 Comparison of both data sets using *procrustes*

The genera of the morphological data can be rotated on the genera of the genetic data to maximum similarity with a significance of 0.015. Whereas, the genera found with both approaches could be rotated to maximum similarity with a significance of 0.229. The species of the morphological data can be rotated on the species of the genetic data to maximum similarity with a significance of 0.001.

Table 6: Comparison of the counts, species number and estimated richness for the species and genera data of each the genetic and the morphological assessment of the lake surface data set.

Lake	Genetic assessment						Morphological assessment					
	species			genera			species			genera		
	count	Species number	estimated richness	count	Species number	estimated richness	count	Species number	estimated richness	count	Species number	estimated richness
13-TY-01	38169	7	3.577	38389	7	3.251	426	20	18.350	426	16	14.76037939
13-TY-02	64862	10	6.332	68215	9	5.735	482	15	13.778	482	15	13.77763001
13-TY-03	22684	9	4.740	24229	9	5.193	506	19	15.893	506	15	13.00722138
13-TY-04	65388	7	2.896	93091	6	4.968	372	17	16.443	372	15	14.46314184
13-TY-05	68144	7	4.346	76496	6	5.878	446	14	13.341	446	10	9.678739671
13-TY-06	35302	8	6.183	39576	8	6.987	370	16	15.444	370	13	12.72186452
13-TY-07	13296	12	5.671	15169	9	5.747	416	18	17.363	416	14	12.55678861
13-TY-08	78437	6	2.017	78979	7	1.771	416	17	15.765	416	13	11.87283513
13-TY-09	311457	6	3.978	318691	8	4.449	406	20	19.403	406	16	15.64334099
13-TY-10	38436	8	3.266	38765	8	3.139	428	17	15.114	428	12	10.55073093
13-TY-11	69623	9	3.390	70157	8	2.345	516	22	19.255	516	17	15.41521744
13-TY-12	16363	10	4.447	16829	8	3.088	566	17	14.542	566	14	12.07198794
13-TY-13	10759	11	5.695	15495	9	5.483	362	15	14.608	362	11	10.62400581
13-TY-14	24111	8	4.227	24373	7	4.282	426	17	15.725	426	13	12.12133113
13-TY-15	14033	7	3.414	14176	7	3.208	440	16	14.622	440	11	9.918412887
13-TY-16	83316	5	3.796	93119	3	2.080	410	16	15.329	410	13	12.39136787
13-TY-17	6250	11	5.593	6717	8	4.260	404	19	17.875	404	15	14.27317581
13-TY-18	77496	10	8.000	77604	7	6.000	496	13	11.249	496	11	9.667113764
13-TY-19	38818	12	4.191	48195	8	4.152	436	14	14.267	436	11	12.01507755
13-TY-20	29875	11	5.801	39425	8	6.533	294	13	13.191	294	11	10.55880218
13-TY-21	88542	10	4.326	89512	7	4.778	498	15	11.000	498	13	10
13-TY-22	17539	11	2.163	18033	8	3.331	384	12	11.634	384	11	8.983772077
13-TY-23	362530	11	4.585	403340	8	4.433	428	15	11.838	428	13	8.980599212
13-TY-24	468936	10	4.839	480895	7	3.754	426	17	14.436	426	13	12.51454341
13-TY-25	132	8	4.895	147	6	4.595	460	12	12.003	460	10	10.33179926
13-TY-26	41931	14	2.529	43194	10	1.753	496	16	13.281	496	13	10.86713493
13-TY-27	14173	13	6.486	14716	10	4.669	520	14	12.959	520	11	10.96157248
13-TY-28	30853	10	7.988	31365	9	5.722	236	11	13.619	236	10	11.76517773
13-TY-29	121997	8	2.634	125947	8	2.634	/	/	/	/	/	/
13-TY-30	194025	8	4.698	197413	8	3.534	406	12	11.820	406	9	10.8235378
13-TY-31	74826	13	5.606	76207	10	5.442	376	12	14.462	376	9	12.66374831
13-TY-32	205835	10	4.225	235689	10	3.342	428	15	16.099	428	13	12.11381071

4.4 Core data

4.4.1 Genetic assessment of the core data

4.4.1.1 Diversity assessment of the tundra core 11-CH-12A

Core 11-CH-12A includes eight species in eight genera with an overall of 2.649,872 counts. The minimum count of 40,790 is found in the sample of 125.5 cm depth, the maximum count of 640,064 in the sample of 111.5 cm depth and the mean count for all samples is 264,987. The genus *Staurosira* holds the maximum amount of 48 sequence types, splitting into 13 types of *Staurosira construens* and 35 sequence types of *Staurosira elliptica*. The genera with the minimum number of sequence types are *Asterionella* and *Nitzschia*. The four most abundant species are *Staurosira elliptica, Cyclotella bodanica, Staurosira construens* and *Aulacoseira subarctica*, leading to the three most abundant and most diverse genera *Staurosira, Cyclotella* and *Aulacoseira*.

The diatom assemblages on species level can be divided into three assemblage groups (AG) indicated by the cluster analysis, which was verified by the broken-stick model (Figure 10). AG I is ranging from 0.5 cm to 42.5 cm, covering an age range of approximately 0 to 2000 years B.P., in which no raphid diatoms occur. The samples hold very few centric diatoms with the exception of *Aulacoseira distans* occurring with approximately 20 % in sample 42.5 cm. Araphid diatoms have a mean occurrence of approximately 80 % in all samples. AG II is ranging from 55.5 cm to 69.5 cm, covering an age range of approximately 2000 to 3000 years B.P. In this assemblage group centric diatoms make approximately 73 % of each sample, while araphid diatoms have an average of approximately 18 % in all samples. AG II is different from the other two AGs and indicates a transition zone. AG III is ranging from 83.5 cm to 125.5 cm and is covering an age range of approximately 3000 to 6700 years B.P. Sample 111.5 cm consists completely of *Cyclotella bodanica*, while the other samples only include araphid diatoms.

All samples, except the ones taken in the depth of 123.5 cm, 111.5 cm, 55.5 cm and 0.5 cm depth have a sufficient sample size to represent the entire species diversity (Figure 14A). The species richness varies between three and seven species.

The genera give slightly different information about the core (Appendix 13). Here, the most prominent finding is the presence of raphid diatom genera, which were absent

from the species data set. This is due to the fact that not all sequences could be identified to species level.

Also this data set could be divided into three AGs, with the same ranges as for the species data. AG I now also holds raphid diatoms, mainly of genus *Pinnularia*. AG II is similar to the group described for the species except for the sample at 55.5 cm depth that now includes the genus *Gomphonema*. Also in AG III raphid diatoms occur with approximately 20 % in sample 83.5 cm and with approximately 65 % in sample 125.5 cm. All samples despite sample 125.5 cm have a sample size large enough to represent the entire genera diversity. The genera richness varies between four and eight genera (Appendix 23A).

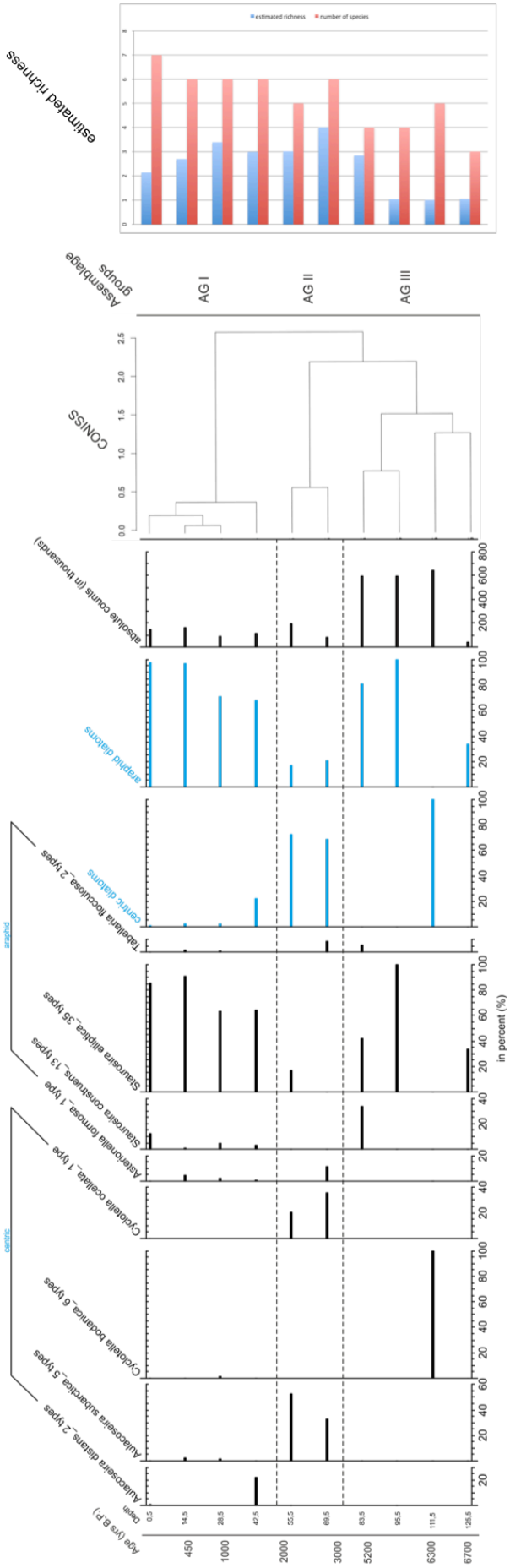


Figure 10: The diatom species of sequence types of core 11-CH-12A and the assemblage groups identified by CONISS analysis, as well as the estimated richness and the identified number of species.

4.4.1.2 Diversity assessment of the light taiga core 11-CH-06D

The sequence data set consists of nine species in seven genera, in a total of 1.881,377 counts. The minimum count of 25 was in sample depth 123.5 cm and the maximum count of 552,170 in 72.5 cm sample depth and the mean number of counts is 171,034. The most diverse genus with 46 sequence types is *Staurosira*, splitting into 10 types of *Staurosira construens* and 36 types of *Staurosira elliptica*. The genus with the minimum number of sequence types is *Navicula* with only one. The four most abundant species are *Staurosira construens, S. elliptica, Navicula radiosa* and *Tabellaria flocculosa*, leading into the three most abundant genera *Staurosira, Navicula* and *Tabellaria*.

The diatom assemblage groups on species level of the core can be divided into two verified groups (Figure 11). AG I is ranging from 0.5 – 123.5 cm and covering an age range of approximately 0 to 8400 years B.P. Araphid diatoms dominate all samples with a mean of approximately 72 %. Centric diatoms occur in percentages between 5-12 % in samples 0.5 cm, 16.5 cm, 28.5 cm and 58.5 cm. Raphid diatoms occur with almost 10 % in sample 123.5 cm and approximately 5 % in samples 29 cm and 59 cm. AG II is ranging from 140 cm to 155.5 cm and covering and age range of approximately 8400 to 9300 years B.P. Only the samples in the depths of 58.5 cm, 93.5 cm, 123.5 cm, and 155.5 cm have an insufficient sample size to represent the entire species diversity (Figure 14C). The species richness varies between two and seven species.

On genus level the ranges of both AGs are similar to the species data (Appendix 15). However, the two major differences can be found: first, in AG I raphid diatoms occur with a mean of approximately 30 % in all samples, and second the sample at 93 cm depth is composed exclusively of 80 % *Pinnularia* and 20 % *Stauroneis*. Still, for all other samples araphid diatoms are most abundant taxa and the appearance of centric diatoms is similar to the data on species level. AG II ranges from 140 cm to 155.5 cm depth. Sample 140 cm consists to 100 % of raphid diatoms, while sample 155.5 cm holds approximately 20 % raphid diatoms, 65 % araphid diatoms and 15 % centric diatoms. Only the samples in the depths of 40.5 cm, 123.5 cm, 155.5 cm, 107.5 cm and 58.5 cm have an insufficient sample size to represent the entire genera diversity (Appendix 23C). The genus richness varies between four and seven genera.

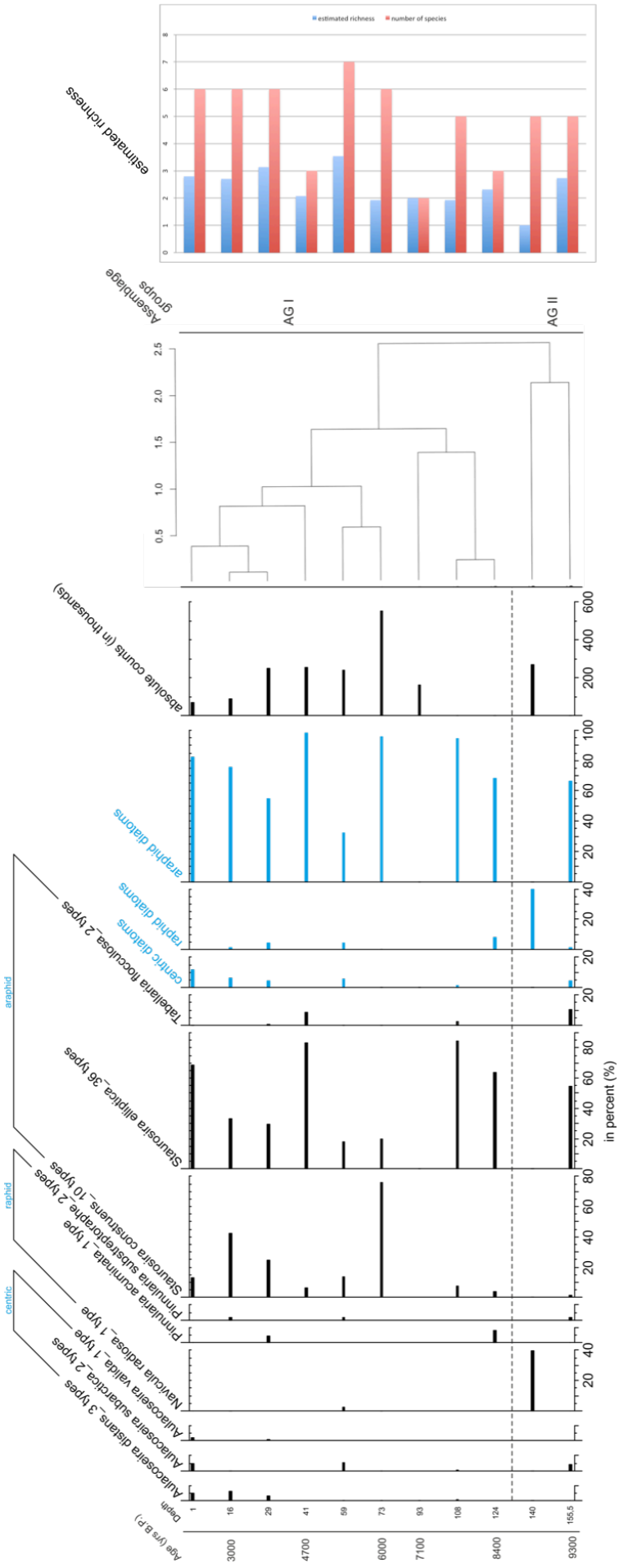


Figure 11: The diatom species of sequence types of core 11-CH-06D and the assemblage groups identified by CONISS analysis, as well as the estimated richness and the identified number of species.

4.4.2 Morphological assessment of sediment cores

4.4.2.1 Diversity assessment of the tundra core 11-CH-12A

The morphological data set of core 11-CH-12A includes 15 species in 10 genera, with an overall of 3,128 counts. The minimum count of 74 was in sample depth 95.5 cm, the maximum count of 460 in sample depth 28.5 cm and the mean number of counts is 312. The most abundant species are *Staurosirella pinnata, Cyclotella iris, Staurosira venter* and *Pseudostaurosira brevistriata*. Thus, the most dominant genera are *Staurosirella, Staurosira, Pseudostaurosira* and *Cyclotella*.

On the species level, two assemblage groups could be distinguished with AG I ranging from 0.5 cm to 42.5 cm, with an age range of approximately 0 to 2000 years B.P. and AG II from 55.5 cm to 124.5 cm, with an age range of approximately 2000 to 6700 years B.P. (Figure 12). In AG I araphid diatoms dominate all samples with an average of approximately 74 %. Centric diatoms occur with approximately 40 % in sample 0.5 cm and 30 % in sample 42.5 cm. Raphid diatoms hold an average of approximately 13 % in all samples. In AG II araphid diatoms occur with a mean of 60 % in all samples. Centric diatoms are missing in sample 83.5 cm and 124.5 cm, for the other samples the mean occurrence is approximately 53 %. Raphid diatoms occur with approximately 10 % in sample 83.5 cm and 17% in sample 124.5 cm. In sample 95.5 cm there are no raphid diatoms and in the others they appear with less than 5 %. Although, the depth from 55.5 cm to 69.5 cm could not be verified as a third AG, it is indicating a transition zone. The sample sizes were large enough for all samples despite of the sample in the depth of 95.5 cm to represent the entire species diversity (Figure 14B). The species richness varies between seven and twelve species.

On the genus level, however, the AGs could be distinguished with AG I, which is similar to AG I of the species data ranging from 0.5 cm to 42.5 cm, AG II ranging from 55.5 cm to 69.5 cm (2000 to 3000 years B.P.), which is composed of centric diatoms that occur in both samples with almost 60 %, raphid diatoms with less than 5 % and araphid diatoms with almost 40 %, and AG III ranging from 83.5 cm to 124.5 cm (5200 to 6700 years B.P.), with araphid diatoms occurring with a mean of approximately 70 % and raphid diatoms do not appear in sample 95.5 cm, but in the others with an mean of 10 % (Appendix 14). Centric diatoms hold approximately 15 % in sample 95.5 cm and

approximately 75 % in sample 111.5 cm. AG II indicates a transition zone from AG III to AG I. For all samples the sample size was sufficient, despite for the one in the depth of 95.5 cm (Appendix 23B). The genus richness varies between six and nine genera.

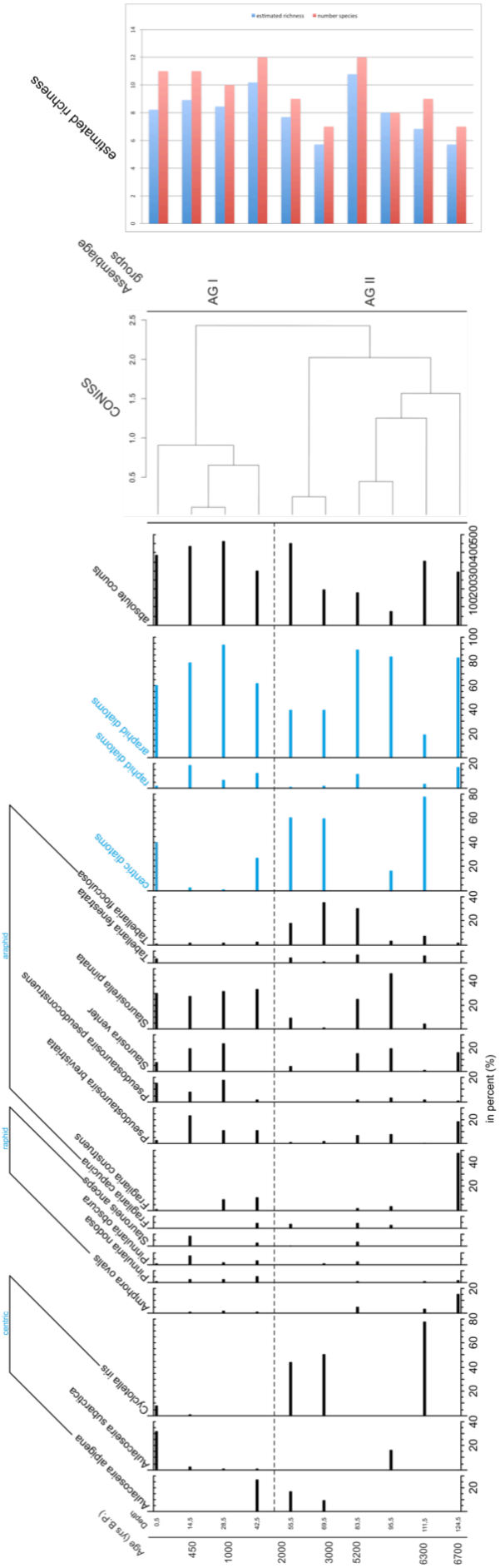


Figure 12: The species identified by light microscopy of core 11-CH-12A and the assemblage groups identified by CONISS analysis, as well as the estimated richness and the number of identified species.

4.4.2.2 Diversity assessment of the light taiga core 11-CH-06D

The data set of core 11-CH-06D includes 20 species in 14 genera. The total count is 2,196, with the minimum count of 14 in sample 123.5 cm, the maximum count of 406 in sample 40.5 cm and a mean count of 199 for all samples. The four most abundant species are *Aulacoseira distans, Staurosirella pinnata, Tabellaria flocculosa* and *Staurosira venter*. Thus, the most abundant genera are *Aulacoseira, Staurosirella, Staurosira* and *Tabellaria*.

