

UNIVERSITY OF APPLIED SCIENCE

**Genetic and morphologic characterisation of possible cryptic species
of the diatom *Fragilariopsis kerguelensis*.**

BACHELOR THESIS
MARITIME TECHNOLOGIES

In cooperation with Alfred-Wegener-Institut



BARBARA GLEMSEK
MATR.NR: 33017

First examiner: Prof. Dr. Stefan Wittke HS Bremerhaven
Second examiner: Dr. Bánk Beszteri AWI

Bremerhaven, June 6, 2018

Abstract

The diatom species *Fragilariopsis kerguelensis* is endemic to the Southern Ocean where it plays a key role in the ocean silica cycle due to its heavily silicified cell walls. Frustules from dead cells can sink to the ocean floor and therefore contribute about 90% to the diatom frustules making up the Antarctic opal belt, a band of heavily silicified sediments below the Antarctic Circumpolar Current. Recent studies showed that two morphotypes of *F. kerguelensis* can be found in core samples originating from the Southern Ocean. The morphotypes can be distinguished by the morphometric descriptor rectangularity and the abundance of the morphotypes can be linked to the origin of the samples from glacial or interglacial periods.

It is, however, so far unknown whether these morphotypes also occur in current diatom assemblages of the Southern Ocean; if yes, how their occurrence is influenced by environmental conditions; and whether they represent different species, or rather phenotypic plasticity within a single species.

For answering these questions diatom valves in water and sediment samples from the Southern Ocean, preserved on microscopic slides were analysed morphologically using a semi automated morphometry workflow. Histograms of the rectangularity were then plotted which all showed a bimodal distribution proving the existence of the two morphotypes in recent samples.

In the next step the biogeographic distribution pattern of the two morphotypes was assessed by plotting the rectangularity distribution at each sample station. Whereas one morphotype was dominant in the northernmost samples, its dominance decreased towards the south, and the other morphotype became dominant in the southernmost locations investigated. This pattern could also be linked to the Sea Surface Temperature with a regression.

After the previous findings the question arose if the morphotypes could be cryptic species or if they occur due to phenotypic plasticity within a single species. To answer this a genetic assessment with diatom strains showing different rectangularity values, isolated from the Southern Ocean and kept in live cultures at the Alfred-Wegener-Institut was done. Their ribosomal internal transcribed spacer (ITS) regions were amplified using PCR and then analysed with Sanger sequencing. The results made it possible to identify three potential species based on their genetic differences. One semi-cryptic species was solely made up of morphotype B whereas morphotype A was divided in two genetic clusters representing two cryptic species.

Mating experiments were carried out to uncover possible reproductive barriers between the cryptic species. The results showed, that strains being assigned the same cryptic species commonly sexually reproduce, whereas sexual reproduction between strains from different cryptic species was only observed in a few exceptions.

In summary, the two investigated morphotypes of *Fragilariopsis kerguelensis* can be found in the Southern Ocean today, with one dominating at higher, the other at lower Sea Surface Temperatures. Comparison of ribosomal ITS sequences and mating experiments indicate that the two morphotypes seem to belong to three different species, one semi-cryptic species belonging to one and two cryptic species to the other morphotype.

Contents

Abstract	ii
1 Introduction	1
1.1 Structure and life cycle of diatoms	1
1.2 Role of diatoms in the Southern Ocean	2
1.3 Methods for species delimitation in diatoms	3
1.4 Aims of the thesis	4
2 Materials and methods	6
2.1 Morphologic characterisation	6
2.2 Analysis of ITS sequences	9
2.2.1 DNA extraction	9
2.2.2 PCR	9
2.2.3 Purification of PCR products	10
2.2.4 Sequencing	10
2.2.5 Assembly of DNA sequences	11
2.2.6 Alignment of DNA sequences	12
2.3 Mating experiments	12
2.4 Statistical analysis	12
2.4.1 Evaluation of rectangularity classes	12
2.4.2 Regression over percentage low rectangularity class valves and the Sea Surface Temperature	12
2.4.3 Significant differences in rectangularity between ITS groups	13
3 Results	15
3.1 Morphologic characterisation by valve rectangularity	15
3.2 Dependence of valve rectangularity on water masses	17
3.3 Dependences of morphotypes on latitude	20
3.4 Dependence of morphotypes on Sea Surface Temperature	21
3.5 Analysis of ITS sequences	23
3.6 Mating experiments	25
3.7 Statistical analysis of rectangularity within ITS groups	26
4 Discussion	28
4.1 Can the two morphotypes be found in recent water samples from the Southern Ocean?	28
4.2 Is there a biogeographic distribution pattern for the morphotypes?	28
4.3 Are the two morphotypes cryptic species?	30
5 Conclusion and outlook	36
Bibliography	37

Appendix	40
A Additional results	41
A.1 Morphology	41
A.2 Genetic characterisation	43
B Acknowledgements	45
C Declaration of Originality	46
D Data medium	47

1 Introduction

Diatoms are a group of unicellular photosynthetic protists. They can be found in any wet, moist or aquatic environment, in lakes or rivers, as well as in the ocean or in damp patches of moss. Pelagic species living freely floating in the water column can be found as well as benthic species living attached to a substrate. There are about 30 000 [1] known species of diatoms, but estimates range up to 10 000 000 species [2], making diatoms one of the most species-rich group of algae. In the oceans they form a major part of the phytoplankton, therefore they are important primary producers, contributing about 50-70% [3] to global primary production and thus forming the base of the aquatic food web.

Due to their abundance diatoms also play an important role in the ocean carbon cycle. During diatom blooms in the photic zone (ocean zone where light is available) CO_2 is assimilated by diatoms cells and converted to organic matter by photosynthesis. The CO_2 originates from the atmosphere and is taken up by the ocean water to reach an equilibrium of CO_2 between the ocean surface layer and the atmosphere. Dead diatom cells can sink and transport organic carbon out of the surface layer into the deep sea. This results in a deficit of CO_2 in the surface layer, leading to a further uptake of CO_2 from the atmosphere [4].

1.1 Structure and life cycle of diatoms

A unique feature of diatoms is their hardened cell wall, reinforced by silica, called frustule. The structure of a frustule is shown in figure 1.1.

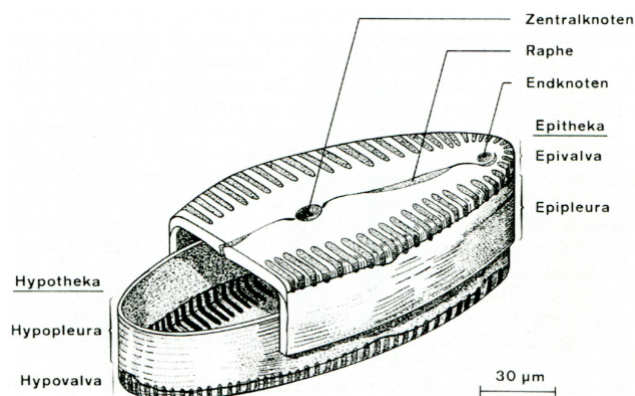


Fig. 1.1: General structure of a diatom frustule [5].

In general it can be described as resembling a Petri dish with two parts that fit into each other. The larger part is called epitheca and the smaller one hypotheca. Each theca is built of a terminal part called valve and a rim called epicingulum (or epipleura) in case of the epitheca, and hypocingulum (or hypopleura) in case of the hypotheca. The two rims together are called girdle and are composed of several girdle bands [6]. Because of the silica frustule the live cycle of diatoms differs from other microalgae. Usually diatoms reproduce by vegetative cell division (mitosis). When a cell divides, each newly formed cell receives one half of the parental cell which will form the epitheca in the new cell, and rebuilds a new hypotheca.

This process is shown in figure 1.2. The cell highlighted in green in figure 1.2 is the parental cell which divides, to form the two daughter cells. Each daughter cell forms a new hypotheca shown in yellow in figure 1.2. In the second generation the newly formed hypothecae are shown in red.

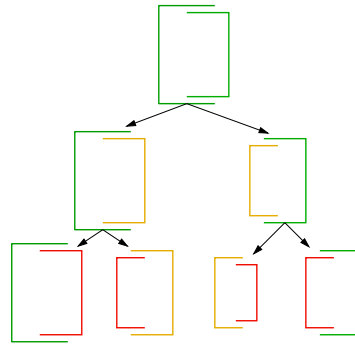


Fig. 1.2: Schematic representation of cell division and reduction of cell size.

This division scheme leads to a progressive reduction of the average cell size in the population due to the fact that one daughter cell is always smaller than the parental cell [7].

This creates a problem, because once the cells have reached a minimum size, the restoration of the original cell size is necessary to ensure survival. Sexual reproduction is necessary in most diatom species to restore the cell size. In order to be able to reproduce sexually the diatom cell has to undergo meiosis I and II to produce haploid gametes which are then released from the frustule [8]. The gametes conjugate and form a zygote. The zygote starts growing and is then referred to as auxospore. The auxospore is stabilised by thin silica bands across the transversal axis, allowing a longitudinal expansion. When the maximum elongation is reached, a new silica frustule is built within the auxospore [7], enclosing the newly built so called initial cell. This initial cell is back to the largest possible size and can reproduce by vegetative cell division until the minimum possible cell size is reached and sexual reproduction is needed again.

1.2 Role of diatoms in the Southern Ocean

Due to the silica frustule diatoms can only grow in habitats that offer silica as an available nutrient. Most diatoms need roughly as much silica as nitrate [4], but some species stand out due to a heavily silicified frustule and for this reason they have a higher demand of silica [4]. One of these species is *Fragilariopsis kerguelensis* (*F. kerguelensis*), a diatom endemic to the Southern Ocean which is the key species in this study and is one of the dominant diatoms in the Southern Ocean. In their normal vegetative state the cells of *F. kerguelensis* chain up forming long ribbon shaped colonies, which break up for sexual reproduction where the cells separate. The cells are easy to distinguish from other diatoms under the microscope due to their heavily silicified frustules.

They sink to the ocean floor after the diatom's death because the heavily silicified frustules are hard to dissolve. Therefore frustules of *F. kerguelensis* make up about 90% of the Antarctic Opal Belt [9]. This is a band of sediments which is extremely rich in silica, located on the ocean floor in the area below the Antarctic Circumpolar Current (ACC), between the Polar Frontal Zone and the winter sea ice edge [10], in the Southern Ocean.

Due to the high abundance in the Antarctic Opal Belt *F. kerguelensis* plays an important role in the paleoenvironmental reconstruction of the Southern Ocean [9]. In this field of study, the aim of the researchers is it to reconstruct the environmental parameters of the oceans from the past. This is done using a transfer function, which is based on the assumption that the composition of species living in the surface waters is preserved in the sediments. Each composition of species can be linked to certain environmental conditions and therefore the study of sediment cores can

reveal the environmental parameters of the past [10].

But not only the abundance of *F. kerguelensis* frustules can reveal information about the past. Research by Cortese et al. [11] shows that the valve size of *F. kerguelensis* changes in sediment cores between glacial and interglacial times.

A more recent study by Kloster et al. [12] shows a significant change in the morphometric descriptor rectangularity, depending on the origin of the sediment sample from a glacial or interglacial period. The rectangularity is a morphometric descriptor in diatom research and can be described as the ratio of the valve area to the area of its enclosing rectangle [13].

The study by Kloster et al. [12] revealed, that two morphotypes of *F. kerguelensis* can be distinguished by the rectangularity in samples originating from a sediment core. These findings raised the question whether these two morphotypes can also be found in water samples and if they could be cryptic species. This led to this thesis and the attempt to answer the following research questions:

- 1. Are the two morphotypes found in sediment samples detectable in recent water samples from the Southern Ocean?**
- 2. If the two morphotypes can be found in the Southern Ocean, is there a pattern detectable in their geographic distribution?**
- 3. Are the two morphotypes cryptic species or are the differences in morphology due to phenotypic plasticity?**

To answer the first research question a morphologic characterisation of diatom valves was done with a semiautomatic workflow, also used by Kloster et al. when discovering the two morphotypes. Data gained from this was also used for the assessment of the biogeographic distribution. To answer the last research questions several methods were used which are shortly described below.

1.3 Methods for species delimitation in diatoms

To delimit diatoms species it is common to use several approaches because using only a single approach might not lead to a clear result. The three methods used the most are described below. The first method is the delimitation by morphology which is the oldest and most commonly used method. The approach is still popular today, because samples can easily be prepared and preserved for a long time on microscopic slides [14]. The development of electron microscopy in 1931 [15] made this method more precise due to increased resolution. Nevertheless frustule morphology alone is not always sufficient for species identification.

In cases where species can not be differentiated by morphology they are called cryptic species. These species then differ in habitat, physiology or life cycle characteristics [16], and are most commonly discovered using molecular methods.

Another obstacle for delimitation by morphology is phenotypic plasticity, which describes the ability of organisms to display different phenotypes depending on the environmental conditions [17]. This is particularly important when the species distribution range spans different environmental conditions, which suggest a possibility for phenotypic plasticity.

The classic approach of morphologic characterisation is the comparison of details that can be seen under the microscope and the delimitation due to these details. Additional information can

be gained with a quantitative approach where morphologic features are expressed in numbers, for example with the rectangularity value of the frustules.

The second method is the genetic delimitation, which gained popularity in recent decades, when it became more available. By using the Polymerase Chain Reaction (PCR) any part of the genome can be amplified and sequenced. In diatoms one of the most used molecular markers is a region of the ribosomal operon, the internal transcribed spacer (ITS) [14].

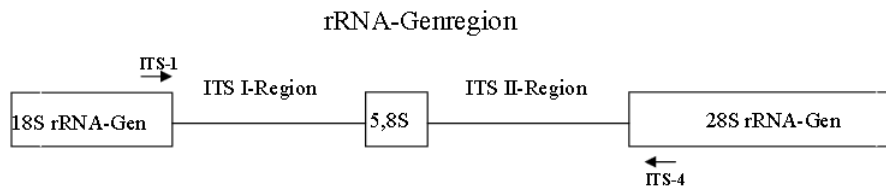


Fig. 1.3: Structure of the internal transcribed spacer region (ITS) of the ribosomal DNA. Standard primers ITS1 and ITS4 are depicted by arrows [18].

The ITS regions are situated between the 18S and 28S genes on the ribosomal DNA (see figure 1.3). During post-translational processing the ITS regions are cut out and degraded to nucleotides [19]. ITS1 and ITS4 primer pairs for any organism are available [14], and because of their multi-copy nature they are easily amplified. They also show high variation even between closely related species, making them a good marker for species identification.

The third method for species delimitation is the exploration of mating compatibility between possible different species. This approach can be related to the biological species concept, which states that a group of individuals that can breed together belong to one species, whilst they are incapable to interbreed with organisms belonging to an other species [20]. To test this, mating experiments are used to reveal reproductive barriers between the individuals, marking individuals incapable of interbreeding as different species.

1.4 Aims of the thesis

The aims of the thesis were to answer the research questions stated in section 1.2 and to evaluate the quality of the obtained results.

If this study proves that the species *F. kerguelensis* can be separated into cryptic species, it could help to improve paleoenvironmental studies which use this diatom as a proxy. The models that are drawn up based on sediment samples could be refined by also taking the spread of each cryptic species and its abundance into account.

To answer the research questions three different approaches were used. A morphologic characterisation of diatom valves as described by Kloster et al. [21] to answer the first questions regarding the two morphotypes as well as their distribution pattern. All three methods for species delimitation which are described above were used to answer the question if the morphotypes are cryptic species.