On the species level, two assemblage groups could be distinguished with AG I ranging from 0.5 cm to 72.5 cm, with an age range of approximately 0 to 6000 years B.P. and AG II ranging from 93.5 cm to 155.5 cm, with an age range of approximately 7100 to 9300 years B.P. (Figure 13). In AG I centric diatoms occur with a mean of approximately 38 % in all samples. Except for sample 58.5 cm, araphid diatoms occur with a mean of approximately 51 %. Raphid diatoms appear with a mean of approximately 17 % in all samples. AG II includes no centric diatoms, raphid diatoms dominate sample 107.5 cm with 90 % and sample 123.5 cm with 100 %, while araphid diatoms dominate the other samples. The differences between the two AGs indicate a transition zone between 6000 and 7100 years B.P. The sample sizes were large enough for all samples despite the samples in the depth of 93.5 cm and 123.5 cm to represent the entire species diversity (Figure 14D). The species richness varies between three and seventeen species.

On the genus level, AG I is ranging from 0.5 cm to 93.5 cm (0 to 7100 years B.P.), with araphid diatoms dominating with a mean of approximately 53 %, except sample 58.5 cm with 80 % of centric diatoms. The mean of centric diatoms in other samples is approximately 26 % and raphid diatoms occur with a mean of approximately 19 %. AG II is ranging from 107.5 cm to 155.5 cm, with an age range of approximately 7800 to 8400 years B.P. It includes almost no centric diatoms with only less than 5 % in sample 107.5 cm. Raphid diatoms occur with an mean of approximately 71 %, dominating in samples 107.5 cm and 123.5 cm. Araphid diatoms are not present in sample 123.5 cm and have a mean of approximately 55 % in samples 139.5 cm and 155.5 cm (Appendix 16). The transition zone indicated by the genus level occurring around 7100 years B.P. differs slightly from the transition zone indicated by the species. For all samples the

sample size was sufficient, despite for the ones in the depth of 93.5 cm and 123.5 cm (Appendix 23D). The genus richness varies between two and thirteen genera.

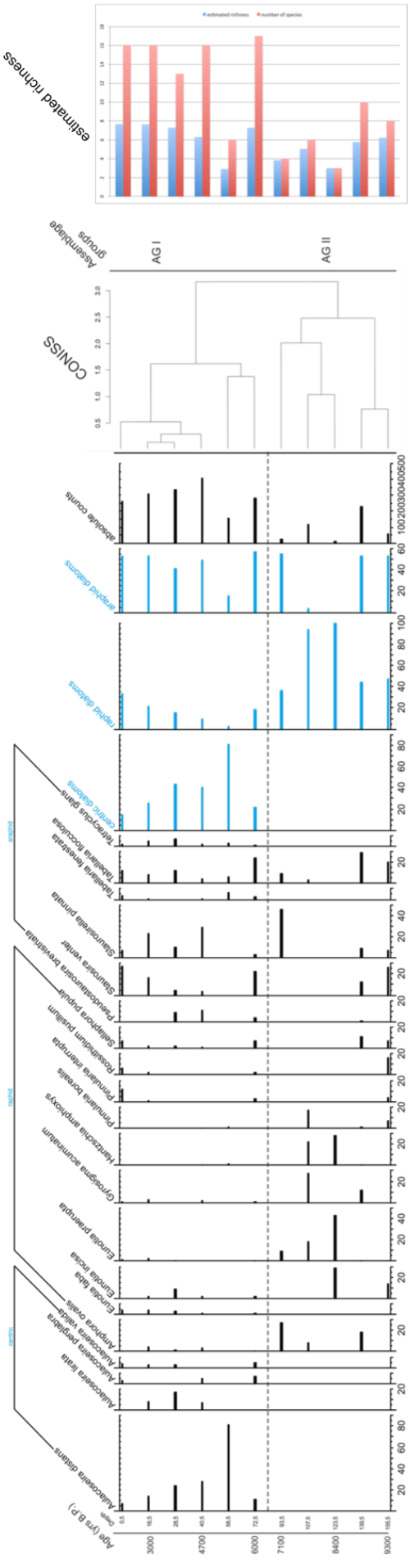


Figure 13: Diatom species identfied by light microscopy of core 11-CH-06D and the assemblage groups identified by CONISS analysis, as well as the estimated richness and the number of identified species.

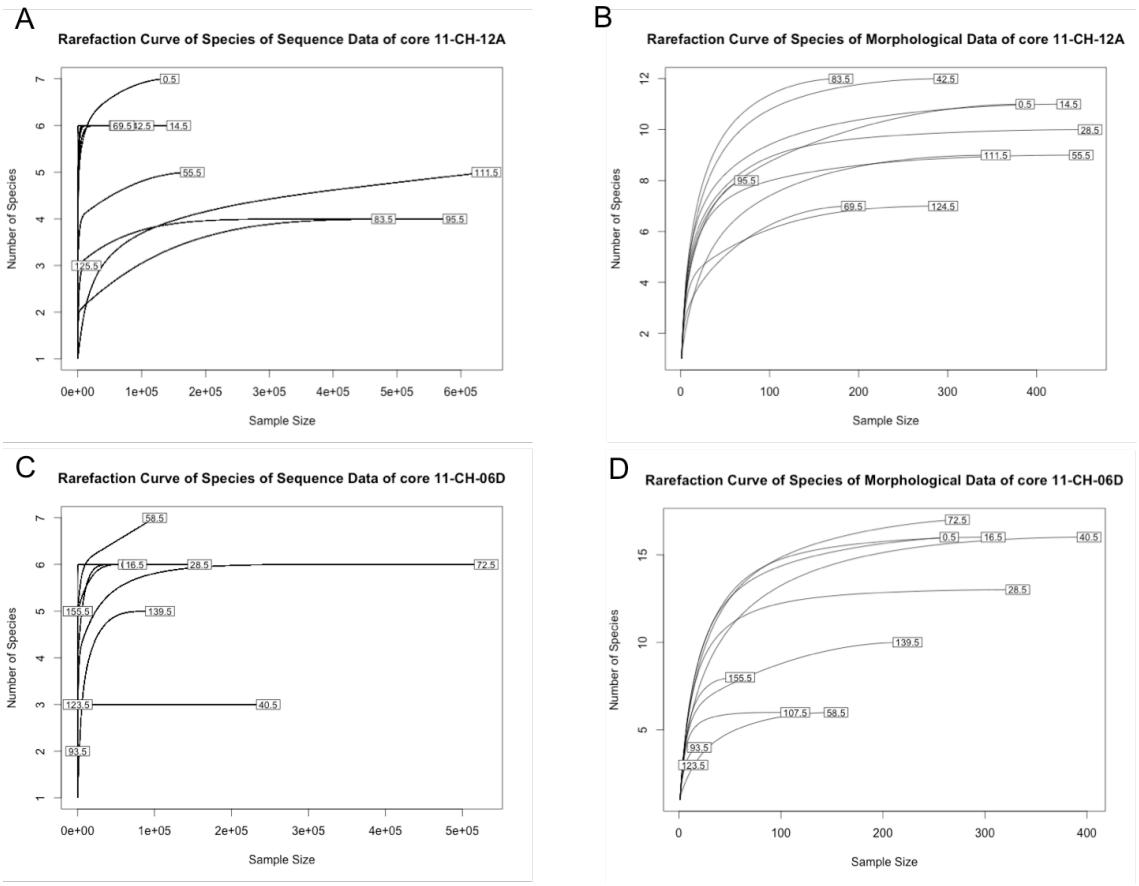


Figure 14: Rarefaction curves of the species data. A - Rarefaction curve of the genetic data of core 11-CH-12A, B - Rarefaction curve of the morphological data of core 11-CH-12A, C - Rarefaction curve of the genetic data of core 11-CH-06D, D - Rarefaction curve of the morphological data of core 11-CH-06D.

4.5 Comparison of the diversity assessment of sediment cores

4.5.1 Comparison between genetic and morphological data of core 11-CH-12A

Using the morphological approach 15 species in 10 genera were identified, which is more than retrieved using the genetic approach with only eight species and genera. Of course the total amount of counts varies strongly with over 2.6 million sequence counts and 3,128 counts of valves using light microscopy. The samples holding the minimum counts vary as well, with sample 125.5 cm in the genetic approach and sample 95.5 cm in the morphological approach. The same for the samples with the maximum counts, in the genetic approach sample 111.5 cm holds the maximum and in the morphological approach sample 28.5 cm (Table 7). Similar are the most abundant genera with *Staurosira* and *Cyclotella* in the genetic and the *Staurosira/Staurosirella* group and *Cyclotella* in the morphological approach. The major difference can be

found with the third most abundant genus, which is *Aulacoseira* in the genetic approach and *Pseudostaurosira* in the morphological approach.

While for the morphological data raphid diatoms can be identified to the species level, this was not the case for the sequence data. The indicated transition zones are similar in the taxonomic levels, despite in the morphological assessment of core 11-CH-06D. Additionally, the indicated transition zones of the genetic analysis are different from the morphological assessment for both sediment cores. Furthermore, in the genetic assessment the sample sizes of four samples were insufficient to represent the entire species diversity, but in the morphological assessment all samples, except one, had a sufficient sample size. In both approaches only one sample had a sample size too small to represent the entire genera diversity.

Table 7: The counts, species number and estimated richness of species and genera for all samples for both genetic and morphological assessment.

11-CH-12 A

sample (depth)	Genetic assessment						sample (depth)	Morphological assessment					
	species			genera				species			genera		
	count	Species number	estimated richness	count	Species number	estimated richness		count	Species number	estimated richness	count	Species number	estimated richness
0.5	143774	7	2.142	146832	7	6.469	0.5	387	12	8.217	387	8	6.886
14.5	157917	6	2.694	158894	7	6.997	14.5	436	11	8.917	436	9	7.073
28.5	68258	6	3.387	93019	8	8.000	28.5	460	10	8.448	460	8	6.782
42.5	100277	6	2.983	110906	6	6.000	42.5	298	12	10.184	298	8	7.159
55.5	179145	5	3.009	199039	6	5.830	55.5	450	9	7.688	450	8	6.713
69.5	69846	6	4.003	78230	7	6.994	69.5	194	7	5.712	194	6	5.094
83.5	478966	4	2.839	591238	6	5.154	83.5	180	12	10.778	180	8	7.904
95.5	590832	4	1.046	590860	4	2.632	95.5	74	8	8.000	74	6	6.000
111.5	640033	5	1.004	640064	6	3.586	111.5	354	9	6.832	354	7	5.616
125.5	13609	3	1.055	40790	4	4.000	124.5	295	7	5.708	295	6	5.268

4.5.2 Comparison between genetic and morphological data of core 11-CH-06D

The genetic approach led to nine species in seven genera, compared to 20 species in 14 genera with the morphological approach. Again the total counts varied extremely with over 1.8 million counts of sequences and 2,196 counts with light microscopy. However, in both approaches the sample with the minimum number of counts was sample 123.5 cm, respective 124 cm (Table 8). The sample with the maximum number of counts varies between both approaches. *Navicula* is an abundant genus in the genetic assessment, while in the morphological assessment *Aulacoseira* is more abundant. *Tabellaria* and *Staurosira* and accordingly *Tabellaria* and the *Staurosira/Staurosirella* group are abundant genera in both assessments.

The sample size of four samples in the genetic approach is insufficient, while for the morphological approach only two sample sizes are too small. Also the sample sizes of five genetic samples have an insufficient sample size, but only two of the morphological samples.

Table 8: The counts, species number and estimated richness for species and genera for all samples both for the genetic and the morphological assessment.

11-CH-06D

sample (depth)	Genetic assessment						sample (depth)	Morphological assessment					
	species			genera				species			genera		
	count	Species number	estimated richness	count	Species number	estimated richness		count	Species number	estimated richness	count	Species number	estimated richness
0.5	64975	6	2.798	69290	6	2.650	0.5	265	16	7.660	265	12	6.666
16.5	74212	6	2.704	88700	6	4.044	16.5	308	16	7.622	308	11	6.348
28.5	158336	6	3.136	249188	6	4.034	28.5	332	13	7.279	332	9	5.705
40.5	247658	3	2.073	252459	4	2.274	40.5	402	16	6.302	406	12	5.039
58.5	100236	7	3.538	240863	7	4.066	58.5	154	6	2.920	156	5	2.526
72.5	530928	6	1.926	552170	7	1.871	72.5	273	17	7.278	283	13	6.171
93.5	10	2	2.000	159080	4	1.996	93.5	20	4	3.842	22	5	4.636
107.5	180	5	1.927	188	6	2.609	107.5	114	6	5.034	118	8	5.440
123.5	19	3	2.316	25	4	4.000	123.5	14	3	3.000	14	2	2.000
139.5	106990	6	1.006	269346	6	3.223	139.5	224	10	5.768	232	11	6.095
155.5	49	5	2.733	68	5	4.797	155.5	60	8	6.221	60	7	5.953

5 Discussion

5.1 Comparison of the diatom richness and composition of the genetic and the morphological data of the surface sediments data set

The underlying question of this work is if genetic analysis with DNA metabarcoding and Next Generation Sequencing can give comparable results to the classical approach of morphological identification by light microscopy. In order to find an answer the analysis of the similarity of the different data sets is crucial. Therefore, comparisons of diatom composition, richness and the similarity of the data were examined on species and genus levels for both the genetic and the morphological assessment. The findings of this study demonstrate that the richness of the genetic and the morphological surface data set are different. This is also reflected by the outcome of the rarefaction curves. Only four out of 32 samples of the genetic data reached a sufficient sample size to represent the entire species diversity and six out of 32 samples for the genera diversity. In comparison, sample sizes of the morphological data were all sufficient. These results are similar to those reported by Jahn et al. (2007). They addressed the fact that they found a similar number of hits being detected by both genetic and morphological approaches, but the comparison of species compositions per sample remained difficult. The retrieved richness is of importance for a comparison of two methods used in biodiversity analyses. Similar to the results of Stoof-Leichsenring et al. (2012) not all morphological identified species could be assigned with the genetic survey. Although they suggested this problem might be overcome by the use of a NGS approach.

The genetically assigned species *Staurosira construens* and *S. elliptica* and the morphologically identified species *Staurosira venter, Staurosirella pinnata, Pseudostaurosira pseudoconstruens* and *P. brevistriata* were the most dominant species throughout the surface data. This supports the common composition of diatom communities in Arctic lakes, which is dominated by small benthic fragilarioid taxa (Biskaborn et al., 2012; Pestryakova et al., 2012), and is also similar to diatom analyses from permafrost regions in North America and Northern Europe (Weckström et al., 1997; Laing & Smol, 2000; Rühland et al., 2003).

So far most studies on diatom diversities in Arctic thermokarst lakes used the morphological approach (e.g. Biskaborn et al., 2012; Pestryakova et al., 2012), while Stoof-Leichsenring et al. (2014) provided first metabarcoding results for freshwater diatoms in thermokarst lakes in this ecotone. The general division in species and genus data was due to the fact that some sequences could only be assigned to genera and not to species level. This is based on the quality and integrity of the reference database. Thus, the same procedure was done for the morphological data, but there all counted diatom valves could be identified to species or subspecies level. So the genera do not give any further information on diversity. The same taxonomic level for genetic and morphological analysis is thus difficult to define. The assignment of sequences to a taxonomic unit is always dependent on references and the reference description in the database. One example for confusing reference description is *Staurosira elliptica* (database entry) and the likely synonym *Staurosirella pinnata* as the morphological described species. Because of the unclear reference description and the lack of a proper morphological description of all *Staurosira/Staurosirella* species, a direct comparison between the genetic and the morphological approach can be difficult.

The diversities gained through metabarcoding studies are all reliant on the used marker and the primer specificity. So far no consent on an appropriate diatom barcode has been made, but several markers are tested and used accordingly (MacGillivary and Kaczmarska, 2011). Generally, there are four types of markers: derived from nuclear, ribosomal, mitochondrial and chloroplast genomes (e.g. 18S rRNA, cox1 and rbcL) (Moritz & Cicero, 2004; Evans et al., 2007; Medinger et al., 2010; Hamsher et al., 2011; Zimmermann et al., 2011; Epp et al., 2012; Kermarrec et al., 2013; Stoof-Leichsenring et al., 2014). The 18S region is less variable than cox1 and rbcL, while the ITS region (rRNA) is often used for within-population genetic variation, because of its high variability (Evans et al., 2007). Zimmermann et al. (2011) used the V4 subregion of the 18S rRNA gene as a barcoding marker on diatom cultures, while Evans et al. (2007) propose the use of the cox1 marker, reporting the intraspecific divergence of cox1 to be equal or greater than that of rbcL, despite the much shorter sequence length of cox1. For some species, such as *Sellaphora blackfordensis* cox1 shows more bp differences than rbcL (Evans et al., 2007). In this study a short and variable sequence of

76 bp of the rbcL gene (rbcL_76) was used. The underlain primer pair was designed and successfully tested for specificity by Stoof-Leichsenring et al. (2012). The primer is known to reliably amplify a variety of diatoms in different ecological habitats, e.g. freshwater lakes in Siberia (Stoof-Leichsenring et al., 2014) and Kenya (Stoof-Leichsenring et al., 2012). Due to the fact that rbcL as part of the chloroplast genome occurs in multiple copies per cell, the amplification probability, especially in older samples, increases (Stoof-Leichsenring et al. 2014), because the amount of the recovered rbcL DNA depends on both the number of genomes per chloroplasts per cell, which varies between diatom species (Stoof-Leichsenring et al., 2012). This fact is important for the analysis of highly degraded aDNA in environmental samples. Furthermore, because of its specificity it is unlikely to amplify non-specific products (Evans et al., 2007; Stoof-Leichsenring et al., 2014). Another supporting fact for the use of rbcL on environmental sedimentary samples is the good reference database (Kermarrec et al., 2013).

The rbcL marker used in this study has a great universality and can distinguish between all species (Hamsher et al., 2011; Kermarrec et al., 2013). For some genera the rbcL_76 marker revealed a large intra-specific diversity, e.g. in *Staurosira* (Stoof-Leichsenring et al., 2014), whereas in other genera the resolution of this marker is much weaker (e.g. *Sellaphora*, Evans et al. 2007). This could also be observed during this study, as assigned *Staurosira* sequences were most diverse, while other genera were assigned to much fewer sequences.

The genetic lineages that could be assigned showed different genetic diversities in distinct diatom genera. A larger diversity was retrieved from the genetic lineages than could be identified by morphological analysis. But in several cases the reference database and the used marker were insufficient in their resolution. Therefore, in some genera a variety of genetic lineages could be identified, but not assigned to a similar number of species, whereas several species could be identified by the morphological analysis, but were not assigned to any sequence. Generally, the similarity of both data sets was significant, supporting the conclusion that even though the sample sizes for genetic analysis need to be improved, the results are valuable. This offers the opportunity to use large metabarcoding approaches instead of the classic morphological method. The genetic analysis through metabarcoding makes it possible

to process a large number of samples with an identical treatment. On the other hand the morphological determination is for now by far cheaper. But to process a similar number of samples it is also more time consuming. The fast development in the field of sequencing technologies promises to become much cheaper in the forthcoming years. Therefore, the genetic assessment offers a reliable, quick and sensitive tool to analyze especially dominant species.

5.2 Genetic and morphological relation with environmental parameters of the surface data set

Diatoms are influenced by environmental parameters determining the lake-water chemistry. Thus, they are widely used as bioindicators (Charles, 1985; Pienitz et al., 1995; Weckström et al., 1997; Smol & Cumming, 2000; Schmidt et al., 2004) and to observe and validate climate change (Rühland et al., 2015). In this study four parameters were identified explaining a significant portion of the variance of the diatom assemblages. Within each of the data sets the four parameters were identical among species and genera, however they partially differed among the two methods. For the genetic data DOC, maximal depth, calcium and sulfate were identified, and for the morphological data DOC, maximal depth, conductivity and HCO_3^-. Those were also the factors with the largest variability across the lake transect. The vegetation type with the strongest influence was single tree tundra for both assessments, meaning that this vegetation type was the most significant in explaining a significant portion of the variance. The tundra vegetation type additionally explained the morphological data. Factors related to water chemistry and lake depth were identified by RDA to be the most important drivers affecting the diatom assemblages. In the genetic data set, DOC was found to account for the largest variation in diatom compositions. As for the morphological data set it was maximal depth. The tundra vegetation was identified to explain the largest variation in the diatom composition in the genetic and the morphological data sets.

The ions calcium and sulfate being significant variables for the genetic data set are also factors influencing the conductivity of lake-water, which is a significant variable for the morphological data set. Hence, the main differences between the explaining variables

of both data sets are in the explained portion of variance of ion concentrations. The identified significant environmental parameters influencing diatom communities in circumpolar treeline lakes (refer to Laing and Smol, 2000; Rühland et al., 2003) and alpine lakes (Schmidt et al., 2004) are similar to those identified in this study. Generally, gradients of e.g. DOC or conductivity are observed from tundra to taiga lakes across the treeline ecotone (Duff et al., 1999; Rühland, 2001; Herzschuh et al., 2013). Nonetheless, a significant difference between the lakes of the distinct vegetation zones (i.e. tundra, single tree tundra, forest tundra and light taiga) could neither be observed in the genetic nor in the morphological data. This may be due to the fact that most identified significant variables are ion concentrations. Due to the fact that the study area is located close to the Arctic Ocean, tundra lakes may be strongly influenced by ion input through ocean water vapor. This implements that in tundra areas close to the sea, conductivity measures might be similar to those of the southern forests, which is supported by the conducted measurements of conductivity in tundra lakes. The influence of ocean proximity on lakes is also mentioned by Laing and Smol (2000) and could explain the non-significant results of the performed analysis. Similar to the study of a large lake transect in Northern Yakutsk from Pestryakova et al. (2012) the water depth was identified as an important factor affecting diatom compositions. As Pestryakova et al. (2012) suggested this may explain the high levels of diversity and large amounts of bottom-dwelling, i.e. benthic diatoms. For these species water transparency to the lake floor is necessary for their photoautotrophic life and may not be available in lakes deeper than three meters (Pestryakova et al., 2012). Anyhow, the most abundant species in both assessments were benthic diatoms and they occurred even in very deep lakes. Furthermore, shallow lakes have a better nutrient supply, because they warm up easily and show no stratification (Pestryakova et al., 2012). Besides, the results presented in this thesis indicate slight differences of influencing factors between lakes of different vegetation zones.

The identified physico-chemical characteristics of lakes in this study that are most strongly related to the composition of diatoms in surface sediments can be used to produce predictive indices and equations for paleoreconstructions (Charles, 1985). A detailed explanation of the mathematical technique for the reconstruction of a given

environmental parameter by use of diatoms is given e.g. by Gasse (1987). Environmental factors used for Holocene reconstructions based on transfer-functions for the Arctic and Alpine regions are e.g. alkalinity (concentration of HCO_3^-)(Laing et al., 1999; Herzschuh et al., 2013), and DOC (Pienitz & Smol, 1993; Pienitz et al., 1999). Other studies from the southern Bolivian Altiplano and from lakes in Greenland developed diatom-conductivity models (Sylvestre et al., 2001; Ryves et al., 2002). Hence, a development of functions based on HCO_3^-, conductivity or DOC could be promising for the studied transect of this project.

5.3 Diatom composition of the temporal data sets

The aim of this thesis was also to analyze the temporal change of diatom communities of two sediment cores on different taxonomic levels and to compare the genetic and the morphological approach, with the main focus on the comparability of the methods. The major findings of the temporal analysis of the tundra core 11-CH-12A and the taiga core 11-CH-06D are that the turnover observed in core 11-CH-12A is very similar to the recently published description of Stoof-Leichsenring et al. (2015). The core could be divided into three assemblage groups (AGs) showing shifts in diatom communities. A similar trend could be observed for core 11-CH-06D. The indicated transition zone differs in the age range and depth between the two methodologies. Anyway, the transition zone seems to be in a depth not directly covered by the samples, thus a further and more detailed analysis of this part of the core would be recommended.

In general, the genetic and the morphological approach showed similar results, especially the analysis on genus level of core 11-CH-12A. Furthermore, the comparability of the results of the two applied methods, specifically for ancient samples could be proven. So far, most studies on diatom records used the classic morphological approach and observed, that fossil diatom assemblages of lakes in Arctic environments are mainly dominated by small benthic fragilarioid taxa (Tamsin E. Laing et al. 1999; K. Rühland et al. 2003; Biskaborn et al. 2012; Pestryakova et al. 2012). This general observation holds true for the data sets used in this thesis. The interspecific shifts of *Staurosira* sp. throughout core 11-CH-12A and those of *Aulacoseira* sp. throughout both cores indicate turnovers in diatom diversity and

possible changes in the environment. For core 11-CH-12A only two assemblage groups were proposed for the species identified with the morphological approach and three AGs for the genera. Generally, for the species a third group could be identified but not verified through the broken-stick analysis. Overall, the trends and assemblage groups described and analyzed are quite similar for both methodologies. For core 11-CH-06D the number of AGs is on both taxonomic levels and for both methods the same. But the groups differ between both approaches, which is due to the verification of AGs. With only 11 samples of 11-CH-06D it is not very convincing to assume too many AGs. Anyway, a third group could be proposed from 107.5 cm to 123.5 cm of both taxonomic levels of the genetic assessment, indicating a transition zone from the described and very distinct AG II to AG I. This possible transition zone is also indicated through cluster analysis. Both assessments indicated a major shift between 7100 and 6000 years B.P., which may be due to significant changes in the environment. A step further would be to use the conducted core data together with the implications from the surface data to develop transfer-functions. For example the relative high abundance of *Aulacoseira* sp. in core 11-CH-06D suggests that it has been mostly located in more or less forested areas, because of the similarity between the overall core diversity and the diatom diversity found in forested areas of the surface transect. The use of transfer-functions would also help to explain the turnovers indicated in the cores and could be used to infer climate changes in the Holocene similar to e.g. Laing et al. (1999), Korhola et al. (2000), Coolen et al. (2004), and Pestryakova et al. (2012).

The used marker proved to be very feasible for the retrieval of ancient DNA (aDNA) sequences, as also shown by Stoof-Leichsenring et al. (2012, 2014, 2015). The difficulties with environmental DNA and specifically aDNA are preservation, damage, retrieval and contamination risks during recovery and further analyses. A complete analysis is hindered through incomplete reference databases (Hofreiter et al., 2001; Anderson-Carpenter et al., 2011; Pedersen et al., 2015). The incomplete reference databases were a problem especially in the old samples of the sediment cores. The recent diversities are not necessarily similar to those 7000 years B.P. and morphologically found frustules are often very difficult to identify and determine, because they are not described yet (Pestryakova, personal communication). Thus, the

lack of references in the database for sequences retrieved from old samples is a general problem of the rbcL marker.

A limitation to the results of this thesis might be due to stringency of the applied thresholds, whereas species dominating only one sample but being absent or present in very little numbers in the other samples were discarded from the analysis, in order to reduce the data set and to run the analyses on abundant taxa. This was the case for both assigned sequence counts of the genetic approach and valve counts of the morphological approach. One such case is *Ellerbeckia arenaria*, which dominates the sample 123.5 cm of the sediment core 11-CH-06D in the morphological data set, but due to the applied threshold it was deleted from the data used for the analyses. While in core 11-CH-06D the sample 123.5 cm of the morphological assessment was almost solely represented by *Ellerbeckia arenaria*, no sequences could be assigned to this species. The embl database release used for the analysis with OBItools (EMBL Release 117 September 2013) did not contain any reference sequences of *Ellerbeckia* sp., while the new release from June 2015 (EMBL Release 124) does contain references. The repeated analysis of the clone data revealed that in sample 123.5 cm *Ellerbeckia* sp. is present. Anyhow, a further analysis of the Illumina sequences of this sample did not reveal hidden *Ellerbeckia* sp. sequences. There are two approaches to this discrepancy between clone sequences and Illumina sequences, either practical or methodological.

The first is based on the fact that *E. arenaria* has robust frustules and is typical in unstable environments (Schmid and Crawford, 2001). It is known to grow in great quantities in subaerial habitats, like mosses (Schmid and Crawford, 2001). This ecological preference makes *E. arenaria* an indicator for shallow and unstable water bodies. The high abundance of this species in the sample depth of 123.5 cm implies drainage or the origin of the lake. Because of the subaerial habitat, *E. arenaria* individuals are more exposed to radiation, heat and other environmental factors that play a crucial role in DNA destruction (Hofreiter et al. 2001; Pedersen et al. 2015). Therefore, it is possible that the robust and large frustules are very well conserved in the lake archive, but the DNA might be too degraded to occur in sufficient amounts. Thus, the *E. arenaria* assigned sequences retrieved from the clones of the pre-check might be just a rare catch. The latter i.e. methodological approach is based on possible weaknesses of the experimental design. It is possible that the DNA extraction

procedure is insufficient and therefore only low amounts of the required DNA are retrieved. Because the DNA extracts of sedimentary DNA always consist of a mixed sample of diverse species, larger amounts of DNA of other species might always outreach the low DNA content. Possibilities to overcome these problems might be a change in the extraction protocol. So far the morphological analysis showed almost always intact frustules of *E. arenaria*. In general, sample 123.5 cm of core 11-CH-06D belongs to the samples with a comparably low DNA content.

This particular example shows the not-yet overcome weaknesses of metabarcoding analysis, relying on an existing reference database and the particular issues with recovery of eDNA, and specifically aDNA. The morphological approach does not have to deal with these limitations and is very useful and important in such cases. Taking everything into account, the results of the sediment core analyses were comparable, showing similar diatom assemblage groups, despite the lack of reference sequences or morphological descriptions of the diatoms.

6 Conclusion

The circumpolar treeline region is vast and encompasses environmental and geologic differences. Because diatoms are sensitive for environmental changes and respond rapidly to them, they are valuable bioindicators for the arctic region. Thus, lakes across the treeline are an important starting point to better understand the basic limnological trends and characteristics and which variables affect diatom composition and growth in such systems. With the relatively new genetic approach, a much more detailed insight into diatom diversity is possible. Thus, genetic and morphologic differences and patterns can be assessed, compared and combined in recent samples and may help interpreting species turnover in ancient samples. The big difference between the two assessed methodologies is the large difference in data availability. With the NGS technology a very large amount of data can be produced, while the morphological identification of taxa using light microscopy usually counts 500 valves per sample. However, this study showed that even though the genetic and the morphological approach are two very different methods, the obtained results are still similar and comparable. The differentiation between species and genus level showed that a very deep taxonomic identification of the studied species is not absolutely necessary. Especially, the analysis with the influencing environmental factors showed that there is no difference between the focused taxonomic levels. Furthermore, similar results were obtained with both assessed methods. Also for the sediment cores similar diatom assemblage groups could be identified with both approaches. Of course, for diversity analyses a distinct and deep taxonomic determination is important, but also dependent on the underlying research question. The genetic approach could also provide information about short-term biodiversity changes on recent samples compared to long-term biodiversity changes inferred from ancient samples of sediment cores. This could be used to reconstruct past biodiversity and possibly distribution patterns. The advantage of the genetic approach on biodiversity research is the possibility of tracking genetic lineages, whereas the morphological approach might be used for biodiversity reconstructions of the present and past, but does not give any information about the genetic lineages. However, given that limnological

studies in the Siberian arctic treeline region are still relatively rare, continuing research in such remote regions is important to contribute to our understanding of the natural variability of these sensitive freshwater ecosystems and their response to environmental changes. To assess possible impacts and changes in these environments an evaluation of the present is necessary. Only then predictions of climate change effects on these vulnerable regions and paleoreconstructions can be made. Depending on the research focus both methods could be used alone or in combination to facilitate each other.

7 References

ACIA, 2004. Arctic Climate Impact Assessment - Scientific Report. *Cambridge University Press*, p.1046. Available at: http://amap.no/acia/.

Anderson, N.J. & Rippey, B., 1994. Monitoring lake recovery from point-source eutrophication: the use of diatom-inferred epilimnetic total phosphorus and sediment chemistry. *Freshwater Biology*, 32(3), pp.625–639. Available at: http://doi.wiley.com/10.1111/j.1365-2427.1994.tb01153.x.

Anderson-Carpenter, L.L. et al., 2011. Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. *BMC evolutionary biology*, 11(1), p.30. Available at: http://www.biomedcentral.com/1471-2148/11/30.

Andreev, A.A. et al., 2002. Late Pleistocene and Holocene vegetation and climate on the Taymyr Lowland, northern Siberia. *Quaternary Research*, 57(1), pp.138–150. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0033589401923026.

Andreev, A.A. & Klimanov, V.A., 2000. Quantitative Holocene climatic reconstruction from Arctic Russia. *Journal of Paleolimnology*, 24(1), pp.81–91.

Appleby, P.G. & Piliposyan, G.T., 2012. *Radiometric Dating of lake sediment cores from four sites in the vicinity of Chatanga, Northern Siberia*,

De Barba, M. et al., 2014. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Molecular Ecology Resources*, 14(2), pp.306–323.

Battarbee, R.W. et al., 2001. Diatoms. In *Smol, J. P., Birks, H. J. B., Last, W. M. (Eds.). Tracking environmental change using lake sediments. Terrestrial, algal and siliceous indicators. Vol 3. Kluwer Academic Publisher, New York*. pp. 155–202.

Bennett, K., 1996. Determination of the number of zones in a biostratigraphical sequence. *New Phytologist*, 132(1), pp.155–170. Available at: http://doi.wiley.com/10.1111/j.1469-8137.1996.tb04521.x\nhttp://onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.1996.tb04521.x/abstract.

Binladen, J. et al., 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE*, 2(2), pp.1–9.

Biskaborn, B.K. et al., 2012. Environmental variability in northeastern Siberia during the last ~13,300yr inferred from lake diatoms and sediment-geochemical parameters. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 329-330, pp.22–36. Available at: http://dx.doi.org/10.1016/j.palaeo.2012.02.003.

7 References

Blok, D. et al., 2010. Shrub expansion may reduce summer permafrost thaw in Siberian tundra. *Global Change Biology*, 16(4), pp.1296–1305. Available at: http://doi.wiley.com/10.1111/j.1365-2486.2009.02110.x [Accessed November 14, 2014].

Borcard, D., Gillet, F. & Legendre, P., 2011. *Numerical Ecology with R* R. Gentleman, K. Hornik, & G. G. Parmigiani, eds., Springer New York, Dordrecht, London, Heidelberg.

Burkhardt, S. et al., 2001. CO2 and HCO3- uptake in marine diatoms acclimated to different CO2 concentrations. *Limnology and Oceanography*, 46(6), pp.1378–1391.

Charles, D.F., 1985. Relationships between Surface Sediment Diatom Assemblages and Lakewater Characteristics in Adirondack Lakes. *Ecology*, 66(3), pp.994–1011.

Coissac, E., 2012. OligoTag: a Program for Designing Sets of Tags for Next-Generation Sequencing of Multiplexed Samples. In F. Pompanon & A. Bonin, eds. *Data Production and Analysis in Population Genomics*. New York: Humana Press, pp. 13–31.

Coolen, M.J.L. et al., 2006. Ancient DNA derived from alkenone-biosynthesizing haptophytes and other algae in Holocene sediments from the Black Sea. *Paleoceanography*, 21(1), pp.1–17.

Coolen, M.J.L. et al., 2004. Combined DNA and lipid analyses of sediments reveal changes in Holocene haptophyte and diatom populations in an Antarctic lake. *Earth and Planetary Science Letters*, 223(1-2), pp.225–239.

Cooper, A. & Poinar, H.N., 2000. Ancient DNA. *Science*, 289, p.1139.

Duff, K.E. et al., 1999. Limnological characteristics of lakes located across arctic treeline in northern Russia. *Hydrobiologia*, 391, pp.205–222.

Epp, L.S. et al., 2015. Lake sediment multi-taxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. *Quaternary Science Reviews*, 117(0318), pp.152–163. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0277379115001341.

Epp, L.S. et al., 2012. New environmental metabarcodes for analysing soil DNA: Potential for studying past and present ecosystems. *Molecular Ecology*, 21(8), pp.1821–1833.

Esper, J. & Schweingruber, F.H., 2004. Large-scale treeline changes recorded in Siberia. *Geophysical Research Letters*, 31(6), p.L06202. Available at: http://doi.wiley.com/10.1029/2003GL019178 [Accessed November 23, 2014].

7 References

Evans, K.M., Wortley, A.H. & Mann, D.G., 2007. An Assessment of Potential Diatom "Barcode" Genes (cox1, rbcL, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in Sellaphora (Bacillariophyta). *Protist*, 158(3), pp.349–364.

Flower, R.J. & Battarbee, R.W., 1983. Diatom evidence for recent acidification of two Scottish lochs. *Nature*, 305.

Fritz, S.C. et al., 1991. Reconstruction of past changes in salinity and climate using a diatom-based transfer function. *Nature*, 352. Available at: http://discovery.ucl.ac.uk/44102/.

Frost, G. V. & Epstein, H.E., 2014. Tall shrub and tree expansion in Siberian tundra ecotones since the 1960s. *Global Change Biology*, 20(4), pp.1264–1277.

Gasse, F. et al., 1997. Diatom-inferred salinity in palaeolakes: An indirect tracer of climate change. *Quaternary Science Reviews*, 16(6), pp.547–563.

Gasse, F., 1987. Diatoms for reconstructing palaeoenvironments and paleohydrology in tropical semi-arid zones - Example of some lakes from Niger since 12 000 BP. *Hydrobiologia*, 154(1), pp.127–163.

Gilbert, M.T.P. et al., 2005. Assessing ancient DNA studies. *Trends in Ecology and Evolution*, 20(10), pp.541–544.

Grimm, E.C., 1987. CONISS: a FORTRAN 77 program for stratigraphically constrained cluster analysis by the method of incremental sum of squares. *Computers & Geosciences*, 13(1), pp.13–35.

Hahne, J. & Melles, M., 1998. Climate and Vegetation History of the Taymyr Peninsula since Middle Weichselian Time - Palynological Evidence from Lake Sediments. In *Land-Ocean Systems in the Siberian Arctic: Dynamics and History*.

Hamsher, S.E. et al., 2011. Barcoding diatoms: exploring alternatives to COI-5P. *Protist*, 162(3), pp.405–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21239228 [Accessed September 24, 2014].

Herzschuh, U. et al., 2013. Siberian larch forests and the ion content of thaw lakes form a geochemically functional entity. *Nature communications*, 4, p.2408. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24005763 [Accessed November 23, 2014].

Hobbs, W.O. et al., 2010. Quantifying recent ecological changes in remote lakes of North America and greenland using sediment diatom assemblages. *PLoS ONE*, 5(4).

Hofreiter, M. et al., 2001. Ancient DNA. *Nature reviews. Genetics*, 2(5), pp.353–359.

7 References

Hongve, D., 1999. Production of dissolved organic carbon in forested catchments. *Journal of Hydrology*, 224(3-4), pp.91–99.

IPCC, 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* T. F. Stocker et al., eds., Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Jahn, R. et al., 2007. Diatoms and DNA barcoding: a pilot study on an environmental sample. *Proceedings of the 1st Central European Diatom Meeting 2007*, 113(May 2006), pp.63–68. Available at: http://www.bgbm.org/bgbmpress/otherpubl/cediatom/contents.htm.

Jiang, H. et al., 2004. Diatoms from the surface sediments of the South China Sea and their relationships to modern hydrography. *Marine Micropaleontology*, 53, pp.279–292.

Jiang, H. et al., 2002. Late-Holocene summer sea-surface temperatures based on a diatom record from the north Icelandic shelf. *The Holocene*, 2, pp.137–147.

Jones, A. et al., 2010. *Soil Atlas of the Northern Circumpolar Region*, JCR - European Commission.

Kearse, M. et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), pp.1647–1649.

Kermarrec, L. et al., 2013. Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: A test for freshwater diatoms. *Molecular Ecology Resources*, 13(4), pp.607–619.

Korhola, A. et al., 2000. A quantitative Holocene climatic record from diatoms in northern Fennoscandia. *Quaternary Research*, 54, pp.284–294. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0033589400921537.

Krings, M. et al., 1997. Neandertal DNA sequences and the origin of modern humans. *Cell*, 90(1), pp.19–30.

Laing, T.E., M. Rühland, K. & Smol, J.P., 1999. Past environmental and climatic changes related to tree-line shifts inferred from fossil diatoms from a lake near the Lena River Delta, Siberia. *The Holocene*, 9(5), pp.547–557.

Laing, T.E., Pienitz, R. & Smol, J.P., 1999. Freshwater Diatom Assemblages From 23 Lakes Located Near Norilsk, Siberia: a Comparison With Assemblages From Other Circumpolar Treeline Regions. *Diatom Research*, 14(2), pp.285–305.

Laing, T.E. & Smol, J.P., 2000. Factors influencing diatom distributions in circumpolar treeline. *Journal of Phycology*, 1048, pp.1035–1048.

7 References

Lauterborn, R., 1896. *Untersuchungen über Bau, Kernteilung und Bewegung der Diatomeen*, Leipzig.

MacDonald, G.M., Kremenetski, K. V & Beilman, D.W., 2008. Climate change and the northern Russian treeline zone. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 363(1501), pp.2285–2299.

MacGillivary, M.L. & Kaczmarska, I., 2011. Survey of the efficacy of a short fragment of the rbcL gene as a supplemental DNA barcode for diatoms. *Journal of Eukaryotic Microbiology*, 58(6), pp.529–536.

Mason, V.C. et al., 2011. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Research*, 21(10), pp.1695–1704.

Matveyeva, N. V., 1994. Floristic classification and ecology of tundra vegetation of the Taymyr Peninsula, northern Siberia. *Journal of Vegetation Science*, 5(6), pp.813–828. Available at: http://doi.wiley.com/10.2307/3236196.

Medinger, R. et al., 2010. Diversity in a hidden world: Potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Molecular Ecology*, 19(SUPPL. 1), pp.32–40.