Figure 1.4 shows the experimental structure of the thesis. The structure is separated into three parts representing the three approaches used.

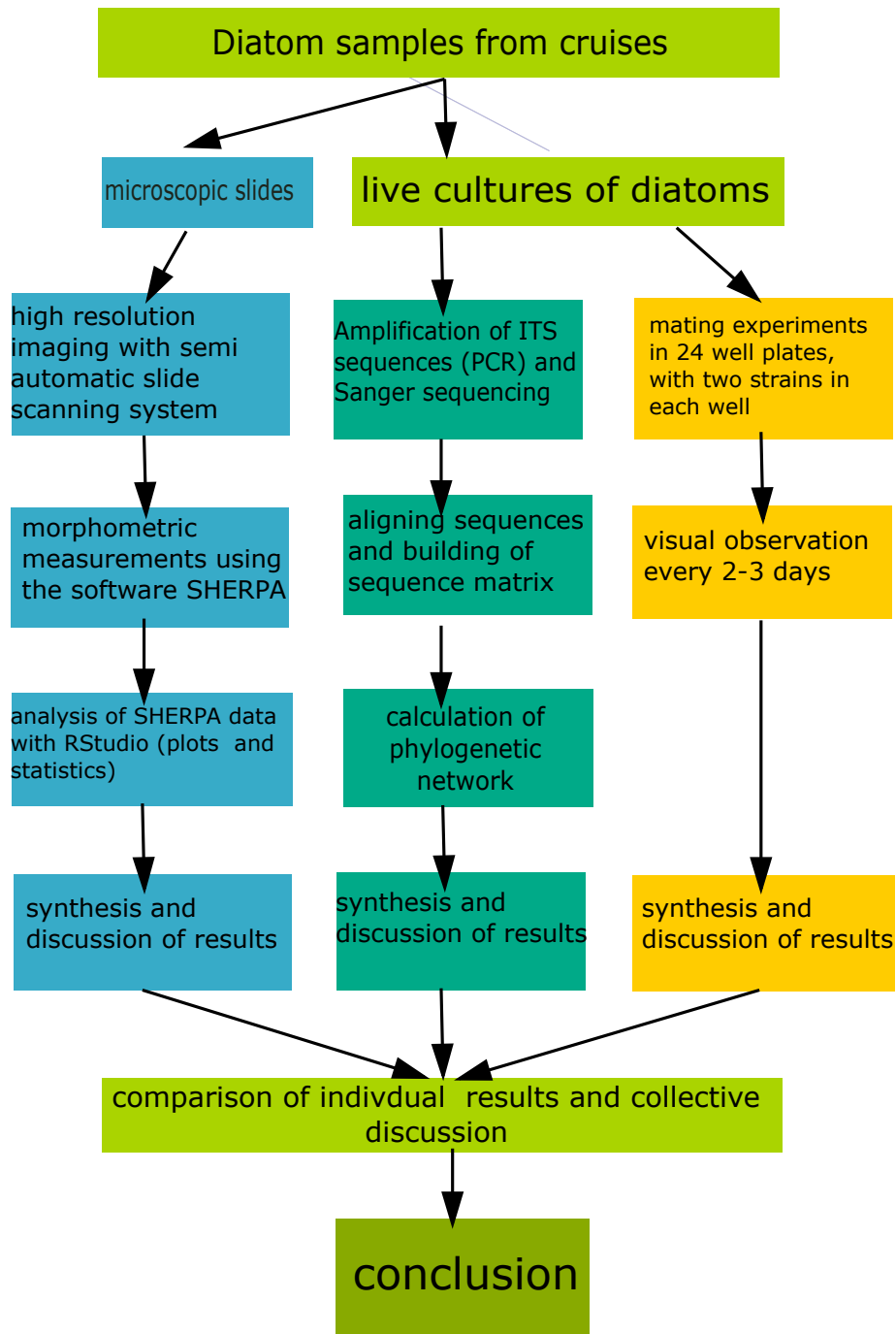


Fig. 1.4: Experimental structure of the thesis. The three parts represent the three approaches used to answer the research questions.

2 Materials and methods

2.1 Morphologic characterisation

For the morphologic characterisation of the key diatom species, valves preserved on microscopic slides were studied.

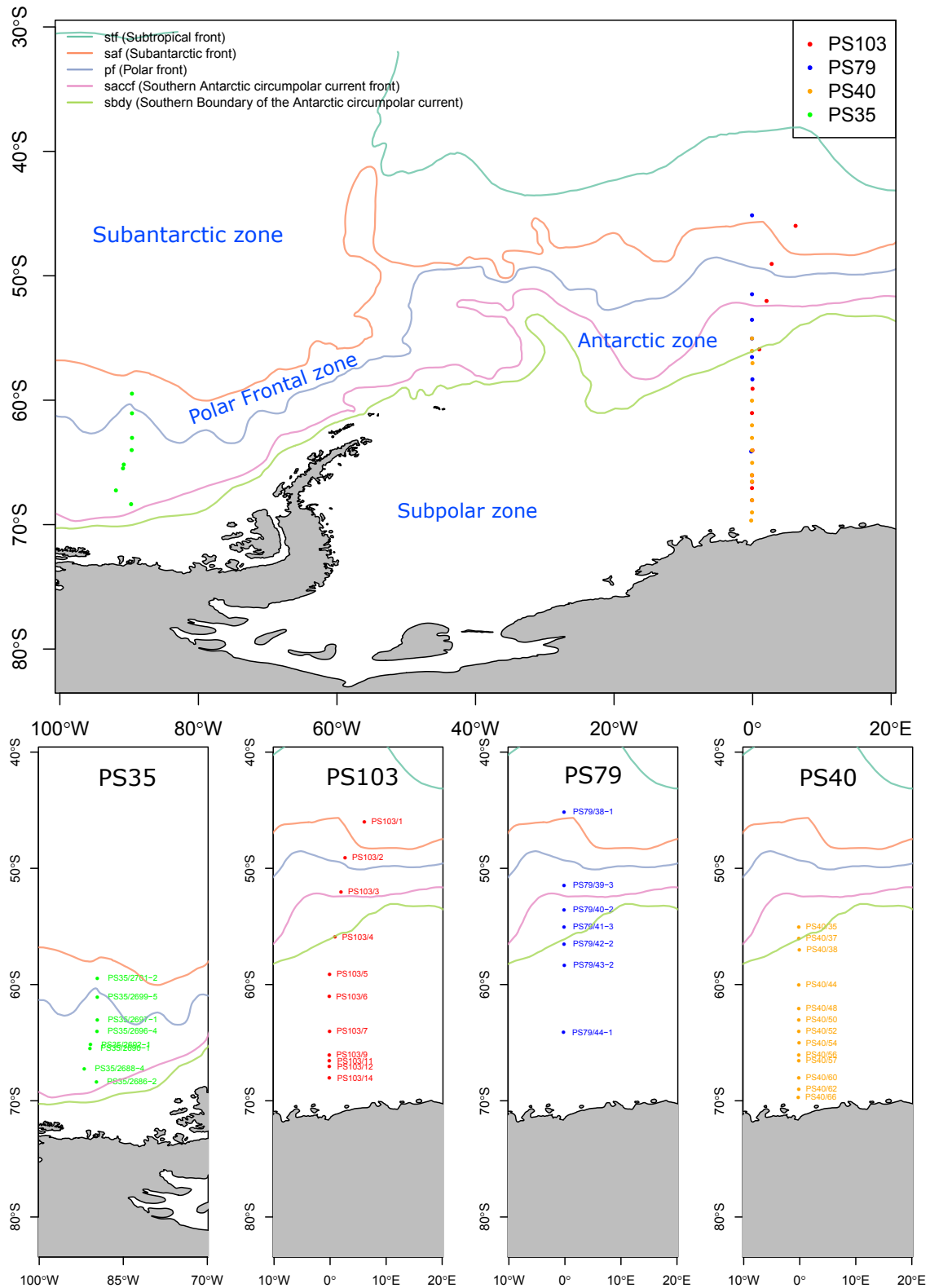






Fig. 2.1: Map of the Southern Ocean showing an overview of the four transects as well as a closer look at the transects, with the sample station names included. The ocean fronts and water masses are also shown. Colour coding according to table 2.1.

In total 39 microscopic slides were analysed. The samples preserved on the slides came from sample stations in the Southern Ocean which are the depicted in figure 2.1. The samples came from four transects in the Southern Ocean and were collected during four Polarstern cruises in 1995, 1996, 2011 and 2016 (see table 2.1). The colour given to each cruise or transect in table 2.1 refer to the one used in figure 2.1 and will be maintained throughout the whole thesis.

Tab. 2.1: Name of the Polarstern cruises, their date, the number of samples taken as well as the sample type. The colours refer to figure 2.1 and are maintained throughout the whole thesis.

Cruise name	Date	Number of sample stations	Sample type	Colour
PS103	December 2016	11	net catches	
PS79 (ANTXXVIII/2)	December 2011	7	net catches	
PS40 (ANTXIII/4)	April 1996	13	net catches	
PS35 (ANTXII/4)	April 1995	8	sediment samples	

Samples were analysed using the workflow and materials described by Kloster et al. [21]. A schematic overview of the workflow is shown in figure 2.2, with the numbers in the figure relating to the steps described in the following.

As step 1 permanent slides from the samples were prepared, which were scanned in step 2 at a low resolution over a large area, creating an overview image of the slide. A Metafer slide scanning system (MetaSystems, Altlussheim, Germany) with a CoolCube 1m monochrome camera (MetaSystems) and a ZEISS Plan-NEOFLUAR 20x/0.5 objective was used to take images. In step 3 these images were stitched together to create a virtual slide. This was done using the VSlide software (MetaSystems, Altlussheim, Germany). In step 4 the low resolution images from step 2 were run through the software SHERPA version 1.1c [22] which located the diatom valves of interest based on given valve shape templates. Identified valves were manually controlled and selected. In step 5 the positions of the valves found by SHERPA were imported into the virtual slide as a prerequisite for the next step. In step 6 a high resolution scan of the selected valves was made using a ZEISS Plan-APOCHROMAT 63x/1.4 objective with oil-immersion. This could be done automatically, because the virtual slide from step 5 contained the X and Y positions of the valves on the slide. At each position 20 different focus levels were captured and stacked to one high resolution image, where valve outlines as well as valve interior structures were clearly visible. In step 7 the stacked images were analysed with SHERPA to measure the valves. In step 8 the resulting morphometric data was exported as a .csv file and was analysed using the R version 3.4.3 software [23].

Depending on the preserved sample between 30 and 300 valves were analysed per slide. Only complete and well preserved valves could be used in the analysis.

When analysing samples with a high density of well preserved valves the workflow was shortened and only steps 6-8 were carried out, because it was faster to mark suitable valves by hand for the high resolution scan.

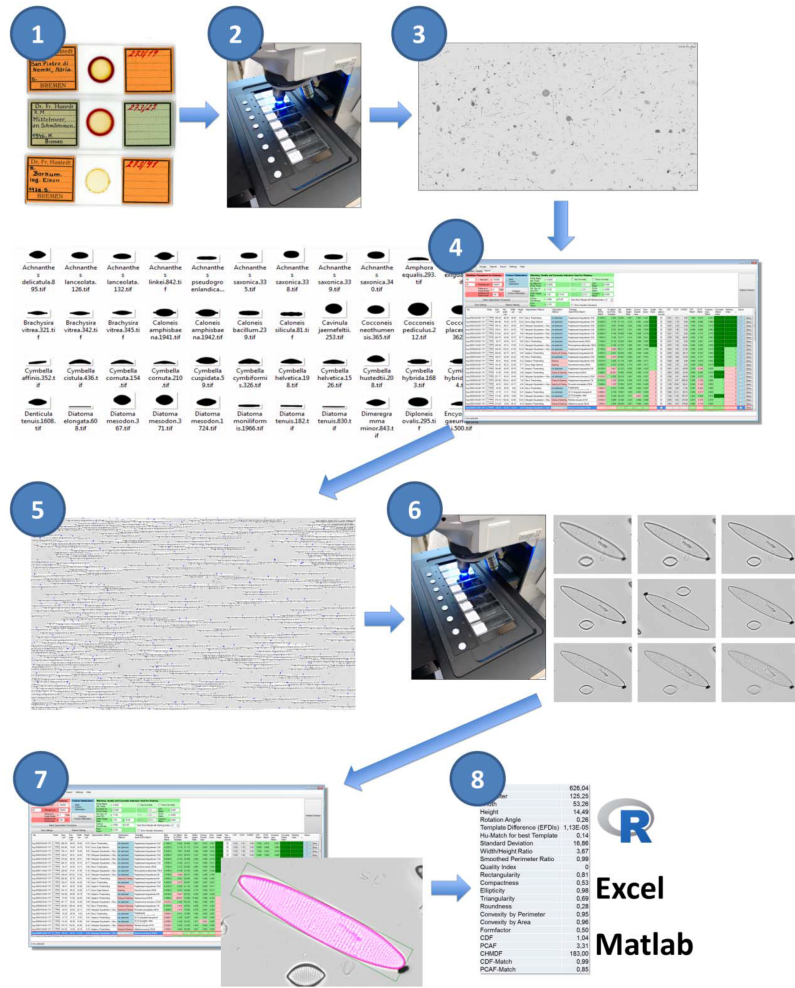


Fig. 2.2: Schematic overview of the workflow for the morphometric analysis developed by M. Kloster et al. [21]. In step 1 microscopic slides are prepared, which are scanned in step 2 at a low resolution. In step 3 a virtual slide of the scanned area is created. In step 4 low resolution pictures are scanned for diatom valves. The positions of the valves are imported to the virtual slide in step 5 and in step 6 a high resolution scan of the valves is done. The valves on the obtained images are measured with SHERPA in step 7 and in step 8 the measured data is exported as a .csv file.

The morphological descriptor of interest which the data was analysed for was the rectangularity. It can be described as the ratio of the valve area to the area of its enclosing rectangle [13]. In figure 2.3 the enclosing rectangle is depicted in green and the valve area is depicted in pink. The rectangularity R is then given by equation 2.1

$$R = \frac{area_{valve}}{area_{enclosing\ rectangle}} \tag{2.1}$$

and can have a value between close to 0 and 1.

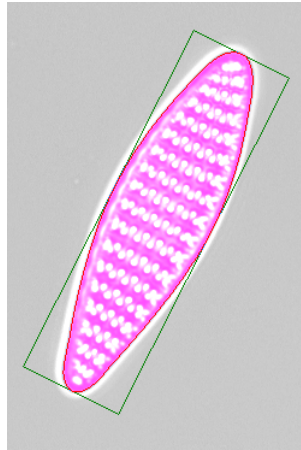


Fig. 2.3: Diatom valve marked in pink with enclosing rectangle shown in green.

2.2 Analysis of ITS sequences

Experiments were carried out using diatom strains cultivated in the Alfred-Wegener-Institut (AWI) laboratories. The diatoms were originally isolated from the Southern Ocean on Polarstern cruise PS103 in December 2016 [24]. Single cells or chains of cells were picked out on board of Polarstern and formed the new strains. Due to this, the strains can be viewed as monoclonal because each cell originates from a single cell. Strains were then transferred to AWI and were kept in cold storage at 4°C in 50 ml cell culture bottles in a day to night rhythm of 16/8 hours. Growth medium F/2 [25] was used for cultivation. Cells were transferred to new cell culture bottles approximately every five weeks. All strains used in this study were transferred to new cell culture bottles the same amount of times.

The sample stations of the Polarstern cruise PS103 are illustrated in figure 2.1. For clear identification strains were assigned a name consisting of the sample station number and a serial number. For example the name "strain 01-09" identifies the ninth strain picked from sample station one. 7 further strains from other diatom species were sequenced in addition to the 28 *F. kerguelensis* strains. This was done to form outgroups for the following phylogenetic analysis.

2.2.1 DNA extraction

Diatoms were separated from the F/2 medium by vacuum filtration. Polycarbonate membrane filters (Whatman, Little Chalfont, United Kingdom) were used with a pore diameter of 5 µm. The diatoms were rinsed of the filter with 700 µl SL1 buffer from the NucleoSpin® Soil kit from Machery-Nagel and were stored in the buffer at -20°C until DNA extraction. The extraction was done with the NucleoSpin® Soil kit from Machery-Nagel according to the manufacturers extraction protocol [26].

To affirm that the extraction worked, the DNA content and the quality was checked with a NanoDrop spectrometer ND-1000 (NanoDrop Technologies, USA). The 260/280 absorbance ratio was measured, which reveals the purity of the DNA and any possible contaminations. The DNA content in ng/µl was also calculated based on an absorbance at 260 nm [27].

2.2.2 PCR

The ITS regions of the DNA were amplified using PCR. Standard primers ITS1 and ITS4 [28] were used for amplification. Reagents used for the PCR reaction are summarized in table 2.2.

Tab. 2.2: Reagents for one PCR reaction

Reagent	Volume [μ l]
5xbuffer	10.0
dNTP	0.5
Primer ITS1	2.5
Primer ITS4	2.5
Taq Polymerase	0.5
Water	33.0
DNA template	1.0

The temperature profile used for the reaction is shown in table 2.3.

Tab. 2.3: Temperature profile of PCR reaction after [16]

Temperature [$^{\circ}$ C]	Time [s]	Number of cycles
94.0	120	1
94.0	35	} 40 cycles
46.2	35	
72.0	60	
72.0	300	1

Gel electrophoresis was used for reasons of quality control. The gel had an agarose concentration of 1% and DNA was stained using HDGreenTM DNA-Dye (INTAS Science Imaging Instruments GmbH, Göttingen, Germany). A Mini-Sub[®] Cell GT Systems (BIO-RAD, Munich, Germany) gel electrophoresis chamber was used with a buffer volume of approx. 270 ml and a 7 x 10 cm gel tray. A voltage of 90 V supplied by a PowerPacTM Basic Power Supply (BIO-RAD, Munich, Germany) was used for the fractionation of DNA fragments. Fractionation was left to run between 30 and 45 minutes. Gels were viewed under UV light with an INTAS Gel iX imager (INTAS Science Imaging Instruments GmbH, Göttingen, Germany) and INTAS-Capture-software. Fragment size of the samples was cross checked with the fragment size of the QX DNA Size Marker 250bp–4kb v2.0 (Qiagen, Hilden, Germany). When the expected fragments were clearly visible at about 700 bp, the sample was used for further analysis.

2.2.3 Purification of PCR products

Any remaining free nucleotides and primers had to be removed from the PCR products before sequencing. For this the High Pure PCR Cleanup Micro Kit (Roche, Mannheim, Germany) was used, according to the manufacturer's protocol.

2.2.4 Sequencing

Reagents used for the sequencing reaction are summarized in table 2.4.

Tab. 2.4: Reagents used for one sequencing reaction

Reagent	Volume [μ l]
Water	6.2
BigDye Buffer	1.5
BigDye	0.3
Primer forward or reverse	1.0
Purified PCR product	1.0

Each reaction was only done with one primer in one direction. So each PCR sample was split into two samples. The primers used were the same ones as in the PCR reaction. The number of cycles and temperature profile for the reaction are shown in table 2.5.

Tab. 2.5: Temperature profile of the sequencing reaction

Temperature [°C]	Time [s]	Number of cycles
96.0	60	1
96.0	10	} 25 cycles
50.0	5	
60.0	240	
10.0	hold	1

Following the sequencing reaction the products had to be cleaned of any remaining nucleotides and primers. The Agencourt CleanSEQ - Dye Terminator Removal (Beckman Coulter, Krefeld, Germany) was used for purification. The purified samples were analysed with an Applied Biosystems® 3130xl Genetic Analyzer (Thermo Fisher Scientific, Waltham, USA).

2.2.5 Assembly of DNA sequences

Sequencer reads were assembled using the Staden Packages software (Whitehead Institute, Cambridge, USA) programs Pregap4 version 1.6-r as well as Gap4 version 11.2-r [29]. Attention had to be paid to the orientation of the reads, which in some cases had to be changed manually. In case of a disagreement in a position, the electropherograms were consulted and if needed the consensus was altered.

In case of an ambiguity the final fasta file had to be altered by hand.

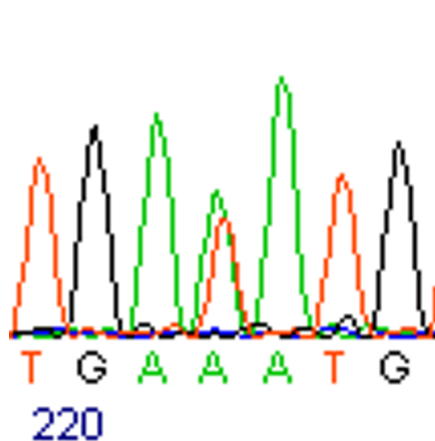


Fig. 2.4: Clipping from an electropherogram showing an ambiguity of A and T at position 222, with both peaks at about the same height.

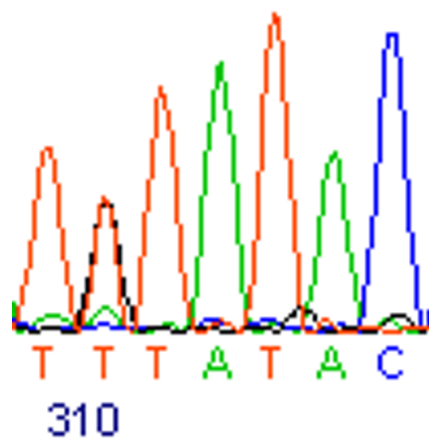


Fig. 2.5: Clipping from an electropherogram showing an ambiguity of T and G at position 310, with both peaks overlapping.

Ambiguities mean that two versions (alleles) of a gene exist in the genome. When they are both amplified and sequenced in the same sample, peaks lying on top of each other in the electropherogram (see figures 2.4 and 2.5) reveal the disagreement in the alleles and their position. Figure 2.4 shows that there are two versions of the sequenced DNA region depicted, one having an A at position 222 and the other a T. This would be noted as a W in the fasta file. Whereas in figure 2.5 one version has a G and the other a T at position 310, noted as K in the fasta file.

2.2.6 Alignment of DNA sequences

After the reads were assembled, they had to be aligned to show similarities and disparities. This was done using the software Mega7 version 7.0 [30] and the algorithm MUSCLE [31]. Disparities in the sequence matrix were cross-checked with the elektropherograms of the reads to rule out any possible personal mistakes.

From the sequence matrix variable positions were extracted into a file in the nexus format. This file was imported into the popART version 1.7 software (Population Analysis with Reticulate Trees) [32] and a minimum spanning network was calculated showing the relationship between the ITS types.

2.3 Mating experiments

Mating experiments were carried out to explore sexual compatibility between the morphotypes. The diatom strains used were the same ones as described in section 2.2. The experiments were conducted in CytoOne[®] Multiwell Plates (STARLAB GmbH, Hamburg, Germany) with 24 wells. Each well was filled with 1500 μ l F/2 medium and 500 μ l diatom culture. The 500 μ l diatom culture was composed from two 250 μ l culture samples from two strains. The plates were sealed with parafilm to avoid evaporation and incubated at 4°C with a light to dark rhythm of 16 to 8 hours. The plates were checked in regular intervals of two days using an inverted microscope Axioplan 35 (ZEISS, Oberkochen Germany) equipped with a ZEISS Plan-NEOFLUAR 20x/0.5 objective. Plates were checked for 2 to 3 weeks depending on the growth of the cells and their vitality. Findings were written down and pictures were taken using a ColorView III camera (Olympus, Hamburg, Germany). Plates were checked for cells detaching from chains which indicated the first step of sexual reproduction as well as auxospores which show that sexual reproduction was successful.

In total 125 mating experiments were carried out in the time period from December 2017 to April 2018.

2.4 Statistical analysis

2.4.1 Evaluation of rectangularity classes

The distributions of valve rectangularity was approximated by a normal mixture model, which was fitted by an EM algorithm, which is part of the R package `mixtools`[33]. For each normal distribution the mean and the standard deviation was calculated.

2.4.2 Regression over percentage low rectangularity class valves and the Sea Surface Temperature

To describe the relationship between the Sea Surface Temperature (SST) and the percentage of valves belonging to the class low rectangularity a regression was applied. The following function describing logistic growth was applied to linearise, because first visual observation revealed that the relationship was not linear

$$y(t) = \frac{y_0 S}{y_0 + (S - y_0) \exp(-kSt)} \quad (2.2)$$

In equation 2.2 t is the time, y_0 is the size of the population at $t=0$, S is the saturation value of the population and k is a growth factor. In this case t is the SST and is therefore later referred to as x . The first step to fit the regression was to try and linearise the given data. The y values were transformed as follows

$$y^T = \ln\left(\frac{1}{y} - \frac{1}{S}\right) \quad (2.3)$$

with S being set to the value of 100 because the percentage of valves in a sample can be a maximum of 100%. The transformation was done in this form because the equation 2.2 can be transformed to

$$\ln\left(\frac{1}{y(t)} - \frac{1}{S}\right) = \ln\left(\frac{S - y_0}{S y_0}\right) - kSt \quad (2.4)$$

and in equation 2.4 the left side of the equation is similar to the transformation used in equation 2.3. The transformed data was already closer to a linearised form but the x values were also transformed in the following form to improve linearisation

$$x^T = \frac{1}{\exp(x)} \quad (2.5)$$

After the transformation of the x values the data was in a linear form and a linear regression could be fitted using the `lm` function from R. From the linear regression the intercept (b) and the slope (m) could be extracted and the parameters for the logistic growth function could be calculated with

$$k = \frac{-m}{S} \quad (2.6) \quad y_0 = \frac{S}{1 + S \cdot \exp(b)} \quad (2.7)$$

After the parameters were known the final function of the regression could be calculated.

2.4.3 Significant differences in rectangularity between ITS groups

The analysis of the ITS sequences of the strains defined several groups which share the same ITS sequence. A statistical analysis should show if the rectangularity values of the ITS groups differ significantly from each other. To do this a mixed effect model was needed, because the data was hierarchically structured. As one possibility for the analysis each individually measured valve could be assigned to an ITS group and the statistical analysis could be done (see column a) in figure 2.6). This approach is biased, since the number of measured valves differed in the ITS groups making this method statistically incorrect. Also would the origin of the valves from different strains be ignored. This is questionable, because a higher similarity within the strains is expected due to their monoclonal nature.

The mean rectangularity values within the strains were calculated (see column b) in figure 2.6) and sorted into ITS groups (see column c)). If the analysis were to be done with the mean values of the rectangularity would the fact that for each mean rectangularity value several valves were measured be withheld from the analysis.

A mixed effect model was used to address all of the problems mentioned above. The model was implemented using the R packages `lme4` [34], `lmerTest` [35] and `arm` [36].

The package `lme4` was used to fit the model and the package `lmerTest` was used to extract the

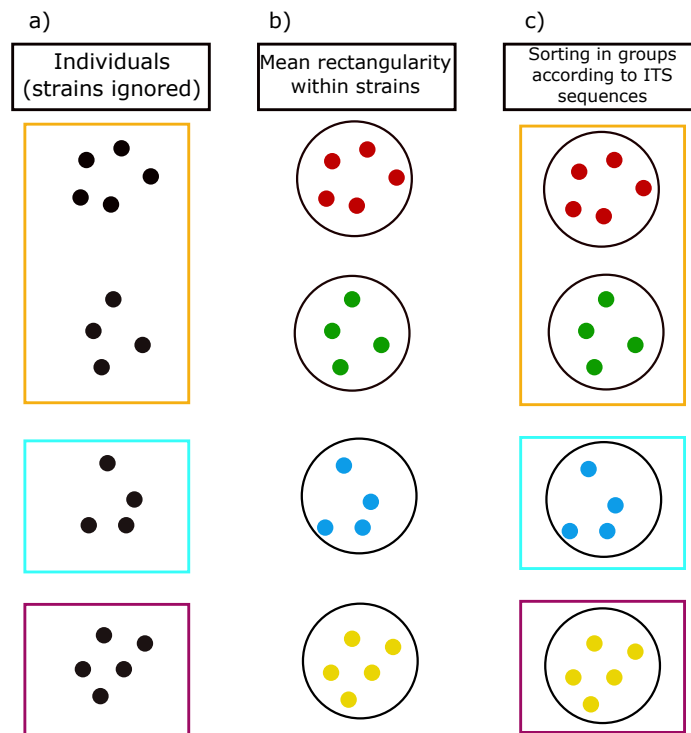


Fig. 2.6: Schematic display of the nested structure of the data, individual valves are represented by dots. Dots of the same colour represent one strain. Coloured rectangles represent different ITS groups. Circles enclose valves for which mean rectangularity was calculated.

p -values of the model. The package `arm` was used to extract the standard deviations from the model.

Two mixed effect models were calculated. One where the ITS groups and strains were seen as random factors and no fixed factor was defined. In this case no p -values could be obtained and the standard deviations were used to visualise the results.

In a second model the ITS groups were set as a fixed effect and therefore p -values could be calculated, to reveal if there were any significant differences in the rectangularity values between the ITS groups.

3 Results

3.1 Morphologic characterisation by valve rectangularity

To investigate if the two morphotypes found by Kloster et al. [12] in core samples can be found in recent water samples rectangularity distributions were plotted. Each of the four distributions in figure 3.1 depicts the pooled data for one transect (PS103, PS79, PS40 and PS35), meaning that the valves that were measured at each station in one transect were pooled and plotted in one distribution.

These distributions are shown in figure 3.1.