Möller, P., Bolshiyanov, D.Y. & Bergsten, H., 1999. Weichselian geology and palaeoenvironmental history of the central Taymyr Peninsula, Siberia, indicating no glaciation during the last global glacial maximum. *Boreas*, 28(1), pp.92–114. Available at: http://dx.doi.org/10.1111/j.1502-3885.1999.tb00208.x\nO:\DATED\Papers in database NB. Brit papers elsewhere\M?ller_et_al_1999_Boreas_TaymyrNoGlaciation.pdf.

Moritz, C. & Cicero, C., 2004. DNA barcoding: Promise and pitfalls. *PLoS Biology*, 2(10).

Myers-Smith, I.H. et al., 2011. Shrub expansion in tundra ecosystems: dynamics, impacts and research priorities. *Environmental Research Letters*, 6(4), pp.1–15. Available at: http://stacks.iop.org/1748-9326/6/i=4/a=045509?key=crossref.3e2b3e3172ca6f9fc39a607c137b880f [Accessed November 23, 2014].

Naito, A.T. & Cairns, D.M., 2011. Patterns and processes of global shrub expansion. *Progress in Physical Geography*, 35(4), pp.423–442. Available at: http://ppg.sagepub.com/cgi/doi/10.1177/0309133311403538 [Accessed November 23, 2014].

Oksanen, J. et al., 2015. vegan: Community Ecology Package.

Pääbo, S. et al., 2004. Genetic analyses from ancient DNA. *Annual review of genetics*, 38, pp.645–679.

Pedersen, M.W. et al., 2015. Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society*, 370(20130383).

Pestryakova, L. a. et al., 2012. Present-day variability and Holocene dynamics of permafrost-affected lakes in central Yakutia (Eastern Siberia) inferred from diatom records. *Quaternary Science Reviews*, 51, pp.56–70. Available at: http://dx.doi.org/10.1016/j.quascirev.2012.06.020.

Pienitz, R. & Smol, J.P., 1993. Diatom assemblages and their relationship to environmental variables in lakes from the boreal forest-tundra ecotone near Yellowknife, Northwest Territories, Canada. *Hydrobiologia*, 269-270(1), pp.391–404.

Pienitz, R., Smol, J.P. & Birks, H.J.B., 1995. Assessment of Fresh-Water Diatoms As Quantitative Indicators of Past Climatic-Change in the Yukon and Northwest-Territories, Canada. *Journal of Paleolimnology*, 13, pp.21–49. Available at: http://discovery.ucl.ac.uk/165814/.

Pienitz, R., Smol, J.P. & Lean, D.R., 1997. Physical and chemical limnology of 24 lakes located between Yellowknife and Contwoyto Lake, Northwest Territories (Canada). *Canadian Journal of Fisheries and Aquatic Sciences*, 54(2), pp.347–358.

Pienitz, R., Smol, J.P. & MacDonald, G.M., 1999. Paleolimnological reconstruction of holocene climatic trends from two boreal treeline lakes, Northwest Territories, Canada. *Arctic and Alpine Research*, 31(1), pp.82–93.

Pietramellara, G. et al., 2009. Extracellular DNA in soil and sediment: Fate and ecological relevance. *Biology and Fertility of Soils*, 45(3), pp.219–235.

Post, E. et al., 2009. Ecological dynamics across the Arctic associated with recent climate change. *Science (New York, N.Y.)*, 325(5946), pp.1355–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19745143 [Accessed September 29, 2014].

Puusepp, L. & Kangur, M., 2010. Linking diatom community dynamics to changes in terrestrial vegetation: a palaeolimnological case study of Lake Ķūži, Vidzeme Heights (Central Latvia). *Estonian Journal of Ecology*, 59(4), p.259.

Rawlence, N.J. et al., 2014. Using palaeoenvironmental DNA to reconstruct past environments: progress and prospects. *Journal of Quaternary Science*, 29, pp.610–626.

Roberts, D. & McMinn, a., 1998. A weighted-averaging regression and calibration model for inferring lakewater salinity from fossil diatom assemblages in saline lakes of the Vestfold Hills: A new tool for interpreting Holocene lake histories in Antarctica. *Journal of Paleolimnology*, 19(2), pp.99–113.

Rouse, W.R. et al., 1997. Effects of climate change on the freshwaters of arctic and subarctic North America. *Hydrological Processes*, 11(8), pp.873–902.

Rudaya, N. et al., 2009. Holocene environments and climate in the Mongolian Altai reconstructed from the Hoton-Nur pollen and diatom records: a step towards better understanding climate dynamics in Central Asia. *Quaternary Science Reviews*, 28(5-6), pp.540–554. Available at: http://dx.doi.org/10.1016/j.quascirev.2008.10.013.

Rühland, K., Priesnitz, A. & Smol, J.P., 2003. Paleolimnological Evidence from Diatoms for Recent Environmental Changes in 50 Lakes across Canadian Arctic Treeline. *Arctic, Antarctic, and Alpine Research*, 35(1), pp.110–123.

Rühland, K.M., 2001. *Diatom assemblage shifts relative to changes in environmental and climatic conditions in the circumpolar treeline regions of the Canadian and Siberian Arctic*.

Rühland, K.M. et al., 2003. Limnological characteristics of 56 lakes in the Central Canadian Arctic Treeline Region. *Journal of Limnology*, 62(1), pp.9–27.

Rühland, K.M., Paterson, A.M. & Smol, J.P., 2015. Lake diatom responses to warming: reviewing the evidence. *Journal of Paleolimnology*. Available at: http://link.springer.com/10.1007/s10933-015-9837-3.

Rühland, K.M., Smol, J.P. & Pienitz, R., 2003. Ecology and spatial distributions of surface-sediment diatoms from 77 lakes in the subarctic Canadian treeline region. *Canadian Journal of Botany*, 81(1), pp.57–73.

Ryves, D.B., Mcgowan, S. & Anderson, N.J., 2002. Development and evaluation of a diatom-conductivity model from lakes in West Greenland. *Freshwater Biology*, 47(5), pp.995–1014.

Schmid, A.-M.M. & Crawford, R.M., 2001. (Bacillariophyceae): formation of auxospores and initial cells. *European Journal of Phycology*, 36(4), pp.307–320.

Schmidt, R. et al., 2004. Fragilaria and Staurosira (Bacillariophyceae) from sediment surfaces of 40 lakes in the Austrian Alps in relation to environmental variables, and their potential for palaeoclimatology. *Journal of Limnology*, 63(2), pp.171–189.

Smol, J.P. & Cumming, B.F., 2000. Tracking long-term changes in climate using algal indicators in lake sediments. *Journal of Phycology*, 36(6), pp.986–1011.

Steve Juggins, 2015. rioja: Analysis of Quaternary Science Data. , p.R package version 0.9–5.

Stoof-Leichsenring, K.R. et al., 2014. A combined paleolimnological/genetic analysis of diatoms reveals divergent evolutionary lineages of Staurosira and Staurosirella (Bacillariophyta) in Siberian lake sediments along a latitudinal transect. *Journal of Paleolimnology*, 52(1-2), pp.77–93. Available at:

http://link.springer.com/10.1007/s10933-014-9779-1 [Accessed September 8, 2014].

Stoof-Leichsenring, K.R. et al., 2015. Genetic data from algae sedimentary DNA reflect the influence of environment over geography. *Scientific Reports*.

Stoof-Leichsenring, K.R. et al., 2012. Hidden diversity in diatoms of Kenyan Lake Naivasha: A genetic approach detects temporal variation. *Molecular Ecology*, 21(8), pp.1918–1930.

Sturm, M., Racine, C. & Tape, K., 2001. Increasing shrub abundance in the Arctic. *Nature*, 411, p.546.

Sylvestre, F., Servant-Vildary, S. & Roux, M., 2001. Diatom-based ionic concentration and salinity models from the south Bolivian Altiplano (15-23°S). *Journal of Paleolimnology*, 25(3), pp.279–295.

Weckström, J., Korhola, A. & Blom, T., 1997. Diatoms as quantitative indicators of pH and water temperature in subarctic Fennoscandian lakes. *Hydrobiologia*, 347, pp.171–184.

Whitmore, T.J., 1989. Florida diatom assemblages as indicators of trophic state and pH. *Limnology and Oceanography*, 34(5), pp.882–895.

Yang, X. et al., 2003. Diatom-based conductivity and water-level inference models from eastern Tibetan (Qinghai-Xizang) Plateau lakes. *Journal of Paleolimnology*, 30(1), pp.1–19.

Zimmermann, J., Jahn, R. & Gemeinholzer, B., 2011. Barcoding diatoms: Evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity and Evolution*, 11(3), pp.173–192.

Appendix

A.1 Material

A.1.1 Chemicals and buffers

10x PCR reaction buffer	Invitrogen/Life Corp., Carlsbad, CA, USA
10x TAE buffer	AppliChem GmbH, Darmstadt, Germany
6x DNA loading dye	Thermo Scientific, Dreieich, Germany
BSA (UV, 20 mg/ml)	VWR International, Darmstadt, Germany
DNA ExitusPlus	AppliChem GmbH, Darmstadt, Germany
dNTP mix, 2.5 mmol/l	Qiagen, Hilden, Germany
GelRed DNA staining	Biotium, Hayward, CA, USA
H_2O (DEPC treated)	GBiosciences, St. Louis, MO, USA
Kanamycin A, 50 mg/ml	Sigma-Aldrich, St. Louis, MO, USA
LB agar	Sigma-Aldrich, St. Louis, MO, USA
$MgSO_4$ (UV, 50 mmol/l)	Invitrogen/Life Corp., Carlsbad, CA, USA
Mucasol	Merz, Frankfurt, Germany
Naphrax	Robert Charles Laboratories Ltd, Milton Ernest, UK
O'range Ruler 50 bp DNA ladder	Thermo Scientific, Dreieich, Germany
Platinum® Taq DNA Polymerase HiFi (5U/µl)	Invitrogen/Life Corp., Carlsbad, CA, USA
Rotigarose® agarose	Carl Roth, Karlsruhe, Germany
SOC culture medium	Invitrogen/Life Corp., Carlsbad, CA, USA

A.1.2 Kits and other materials

50 ml Falcon® tubes	Becton Dickinson Labware, Franklin Lakes, NJ, USA
8-well RealTime PCR tubes	Brand GmbH & CO KG, Wertheim, Germany
MinElute® PCR Purification Kit	Qiagen, Hilden, Germany
PeqLab Electroporation Cuvettes, sterile	PeqLab GmbH, Erlangen, Germany

PowerMax Soil DNA Isolation Kit	MoBio Labs, Carlsbad, CA, USA
Qubit® dsDNA BR Assay Kit	Invitrogen/Life Corp., Eugene, Oregon, USA
Safe-Lock Tubes 2.0 ml	Eppendorf AG, Hamburg, Germany
TOPO® TA Cloning® Kit for Sequencing	Invitrogen/Life Corp., Carlsbad, CA, USA

A.1.3 Laboratory equipment

AirClean 600 PCR Workstation	Starlab GmbH, Hamburg, Germany
Biometra BioDocAnalyze®	Analytik Jena, Jena, Germany
Bio-Rad Sub-Cell® Model 96	Bio-Rad, Munich, Germany
CL-1000 Ultraviolet Crosslinker	UVP, LLC; Upland, CA, USA
Cooled incubator KB (E2)	Binder GmbH, Tuttlingen, Germany
Cryo Thermomixer comfort	Eppendorf AG, Hamburg, Germany
ELMI Centrifuge CM-6MT SkyLine	ELMI Ltd., Riga, Latvia
FastPrep®-24 homogenizer	MP Biomedicals, Irvine, CA, USA
GFL-7601 incubator	GFL Gesellschaft für Labortechnik mbH, Burgwedel, Germany
Heraeus BioFugeTM PicoTM centrifuge	Thermo Scientific, Dreieich, Germany
Heraeus Fresco 17 Centrifuge	Thermo Scientific, Dreieich, Germany
Heraeus HERAsafe HS12 workbench	Thermo Scientific, Dreieich, Germany
MicroPulserTM Electroporator	Bio-Rad, Munich, Germany
Mini Centrifuge Spectrafuge	neoLab, Heidelberg, Germany
PCR hood: DNA/RNA UV Cleaner	biosan, Riga, Latvia
PerfectSpin 24 Plus peqlab	VWR International GmbH, Erlangen, Germany
PowerPac basic	Bio-Rad, Munich, Germany
Qubit® 2.0 Fluorometer	Invitrogen/Life Corp., Carlsbad, CA, USA
Sartorius BL310 scale	Sartorius AG, Göttingen, Germany
Sigma 6K15 high capacity refrigerated Centrifuge	SIGMA Laboratory Centrifuges, Osterode am Harz, Germany
Sprout® small table centrifuge	Heathrow Scientific® LLC, Illinois, USA
Thermal cycler: Biometra Professional	Analytik Jena, Jena, Germany
Thermal cycler: Bio-Rad iCycler®	Bio-Rad, Munich, Germany
Thermal cycler: Techne® TC-Plus	Bibby Scientific Ltd, Stone, Staffordshire, UK
Vortex-Genie 2	MoBio Labs, Carlsbad, CA, USA

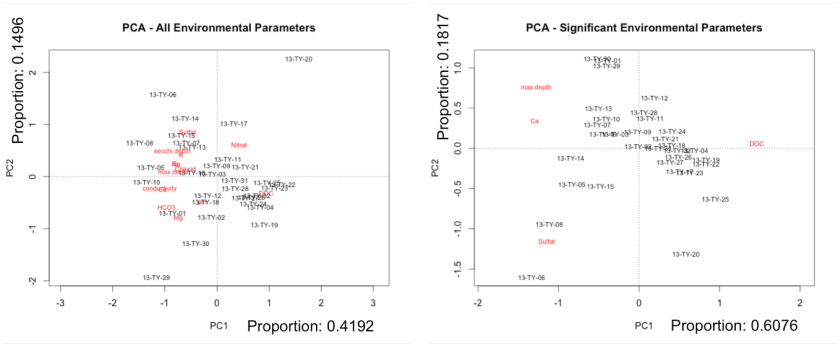
A.2 Preliminary statistics

A.2.1 Detrended Correspondence Analysis

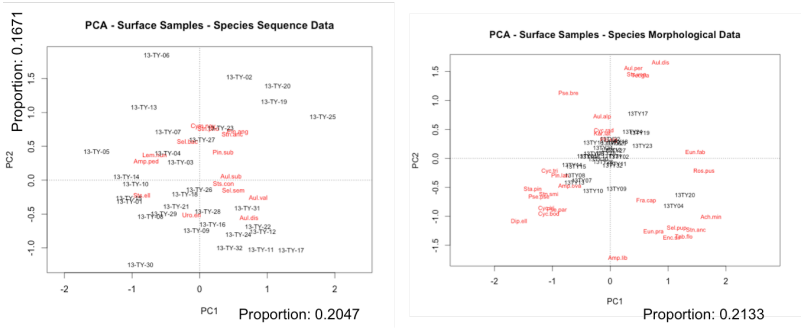
Appendix 1: Length of the first axis for each data set of the DCA.

Data	Axis length
Sequence Species	1.7098
Sequence Genera	1.3612
Morphology Species	1.7968
Morphology Genera	1.3286

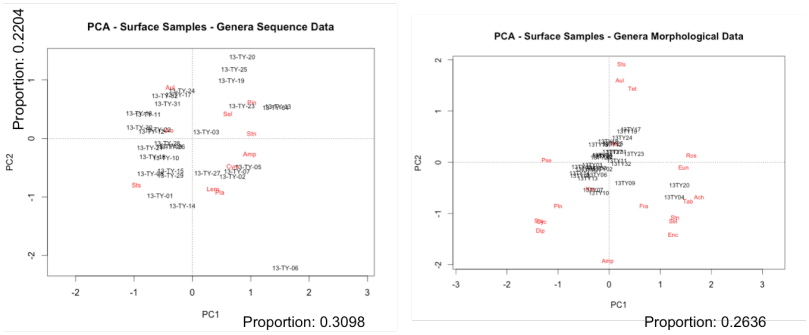
A.2.2 Principal Component Analysis



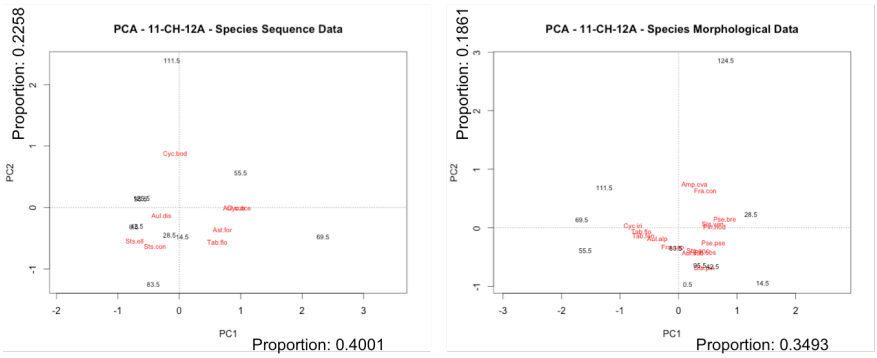
Appendix 2: PCA of the sampled lakes and the environmental parameters, as well as with the vegetation types.



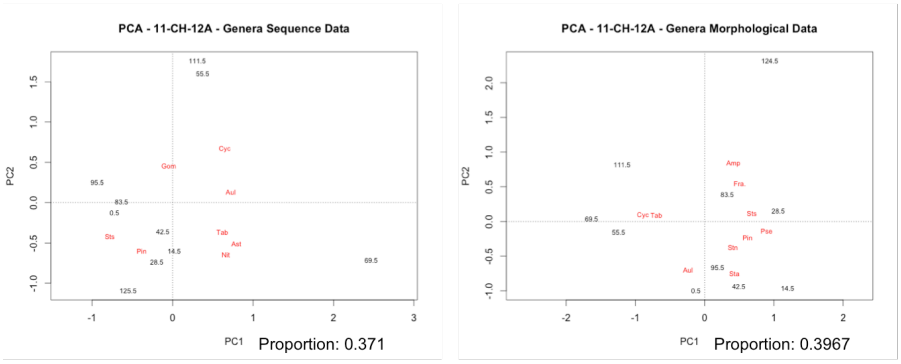
Appendix 3: PCAs of the species retrieved from the surface samples of both approaches.



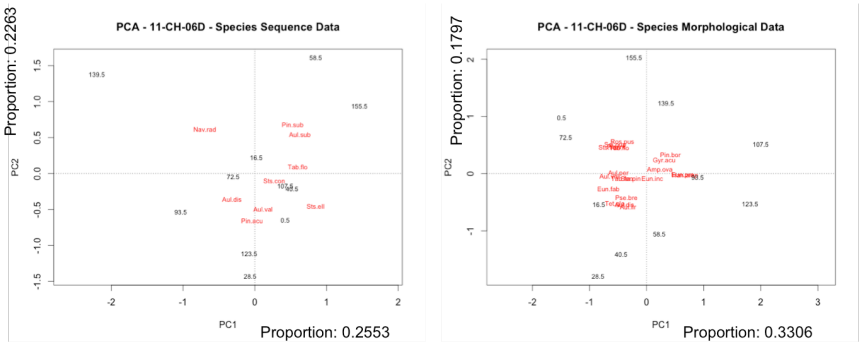
Appendix 4: PCAs of the genera retrieved from the surface samples of both methods.



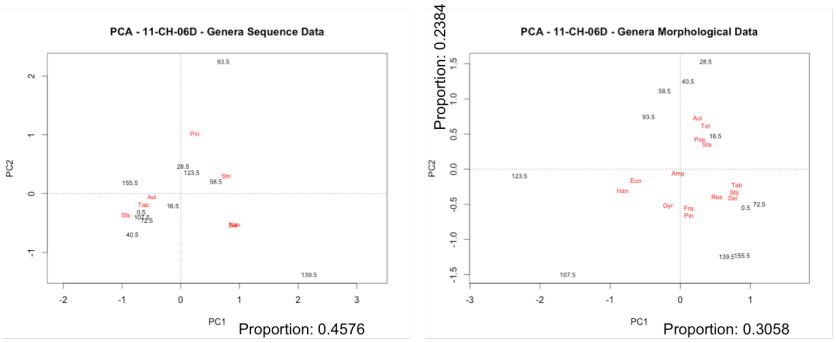
Appendix 5: PCAs of the species data of core 11-CH-12A on both methods.



Appendix 6: PCAs of the genera data of core 11-CH-12A on both methods.

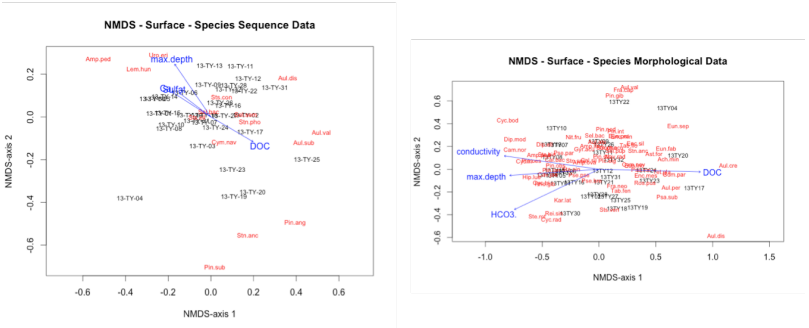


Appendix 7: PCAs of the species data of core 11-CH-06D on both methods.