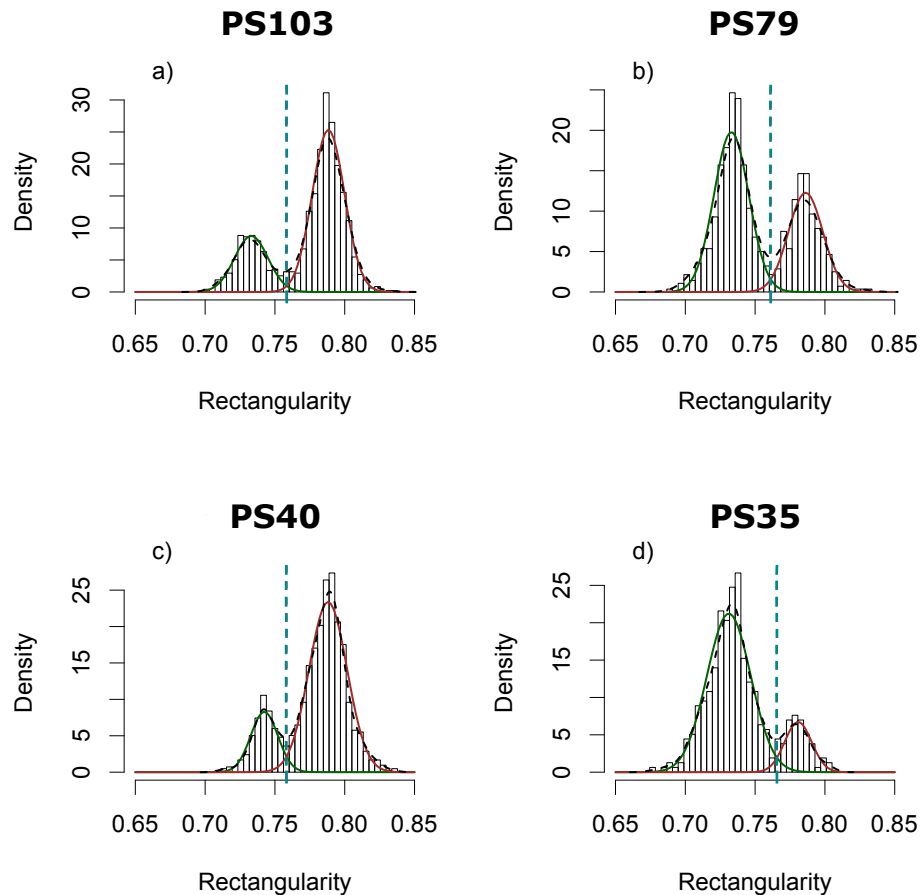


Fig. 3.1: Histograms of pooled rectangularity values of the valves for the four transects. The dashed black curve depicts the density estimation, the coloured curves show the two modelled normal distributions; green marks class low rectangularity, red marks class high rectangularity. The turquoise dashed line marks their intersection point.

All transects showed a bimodal distribution which was approximated by a normal mixture model, summing up two normal distributions of different means (see section 2.4.3).

The normal distributions with the lower mean (figure 3.1 green curves) are in the following referred to as class low rectangularity, the ones with the higher means (figure 3.1 red curves) as class high rectangularity.

Figures 3.1 a) and c) both show a larger contribution of the class high rectangularity, showing that these are more abundant in the transects PS103 and PS40.

Figures 3.1 b) and d) show a larger contribution of the class low rectangularity, showing that these are more abundant in the transects PS79 and PS35.

Table 3.1 shows the key values that were approximated by the normal mixture model.

Tab. 3.1: Key values for the rectangularity classes. Values were approximated by the normal mixture model used to approximate the distributions in figure 3.1.

Transect	class low rectangularity		class high rectangularity		intersection point
	mean	s.d.	mean	s.d.	
PS103	0.7300	0.0120	0.7880	0.0116	0.7585
PS79	0.7330	0.0125	0.7860	0.0123	0.7612
PS40	0.7420	0.0097	0.7880	0.0136	0.7583
PS35	0.7310	0.0158	0.7810	0.0096	0.7657

In the low rectangularity class the means of rectangularity show similar results, with only PS40 being about 0.01 higher than the other three values. In the high rectangularity class the means of the rectangularity are highly similar, only showing differences in the third decimal digit.

Table 3.1 shows the rectangularity classes persisting also in recent water samples from the Southern Ocean. To be able to assign each measured valve clearly to one of the two classes, a threshold separating the classes was defined by their intersection point, which on average for the four transects was 0.7609.

The two rectangularity classes define two morphotypes. Figure 3.2 shows a valve belonging to the low rectangularity class, in the following referred to as morphotype A. Figure 3.3 shows a valve from the high rectangularity class, in the following referred to as morphotype B.

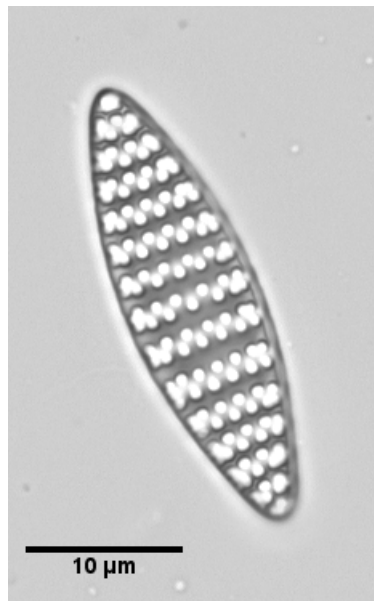


Fig. 3.2: Valve of morphotype A with a low rectangularity value.

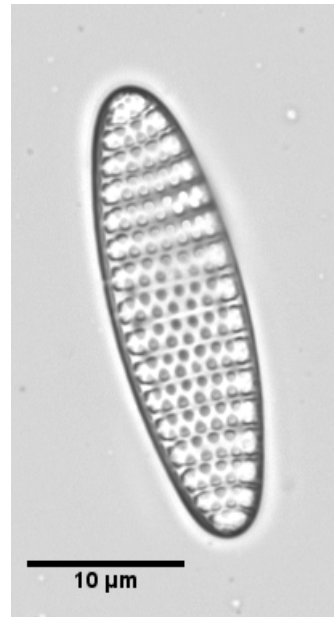


Fig. 3.3: Valve of morphotype B with a high rectangularity value.

Both images were taken using the Metafer slide scanning system and objective described in step 6 in section 2.1. The scale bar provides information about the size of the valves showing that they are about the same size. The difference between the morphotypes is visible in the valve shape. Morphotype A has pointy tips and an overall more lanceolate shape, whereas morphotype B has rounded tips and resembles overall a more oval shape.

3.2 Dependence of valve rectangularity on water masses

After the rectangularity distributions showed the presence of the two morphotypes, the next step was to investigate how the occurrence of the morphotypes is distributed throughout the Southern Ocean.

For this, rectangularity of the measured valves at individual sample stations as opposed to pooled stations in the previous section was investigated. The sample stations were grouped according to their positions in the different water masses of the Southern Ocean. For each water mass, representative stations were chosen. The histograms of the other stations are included in the appendix (see figure A.1 and A.2).

The water masses of the Southern Ocean are separated by the ocean fronts which, as well as the water masses, can be seen in figure 2.1. An ocean front is an area where the properties of the water change substantially. Water masses with homogeneous properties lie between the fronts. The northernmost sample stations were within the Subantarctic Zone, a water mass delimited by the Subtropical Front and the Subantarctic Front. Figure 3.4 shows the rectangularity histograms from samples taken at these stations.

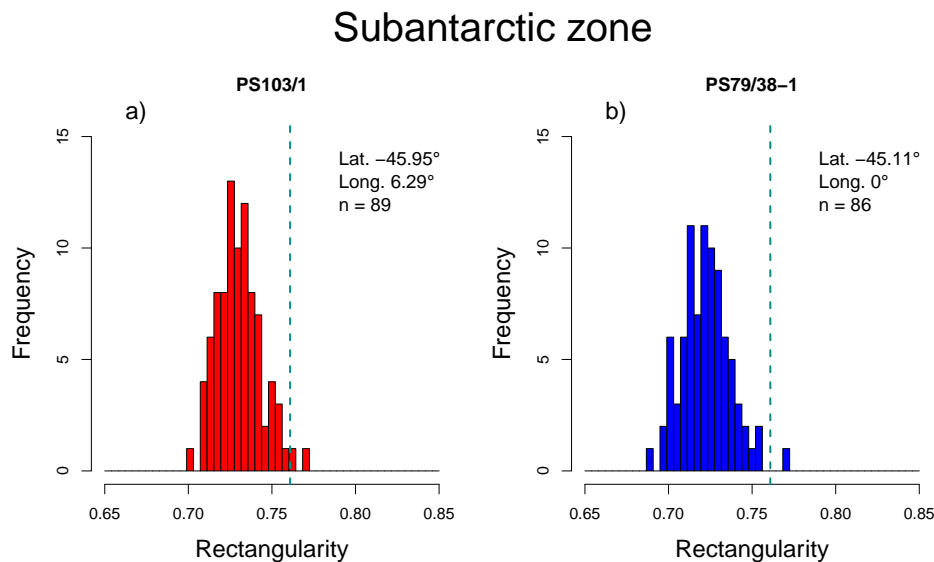


Fig. 3.4: Histograms from two sample stations in the Subantarctic zone, showing the frequency of the rectangularity values. a) transect PS103 and b) transect PS79. The coordinates of the sample station, the sample size and the threshold to separate the rectangularity classes are displayed in the plot.

Both are net catches and show a similar pattern of an unimodal distribution with a peak in the lower rectangularity class. They both have outliers in the high rectangularity class, but they both show that morphotype A is dominant at these sample stations in the Subantarctic zone.

Figure 3.5 shows the sample stations within the Polar Frontal zone. This water mass is delimited by the Subantarctic Front and the Polar Front. For transect PS103 samples were net catches, for transect PS35 sediment was sampled. Both transects show a large peak in the low rectangularity class. At sample station PS35/2701-2 the high rectangularity class also contributes substantially, but both rectangularity histograms show the dominance of morphotype A at these sample stations in the Polar Frontal Zone.

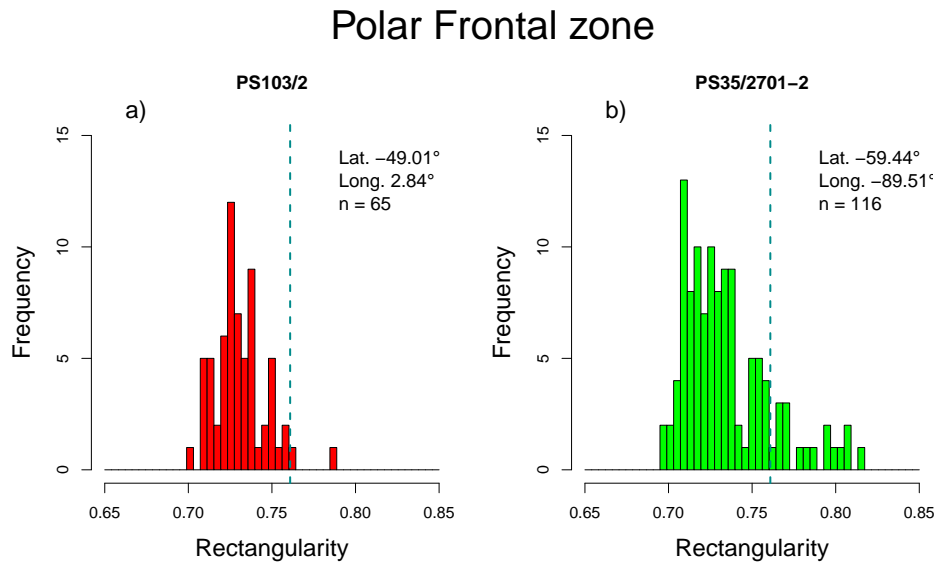


Fig. 3.5: Two sample stations in the Polar Frontal zone. a) transect PS103 and b) transect PS35. The coordinates of the sample station, the sample size and the threshold to separate the rectangularity classes are displayed in the plot.

Figure 3.6 shows histograms from sample stations located in the Antarctic zone. This water mass is delimited by the Polar Front and the Southern Boundary of the ACC.

All samples were net catches with the exception of h)- k), which were sediment samples taken on PS35.

The majority of the histograms show a mostly unimodal distribution with a peak in the low rectangularity class, showing a clear dominance of morphotype A at these sample stations.

In contrast, figures 3.6 b), j) and k) all show bimodal distributions. The histogram b) from sample station PS103/4 and k) from sample station PS35/2686-2 have a large peak in the high rectangularity class and a small peak in the low rectangularity class, showing that morphotype B dominates at these sample stations.

The histogram j) from sample station PS35/2690-1 has a larger peak in the low rectangularity class showing that morphotype A dominates the population, but also morphotype B can be found at an increased frequency.

Both histograms showing the dominance of morphotype B come from sample stations at the southern end of the Antarctic Zone, which show that a change in the population composition can be observed in the transects from north to south. Overall, a change in the composition of the samples from north to south was observed, with samples from the north in general being richer in morphotype A valves and samples from the south being richer with morphotype B valves.

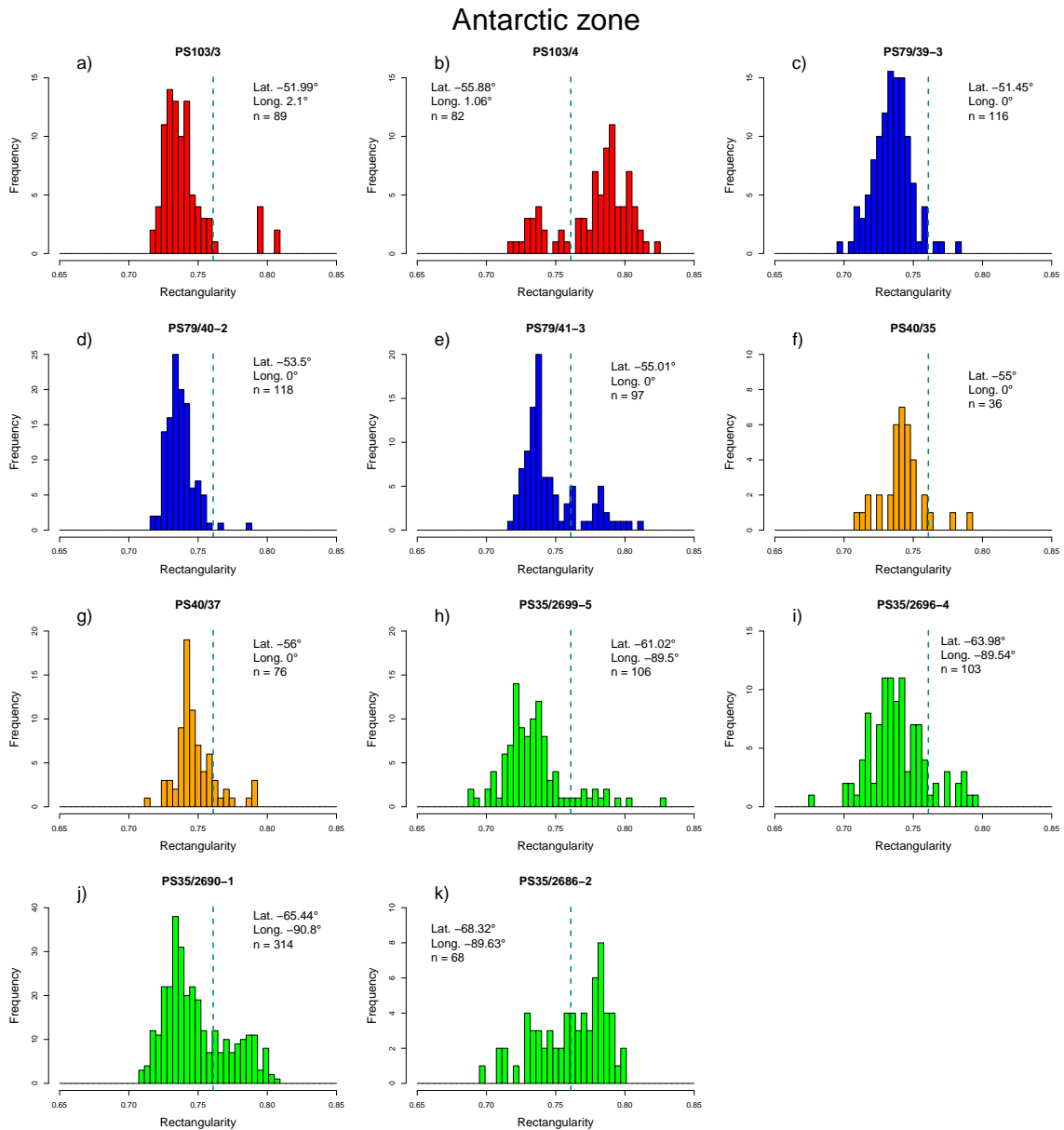


Fig. 3.6: Sample stations in the Antarctic Zone. a) and b) transect PS 103, c)- e) transect PS79, f) and g) transect PS40 and h)- k) transect PS35. The coordinates of the sample station, the sample size and the threshold to separate the rectangularity classes are displayed in the plot.