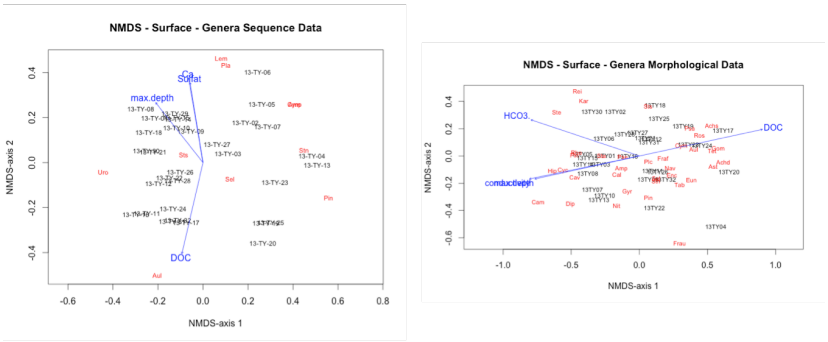
Appendix 8: PCAs of the genera data of core 11-CH-06D on both methods.

A.2.3 Non-metric Multidimensional Scaling



k=3

Appendix 9: NMDS analysis of the species data of both methods.

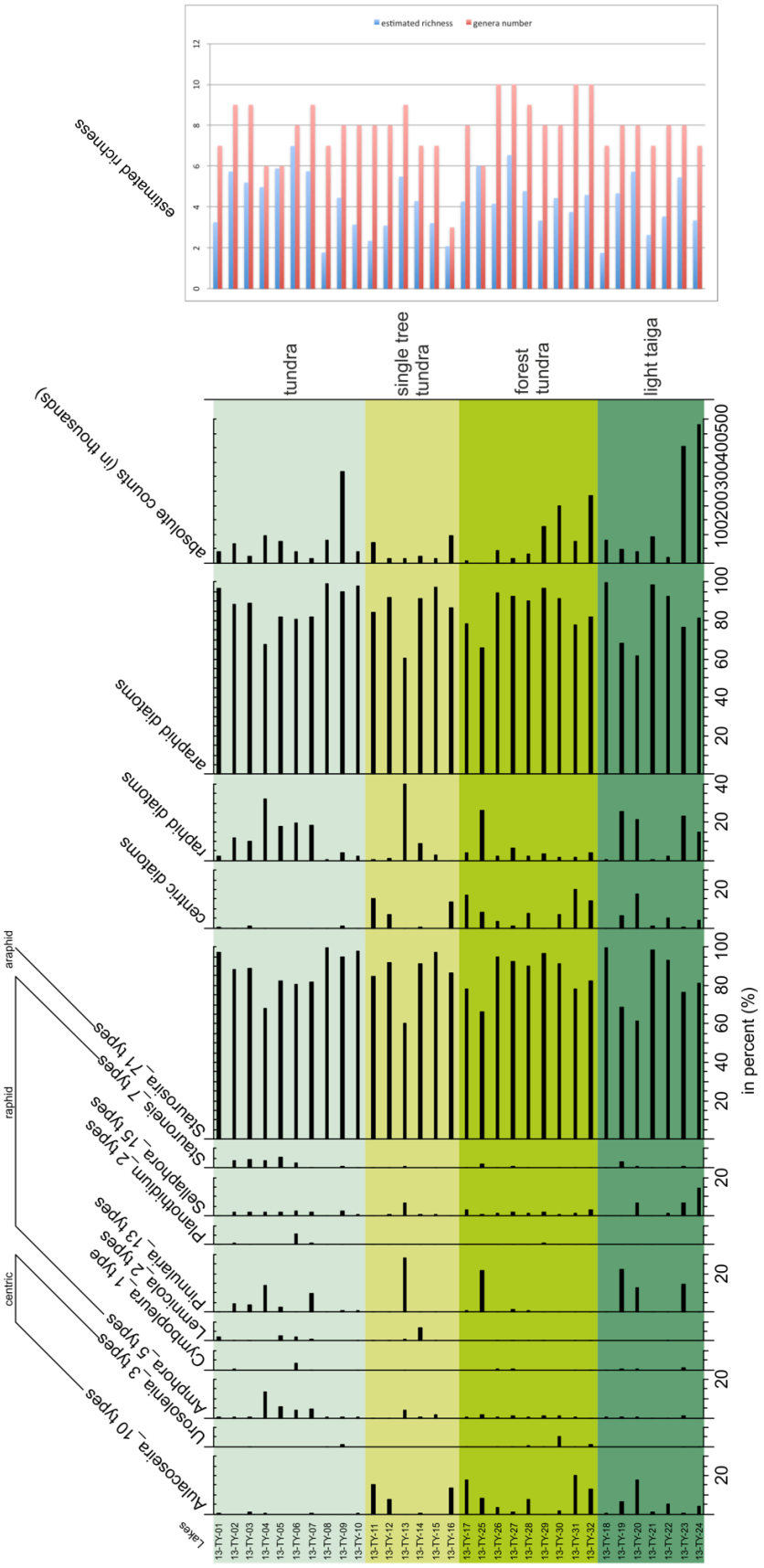


k=3

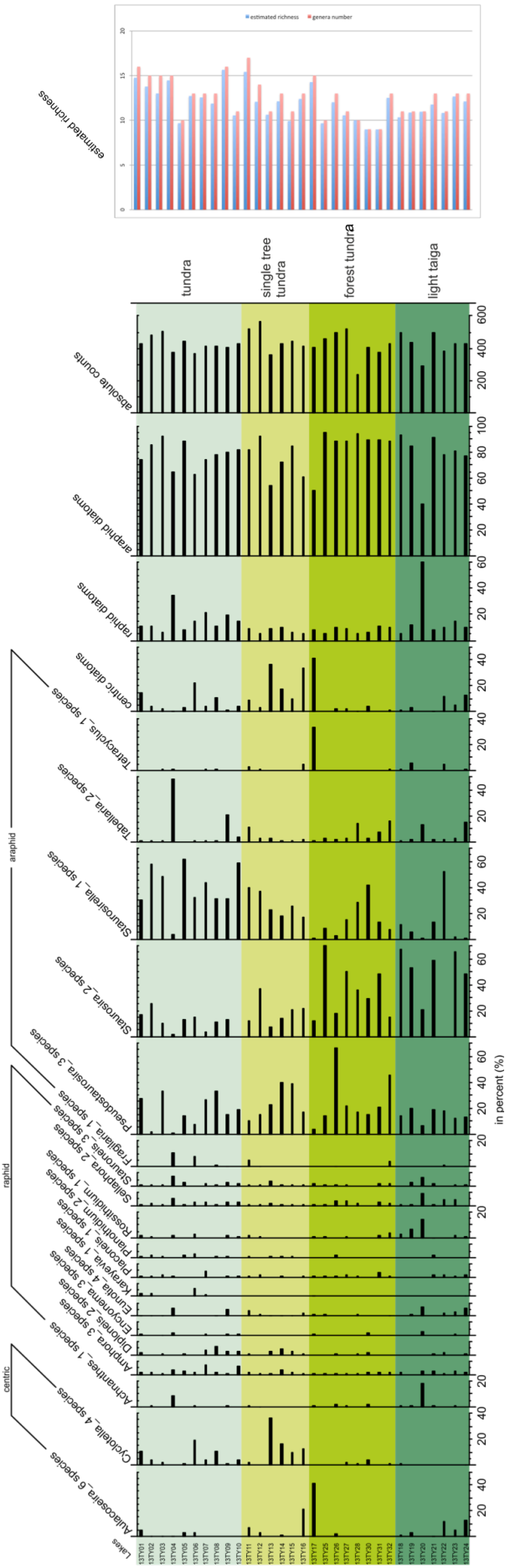
Appendix 10: NMDS analysis of the genera data of both methods.

A.3 Stratigraphic plots of the genera

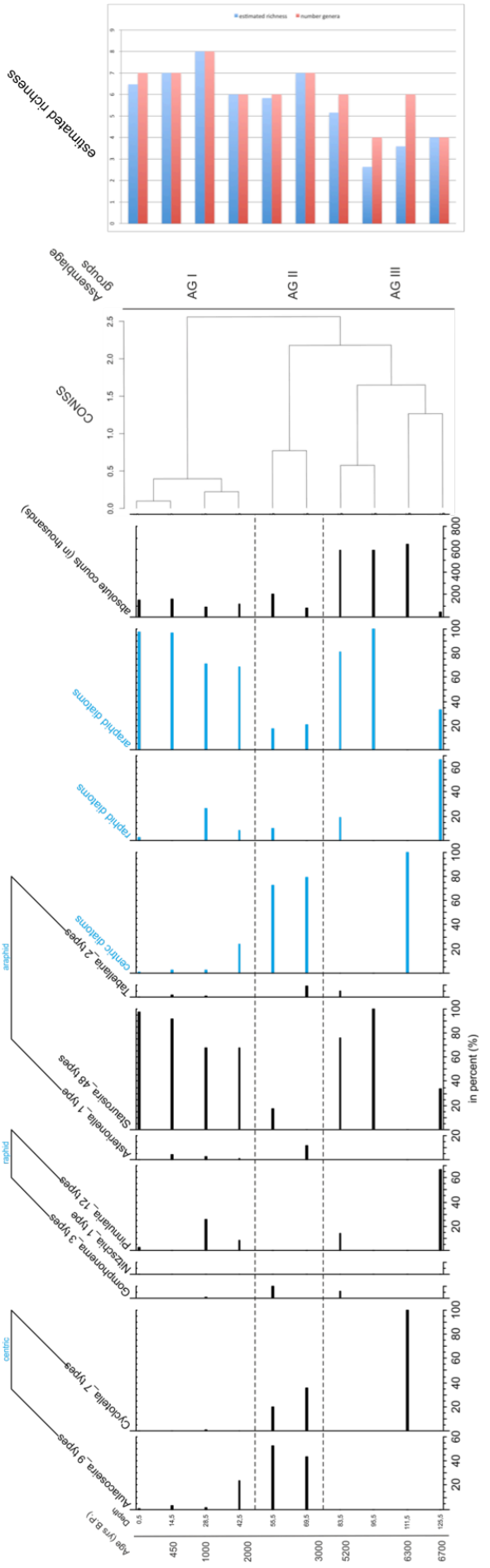
The following pages show the stratigraphic plots of the genera for both methods.



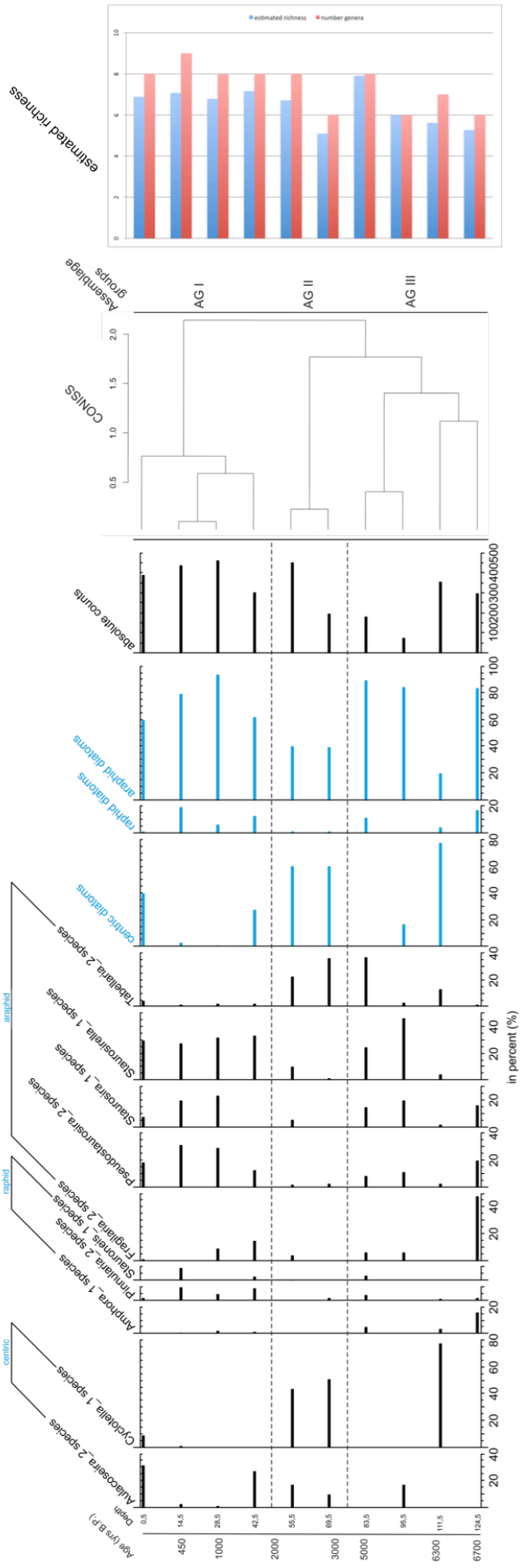
Appendix 11: Diatom sequence types identified to genus level of the lake transect from north to south, in the four vegetation zones tundra, single tree tundra, forest tundra and light taiga.



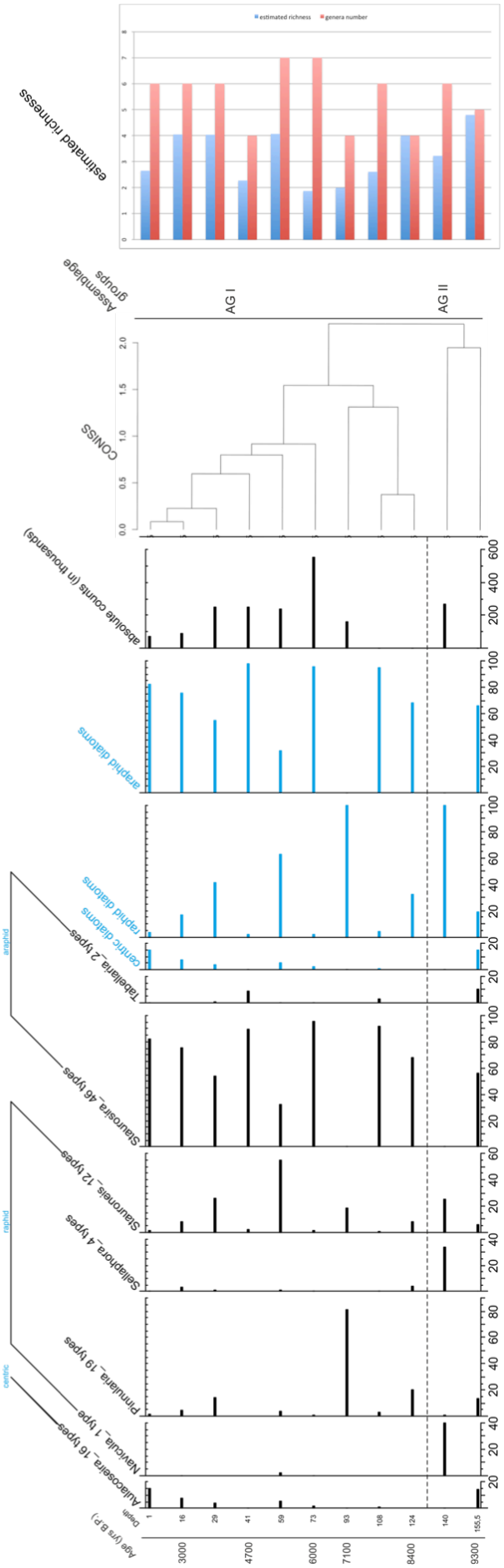
Appendix 12: Diatoms identified to genus level by light microscopy of the lake transect from north to south, in the four vegetation zones tundra, single tree tundra, forest tundra and light taiga.



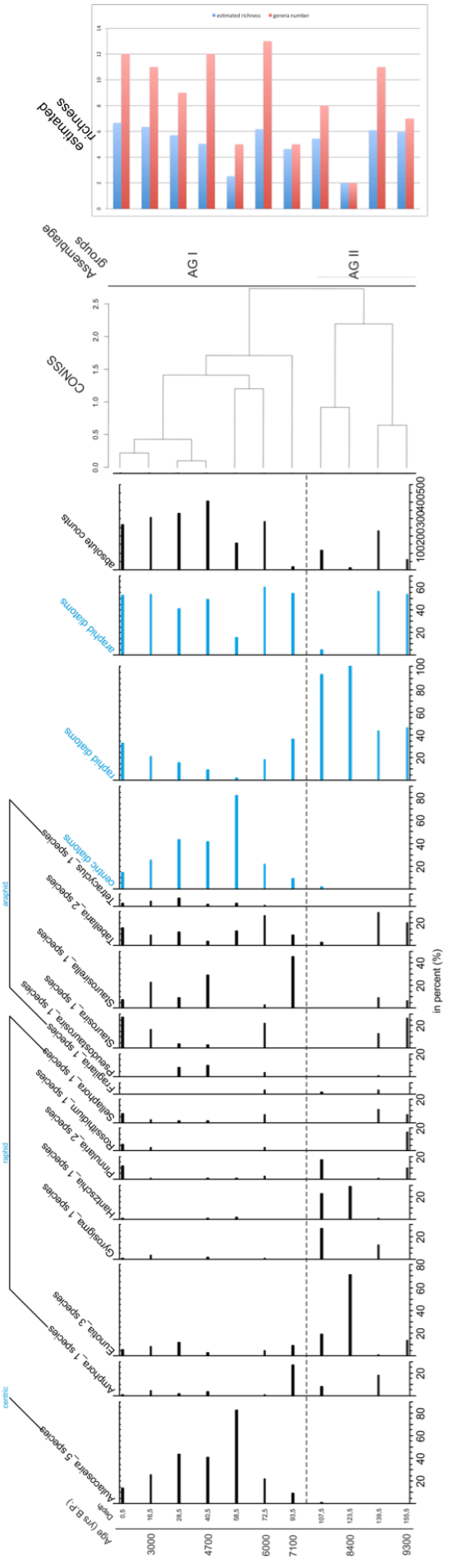
Appendix 13: Diatom sequence types identified to genus level of core 11-CH-12A with assemblage groups identified by CONISS analysis and the estimated richness.



Appendix 14: Diatoms identified to genus level by light microscopy of core 11-CH-12A with assemblage groups identified by CONISS analysis and estimated richness.

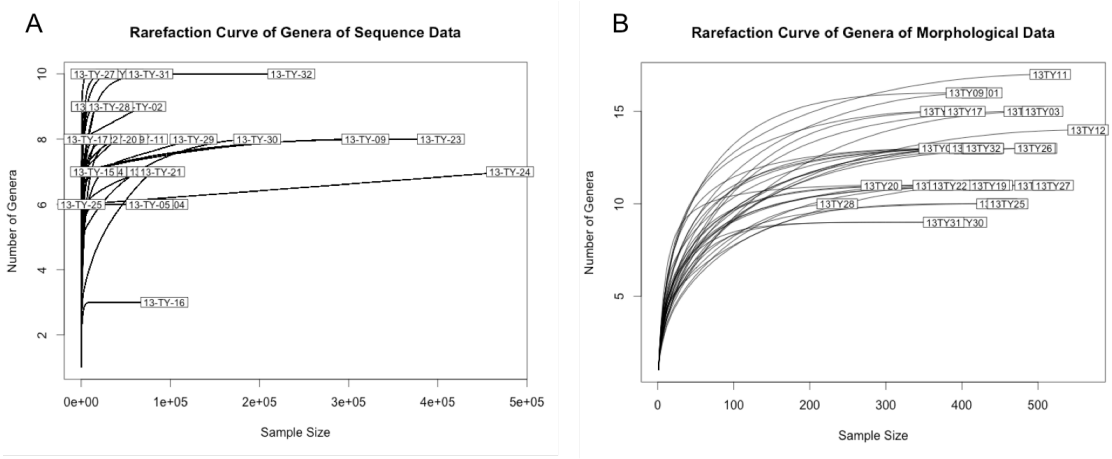


Appendix 15: Diatom sequence types identified to genus level of core 11-CH-06D, assemblage groups identified by CONISS analysis and estimated richness.



Appendix 16: Diatoms identified to genus level by light microscopy of core 11-CH-06D, assemblage groups identified by CONISS analysis and estimated richness.

A.4 Rarefaction curves of the genera of the surface data



Appendix 17: Rarefaction curves of the genera retrieved from the surface samples of both methods.

A.5 Environmental parameters

The environmental parameters of the studied lakes were measured during the expeditions and subsequently in the laboratories of the AWI Potsdam. The preparation of the samples was done in the field by the expedition team and Antje Eulenburg (AWI Potsdam) did further measurements in the laboratory. The data were kindly at my disposal and I used them in the statistical analysis of the diatom surface data set. All measured parameters are shown in Appendix 12.