Figure 3.7 shows results from sample stations, which are located in the Subpolar zone, east of the Antarctic Peninsula also called Weddel Sea. The water masses south of the Southern Boundary of the ACC are called Subpolar zone. Figures 3.7 a) - f) show data from transects where all samples were net catches.

Three histograms show an unimodal distribution with the peak in the high rectangularity class, showing a dominance of morphotype B at these sample stations.

Three histograms show a bimodal distribution. Histogram b) from sample station PS103/4 from the southern end of the Weddel Sea. This shows a large peak in the high rectangularity class and a smaller peak in the class low rectangularity. This shows that morphotype B is more frequent at this sample station, but morphotype A also contributes substantially. From the northern end of the Weddel Sea two histograms show a bimodal distribution. Histogram c) from sample station PS79/42-2 shows a clear dominance of morphotype B and only a small peak in the low rectangu-

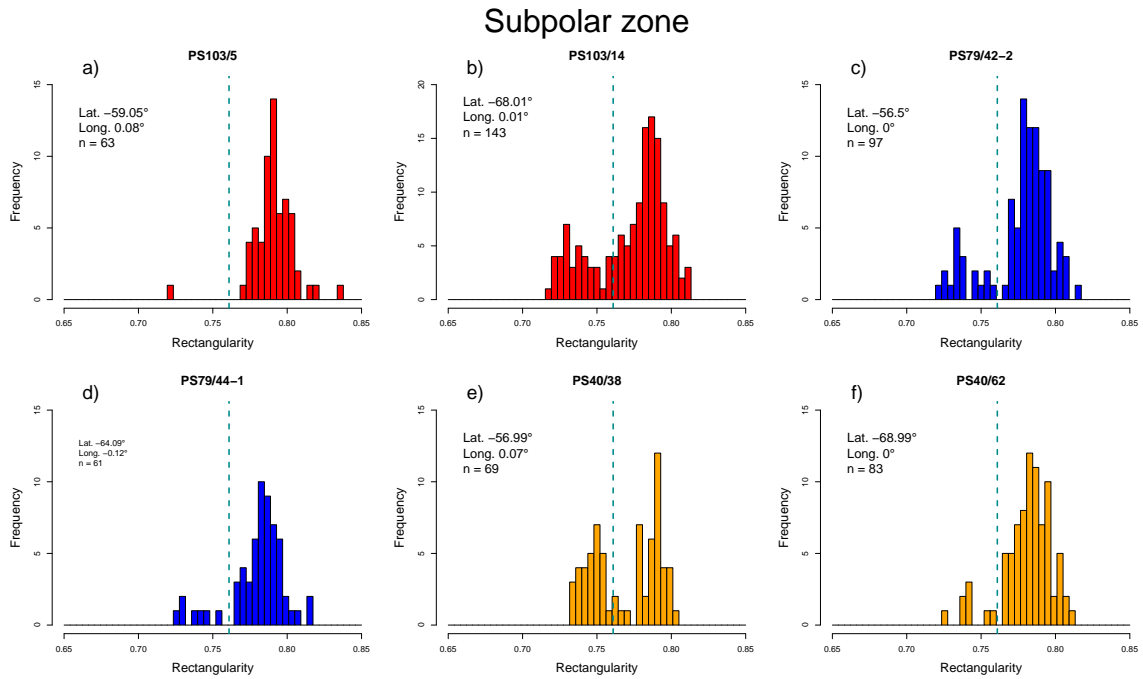


Fig. 3.7: Histograms from sample stations in the Weddel Sea. a) and b) transect PS103, c) and d) transect PS79, e) and f) transect PS40. The coordinates of the sample station, the sample size and the threshold to separate the rectangularity classes are displayed in the plot.

larity class. Histogram e) from sample station PS40/38 shows a slightly higher contribution of morphotype B than of morphotype A.

Overall, all sample stations in the Weddel sea are dominated by morphotype B. Morphotype A can be found in low frequencies in nearly all sample stations, with a tendency to become less frequent in the south.

Looking at all the sample stations, a shift in the population of each transect can be observed. Stations in the north in general show a dominance of morphotype A. The further south the sample stations are the more frequent morphotype B becomes in the samples until it dominates the population in the south.

3.3 Dependences of morphotypes on latitude

In the last section it was shown, that a change in the population from north to south can be observed in the samples. The percentage of valves belonging to morphotype A was calculated and plotted against the latitude to investigate this pattern more closely.

Figure 3.8 shows, that for the three transects from the Atlantic (PS 103, PS79 and PS40) the percentage of morphotype A shows a very similar pattern. In all three of these transects the dominance of morphotype A shifts to the dominance of morphotype B for latitudes lower than -55° south. In the Atlantic, where the three transects originate from, -55° south is about the southern end of the Antarctic Zone (see figure 2.1).

In contrast to this, transect PS35 in the southern Pacific shows a very different pattern: Here data points lie between -70° and -60° south where all but one sampling station show a dominance of morphotype A. At the southernmost sampling station, morphotype B dominates by about 60%.

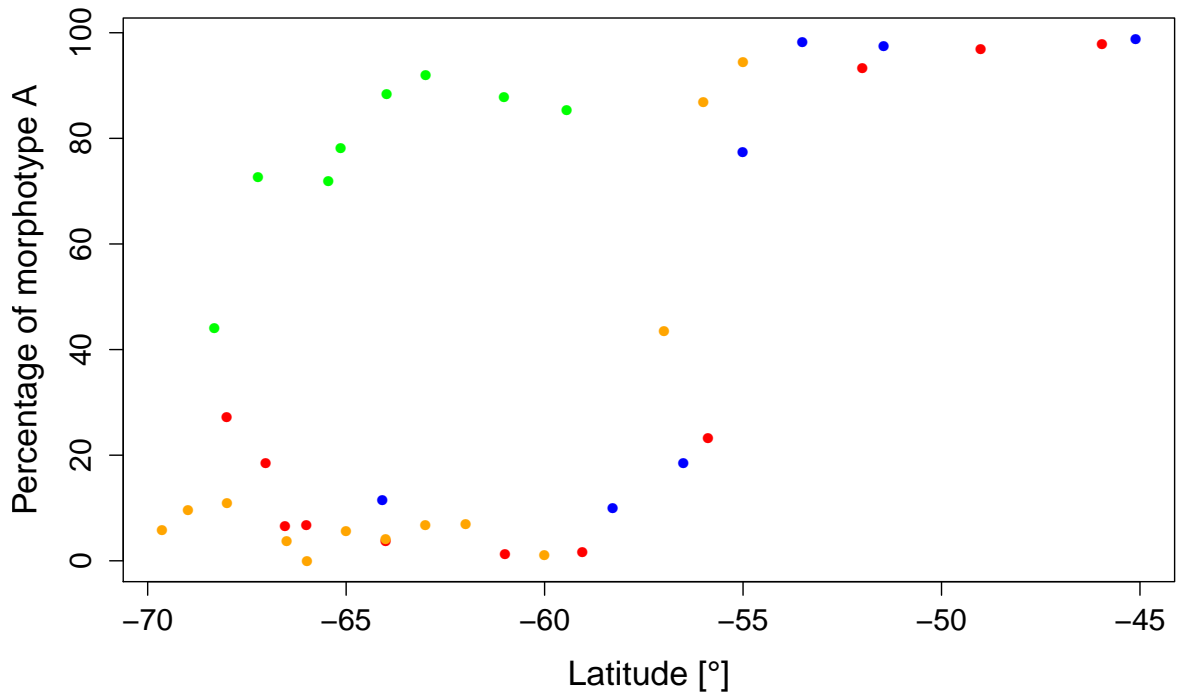


Fig. 3.8: Percentage of morphotype A in relation to the latitude. Sample stations are coloured according to the scheme given in table 2.1.

3.4 Dependence of morphotypes on Sea Surface Temperature

After the plot of the percentage of morphotype A over the latitude did not show a similar pattern for all transects the percentage of morphotype A was plotted over the SST. This was done because the position of the water masses in the Pacific is different than in the Atlantic and therefore the plot over the latitude showed different patterns for the transects.

Figure 3.10 visualises that all transects show the same pattern. For SST above approx. -1°C values of morphotype A dominate, for SST below approx. -1°C values of morphotype B dominate.

Because in previous studies *F. kerguelensis* was proposed as a proxy for SST a regression was applied to find a function describing the relationship between the frequency of the morphotypes and the SST. This regression was modelled on the logistic growth function as described in section 2.4.2. Figure 3.9 shows the linearised data and the linear regression that was fitted through the data. The final regression was fitted to

$$y = \frac{1}{e^{-7.713 + \frac{0.993}{e^x}} + \frac{1}{100}} \quad (3.1)$$

with an R^2 of 0.8402 and can be seen in figure 3.10. This plot shows that morphotype B can be more frequently found in colder water than morphotype A.

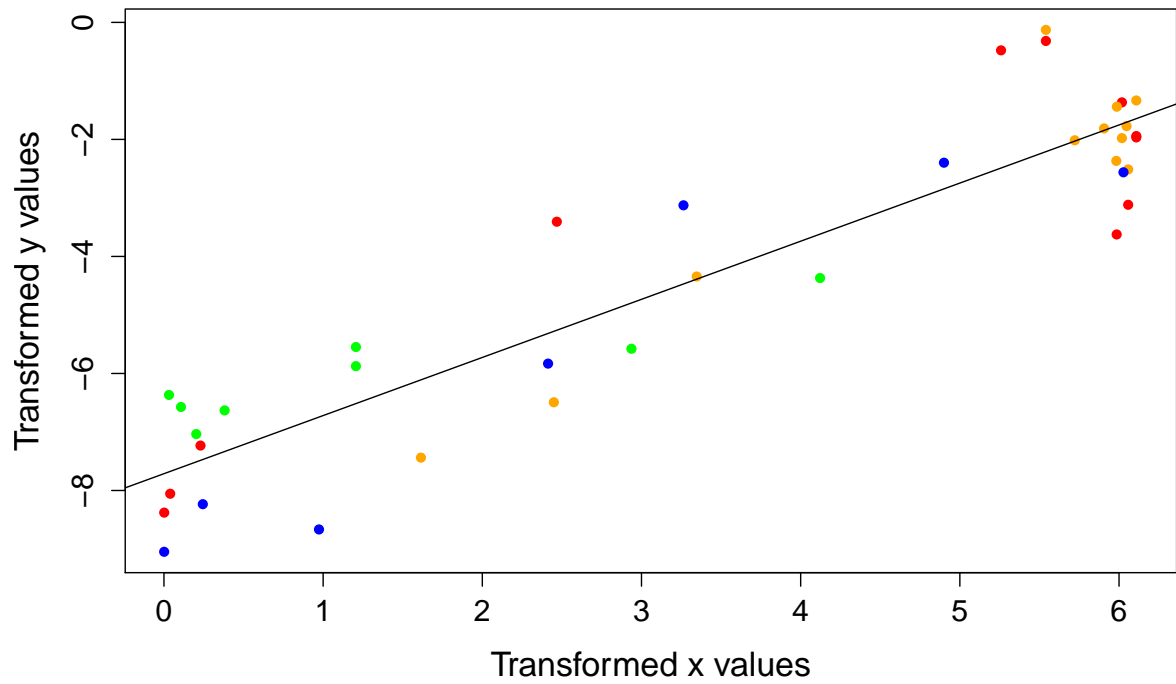


Fig. 3.9: Linearised data and the linear regression which was fitted through the transformed data. Colours represent sample stations as given in table 2.1.

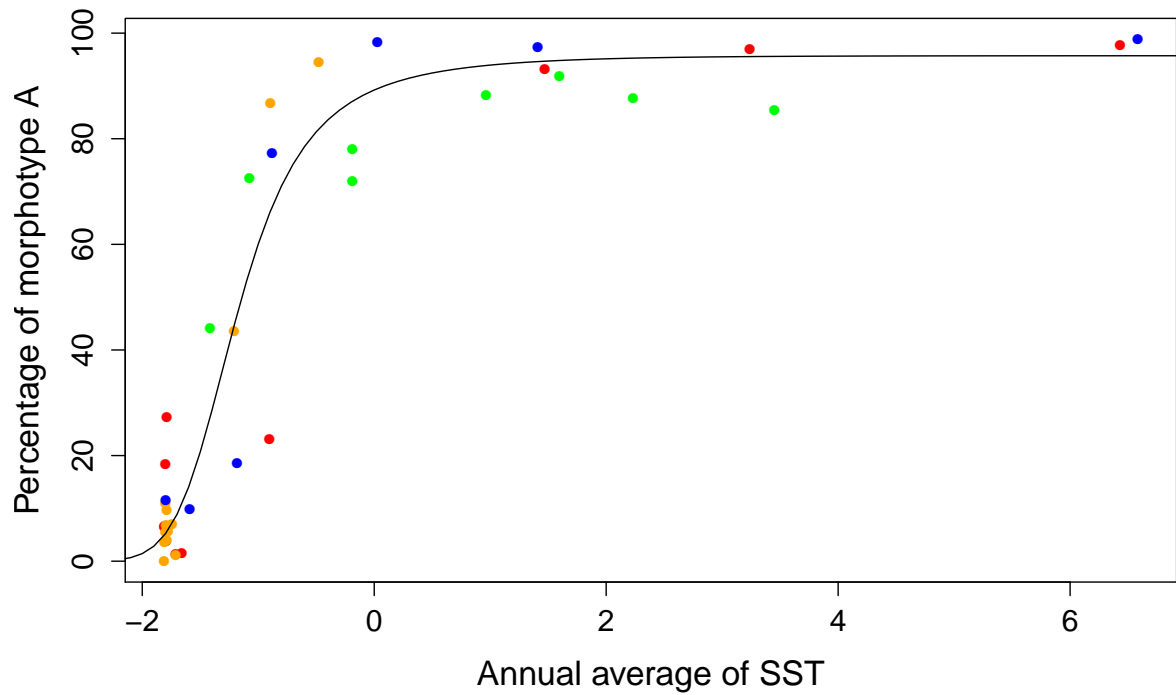


Fig. 3.10: Percentage of morphotype A in relation to the SST. Sample stations are coloured according to the scheme given in table 2.1.

3.5 Analysis of ITS sequences

After the previous results showed the presence of two morphotypes the next step was to investigate if the genotypes also differ or only the phenotypes. For this a genetic assessment with diatom strains originating from sample stations one and fourteen from transect PS103 was done. An exemplary image showing the result of the gel electrophoresis is added in the appendix (see figure A.3). Gel electrophoresis was used as quality control for the PCR before the samples were sequenced.

In total, ITS sequences of 35 diatom strains were sequenced and aligned: 28 strains of *F. kerguelensis* (thirteen from sample station one in the north and fifteen from sample station fourteen in the south), two strains of *Fragilariopsis sublinearis*, four strains of *Fragilariopsis ritscheri* and one strain of *Fragilariopsis obliquecostata*. The alignment spanned 740 base pairs.

From these 740 sites 19 were variable, meaning DNA sequences differed from one another. 16 out of the 19 variable sites differentiated the strains of *F. sublinearis*, *F. ritscheri* and *F. obliquecostata* from the *F. kerguelensis* strains. A table showing all variable sites is added in the appendix (see table A.1).

The 28 *F. kerguelensis* strains only differed in three sites, which are shown in table 3.2.

Tab. 3.2: Sites in alignment, where *F. kerguelensis* strains differed. Mean rectangularity of the strain and standard deviation of the rectangularity is shown, light blue rectangularity values belong to morphotype B. Groups based on the genetic data are marked. Bases are illustrated in different colours, with green and yellow marking ambiguities.

Strain	Rectangularity	s.d.	Position in alignment			Group	
			482	572	704		
01-07	0.754	0.0130	T	T	G	I	
01-20	0.736	0.0090	T	T	G		
01-59	0.767	0.0148	T	T	G		
14-14	0.781	0.0143	T	T	G		
14-18	0.768	0.0163	T	T	G		
14-19	0.755	0.0070	T	T	G		
14-24	0.760	0.0090	T	T	G		
14-27	0.755	0.0085	T	T	G		
14-23	0.744	0.0077	T	W	G	A-1	
14-01	0.802	0.0061	T	A	G	II	
14-02	0.789	0.0092	T	A	G		
14-04	0.793	0.0133	T	A	G		
14-09	0.787	0.0089	T	A	G		
14-17	0.815	0.0091	T	A	G		
14-25	0.769	0.0067	T	A	G		
14-26	0.788	0.0085	T	A	G		
14-29	0.812	0.0087	T	A	G		
14-30	0.784	0.0066	T	A	G		
01-60	0.750	0.0099	K	A	C	A-2	
01-51	0.744	0.0129	K	A	C		
01-52	0.740	0.0071	K	A	C	III	
01-09	0.774	0.0270	G	A	C		
01-14	0.729	0.0088	G	A	C		
01-48	0.733	0.0067	G	A	C		
01-54	0.752	0.0077	G	A	C		
01-56	0.746	0.0080	G	A	C		
01-58	0.762	0.0103	G	A	C		
01-61	0.764	0.0104	G	A	C		

The different sites show a separation of the strains into three major groups. Group I has the bases T, T and G in the three variable positions. The group includes eight strains, three from station 1 in the north and five strains from station 14 in the south.

Group II consists of 9 strains which have the base combination T, A and G. All strains in this group come from sample station 14 in the south and share a high valve rectangularity.

Group III consists of seven strains and has the base combination G, A and C. All strains in this group come from sample station 1 in the north.

Four strains could not be assigned to one of the major groups, because they exhibit ambiguities at the variable positions, which is due to the fact that the alleles of the gene differ from each other at these positions. Group A-1 can have a T or A at position 572, depicted by a W.

Group A-2 exhibits a T or G at position 482, depicted as a K.

Based on this alignment a minimum spanning phylogenetic network was calculated (see figure 3.11).

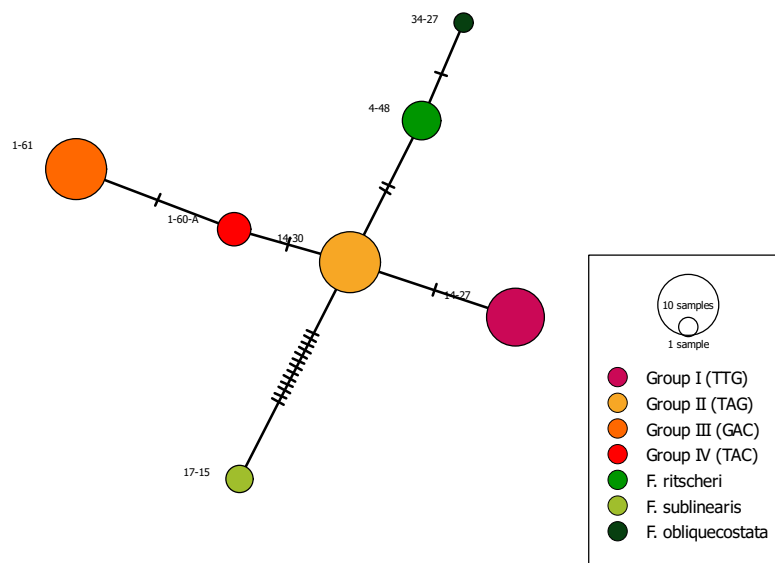


Fig. 3.11: Minimum spanning network calculated on genetic data. Each dot represents a group of strains with the same sequences, matching the groups described above. Hatch marks represent mutations differentiating the groups. Green dots depict outgroups, dots in orange, yellow and red represent *F. kerguelensis*.

This was done to investigate how close the groups formed by the different ITS sequences are related to each other. And how close they are related to the other species.

To do so, ambiguities had to be resolved manually. Each strain containing an ambiguity was split into its two possible combinations. For example was strain 14-23 with the sequence T, W, G split into T, T, G noted as 14-23 and its allele T, A, G noted as 14-23-A.

In figure 3.11 the green dots depict the diatom species not being *F. kerguelensis*, which were sequenced to form outgroups to provide information on the phylogenetic distances. For example, the thirteen hatch marks in the line linking *F. sublinearis* to group II show that their ITS sequences differ in thirteen bases. This shows that the relationship between these two species is more distant than between *F. kerguelensis* and *F. ritscheri* or *F. obliquecostata*. The ITS sequences of *F. ritscheri* only differ in two positions from those of *F. kerguelensis*. *F. ritscheri* and *F. obliquecostata* are even more closely related to each other, having only one mutation between them.

The red, orange and yellow dots represent the groups formed by the *F. kerguelensis* strains. The small red dot labelled 1-60-A represents the group that forms when the ambiguities are dissolved (group IV). All groups formed by *F. kerguelensis* strains only differ by one single mutation.

The outgroups *F. sublinearis* and *F. ritscheri* fork of group II, this indicates that T, A, G is the oldest sequence from all observed ones, because the separation of *F. sublinearis* and *F. ritscheri* happened more likely earlier than the separation of the *F. kerguelensis* population.

3.6 Mating experiments

To investigate if there are reproductive barrier between the ITS groups discovered in the genetic assessment mating experiments were conducted. The presence of reproductive barriers would indicate different species. Figure 3.12 shows results of mating experiments from all diatom strains. The same strains as in the genetic characterisation (see section 2.2) were used.

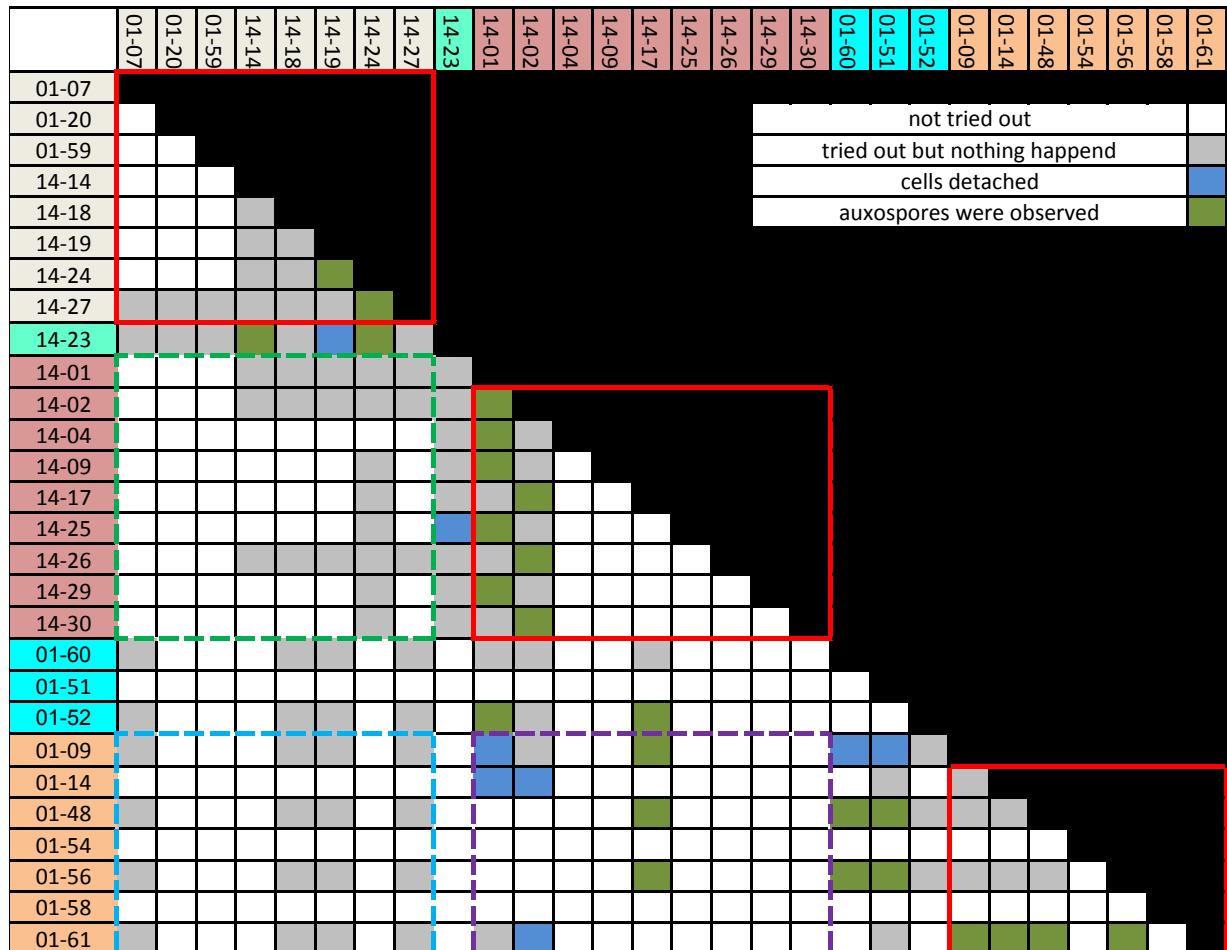


Fig. 3.12: Results of the mating experiments. The red rectangles frame crosses inside the same ITS groups. The red, purple and green rectangle frame crosses in between ITS groups. The strains are colour coded corresponding to the ITS groups from table 3.2. White cells depict crosses not tried out. The grey crosses were tried out but no changes were observed. Blue cells depict crosses where cells detached, green where cells detached and auxospores were observed (successful sexual reproduction).

Sexual reproduction between strains belonging to the same ITS group is possible (see green highlights in red frames in figure 3.12). Sexual reproduction between different ITS groups was observed in only three cases. Between group III and strain 14-17 of group II (green highlights in the purple frame). No auxospores were observed when crossing group I with group II or III (green and turquoise frame).

Auxospores were also observed in crosses where one of the strains exhibits an ambiguity. Strain 14-23 sexually reproduced with two strains belonging to group I, and showed detachment of

cells with strain 14-19 from group I and strain 14-25 from group II.

Strains 01-60 and 01-51, which also exhibit an ambiguity, were able to sexually reproduce with strains belonging to ITS group III. The fourth strain exhibiting an ambiguity (01-51) was able to sexually reproduce with strains belonging to ITS group II.

Sexual reproduction is more frequent between strains from the same ITS group, and sexual reproduction between different groups was only observed between group II and III. Strains exhibiting an ambiguity were only able to sexually reproduce with strains belonging to one specific ITS group.

3.7 Statistical analysis of rectangularity within ITS groups

To investigate if the rectangularity significantly differs between the three largest ITS groups (I, II, III) a statistic analysis with an mixed effect model was done.

The mixed effect model with no fixed effect led to results presented in figure 3.13. Because this model did not return any p -values, a plot was used to show if there are significant differences in the rectangularity values between the ITS groups.

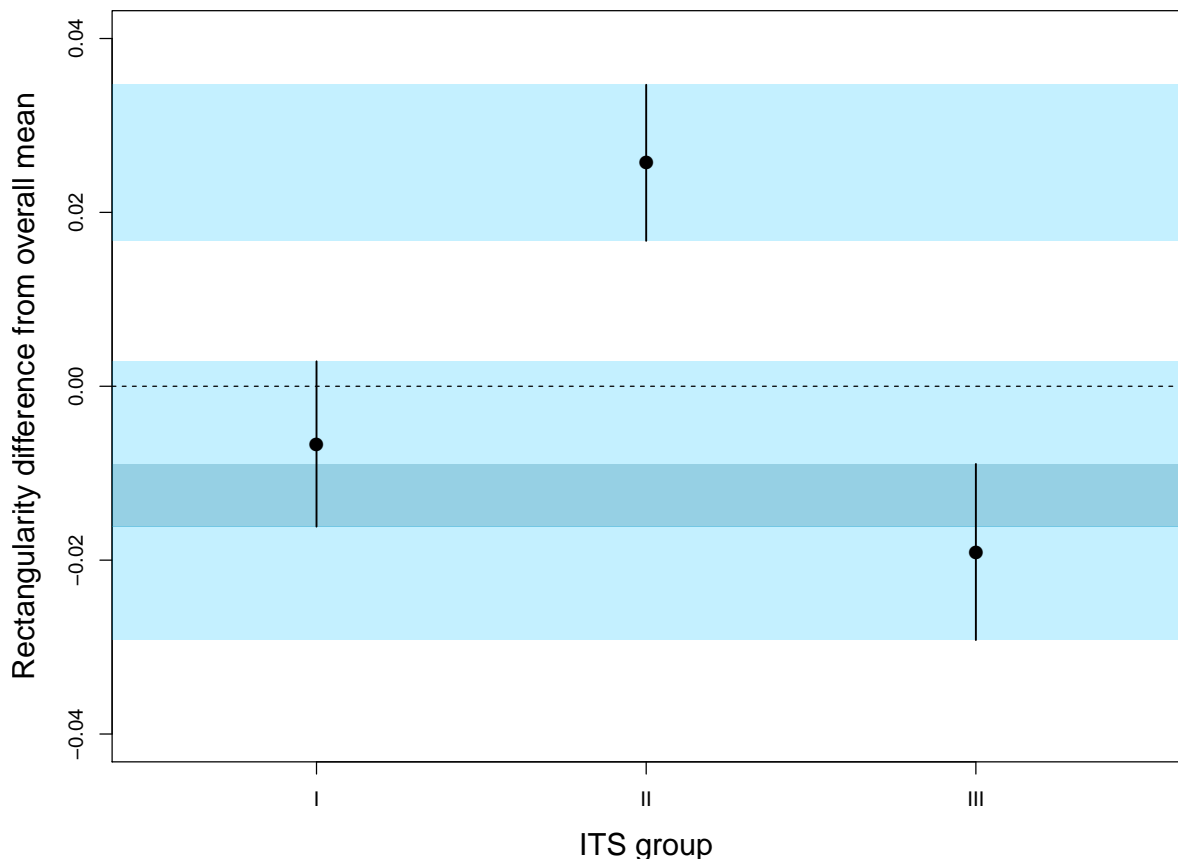


Fig. 3.13: Results obtained from the mixed effect model without fixed effects. The light blue bands represent the 95% confidence interval of groupwise mean rectangularity (marked by a black dot) in relationship to the statistic analysis of the rectangularity. The dashed black line represents the total mean of the rectangularity from all groups. The overlapping of the bands from ITS group I and III shows that they do not differ significantly from each other.

The light blue areas, representing the 95 % confidence interval, of ITS group I and III overlap, which means that the rectangularity values are not significantly different from each other.