Appendix 18: All measured physico-chemical parameters of the lakes used for analysis of modern sediments.

Lake ID	Latitude (°)	Longitude (°)	max. depth	secchi depth (m)	pH	conductivity (µS/cm)	Chlorid	Sulfat	HCO_3^-	Ca	K	Mg	Na	Si	DOC	vegetation type
13-TY-01	72.66776	105.88073	15	3.25	8.49	98	7.0161	0.12	36.30	7.16	0.40	4.74	4.72	1.09	5.29	tundra
13-TY-02	72.55307	105.71745	3.4	1.45	8.53	69.00	0.9321	0.30	38.44	5.90	0.34	3.95	2.31	1.03	4.57	tundra
13-TY-03	72.54794	105.72582	5.6	3.05	8.03	73.00	1.9324	0.33	27.00	5.52	0.22	3.17	1.98	0.75	2.55	tundra
13-TY-04	72.54733	105.74213	1.5	1.5	8.07	50.00	1.0103	0.00	25.02	4.66	0.00	3.00	2.78	0.45	10.37	tundra
13-TY-05	72.54811	105.75047	6.6	5.15	8.33	96.00	3.2877	1.57	44.24	7.98	0.46	4.55	4.34	2.55	2.37	tundra
13-TY-06	72.54158	105.76354	5	4.4	8.05	98	1.1199	5.17	33.10	11.61	0.59	2.13	9.01	1.46	2.33	tundra
13-TY-07	72.5575	105.72706	8.3	6.2	7.85	65.00	1.507	0.51	28.07	6.54	0.85	3.50	10.04	0.47	2.77	tundra
13-TY-08	72.4883	105.64833	9.1	3.6	7.87	124	6.4665	3.44	48.05	9.23	0.66	4.10	7.34	1.94	3.36	tundra
13-TY-09	72.40078	105.43985	7.4	2.7	7.84	79.00	3.5392	0.46	24.86	5.16	0.60	3.19	3.28	0.30	7.24	tundra
13-TY-10	72.40615	105.44247	7.9	5.5	7.85	180.00	33.5431	0.49	32.19	7.41	0.61	5.43	16.96	0.48	4.56	s tree tundra
13-TY-11	72.41184	105.46357	8.3	3.15	7.66	57.00	3.5036	0.28	24.71	4.42	0.27	2.86	2.61	0.85	8.31	s tree tundra
13-TY-12	72.41291	105.44771	5.3	2.5	7.56	90.00	9.6584	0.00	29.59	6.02	0.37	3.94	4.99	0.45	8.28	s tree tundra
13-TY-13	72.39159	105.44349	14.9	7.33	8.00	73.00	3.6232	0.54	27.30	5.35	0.61	3.28	3.96	0.25	3.54	s tree tundra
13-TY-14	72.38663	105.45663	23.1	7.4	7.84	69.00	3.4147	1.82	28.52	5.03	0.47	3.03	3.71	0.62	3.46	s tree tundra
13-TY-15	72.39236	105.46581	10.6	7.7	7.78	81.00	2.8083	1.71	36.76	5.10	0.53	3.45	5.76	0.70	5.17	s tree tundra
13-TY-16	72.18103	104.48785	12.5	4	7.90	69.00	0.9946	1.05	35.85	7.07	0.45	3.90	2.24	1.34	7.58	s tree tundra
13-TY-17	71.4031	102.2826	3	2.2	7.30	40.00	0.4809	0.64	16.17	5.53	1.26	3.26	17.94	1.12	19.35	Forest tundra
13-TY-18	71.09266	100.83102	2.5	2.5	9.11	61.00	0.8063	0.15	33.71	5.33	1.01	3.74	13.52	0.77	9.12	light taiga
13-TY-19	71.10203	100.82722	1.3	1.3	8.40	51.00	0.6818	0.00	28.22	3.94	0.00	3.24	1.80	0.37	11.28	light taiga
13-TY-20	71.11221	100.85285	2	2	7.59	33.00	3.5353	1.33	14.19	2.63	0.20	2.03	1.08	0.15	10.69	light taiga
13-TY-21	71.10669	100.82288	6.4	3.8	8.08	46.00	0.6405	0.31	24.56	3.58	0.26	2.79	2.02	0.46	8.79	light taiga
13-TY-22	71.09708	100.85389	2.5	1.95	7.23	43.00	0.5218	0.22	20.29	3.59	0.00	3.11	0.69	0.25	16.70	light taiga
13-TY-23	71.10383	100.87602	3.1	1.6	8.64	40.00	0.6054	0.35	19.83	3.20	0.22	2.62	0.67	0.16	11.08	light taiga
13-TY-24	71.09591	100.79747	6.1	1.4	7.57	51.00	0.6175	0.31	27.76	4.38	0.23	3.80	0.95	0.27	13.73	light taiga
13-TY-25	71.88648	101.21703	1.2	1.2	8.57	51	3.5895	0.27	19.22	2.85	0.24	2.63	3.94	0.24	12.59	forest tundra
13-TY-26	72.14929	102.05598	2.4	1.5	7.61	59	2.1525	0.30	25.93	5.50	0.21	3.32	1.82	0.13	12.72	Forest tundra
13-TY-27	72.15327	102.07542	2.7	1.8	7.82	59	2.0402	0.38	28.37	5.25	0.22	3.37	1.37	0.21	10.50	Forest tundra
13-TY-28	72.15487	102.10133	8.1	1.4	7.42	58	1.4346	0.29	28.83	5.38	0.25	3.38	1.08	0.70	8.79	Forest tundra
13-TY-29	72.14840	102.11604	7.9	2.2	8.40	123	2.1619	0.15	69.25	12.93	0.55	7.16	7.49	0.48	8.24	Forest tundra
13-TY-30	72.14203	102.0905	18.8	4.4	8.68	81.00	1.0272	0.26	45.15	7.97	0.00	4.60	1.06	0.24	5.98	Forest tundra
13-TY-31	72.13599	102.06801	3.8	1.1	7.39	52	1.4861	0.42	24.71	5.99	0.53	3.82	5.35	0.47	10.84	Forest tundra
13-TY-32	72.12775	102.03437	3.4	1.3	7.59	57	0.9109	0.30	26.85	4.76	0.23	3.11	0.97	0.23	12.33	Forest tundra

Appendix 19: Proportions of the physico-chemical lake characteristics for the genera of the genetic data.

Table of proportion of chemical and morphological lake characteristics for genera of sequence data

	Parameter	single proportion			unique proportion		
		percent	adjusted R^2	p-value	constrained percent	adjusted R^2	p-value
morphological and chemical parameters	max.depth	0.1038373	0.073965	0.015	0.06526474	0.068729	0.05
	secchi.depth	0.1061946	0.076401	0.025			
	pH	0.0482615	0.016537	0.18			
	conductivity	0.0884281	0.058042	0.035			
	Chlorid	0.0547282	0.023219	0.135			
	Sulfat	0.1003541	0.070366	0.025	0.01061686	-0.025387	0.8
	Nitrat	0.0123071	-0.020616	0.8			
	HCO_3^-	0.0428082	0.010902	0.26			
	Ca	0.1069919	0.077225	0.015	0.01642783	-0.015379	0.625
	K	0.0481955	0.016469	0.18			
	Mg	0.0280834	-0.004314	0.51			
	Na	0.0682978	0.037241	0.125			
	Si	0.0682203	0.037161	0.065			
	DOC	0.1214561	0.092171	0.015	0.0482028	0.039345	0.13

Parameter	single proportion		p-value			
	percent	adjusted R^2	tundra	single.tree	forest.tundra	light.taiga
vegetation type	0.1911887	0.1045304	0.01	0.425	/	0.23
veg+cond(sign.variables)	0.04829196	-0.0313436	0.33	0.88	/	0.795

Parameter	single proportion		p-value			
	percent	adjusted R^2	Sulfat	max.depth	Ca	DOC
all significant variables	0.3469781	0.250234	0.055	0.005	0.215	0.01
sign.variables+cond(veg)	0.2040813	0.1143601	0.115	0.01	0.43	0.16

Appendix 20: Proportion of the physico-chemical lake characteristics of the genera of the morphological data.

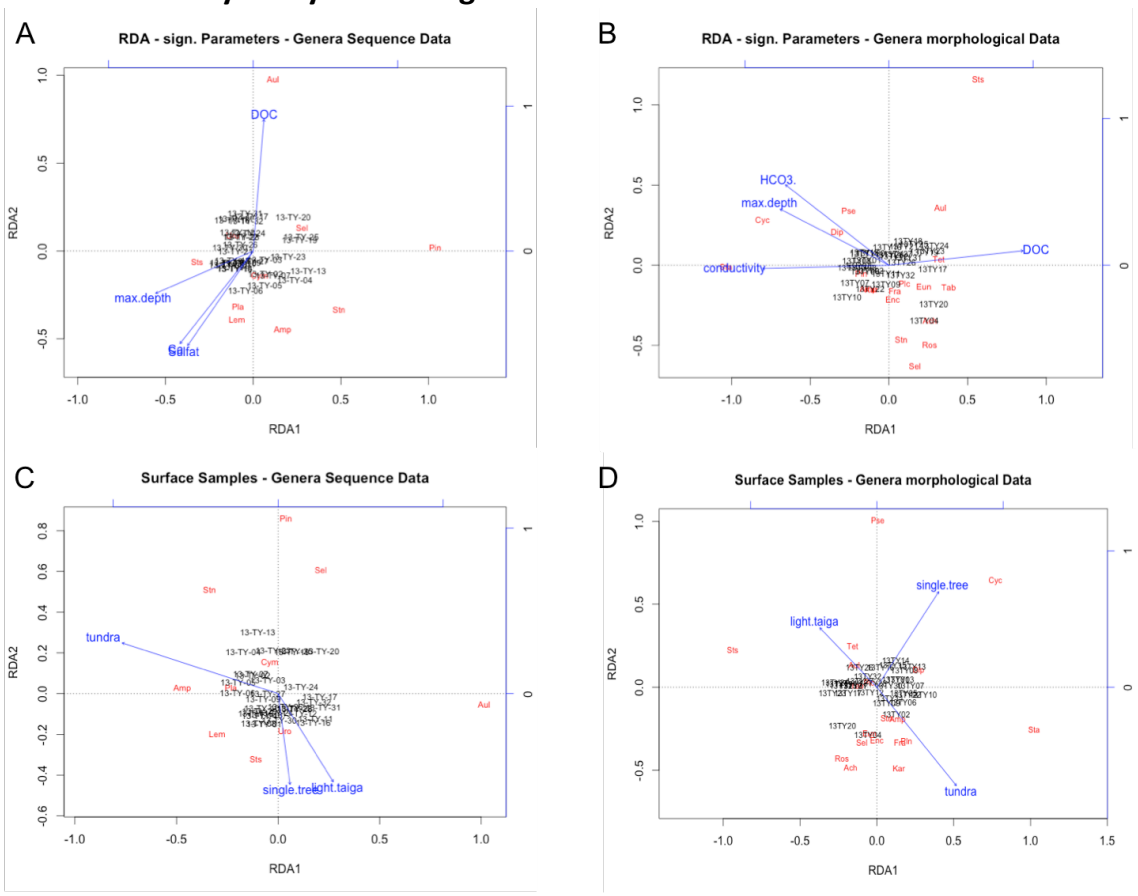
Table of proportion of chemical and morphological lake characteristics for genera of morphological data

	Parameter	single proportion			unique proportion		
		percent	adjusted R^2	p-value	constrained percent	adjusted R^2	p-value
morphological and chemical parameters	max.depth	0.153	0.123798	0.005	0.01828711	-0.006758	0.475
	secchi.depth	0.1808	0.152546	0.005			
	pH	0.0586	0.02614	0.08			
	conductivity	0.1897	0.161784	0.005	0.0216	-0.000933	0.52
	Chlorid	0.083667	0.05207	0.035			
	Sulfat	0.061806	0.029455	0.07			
	Nitrat	0.072041	0.040042	0.095			
	HCO_3^-	0.142088	0.112505	0.01	0.04212049	0.035301	0.08
	Ca	0.138338	0.108626	0.005	0.01049731	-0.020504	0.905
	K	0.041376	0.00832	0.32			
	Mg	0.076854	0.045022	0.03			
	Na	0.047979	0.015151	0.145			
	Si	0.092351	0.061052	0.005	0.05917306	0.065394	0.015
	DOC	0.213869	0.186761	0.005	0.03709838	0.026439	0.175

Parameter	single proportion		p-value			
	percent	adjusted R^2	tundra	single.tree	forest.tundra	light.taiga
vegetation type	0.2057866	0.1206923	0.01	0.01	/	0.34
veg+cond(sign.variables)	0.09651212	0.02439726	0.045	0.185	/	0.9

Parameter	single proportion		p-value			
	percent	adjusted R^2	max.depth	conductivity	HCO3-	DOC
all significant variables	0.325668	0.221924	0.005	0.01	0.15	0.09
sign.variables+cond(veg)	0.2064492	0.1177319	0.015	0.485	0.25	0.005

A.6 Redundancy analysis of the genera data sets



Appendix 21: RDAs of the environmental parameters and the vegetation types for the genera data of both methods. A - RDA environmental parameter with genetic data, B - RDA of environmental parameters with morphological data, C - RDA of vegetation types with genetic data, D - RDA of vegetation types with morphological data.

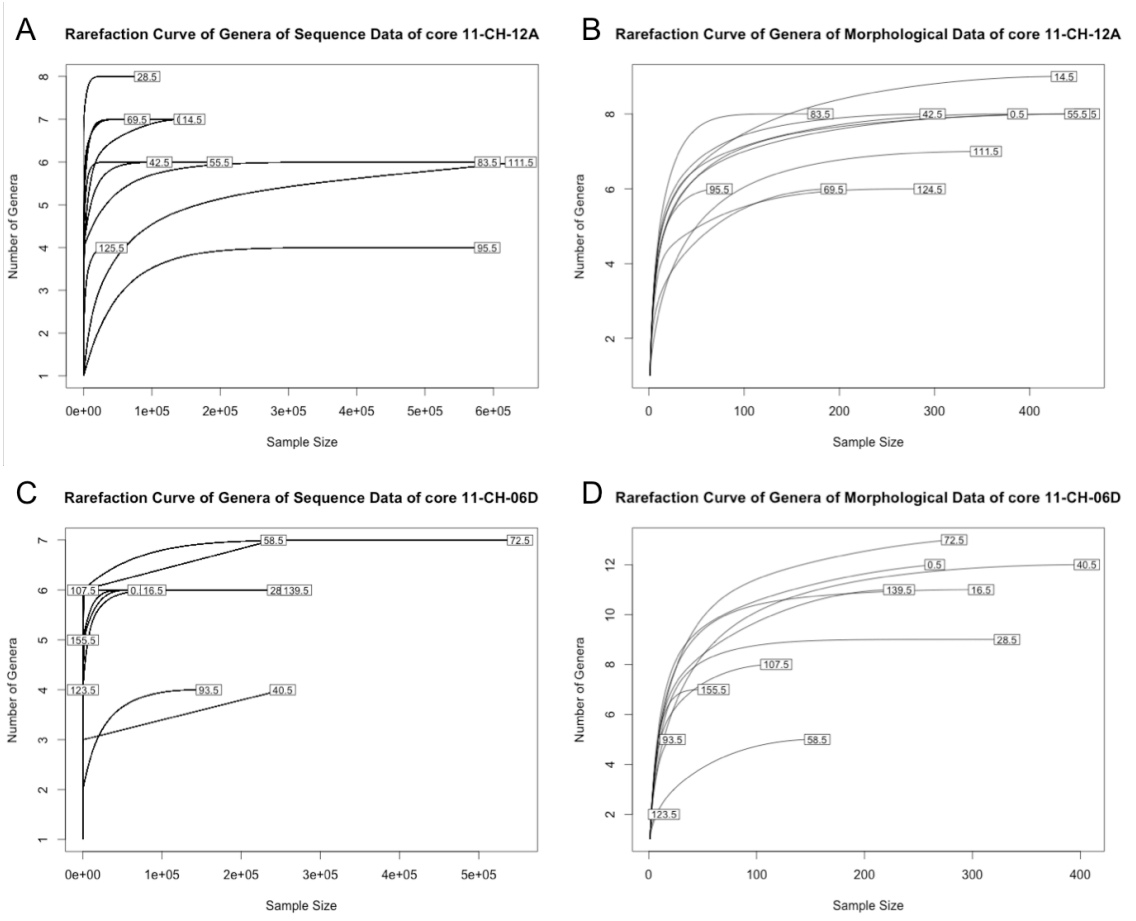
A.7 Diatom counts of core 11-CH-06D

For this Master thesis I studied Arctic diatoms and counted the given samples of core 11-CH-06D. I counted the valves in the three large morphological categories centric, araphid and raphid, but to have comparable numbers the data counted by Luidmila Pestryakova were used for the analyses done in this study.

Appendix 22: Counted diatom valves in three morphological categories.

Slide	centric	araphid	raphid	total
11CH06D; 0-1cm; 1; 10ml; 7k	109	76	315	500
11CH06D; 16-17cm; 2; 2,5ml; 10k	156	113	241	510
11CH06D; 28-29cm; 3; 4,5ml; 10k	135	113	252	500
11CH06D; 40-41cm; 4; 10ml; 10k	178	158	164	500
11CH06D; 72-73cm; 6; 2,5ml; 10k	332	49	119	500
11CH06D; 93-94cm; 7; 10ml; 10k	124	32	344	500
11CH06D; 123-124cm; 9; 10ml; 10k	477	4	19	500

A.8 Rarefaction curves of the genera of the core data



Appendix 23: Rarefaction curves of the genera of both methods for both sediment cores. A - Rarefaction curve of the genetic data of core 11-CH-12A, B - Rarefaction curve of the morphological data of core 11-CH-12A, C - Rarefaction curve of the genetic data of core 11-CH-06D, D - Rarefaction curve of the morphological data of core 11-CH-06D.

A.9 Primer and tag-combinations

Appendix 24: Table of all used primer-tag combinations with color code for modern and old samples.