These findings are confirmed by the second mixed effect model in which the ITS groups were set as a fixed effect. Here ITS group I differed significantly from the total mean with a p -value of $2 \cdot 10^{-16}$. Group II also differed significant from group I with a p -value of $5.65 \cdot 10^{-5}$. Group III did not differ significantly from group I with a p -value of 0.0802.

4 Discussion

4.1 Can the two morphotypes be found in recent water samples from the Southern Ocean?

The rectangularity distributions from the four transects all show a bimodal distribution (see figure 3.1), which define two rectangularity classes. These two classes correspond well to the two morphotypes originally noted by Kloster et al. [12].

The transects PS103 and PS40 show a larger peak in the high rectangularity class whereas the other two transects show a larger peak in the low rectangularity class. These differences are due to the different sample locations in the transects, with PS103 and PS40 having more sample stations in the south. The mean rectangularity values for the classes are highly similar in the transects, showing a consistency in the shape of the two morphotypes. The values are very similar to the ones found by Kloster et al., only showing differences in the second decimal digit [12].

All these findings show, that two morphotypes of *F. kerguelensis* can be found in the Southern Ocean in recent water samples, older water samples and in sediment samples from the ocean floor.

4.2 Is there a biogeographic distribution pattern for the morphotypes?

The first analysis showed, that two morphotypes of *F. kerguelensis* can be found in the Southern Ocean. As a next step the geographic distribution of the morphotypes was analysed for patterns. The rectangularity histograms plotted for each sample station showed a change in the population from north to south, with morphotype A being dominant in the north and morphotype B being dominant in the south. The change in the dominance was observed around the Southern Boundary of the ACC.

The southernmost sample stations in the Antarctic zone, lying on the border to the Subpolar zone are PS103/4 and PS40/37. PS103/4 shows a dominance of morphotype B whereas PS40/37 shows a dominance of morphotype A even though they were taken at similar locations. These differences might be due to the time difference of 20 years between the samples. Ocean fronts are not fixed in their position but can shift over time due to weather conditions. The shift could in this case also be due to the different seasons the samples originate from. PS103/4 was taken in December, which is summer in the southern hemisphere and PS40/37 was taken in April which corresponds to autumn in the southern hemisphere. The difference in the seasons could lead to differences in the position of the ocean fronts or water temperatures.

Therefore it could be possible, that the samples were taken at the same location but originate from two different water masses thus showing different patterns. The change of the dominance at the front suggests that the abundance of the two morphotypes is linked to the water temperature. Water in the Antarctic zone is approx. 2 ° C cold whereas water in the Subpolar zone is approx. -1 ° C cold. This shows that morphotype B prefers colder water more than morphotype A does. Morphotype B dominates the population in the Subpolar zone. The further south the sample stations are the less morphotype A contributes to the samples. The only exception from this pattern is sample station PS103/14, where the histogram shows a bimodal distribution. At this sample station the salinity at the time of the sampling was rather low due to recently melted sea

ice . This indicates that the temperature might not be the only parameter affecting the distribution of the morphotypes.

As a different display of the distribution pattern the percentage of morphotype A at the sample stations was plotted against the latitude. Samples from the South Atlantic (depicted in red, orange and blue in figure 3.8) all show a similar pattern in the plot, with a dominance of morphotype A from -55° to -45° south and a dominance of morphotype B south of -55° south. -55° south can roughly be linked to the Southern Boundary of the ACC in the three transects from the South Atlantic confirming the pattern from the histograms.

Samples originating from the South Pacific (depicted in green in figure 3.8) do not fit the pattern from the Atlantic samples. They do all group between -70° to -60° south, and only the southernmost data point shows a dominance of morphotype B.

The differences can be explained by the fact, that the water masses in the Pacific are spread over different latitudes than in the Atlantic. Looking at the Southern Ocean it can be said that water masses south of the Subtropical front circulate in an unbroken manner around the continent in the ACC. This leads to the distribution of the water masses and their physical parameters, which follow the stream lines of the ACC. The path of the ACC is controlled by bottom ridges in the Southern Ocean and adjoining land masses [37]. Due to this the positions of the stream lines in the different ocean basins differ, leading to the fact that the water masses are distributed over different latitudes in the basins.

This makes a comparison of the data between different ocean basins unsuitable.

The percentage of morphotype A at the sample stations was plotted in a next step against the SST (see figure 3.10) because of the differences between data from the South Pacific and the South Atlantic concerning the latitude. All data points followed one pattern there showing that morphotype B gains dominance at water temperatures of approx. -1°C and colder.

The fact that all transects showed the same pattern demonstrates that samples from the Pacific can be compared to samples from the Atlantic, when it comes to physical parameters. As explained above the water parameters are linked to the flow of the ACC, so that sample in different ocean basins cannot be compared with regard to their latitude, whereas a comparison of water parameters is possible.

Due to the clearly visible pattern a regression was applied for a quantitative description of the correlation. The regression gave a good description of the data with an R^2 of 0.8402. The good fit shows that it could be possible to use the two morphotypes as proxies in paleoenvironmental research in the reconstruction of the SST from sediment samples.

The first study, which indicated, that morphometric descriptors of *F. kerguelensis* valves could have implications for paleoenvironmental studies, was by Cortese and Gersonde [11], who analysed the valve area and the average valve length, and proposed the valve area and the shape of the valve as new proxies.

The possibility for morphologic descriptors of *F. kerguelensis* being a proxy for paleoenvironmental studies was picked up by several researchers. Cortese et al [38] detected larger valve areas of *F. kerguelensis* during glacial periods in core samples. X. Crosta [39] showed that the valve length differs in core sample from the Holocene (current geological epoch) and Shukla and Crosta [40] showed that valve area of *F. kerguelensis* could be linked to the SST and the availability of silica. This shows that morphometric descriptors of *F. kerguelensis* play an important role in paleoenvironmental studies, but that there is still work to be done in finding out which descriptor can

be linked to which environmental parameter. The rectangularity is a rather new proxy first described in 2017 by Kloster et al. [12] in a fossil assemblage, which is similar to the observations made here but this time from recent samples. This study showed that the rectangularity and the SST seem to be linked. This recognition could help establish the rectangularity as a proxy in paleoenvironmental research.

All results showed that there is a biogeographic pattern in the distribution of the two morphotypes of *F. kerguelensis*. Up until this the understanding was that the distribution area of *F. kerguelensis* ranges from the Subtropical front in the north to the Antarctic continent in the south [41]. This study now showed, that each of the morphotypes has different distribution areas and that a general statement of the distribution area of *F. kerguelensis* might not be the ideal solution but rather a statement of the distribution areas of the two morphotypes.

4.3 Are the two morphotypes cryptic species?

The results of the morphologic analysis showed that two morphotypes exist in the Southern Ocean. Morphotype A preferring warmer waters in the north and morphotype B preferring colder waters in the south. The follow up question was if these are cryptic species or if the differences in morphology are due to phenotypic plasticity. The biogeographic distribution pattern could be an indicator for phenotypic plasticity with morphotype B being more frequent in colder waters. More precisely this means that it could be possible that one cell would be able to build morphologically different frustules depending on the water temperature it lives in. This in turn would mean that as soon as the water temperature drops to approx. -1°C the cell would start building frustules from morphotype B.

In this case, the genotypes of the strains should be identical and only the phenotypes should differ.

In case of cryptic species the genotypes should differ and the phenotypes should be the same. In this context it should be noted that the term cryptic species is not completely accurate because the previous results showed the existence of two morphotypes which can be distinguished by their valve shape (see figures 3.2 and 3.3). In a strict sense cryptic species cannot be distinguished by their phenotypes, therefore the correct term here would be semi-cryptic species. This means that the species were classified erroneously in the past due to their very similar morphology which later turned out to be different species. In the following the possible semi-cryptic species will be referred to as different species but at this point this is only a working hypothesis.

A genetic characterisation of *F. kerguelensis* strains was done to assess a possible genetic differentiation.

A similar study assessing the ITS regions of eleven *F. kerguelensis* strains was done by Fuchs et al. [8]. In this study slight differences were found in the ITS regions but without any phylogenetic pattern. However at that time they did not know about the two morphotypes. Therefore this data could not be used in this study because the strains used could not be assigned to one morphotype.

If the genetic assessment showed a congruence between the morphotypes and the DNA sequences this would indicate that the morphotypes are two different species, whereas no difference in the DNA would indicate phenotypic plasticity.

None of this was the case with the genetic analysis, which revealed five patterns for the ITS

region, making the results more complex.

The five patterns found are shown in table 3.2 and are the basis for the five groups marked in the table (I, II, III, A-1 and A-2) which are referred to as ITS groups in the following.

Table 3.2 shows that the ITS group II generally shows higher rectangularity values than the other groups and is solely made up of morphotype B, whereas the other groups in general show lower rectangularity values belonging to morphotype A. This pattern shows several exceptions with relatively high rectangularity values in ITS groups I and III. These exceptions have a rather high rectangularity value marking them as morphotype B but most values are still lower than the rectangularity values from group II.

The mixed effect model (see section 3.7) shows that there is a significant difference in the rectangularity between group II and groups I and III and no significant difference in the rectangularity between groups I and III. These results indicate that each major ITS group (I, II, III) is solely made up of one morphotype. This leads to the hypothesis that morphotype B is solely found in group II whereas groups I and III are made up of morphotype A.

ITS groups A-1 and A-2 could not be part of the mixed effect model due to the low number of strains making a statistical analysis impossible.

The different results from table 3.2, which showed that ITS groups I and III are made up of both morphotypes, could be explained by the fact that the two morphotypes were defined by two overlapping normal distributions (see figure 3.1). A schematic view showing two overlapping normal distributions is depicted in figure 4.1.

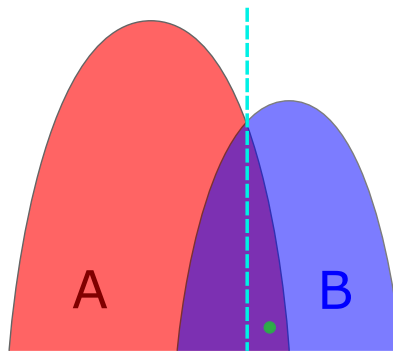


Fig. 4.1: Schematic view of two normal distribution defining the morphotypes. The vertical dashed line marks the threshold set at the intersection of the normal distribution. The green dot depicts an exemplary rectangularity value, which could belong to either morphotype.

The threshold used to assign each rectangularity value to a morphotype was set at the intersection of the normal distributions. An exemplary rectangularity value at the right of the threshold (depicted by a green dot in figure 4.1) would be sorted to morphotype B. This might not be correct for all cases. The exemplary value could belong to morphotype A because it is within the limits of the normal distribution defining morphotype A. This shows that the assignment of values to one morphotype simply by the threshold is not always correct, especially when the rectangularity values lie in the overlapping area of the normal distributions, showing that maybe not all rectangularity values in table 3.2 are assigned to the correct morphotype.

The threshold can only be seen as an orientation tool to assign rectangularity values but not as a definite value separating the morphotypes.

This shows, that the hypothesis, of each ITS group only being made up of one morphotype which is based on the statistic analysis is presumable the most likely one. This is supported by the fact

that in the ITS groups I and III morphotype A is more dominant.

This shows that the findings from the morphological analysis and the genetic assessment correlate with each other. The morphology can give an indication about the allocation of a strain to an ITS group, as well as the ITS group being able to indicate the allocation to a morphotype.

A closer look at the five patterns forming the ITS groups reveals that they are made up of four sequence variations formed by the variable position in the alignment. The five patterns are TTG, TWG, TAG, KAC and GAC (with W and K marking ambiguities as described in section 2.2.5). These five variations can be simplified to four variants by dissolving the ambiguities. These four variants are TTG, TAG, TAC and GAC, with TAC occurring when the ambiguity in the ITS group A-2 (KAC) is dissolved.

Some strains carry two of the variants, marked by an ambiguity. The combination of TTG with TAG can be found as well as TAC and GAC.

The origin of the variants can be explained by mutations or more precisely substitution which means the mutation of only one base in the DNA. Figure 4.2 depicts a gene tree showing a possible evolution of the ITS region leading to the variants that can be found today.

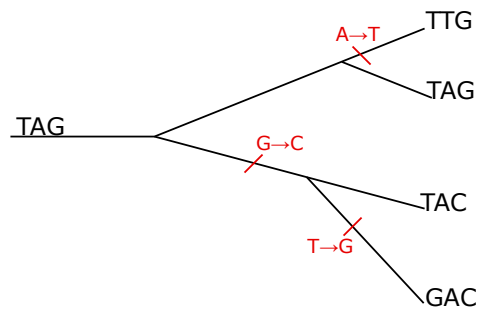


Fig. 4.2: Gene tree showing a possible evolution of the variants. Hatch marks represent mutation, with substitutions marked in red.

The variants are the results of substitutions which occurred over time and are depicted as red hatch marks in figure 4.2. The tree is based on the principle of maximum parsimony meaning that the least possible number of mutations was used in the construction of the tree, leading to only three mutations being necessary to construct this tree.

It was possible to identify the original variable positions in the ITS sequence as TAG with figure 3.11, which shows that the outgroups (*F. sublinearis* and *F. ritscheri*) fork of this sequence. This shows that it is older than the other sequences because the separation of the species forming the outgroups more likely happened earlier than the separation of the *F. kerguelensis* population.

Looking at the tree in figure 4.2 the obvious hypothesis would be, that there are four species, each carrying one of the variants depicted in the tree. However would this be incorrect because it is important to remember that this is a gene tree showing only the evolution of a single gene.

The hypothesis of four species conflicts with the conclusion regarding the species number which can be drawn from table 3.2. These results suggest that there are three species: one made out of strains with the variant TTG and strains carrying the variants TTG and TAG referred to in the following as S1, a second species made up of strains only carrying the variant TAG in the following referred to as S2 and a third species made up of strains carrying only the variant GAC and strains carrying the variants GAC and TAC referred to in the following as S3.

The separation into three species is based on the fact that only a few combinations of the variants

can be found in the strains. For example TAG and GAC never occur together in a strain, otherwise the analysis should have revealed an ITS group with the pattern KAS. If all strains would belong to one species a higher number of variant combinations would be expected because all strains would be able to exchange genetic material freely. This makes the hypothesis of a separation in three species more likely.

The assignment of strain 14-23 to species S1 can be explained with its low rectangularity value. Also due to their also rather low rectangularity values were strains 01-60, 01-51 and 01-52 assigned to S3.

There are two possible explanations for the fact that some species carry more than one variant in the population.

The first one is based on the principle of lineage sorting, which is depicted as a schematic overview in figure 4.3.

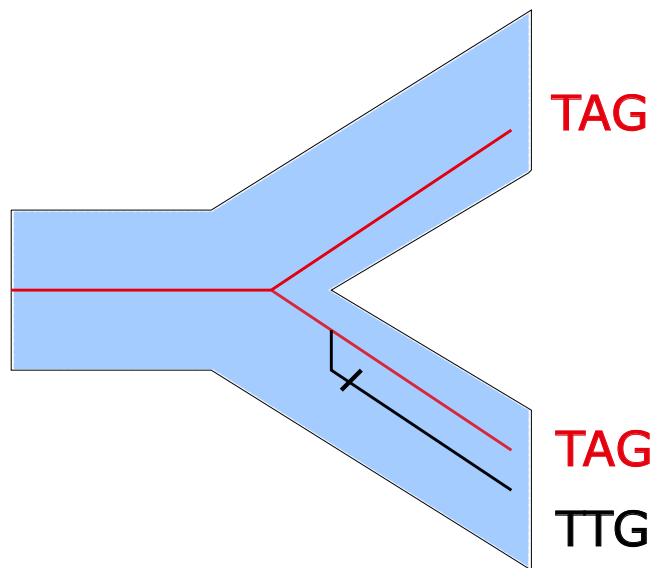


Fig. 4.3: Schematic example of lineage sorting. Red and black lines depict the gene tree whereas the blue shaded tree depicts the species tree. The hatch mark represents a mutation happening.