PRIMERKOMBINATIONEN

aDNA/45

	rbcL808R_N_T11	rbcL808R_N_T12	rbcL808R_N_T13	rbcL808R_N_T14	rbcL808R_N_T15	rbcL808R_N_T16	rbcL808R_N_T17	rbcL808R_N_T18	rbcL808R_N_T19	rbcL808R_N_T20
rbcL705F_N_T09	aLE10	NTC06	ELP028a							
rbcL705F_N_T01	EJL010	aLE12	EKD019	ELP029a						
rbcL705F_N_T03	EJL009	EJL012	aLE14	EKD020	ELP030a					
rbcL705F_N_T06	EKD001	EJL011	EJL015	aLE15	EKD021	ELP031a				
rbcL705F_N_T04	aLE005	EKD002	EJL013	EBN008	aLE16	EKD022	aLE001			
rbcL705F_N_T08		aLE006	EKD003	EKD014	EJL014	aLE19	EKD023	aLE002		
rbcL705F_N_T02			aLE007	EKD004	EKD015	EBN010	aLE20	EKD024	aLE003	
rbcL705F_N_T05				aLE008	NTC05	EKD016	EJL016	NTC01	EKD025	aLE004
rbcL705F_N_T10		aLE018			aLE009	EKD026	EKD017	NTC02	aLE011	NTC07
rbcL705F_N_T07	aLE017				EKD017-2	NTC 08		EKD018	EKD022.2	aLE013

PRIMERKOMBINATIONEN

nur A45

	rbcL808R_N_T21	rbcL808R_N_T22	rbcL808R_N_T23	rbcL808R_N_T24
rbcL705F_N_T27	EJL_008	EBN_009	NTC04	NTC03
rbcL705F_N_T28	EBN_006	EJL_001	EJL_006	EBN_013
rbcL705F_N_T29	EBN_007	EJL_005	EJL_002	EJL007
rbcL705F_N_T30	EJL_004	EBN_011	EBN_012	EJL_003

Primerkombinationen sind in 8er Stripes vorgelegt (1:1 Primer Mix (10uM pro Primer))

Primer Kombination im aDNA (A43)
Primeraliquots Kombination in Genetiklabor A45 (Raum 08)
Primeraliquots Kombination in Genetiklabor A45 (Raum 08)
Primer Kombination nur in Genetiklabor A45
Primer Kombination nur in Genetiklabor A45

A.10 Sequencing data sets

Appendix 25: Table of all retrieved sequences with best identity, best match, count and assigned scientific name.

best_identity	best_match	count	*scientific_name*
0.973684211	HQ912451_1	812552	*Staurosira construens*
0.986842105	HQ828193_1	774642	*Staurosira elliptica*
0.973684211	HQ828193_1	687022	*Staurosira elliptica*
1	DQ514829_1	629716	*Cyclotella bodanica*
0.973684211	HQ828193_1	509093	*Staurosira elliptica*
0.960526316	JQ003572_1	465929	*Nitzschia*
1	AM710446_1	458890	*Pinnularia anglica*
0.973684211	HQ828193_1	437995	*Staurosira elliptica*
0.986842105	AM710493_1	350028	*Cymbellaceae*
0.961038961	HQ828193_1	296341	*Staurosira elliptica*
0.986842105	HQ828193_1	255254	*Staurosira elliptica*
0.960526316	HQ912468_1	210424	*Bacillariophycidae*
0.973684211	HQ828193_1	179646	*Staurosira elliptica*
0.986842105	HQ828193_1	170463	*Staurosira elliptica*
0.960526316	AM710454_1	138731	*Stauroneis*
0.986842105	JN162793_1	138708	*Nitzschia aff. dissipata*
0.960526316	HE978356_1	136090	*Haslea pseudostrearia*
1	EF143282_1	125895	*Sellaphora*
0.960526316	HQ828193_1	114123	*Staurosira elliptica*
1	AM710438_1	112728	*Navicula radiosa*
0.973684211	HQ912451_1	112518	*Staurosira construens*
0.960526316	AY571745_1	92056	*Sellaphora*
0.986842105	AY569592_1	84206	*Aulacoseira subarctica*
0.960526316	HQ828193_1	75883	*Fragilariaceae*
1	HQ912451_1	72179	*Staurosira construens*
0.973684211	DQ514832_1	68232	*Cyclotella ocellata*
1	AY569592_1	67113	*Aulacoseira subarctica*
0.986842105	HQ828193_1	67032	*Staurosira elliptica*
1	JN418657_1	65594	*Pinnularia*
0.986842105	AM710451_1	65138	*Pinnularia*
0.973684211	JQ610172_1	62785	*Planothidium sp. Styx 1*
0.973684211	HQ828193_1	62286	*Staurosira elliptica*
0.973684211	AM710493_1	58898	*Cymbellaceae*
1	EF143280_1	57291	*Sellaphora cf. seminulum*
0.960526316	AY569592_1	53857	*Aulacoseira*
0.960526316	HQ912448_1	47962	*Tabellaria flocculosa*
0.960526316	HQ828193_1	47698	*Staurosira elliptica*
0.961038961	AM710498_1	46211	*Stauroneis*
0.973684211	HQ828193_1	43923	*Staurosira elliptica*
0.960526316	JN418655_1	43195	*Naviculales*
0.960526316	HQ912451_1	41509	*Fragilariaceae*

0.960526316	AM710475_1	41444	*Stauroneis*
1	HQ828193_1	40653	*Staurosira elliptica*
0.960526316	JQ003564_1	39887	*Aulacoseira*
0.960526316	AB052286_1	39185	*Nannochloropsis*
0.960526316	HQ828193_1	39071	*Staurosira elliptica*
0.973684211	AM710458_1	36402	*Pinnularia*
0.960526316	AM710475_1	35754	*Stauroneis*
0.973684211	JQ003564_1	35648	*Aulacoseira distans var. alpigena*
0.986842105	HQ912451_1	35505	*Staurosira construens*
0.973684211	AM710475_1	34317	*Stauroneis*
0.960526316	AY571757_1	33759	*Bacillariophycidae*
0.974025974	AM710445_1	32191	*Pinnularia*
0.960526316	HQ912459_1	31934	*Gomphonema*
1	AB430671_1	31562	*Asterionella formosa*
1	AM710498_1	30290	*Stauroneis phoenicenteron*
1	AY569596_1	29939	*Aulacoseira*
0.986842105	AM710511_1	28400	*Amphora*
0.960526316	HQ828193_1	26766	*Staurosira elliptica*
1	AM710484_1	26196	*Cymbellaceae*
0.986842105	JN418655_1	25060	*Pinnularia*
0.973684211	HQ828193_1	24170	*Staurosira elliptica*
0.960526316	HQ828193_1	24025	*Staurosira elliptica*
0.986842105	HQ912451_1	21784	*Staurosira construens*
0.973684211	JQ003564_1	21750	*Aulacoseira distans var. alpigena*
0.960526316	HQ828193_1	21206	*Staurosira elliptica*
0.986842105	AM710451_1	20673	*Pinnularia*
0.960526316	AY571757_1	20347	*Bacillariophycidae*
0.960526316	HQ828193_1	20074	*Staurosira elliptica*
0.973684211	EF143280_1	19614	*Sellaphora cf. seminulum*
1	EF143272_1	18824	*Sellaphora pupula*
0.960526316	HQ828193_1	18359	*Staurosira elliptica*
0.986842105	AM710511_1	18241	*Amphora*
0.973684211	HQ912451_1	18192	*Staurosira construens*
0.973684211	HQ912451_1	18009	*Staurosira construens*
0.986842105	EF143280_1	17979	*Sellaphora cf. seminulum*
0.986842105	AM710459_1	17451	*Cymbellaceae*
0.986842105	AM710460_1	17118	*Gomphonema*
0.986842105	AM710438_1	16245	*Navicula*
0.960526316	JN418667_1	16185	*Pinnularia*
0.986842105	AM713181_1	16150	*Fragilariaceae*
0.986842105	HQ912448_1	16145	*Tabellaria flocculosa*
0.973684211	AY569592_1	15995	*Aulacoseira subarctica*
0.973684211	HQ828193_1	15993	*Staurosira elliptica*
0.986842105	AM710446_1	15384	*Pinnularia*
0.960526316	FJ002098_1	14985	*Naviculales*

0.973684211	HQ912451_1	14687	*Staurosira construens*
0.960526316	JN162778_1	14233	*Bacillariophycidae*
0.960526316	AF195952_1	14083	*Bacillariophycidae*
0.960526316	HQ828193_1	14026	*Fragilariaceae*
1	HQ912441_1	13806	*Urosolenia eriensis*
0.986842105	AM710475_1	13293	*Stauroneis anceps*
0.973684211	EF143263_1	12789	*Sellaphora*
1	AM710451_1	12122	*Pinnularia*
0.986842105	JN418673_1	12105	*Pinnularia*
0.986842105	AM710446_1	11680	*Pinnularia*
0.986842105	HQ912451_1	11640	*Staurosira construens*
0.986842105	JN418667_1	11575	*Pinnularia acuminata*
0.973684211	HQ828193_1	10804	*Staurosira elliptica*
1	EF143293_1	10358	*Sellaphora laevissima*
0.986842105	AM710479_1	10274	*Navicula capitatoradiata*
0.960526316	HQ912451_1	10131	*Fragilariaceae*
0.986842105	DQ514829_1	10128	*Cyclotella bodanica*
1	HQ828194_1	9940	*Fragilariaceae*
0.986842105	AM710488_1	9925	*Pinnularia*
0.973684211	JQ003564_1	9800	*Aulacoseira distans var. alpigena*
0.960526316	AM710464_1	9640	*Bacillariophycidae*
0.960526316	HQ828193_1	9480	*Staurosira elliptica*
0.960526316	AY569592_1	9319	*Aulacoseira*
0.973684211	HQ828193_1	9308	*Staurosira elliptica*
1	AY569602_1	9274	*Aulacoseira valida*
0.960526316	HQ828193_1	9112	*Fragilariaceae*
0.960526316	HQ912451_1	9040	*Fragilariaceae*
0.973684211	HQ828193_1	8857	*Staurosira elliptica*
0.973684211	HQ828193_1	8704	*Staurosira elliptica*
0.960526316	HQ912451_1	8444	*Fragilariaceae*
0.974025974	HQ828193_1	8103	*Staurosira elliptica*
0.960526316	HQ912451_1	8019	*Fragilariaceae*
0.973684211	AM710451_1	7832	*Pinnularia*
0.960526316	HQ828193_1	7726	*Fragilariaceae*
0.960526316	AM710475_1	7689	*Stauroneis*
0.960526316	AM710475_1	7368	*Stauroneis*
0.986842105	EF143263_1	6972	*Sellaphora*
0.960526316	JN418660_1	6953	*Pinnularia*
0.960526316	HQ828193_1	6914	*Staurosira elliptica*
0.960526316	EF143291_1	6770	*Sellaphora*
0.960526316	EF143280_1	6705	*Naviculales*
0.973684211	EF143280_1	6589	*Sellaphora cf. seminulum*
0.973684211	HQ828193_1	6404	*Staurosira elliptica*
0.986842105	HQ912451_1	6344	*Staurosira construens*
1	HQ912442_1	6310	*Fragilariaceae*

0.973684211	AM710493_1	6289	*Cymbellaceae*
0.973684211	AM710459_1	6269	*Cymbellaceae*
0.960526316	HQ912451_1	6141	*Fragilariaceae*
1	AM710503_1	5948	*Pinnularia substreptoraphe*
0.960526316	HQ828193_1	5899	*Fragilariaceae*
0.960526316	HQ912403_1	5891	*Amphora pediculus*
1	KC736601_1	5879	*Navicula cryptocephala*
1	JN418655_1	5750	*Pinnularia*
0.973684211	HQ912472_1	5532	*Gomphonema*
0.973684211	HQ828193_1	5518	*Staurosira elliptica*
0.973684211	AM710475_1	5488	*Stauroneis*
0.960526316	HQ828193_1	5467	*Staurosira elliptica*
0.973684211	DQ514833_1	5255	*Thalassiosirales*
0.960526316	HQ912451_1	5251	*Fragilariaceae*
0.960526316	HQ912451_1	5193	*Fragilariaceae*
0.973684211	DQ514829_1	5110	*Cyclotella bodanica*
0.986842105	AM710446_1	5054	*Pinnularia*
0.986842105	DQ514829_1	4770	*Cyclotella bodanica*
1	AM710464_1	4718	*Cymbopleura naviculiformis*
0.960526316	EF143291_1	4677	*Bacillariophycidae*
1	AY571745_1	4614	*Sellaphora bacillum*
0.973684211	HQ828196_1	4468	*Fragilariaceae*
0.960526316	HQ912451_1	4393	*Fragilariaceae*
0.986842105	AM710503_1	4350	*Pinnularia substreptoraphe*
0.986842105	JN418663_1	4349	*Caloneis silicula*
0.960526316	HQ828193_1	4336	*Fragilariaceae*
0.960526316	EF143280_1	4252	*Naviculales*
0.973684211	AM710438_1	4251	*Navicula*
0.960526316	HQ828193_1	4220	*Staurosira elliptica*
0.986842105	HQ912451_1	4143	*Staurosira construens*
0.960526316	AM710493_1	4142	*Cymbellaceae*
0.973684211	HQ828193_1	4134	*Staurosira elliptica*
0.960526316	HQ828193_1	4096	*Staurosira elliptica*
1	AM710486_1	4058	*Gomphonema*
0.960526316	JQ003564_1	3954	*Aulacoseira*
0.960526316	FJ002103_1	3940	*Bacillariophycidae*
1	EF143263_1	3838	*Sellaphora laevissima*
0.960526316	JQ610172_1	3833	*Planothidium sp. Styx 1*
0.960526316	AM710475_1	3792	*Stauroneis*
1	AY569594_1	3764	*Aulacoseira subarctica*
0.960526316	JQ003564_1	3677	*Aulacoseira*
0.986842105	HQ828193_1	3659	*Staurosira elliptica*
0.960526316	JN418667_1	3624	*Naviculales*
0.986842105	DQ514829_1	3610	*Cyclotella bodanica*
0.960526316	HQ912490_1	3577	*Lemnicola hungarica*

0.960526316	AM710464_1	3574	*Bacillariophycidae*
1	AM710511_1	3509	*Amphora sp. AT-221.04*
1	AM710447_1	3508	*Pinnularia*
0.960526316	AM710502_1	3369	*Encyonema caespitosum*
0.973684211	HQ912472_1	3313	*Gomphonema*
1	JN418653_1	3307	*Pinnularia sp. 8 CS-2011*
0.973684211	AM710475_1	3264	*Stauroneis*
1	EF143271_1	3259	*Sellaphora*
1	DQ514824_1	3224	*Stephanodiscus*
0.973684211	AM710436_1	3220	*Navicula cryptocephala var. veneta*
0.986842105	JN418655_1	3099	*Pinnularia*
0.960526316	HQ912451_1	3098	*Fragilariaceae*
0.960526316	HQ828193_1	3013	*Fragilariaceae*
0.973684211	HQ912451_1	2948	*Staurosira construens*
0.986842105	EF143271_1	2929	*Sellaphora*
0.986842105	FN557026_1	2880	*Bacillariophycidae*
0.960526316	HQ912451_1	2836	*Fragilariaceae*
0.960526316	AF195952_1	2795	*Phaeodactylum tricornutum*
0.973684211	HQ912490_1	2782	*Lemnicola hungarica*
0.973684211	HQ828193_1	2753	*Staurosira elliptica*
0.960526316	AM710475_1	2743	*Stauroneis*
0.960526316	JN418660_1	2725	*Pinnularia*
0.986842105	HQ912441_1	2663	*Urosolenia eriensis*
0.973684211	HQ828193_1	2608	*Staurosira elliptica*
0.973684211	HQ828193_1	2561	*Staurosira elliptica*
0.986842105	JQ003564_1	2560	*Aulacoseira distans var. alpigena*
1	AM710461_1	2527	*Pinnularia*
0.960526316	KC736602_1	2513	*Bacillariaceae*
0.973684211	KC736602_1	2420	*Bacillariaceae*
0.973684211	EF143271_1	2399	*Sellaphora*
0.973684211	HQ828193_1	2379	*Staurosira elliptica*
0.986842105	HQ828193_1	2368	*Staurosira elliptica*
0.960526316	HQ828193_1	2323	*Staurosira elliptica*
0.960526316	HQ828193_1	2320	*Staurosira elliptica*
0.973684211	JN162793_1	2264	*Nitzschia aff. dissipata*
0.960526316	HQ912451_1	2247	*Fragilariaceae*
0.986842105	AB195670_1	2244	*Eukaryota*
0.986842105	AM710459_1	2196	*Cymbellaceae*
0.960526316	EF143271_1	2190	*Naviculales*
0.960526316	HQ912451_1	2188	*Fragilariaceae*
0.973684211	HQ828193_1	2176	*Staurosira elliptica*
0.973684211	HQ828193_1	2164	*Staurosira elliptica*
0.961038961	HQ828193_1	2155	*Staurosira elliptica*
0.973684211	JN418673_1	2099	*Pinnularia*
0.973684211	AM710446_1	2075	*Pinnularia*

0.960526316	HQ912441_1	2056	*Urosolenia eriensis*
0.960526316	HQ828193_1	2042	*Staurosira elliptica*
0.960526316	HQ828193_1	1944	*Fragilariaceae*
0.960526316	HQ828193_1	1936	*Staurosira elliptica*
0.973684211	AM710438_1	1926	*Bacillariophyta*
0.960526316	HQ828193_1	1882	*Staurosira elliptica*
0.986842105	AM710438_1	1828	*Navicula radiosa*
0.973684211	HQ828193_1	1823	*Staurosira elliptica*
0.960526316	HQ828193_1	1811	*Fragilariaceae*
1	HQ912471_1	1785	*Placoneis elginensis*
0.960526316	HQ828193_1	1775	*Staurosira elliptica*
0.973684211	HQ912451_1	1748	*Staurosira construens*
0.973684211	HQ912451_1	1738	*Staurosira construens*
0.960526316	HQ828193_1	1731	*Staurosira elliptica*
0.973684211	AY569592_1	1718	*Aulacoseira subarctica*
0.986842105	KC911808_1	1702	*Sellaphora*
0.973684211	HQ828193_1	1644	*Staurosira elliptica*
0.960526316	HQ828193_1	1623	*Staurosira elliptica*
0.960526316	HQ828193_1	1609	*Staurosira elliptica*
0.960526316	AM710427_1	1606	*Naviculales*
0.960526316	HQ912451_1	1599	*Fragilariaceae*
0.960526316	AM710486_1	1566	*Bacillariophycidae*
0.986842105	AM710475_1	1564	*Stauroneis anceps*
0.960526316	JQ003572_1	1535	*Nitzschia*
0.973684211	JN418673_1	1484	*Pinnularia*
0.973684211	HQ828193_1	1466	*Staurosira elliptica*
0.973684211	HQ828193_1	1455	*Staurosira elliptica*
0.960526316	JQ354691_1	1425	*Bacillariophycidae*
0.973684211	HQ912451_1	1412	*Staurosira construens*
0.960526316	HQ912451_1	1403	*Fragilariaceae*
0.973684211	JN418673_1	1386	*Pinnularia*
0.986842105	HQ912451_1	1380	*Staurosira construens*
0.960526316	HQ828193_1	1380	*Staurosira elliptica*
0.986842105	EF143280_1	1366	*Sellaphora cf. seminulum*
0.960526316	AM710475_1	1356	*Stauroneis*
0.960526316	AY569592_1	1298	*Aulacoseira*
0.986842105	EF143282_1	1291	*Sellaphora*
0.986842105	HQ912451_1	1228	*Staurosira construens*
0.960526316	HQ828193_1	1221	*Staurosira elliptica*
0.960526316	AM710426_1	1207	*Amphora*
0.960526316	HQ828193_1	1205	*Fragilariaceae*
1	EF143285_1	1180	*Sellaphora pupula*
0.961038961	HQ828193_1	1148	*Staurosira elliptica*
0.960526316	HQ828193_1	1131	*Staurosira elliptica*
0.973684211	HQ828193_1	1125	*Staurosira elliptica*

0.986842105	JN418657_1	1119	*Pinnularia*
0.960526316	HQ828193_1	1117	*Staurosira elliptica*
0.960526316	HQ912451_1	1116	*Fragilariaceae*
0.960526316	HQ828193_1	1115	*Staurosira elliptica*
0.973684211	AM710459_1	1110	*Cymbellaceae*
0.986842105	JN418655_1	1109	*Pinnularia*
0.960526316	HQ828193_1	1108	*Staurosira elliptica*
0.960526316	KC736613_1	1105	*Sellaphora*
0.986842105	DQ514829_1	1085	*Cyclotella bodanica*
0.960526316	HQ912451_1	1068	*Fragilariaceae*
0.960526316	AM710442_1	1067	*Navicula reinhardtii*
0.960526316	HQ828193_1	1064	*Staurosira elliptica*
0.973684211	HQ912451_1	1057	*Staurosira construens*
0.986842105	HQ912441_1	1056	*Urosolenia eriensis*
0.986842105	JN418655_1	1019	*Pinnularia*
0.973684211	JN418660_1	1015	*Pinnularia*
0.986842105	DQ514829_1	1011	*Cyclotella bodanica*
0.973684211	DQ514829_1	1007	*Cyclotella bodanica*
0.960526316	HQ828193_1	1005	*Staurosira elliptica*
0.960526316	HQ912451_1	1001	*Fragilariaceae*
0.973684211	HQ828194_1	1000	*Fragilariaceae*
0.973684211	AY569592_1	1000	*Aulacoseira subarctica*

Appendix 26: Sequencetypes assigned to species level of the lake transect.

	Amp.ped	Aul.dis	Aul.sub	Aul.val	Cym.nav	Lem.hun	Pin.ang	Pin.sub	Sel.bac	Sel.sem	Stn.anc	Stn.pho	Sts.con	Sts.ell	Uro.eri
13-TY-01	88	67	42	0	0	738	0	0	0	14	0	0	2637	34583	0
13-TY-02	0	1	0	0	252	254	1022	0	53	692	1652	834	16004	44098	0
13-TY-03	0	0	228	0	0	49	675	0	71	23	0	57	2405	19172	4
13-TY-04	1	0	19	0	0	0	0	1015	27	0	1237	0	5	63084	0
13-TY-05	2951	0	0	0	0	2195	0	0	31	3	0	182	7170	55612	0
13-TY-06	200	0	0	0	1316	708	0	0	50	233	0	977	4840	26978	0
13-TY-07	111	1	0	2	5	169	460	0	95	40	12	6	2138	10257	0
13-TY-08	4	0	1	0	0	0	4	0	0	0	0	0	5848	72579	1
13-TY-09	0	0	0	1	0	0	0	805	0	5397	0	0	61684	239937	3633
13-TY-10	165	0	11	0	3	53	0	0	276	0	1	0	3702	34225	0
13-TY-11	0	10195	0	103	0	13	6	1	0	100	1	0	14033	45171	0
13-TY-12	0	642	0	161	0	13	2	1	7	94	1	0	4534	10908	0
13-TY-13	396	0	1	0	1	86	35	0	839	81	1	2	2208	7109	0
13-TY-14	100	88	0	0	0	1637	86	1	0	7	0	0	2522	19670	0
13-TY-15	173	0	0	0	0	42	2	3	0	28	0	0	1125	12660	0
13-TY-16	0	1372	1337	0	0	0	0	0	0	16	0	0	18436	62155	0
13-TY-18	0	1	29	0	31	0	53	42	20	136	19	0	13204	63961	0
13-TY-19	0	6	1253	110	30	0	1923	1804	17	31	706	1	3428	29509	0
13-TY-20	0	4	600	62	180	0	2774	709	228	912	249	0	2562	21595	0
13-TY-21	0	0	91	1	0	0	119	6	1	254	1	10	10969	77090	0
13-TY-22	0	360	1	159	0	12	4	2	4	261	4	0	4593	12139	0
13-TY-23	0	4	1273	1	2614	0	37418	0	2717	11411	0	3	32207	274879	3
13-TY-24	0	8640	236	274	0	0	169	214	0	68072	108	0	27613	363472	138
13-TY-17	0	606	18	157	0	0	8	0	1	191	2	4	497	4765	1
13-TY-25	0	6	2	1	0	0	23	0	0	0	1	2	19	78	0
13-TY-26	0	430	2	2	139	38	25	1	21	435	7	10	7068	33746	7
13-TY-27	0	78	4	0	80	62	81	25	21	153	22	45	2245	11350	7
13-TY-28	0	29	2264	1	0	100	0	1	11	208	0	0	5392	22791	56
13-TY-29	0	1	0	0	0	13	53	0	10	68	0	0	15027	106815	10
13-TY-30	1701	1679	6	0	0	0	0	0	1	351	0	0	9004	170712	10571
13-TY-31	0	832	13542	1	10	38	0	31	10	983	1	1	9386	49942	49
13-TY-32	0	1671	14	733	17	45	0	0	54	7731	0	0	41484	151710	2376

Appendix 27: Genera of the sequence data of the surface transect.

	Amp	Aul	Cym	Lem	Pin	Pla	Sel	Stn	Sts	Uro
13-TY-01	90	191	0	738	7	103	40	0	37220	0
13-TY-02	147	1	252	254	3062	565	1141	2691	60102	0
13-TY-03	38	230	0	49	872	7	399	1053	21577	4
13-TY-04	12250	19	0	0	12658	0	1806	3269	63089	0
13-TY-05	4322	0	0	2195	1785	0	1323	4089	62782	0
13-TY-06	1521	0	1316	708	33	2209	994	977	31818	0
13-TY-07	687	4	5	169	1481	101	293	34	12395	0
13-TY-08	412	1	0	0	7	129	2	0	78427	1
13-TY-09	298	2	0	0	2854	761	8032	1490	301621	3633
13-TY-10	223	13	3	53	239	0	292	15	37927	0
13-TY-11	4	10774	0	13	19	16	120	7	59204	0
13-TY-12	2	1215	0	13	37	5	114	1	15442	0
13-TY-13	597	1	1	86	4311	40	1049	93	9317	0
13-TY-14	128	90	0	1637	97	33	196	0	22192	0
13-TY-15	211	0	0	42	23	4	108	3	13785	0
13-TY-16	0	12475	0	0	0	0	53	0	80591	0
13-TY-18	57	32	31	0	121	0	165	33	77165	0
13-TY-19	114	3015	30	0	10583	2	114	1400	32937	0
13-TY-20	168	6952	180	0	4949	1	2637	381	24157	0
13-TY-21	2	781	0	0	288	29	280	73	88059	0
13-TY-22	5	922	0	12	67	3	286	6	16732	0
13-TY-23	3963	1751	2614	0	58343	0	27158	2422	307086	3
13-TY-24	0	19183	0	0	903	1	69378	207	391085	138
13-TY-17	4	1162	0	0	36	10	218	24	5262	1
13-TY-25	2	12	0	0	32	0	1	3	97	0
13-TY-26	19	1393	139	38	125	3	511	145	40814	7
13-TY-27	140	153	80	62	228	50	294	107	13595	7
13-TY-28	110	2313	0	100	159	91	330	23	28183	56
13-TY-29	1036	1	0	13	61	799	2185	0	121842	10
13-TY-30	1701	3354	0	0	2	7	2060	2	179716	10571
13-TY-31	50	15312	10	38	252	98	1040	30	59328	49
13-TY-32	10	30493	17	45	618	26	7966	944	193194	2376

Appendix 28: Species of core 11-CH-12A of the genetic data.

	Ast.for	Aul.dis	Aul.sub	Cyc.bod	Cyc.oce	Sts.con	Sts.ell	Tab.flo
0.5	15	497	253	125	0	17555	125327	2
14.5	6847	0	3621	33	0	998	144121	2297
28.5	2076	0	1370	1010	0	4363	58926	513
42.5	1037	24326	231	67	0	3691	70925	0
55.5	0	2	104751	0	40256	78	34058	0
69.5	8996	0	25890	0	27916	29	176	6839
83.5	0	0	6	0	0	198639	250134	30187
95.5	0	4	4	0	0	992	589832	0
112	1	0	5	639949	0	19	59	0
126	0	0	4	0	0	23	13582	0

Appendix 29: Genera of core 11-CH-12A of the genetic data.

	Ast	Aul	Cyc	Gom	Nit	Pin	Sts	Tab
0.5	15	800	125	0	20	2988	142882	2
14.5	6847	4200	33	0	20	378	145119	2297
28.5	2076	1370	1010	625	20	24116	63289	513
42.5	1037	26097	67	0	20	9069	74616	0
55.5	0	104753	40256	19866	20	8	34136	0
69.5	8996	34203	27916	0	64	7	205	6839
83.5	0	7	0	31788	20	80463	448773	30187
95.5	0	8	0	0	20	8	590824	0
112	1	5	639949	0	20	11	78	0
126	0	4	0	0	20	27161	13605	0

Appendix 30: Species of core 11-CH-06D of the genetic data.

	Aul.dis	Aul.sub	Aul.val	Nav.rad	Pin.acu	Pin.sub	Sts.con	Sts.ell	Tab.flo
1	3379	3360	1371	0	0	0	9119	47742	4
16	5702	10	0	8	0	1266	37783	29443	0
29	8528	0	1547	0	11520	0	61824	73316	1601
41	0	0	0	0	0	0	15430	210634	21594
59	22	13171	0	5681	0	4351	33394	43616	1
73	494	26	0	8	0	0	420284	109237	879
93	7	0	0	0	0	0	0	3	0
108	1	1	0	0	0	0	14	159	5
124	0	0	0	0	2	0	1	16	0
140	4	6	0	106926	0	0	33	21	0
155.5	0	3	0	0	0	1	1	37	7

Appendix 31: Genera of core 11-CH06D of the genetic data.

	Aul	Nav	Pin	Sel	Stn	Sts	Tab
1	10218	0	946	61	1200	56861	4
16	6649	8	4196	3128	7493	67226	0
29	10086	0	34819	3050	64492	135140	1601
41	1	0	0	0	4800	226064	21594
59	13193	5681	9432	2500	133046	77010	1
73	11304	8	3583	856	6019	529521	879
93	10	0	129472	0	29595	3	0
108	2	0	6	1	1	173	5
124	0	0	5	1	2	17	0
140	13	106926	2630	91970	67753	54	0
155.5	10	0	9	0	4	38	7

A.11 Morphological data sets

Appendix 32: Morphological identified species of the lake transect.

	Ach.min	Amp.lib	Amp.ova	Aul.alp	Aul.dis	Aul.per	Cyc.bod	Cyc.iri	Cyc.rad	Cyc.tri	Dip.ell	Enc.sil	Eun.fab	Eun.pra	Fra.cap	Kar.lat	Plc.elg	Pln.lan	Pse.bre	Pse.par	Pse.pse	Ros.pus	Sel.pup	Stn.anc	Stn.smi	Sts.ven	Sta.pin	Tab.flo	Tet.gla
13TY01	0	2	6	20	0	0	0	6	2	36	6	2	0	2	0	10	2	4	50	0	66	8	2	0	2	70	128	2	0
13TY02	4	8	0	2	0	0	0	0	18	0	0	0	2	0	0	10	2	8	0	0	6	4	8	6	0	122	280	2	0
13TY03	0	2	2	0	0	2	0	0	10	0	2	2	0	2	0	0	6	4	102	2	64	0	4	2	4	50	242	2	2
13TY04	32	8	6	2	0	0	0	0	0	0	0	8	4	16	40	0	2	0	2	0	0	6	20	26	0	6	12	180	2
13TY05	0	6	4	0	0	12	0	0	4	0	4	0	0	0	0	0	0	8	30	2	30	0	2	6	4	58	276	0	0
13TY06	2	0	4	2	0	8	0	0	72	0	0	0	0	0	28	22	0	8	20	2	4	10	10	0	0	56	120	2	0
13TY07	0	14	16	0	0	0	10	0	0	8	14	2	0	0	0	4	18	0	28	10	72	0	12	2	6	14	180	4	2
13TY08	0	4	4	0	0	0	42	0	0	4	28	0	0	2	6	0	0	2	18	6	114	0	2	0	2	48	130	2	2
13TY09	4	0	4	0	0	0	0	4	0	0	10	4	6	14	2	0	4	4	4	4	54	8	10	2	8	52	126	82	0
13TY10	0	16	10	0	0	0	0	16	0	2	10	2	0	0	0	0	2	0	8	8	66	4	12	4	2	0	250	16	0
13TY11	2	2	4	34	0	0	0	0	0	10	2	0	8	10	24	0	2	2	26	2	26	10	2	2	2	64	206	60	16
13TY12	2	2	0	18	0	0	0	0	0	0	0	0	0	4	2	0	6	0	58	4	20	6	2	2	4	210	208	16	2
13TY13	0	0	2	0	0	0	84	48	0	0	10	0	0	2	0	0	0	2	24	4	52	0	4	6	8	26	82	8	0
13TY14	0	2	12	6	0	0	2	14	0	52	18	0	0	2	0	0	2	4	34	0	136	0	2	0	2	60	74	4	0
13TY15	0	2	4	0	0	0	14	18	0	10	12	0	2	0	0	0	0	2	18	4	148	0	2	0	2	90	110	2	0
13TY16	0	0	4	86	0	0	0	18	0	34	4	0	4	4	0	0	2	0	42	0	26	0	2	0	2	90	68	6	18
13TY17	4	0	4	0	52	114	0	0	0	0	0	2	4	0	0	2	4	0	4	2	10	4	4	2	4	48	4	2	134
13TY18	4	4	2	0	4	0	0	0	0	4	0	0	0	0	0	0	0	0	24	0	44	12	4	0	0	334	54	4	2
13TY19	4	0	0	0	12	2	0	0	0	0	0	0	2	2	0	0	0	0	54	0	32	30	4	10	0	230	22	8	24
13TY20	52	6	0	0	0	0	0	0	0	0	0	8	16	4	0	0	0	0	4	0	14	42	28	20	0	60	2	38	0
13TY21	2	0	10	2	0	0	0	0	0	0	2	0	0	0	0	0	8	6	34	0	60	2	4	4	4	290	64	6	0
13TY22	0	0	2	40	0	6	0	0	0	0	6	0	8	0	4	0	6	0	70	0	0	0	16	0	0	0	200	8	18
13TY23	4	2	8	0	22	0	0	0	0	0	0	4	14	0	0	0	2	0	40	0	10	6	18	4	0	278	6	10	0
13TY24	0	0	6	28	20	6	0	0	0	0	2	0	18	6	2	0	6	0	30	0	24	4	0	2	0	204	2	62	4
13TY25	0	0	4	0	0	0	0	0	0	0	0	0	4	0	0	0	4	0	18	0	46	4	4	2	2	322	38	12	0
13TY26	8	2	0	12	0	0	0	0	0	0	2	0	0	0	2	0	4	10	32	2	292	0	18	2	2	86	14	8	0
13TY27	6	2	0	0	0	0	0	0	12	0	0	0	0	0	0	0	10	0	84	8	20	4	20	4	2	260	76	12	0
13TY28	0	0	2	0	0	0	0	0	0	2	2	0	0	2	0	0	2	0	18	0	22	0	4	0	0	84	66	32	0
13TY30	6	0	6	0	0	0	0	14	2	0	6	8	0	0	0	0	0	0	30	2	30	0	0	0	0	120	170	12	0
13TY31	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	44	6	28	8	10	2	6	180	48	28	0
13TY32	0	4	4	0	0	0	0	0	0	4	0	0	0	2	18	0	2	0	30	0	164	16	8	0	8	64	32	68	4

Appendix 33: Morphological identified genera of the lake transect.

	Ach	Amp	Aul	Cyc	Dip	Enc	Eun	Frau	Kar	Plc	Pln	Pse	Ros	Sel	Stn	Sts	Sta	Tab	Tet
13TY01	0	8	20	44	6	2	2	0	10	2	4	116	8	2	2	70	128	2	0
13TY02	4	8	2	18	0	0	2	0	10	2	8	6	4	8	6	122	280	2	0
13TY03	0	4	2	10	2	2	2	0	0	6	4	168	0	4	6	50	242	2	2
13TY04	32	14	2	0	0	8	20	40	0	2	0	2	6	20	26	6	12	180	2
13TY05	0	10	12	4	4	0	0	0	0	0	8	62	0	2	10	58	276	0	0
13TY06	2	4	10	72	0	0	0	28	22	0	8	26	10	10	0	56	120	2	0
13TY07	0	30	0	18	14	2	0	0	4	18	0	110	0	12	8	14	180	4	2
13TY08	0	8	0	46	28	0	2	6	0	0	2	138	0	2	2	48	130	2	2
13TY09	4	4	0	4	10	4	20	2	0	4	4	62	8	10	10	52	126	82	0
13TY10	0	26	0	18	10	2	0	0	0	2	0	82	4	12	6	0	250	16	0
13TY11	2	6	34	10	2	0	18	24	0	2	2	54	10	2	4	64	206	60	16
13TY12	2	2	18	0	0	0	4	2	0	6	0	82	6	2	6	210	208	16	2
13TY13	0	2	0	132	10	0	2	0	0	0	2	80	0	4	14	26	82	8	0
13TY14	0	14	6	68	18	0	2	0	0	2	4	170	0	2	2	60	74	4	0
13TY15	0	6	0	42	12	0	2	0	0	0	2	170	0	2	2	90	110	2	0
13TY16	0	4	86	52	4	0	8	0	0	2	0	68	0	2	2	90	68	6	18
13TY17	4	4	166	0	0	2	4	0	2	4	0	16	4	4	6	48	4	2	134
13TY18	4	6	4	4	0	0	0	0	0	0	0	68	12	4	0	334	54	4	2
13TY19	4	0	14	0	0	0	4	0	0	0	0	86	30	4	10	230	22	8	24
13TY20	52	6	0	0	0	8	20	0	0	0	0	18	42	28	20	60	2	38	0
13TY21	2	10	2	0	2	0	0	0	0	8	6	94	2	4	8	290	64	6	0
13TY22	0	2	46	0	6	0	8	4	0	6	0	70	0	16	0	0	200	8	18
13TY23	4	10	22	0	0	4	14	0	0	2	0	50	6	18	4	278	6	10	0
13TY24	0	6	54	0	2	0	24	2	0	6	0	54	4	0	2	204	2	62	4
13TY25	0	4	0	0	0	0	4	0	0	4	0	64	4	4	4	322	38	12	0
13TY26	8	2	12	0	2	0	0	2	0	4	10	326	0	18	4	86	14	8	0
13TY27	6	2	0	12	0	0	0	0	0	10	0	112	4	20	6	260	76	12	0
13TY28	0	2	0	2	2	0	2	0	0	2	0	40	0	4	0	84	66	32	0
13TY30	6	6	0	16	6	8	0	0	0	0	0	62	0	0	0	120	170	12	0
13TY31	0	4	0	0	0	0	0	0	0	12	0	78	8	10	8	180	48	28	0
13TY32	0	8	0	4	0	0	2	18	0	2	0	194	16	8	8	64	32	68	4

Appendix 34: Morphological identified species of core 11-CH-12A.

	Amp.ova	Aul.alp	Aul.sub	Cyc.iri	Fra.cap	Fra.con	Pin.nod	Pin.obs	Pse.bre	Pse.pse	Stn.anc	Sts.ven	Sta.pin	Tab.fen	Tab.flo
0.5	0	0	121	32	0	2	2	2	10	60	0	28	114	14	2
14.5	2	0	8	2	0	0	10	32	102	34	38	84	118	0	6
28.5	8	0	2	0	0	40	12	8	52	80	0	106	144	0	8
42.5	2	78	2	0	12	30	16	10	32	4	8	0	98	0	6
55.5	0	74	0	196	14	0	0	0	4	0	2	20	42	20	78
69.5	0	18	0	98	0	0	0	2	4	0	0	0	2	2	68
83.5	8	0	0	0	8	2	2	4	12	2	6	26	44	12	54
95.5	0	0	12	0	2	2	0	0	6	2	0	14	34	0	2
111.5	10	0	0	274	0	0	2	0	2	4	0	4	14	20	24
124.5	46	0	0	0	0	140	4	0	54	2	0	46	0	0	3

Appendix 35: Morphological identified genera of core 11-CH-12A.

	Amp	Aul	Cyc	Fra.	Pin	Pse	Stn	Sts	Sta	Tab
0.5	0	121	32	2	4	70	0	28	114	16
14.5	2	8	2	0	42	136	38	84	118	6
28.5	8	2	0	40	20	132	0	106	144	8
42.5	2	80	0	42	26	36	8	0	98	6
55.5	0	74	196	14	0	4	2	20	42	98
69.5	0	18	98	0	2	4	0	0	2	70
83.5	8	0	0	10	6	14	6	26	44	66
95.5	0	12	0	4	0	8	0	14	34	2
111.5	10	0	274	0	2	6	0	4	14	44
124.5	46	0	0	140	4	56	0	46	0	3

Appendix 36: Morphological identified species of core 11-CH06D.

	Amp.ova	Aul.dis	Aul.lir	Aul.per	Aul.val	Eun.fab	Eun.inc	Eun.pra	Gyr.acu	Han.amp	Pin.bor	Pin.int	Pse.bre	Ros.pus	Sel.pup	Sts.ven	Sta.pin	Tab.fen	Tab.flo	Tet.gla
0.5	2	20	0	8	10	10	0	4	2	1	0	32	0	16	20	72	20	10	32	6
16.5	14	44	26	0	8	12	6	8	10	0	0	2	0	8	6	52	70	2	26	14
28.5	6	78	56	0	10	10	28	2	0	0	0	0	28	0	6	14	32	0	40	22
40.5	14	114	30	20	0	6	6	0	6	2	2	0	42	0	4	14	118	2	16	6
58.5	0	126	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	10	10	4
72.5	2	30	0	18	14	4	6	2	1	0	0	8	10	8	20	62	10	8	68	2
93.5	6	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	10	0	2	0
107.5	10	0	0	0	0	0	0	22	32	26	20	0	0	0	0	0	0	0	4	0
123.5	0	0	0	0	0	0	4	6	0	4	0	0	0	0	0	0	0	0	0	0
139.5	42	0	0	0	0	0	0	2	28	2	2	0	2	0	26	30	22	0	68	0
155.5	0	0	0	0	0	0	8	0	0	0	4	2	0	10	4	16	4	0	12	0

Appendix 37: Morphological identified genera of core 11-CH-06D.

	Amp	Aul	Eun	Fra.	Gyr	Han	Pin	Pse	Ros	Sel	Sts	Sta	Tab	Tet
0.5	2	38	14	0	2	1	32	0	16	20	72	20	42	6
16.5	14	78	26	0	10	0	2	0	8	6	52	70	28	14
28.5	6	144	40	0	0	0	0	28	0	6	14	32	40	22
40.5	14	168	12	0	6	2	2	42	0	4	14	118	18	6
58.5	0	128	0	0	0	2	2	0	0	0	0	0	20	4
72.5	2	62	12	10	1	0	8	10	8	20	62	10	76	2
93.5	6	2	2	0	0	0	0	0	0	0	0	10	2	0
107.5	10	2	22	2	32	26	20	0	0	0	0	0	4	0
123.5	0	0	10	0	0	4	0	0	0	0	0	0	0	0
139.5	42	0	2	8	28	2	2	2	0	26	30	22	68	0
155.5	0	0	8	0	0	0	6	0	10	4	16	4	12	0

Acknowledgements

First of all I would like to thank my supervisor Professor Ulrike Herzschuh from the Alfred Wegener Institute for the possibility of participating in the expedition to Siberia, the opportunity of making my Master thesis in her working group, her support in the statistical analysis and most of all for her support in my career development.

I thank Professor Ralph Tiedemann from the University of Potsdam for being the second referee.

A special thanks goes to Professor Luidmila Pestryakova from the North-Eastern Federal University of Yakutsk, Russia, for teaching me in diatom determination, sharing her knowledge and her constant enthusiasm during long periods on the microscopes.

I would like to thank Dr. Kathleen Stoof-Leichsenring for being my tutor in the laboratory and her idea for the Master project.

A very special thanks goes to Bastian Niemeyer for his help with R and Tilia, and his remarks and suggestions on the introduction and the results of my thesis. Your comments were very welcome.

I thank Heike Zimmermann for being a wonderful office mate. You were always listening and offering your help. Thanks for the great atmosphere, the chocolate and all the fun we had during the breaks. Your final remarks were really helpful.

I also want to thank Stefan Kruse for his help and the joy he is spreading during the work all the time. You really make one feel positive about everything.

Furthermore, I would like to thank the rest of the AWI working group for the great atmosphere and support.

Finally, I want to thank my family and friends that always support me. A special thanks goes to Sophie who always makes me laugh and to Michele for being my best friend.

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Potsdam

Date Signature