A gene tree (red and black lines) as well as the species tree (blue) is depicted in figure 4.3. In lineage sorting the assumption made is that one variation of the gene is older than both species (shown in red) and is inherited by both species. The second variation of the gene evolved after the formation of both species by mutation in one of the species. One species only carries one variation as no genetic exchange occurred between the species. The fact that the older variant (TAG) does not disappear with the appearance of the new one can in the case of the ITS region be explained with the fact that each individual carries several thousand copies of the region in their genome. Therefore it is likely that the older version also remains in some copies.

The second explanation for two variants in a species could be hybridisation between two species. During sexual reproduction gene material is exchanged which could lead to one species gaining new variations from an other species.

At this point both explanations could be true in the case of the several variants in species S1 and S3.

The next question was how the evolution of the four variants can be brought together with the evolution of the three species. For this purpose the gene tree from figure 4.2 was combined

with a species tree based on the hypothesis of the three species. The combination of both trees is depicted in figure 4.4. This shows that the gene tree (depicted by black lines) and species

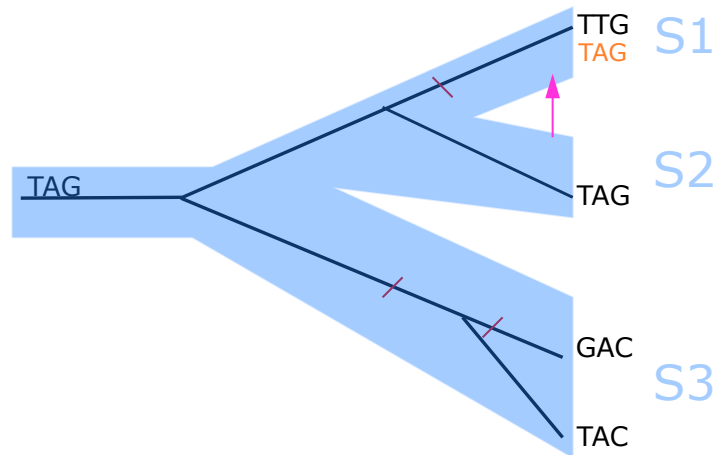


Fig. 4.4: Combination of gene tree and species tree. Black lines depict the gene tree, blue depicts the species tree. Names of species are included with variants from the gene tree marked in black and otherwise acquired variants in orange. The pink arrow indicates a possible hybridisation event.

tree (depicted in blue) do not correspond in all branches. Species S3 has two variants from the gene tree whereas the others have only one. In two species two variants can be found: the ones originating from the gene tree depicted in black and the other depicted in orange.

A possible explanation for the origin of the TAG variant in species S1 is a hybridisation event which took place at some point in the past with species S2, depicted by a pink arrow in figure 4.4. The two variants in species S3 can be explained by mutations, with the second variant evolving after the evolution of species S3 leading to both variants being present in the population today. The tree in figure 4.4 shows a possible explanation of how the evolution of the variants and the species could have happened. It is important to notice that the length of the branches is not from of any importance in this tree and that no statement is made over the course of time of the evolution.

All these results support the hypothesis of the three species, but a species determination based on three variable positions in the ITS region is not a very sound one.

In the next step the mating experiments were analysed for further information. Figure 3.12 shows the results of the mating experiments, showing that successful sexual reproduction in most cases happened between strains belonging to the same ITS group, whereas this was rare among strains from different ITS groups. This pattern also corresponds with the three species, showing that most successful sexual reproduction events happened within the species, pointing to the existence of a reproductive barrier between the species.

A few exceptions to this pattern were observed. Strain 14-17 from species S2 was able to successfully sexually reproduce with strains 01-09, 01-48 and 01-56 from species S3. Strain 01-52 from species S3 was also able to sexually reproduce with strains 14-01 and 14-17 from species S2.

All crosses standing out from the pattern happened between species S2 and S3, leading to the question of why the pattern KAS was never observed in the ITS groups. Strains from species S2 and S3 are obviously able to mate and exchange genetic material, which would lead to the appearance of KAS.

The fact that KAS was never observed in the genetic assessment can be explained by the fact,

that the crosses happened always between strains from sample station one in the north and sample station fourteen in the south. The geographic distance between the stations means that in nature the strains would never or very rarely meet under normal circumstances because their distribution areas are far away and do not overlap. This makes a physiological reproductive barrier between these strains not necessary at the given time. During the further process of evolution this reproductive barrier might develop.

This probably explains why KAS was never found in the genetic assessment because this was done with strains from the Southern Ocean.

Apart from the few exceptions all findings from the mating experiments support the hypothesis that there are three species and that the assignment of the species based on the ITS variants was correct.

It is important to keep in mind that the mating experiments showed clear results, but they were only done once and no replicas were done. Also was only the outcome of the first daughter generation noted. It was never tested if the resulting auxospores are able to sexually reproduce again or if they are infertile. Them being infertile could also show that the parental cells belonged to different species which in this case were able to sexually reproduce but the reproduction barrier takes action in the first daughter generation.

The result of the genetic analysis and the mating experiments show that it is very likely that three species of *F. kerguelensis* can be found. The low number of variations in the ITS sequences makes it not possible to definitely determine the species, but all findings in this thesis support this hypothesis.

After the findings revealing two species consisting of morphotype A (S1 and S3) and one of morphotype B (S2) it can be said that S1 and S3 can be noted as cryptic species because they both are made up of morphotype A, making it impossible to distinguish them by morphology. Species S2 can be noted as a semi-cryptic species because it can be distinguished from S1 and S3 by morphology which was noted in this study.

The genetic assessment showed clear differences in the ITS sequences of the strains which stand in contrast to the results from the study by Fuchs et al. [8]. A possible explanation for this is, that Fuchs et al. only had strains from the northern part of the ACC whereas strains used in this study originated from sample stations all across the Southern Ocean.

It is not unusual that groups of diatoms identified as one species based on the morphology reveal cryptic or semi-cryptic species when reanalysed genetically or with mating experiments. Lundholm et al. [42] described two new species in the genus *Pseudo-nitzschia* after morphological and genetic analyses. Amato et al. [16] showed the presence of cryptic species in the diatoms *P. delicatissima* and *P. pseudodelicatissima*. The frequent discovery of new species or cryptic species in diatoms can be explained with the high similarity or near equality of some diatom frustules, leading to an incorrect classification in the past. With molecular methods on hand today it is easier to detect cryptic species.

5 Conclusion and outlook

In summary it can be said that all research questions could be answered in this study.

The first research question was, if the two morphotypes found in core samples can also be found in recent waters samples? This question can be answered with yes, the two morphotypes previously found in samples originating from a sediment core can also be found in recent water and sediment samples from the Southern Ocean. The morphotypes were found in all of the four analysed transects and could clearly be differentiated from each other.

The second research question was if there is a biogeographic distribution pattern for the morphotypes. It can be answered with yes. There is a clear detectable pattern in the geographic distribution of the morphotypes. It can be linked to the SST with morphotype A being more abundant in samples from the north thus preferring warmer waters. Morphotype B is more abundant in samples from the south showing that it prefers colder waters than morphotype A. The answer to the last research question about the morphotypes being cryptic species or originating from phenotypic plasticity was not as straight forward as the previous ones. It became clear that there are most likely two cryptic species of *F. kerguelensis* corresponding to morphotype A and one semi-cryptic species corresponding to morphotype B.

The results obtained in this study could answer all the given research questions. The answers found could lead to further research questions and thus studies being performed.

Possibilities for further research could include a deeper look into the correlation between the SST and the abundance of the morphotypes. This could be done by analysing more samples, originating from different transects in the ACC. Live cultures containing both morphotypes could also be exposed to different temperatures in the laboratory to see if one of the morphotypes would grow better than the other.

Further research could also be conducted to help differentiate the cryptic species more clearly from one another. This could be achieved by sequencing and comparing different DNA regions which might show more variable positions than the ITS sequences. More mating experiments and replicas of the conducted experiments could help to define the reproductive barriers more closely.

Bibliography

- [1] David G. Mann and Pieter Vanormelingen. "An Inordinate Fondness? The Number, Distributions, and Origins of Diatom Species". In: *Journal of Eukaryotic Microbiology* 60.4 (2013), pp. 414–420. DOI: 10.1111/jeu.12047.
- [2] Trevor A. Norton, Michael Melkonian, and Robert A. Andersen. "Algal biodiversity". In: *Phycologia* 35.4 (1996), pp. 308–326. DOI: 10.2216/i0031-8884-35-4-308.1.
- [3] David M. Nelson et al. "Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation". In: *Global Biogeochemical Cycles* 9.3 (), pp. 359–372. DOI: 10.1029/95GB01070.
- [4] Victor Smetacek. "Diatoms and the Ocean Carbon Cycle". In: *Protist* 150.1 (1999), pp. 25–32. DOI: [https://doi.org/10.1016/S1434-4610\(99\)70006-4](https://doi.org/10.1016/S1434-4610(99)70006-4).
- [5] Karl Esser. *Kryptogamen 1*. Springer Berlin Heidelberg, 2000. DOI: 10.1007/978-3-642-57139-8.
- [6] Simon A. Crawford et al. "NANOSTRUCTURE OF THE DIATOM FRUSTULE AS REVEALED BY ATOMIC FORCE AND SCANNING ELECTRON MICROSCOPY". In: *Journal of Phycology* 37.4 (), pp. 543–554. DOI: 10.1046/j.1529-8817.2001.037004543.x.
- [7] CAROL M. et al. "CHAPTER 3 - PHYTOPLANKTON AND PRIMARY PRODUCTION". In: *Biological Oceanography: An Introduction (Second Edition)*. Ed. by CAROL M. LALLI and TIMOTHY R. PARSONS. Second Edition. Oxford: Butterworth-Heinemann, 1997, pp. 39–73. DOI: <https://doi.org/10.1016/B978-075063384-0/50059-1>.
- [8] Nike Fuchs et al. "Genetic characterization and life cycle of the diatom *Fragilariopsis kerguelensis*". In: *European Journal of Phycology* 48.4 (2013), pp. 411–426. DOI: 10.1080/09670262.2013.849360.
- [9] U. Zielinski and R. Gersonde. "Diatom distribution in Southern Ocean surface sediments (Atlantic sector): Implications for paleoenvironmental reconstructions". In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 129.3 (Apr. 1997), pp. 213–250. ISSN: 0031-0182. URL: <http://www.sciencedirect.com/science/article/pii/S0031018296001307>.
- [10] U. Zielinski et al. "Quaternary surface water temperature estimations: Calibration of a diatom transfer function for the Southern Ocean". In: *Paleoceanography* 13.4 (), pp. 365–383. DOI: 10.1029/98PA01320.
- [11] G. Cortese and R. Gersonde. "Morphometric variability in the diatom *Fragilariopsis kerguelensis*: Implications for Southern Ocean paleoceanography". In: *Earth and Planetary Science Letters* 257.3 (2007), pp. 526–544. ISSN: 0012-821X. DOI: <https://doi.org/10.1016/j.epsl.2007.03.021>.
- [12] M. Kloster et al. "Morphometry of the diatom *Fragilariopsis kerguelensis* from Southern Ocean sediment: High-throughput measurements show second morphotype occurring during glacials". In: *Marine Micropaleontology* (in review).
- [13] *Proceedings of the Thirteenth Diatom Symposium*. Bristol: Biopress Ltd, 1995. ISBN: 0948737352.
- [14] Bánk Beszteri. "Morphometric and molecular investigations of species limits in *Cyclotella meneghiniana* (Bacillariophyceae) and closely related species". In: (Jan. 2005).

- [15] Tim Palucka. *Overview of Electron Microscopy*. URL: https://authors.library.caltech.edu/5456/1/hrst.mit.edu/hrs/materials/public/ElectronMicroscope/EM_HistOverview.htm (visited on 04/17/2018).
- [16] Alberto Amato et al. "Reproductive Isolation among Sympatric Cryptic Species in Marine Diatoms". In: *Protist* 158.2 (2007), pp. 193–207. ISSN: 1434-4610. DOI: <https://doi.org/10.1016/j.protis.2006.10.001>.
- [17] Anurag A. Agrawal. "Phenotypic Plasticity in the Interactions and Evolution of Species". In: *Science* 294.5541 (2001), pp. 321–326. DOI: [10.1126/science.1060701](https://doi.org/10.1126/science.1060701).
- [18] Unterausschuss Methodenentwicklung. *Molekularbiologische Identifizierung von Pilzen mittels ITS-PCR und nachfolgender Sequenzierung. PCR und nachfolgender Sequenzierung*. Bund/Länder-Arbeitsgemeinschaft Gentechnik (LAG). URL: <http://www.lag-gentechnik.de/dokumente/uam-methoden/028.pdf> (visited on 05/30/2018).
- [19] Conrad L. Schoch et al. "Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi". In: *Proceedings of the National Academy of Sciences* 109.16 (2012), pp. 6241–6246. DOI: [10.1073/pnas.1117018109](https://doi.org/10.1073/pnas.1117018109).
- [20] Ernst Mayr. *Animal Species and Evolution*. Cambridge: Belknap Press, 1963. ISBN: 0674037502.
- [21] Michael Kloster et al. "Large-Scale Permanent Slide Imaging and Image Analysis for Diatom Morphometrics". In: *Applied Sciences* 7.4 (Mar. 2017), p. 330. DOI: [10.3390/app7040330](https://doi.org/10.3390/app7040330).
- [22] Michael Kloster, Gerhard Kauer, and Bánk Beszteri. "SHERPA: an image segmentation and outline feature extraction tool for diatoms and other objects". In: *BMC Bioinformatics* 15.1 (June 2014), p. 218. DOI: [10.1186/1471-2105-15-218](https://doi.org/10.1186/1471-2105-15-218).
- [23] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2017. URL: <https://www.R-project.org/>.
- [24] Olaf Boebel. *Station list and links to master tracks in different resolutions of POLARSTERN cruise PS103 (ANT-XXXII/2), Cape Town - Punta Arenas, 2016-12-16 - 2017-02-03*. data set. Master track der Polarsternfahrt PS103. Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, Bremerhaven, 2017. DOI: [10.1594/PANGAEA.875075](https://doi.org/10.1594/PANGAEA.875075).
- [25] Robert R. L. Guillard. "The Culture of Marine Invertebrate Animals". In: (Jan. 1975), pp. 26–60.
- [26] Machery-Nagel. *Genomic DNA from soil*. Machery-Nagel, 2017. URL: http://www.mn-net.com/Portals/8/attachments/Redakteure_Bio/Protocols/Genomic%20DNA/UM_gDNASoil.pdf (visited on 02/12/2018).
- [27] Inc NanoDrop Technologies. *ND-1000 Spectrophotometer V3.3 User's Manual*. 2007. URL: <http://ipmb.sinica.edu.tw/microarray/index.files/nd-1000-users-manual.pdf> (visited on 05/09/2018).
- [28] White et al. *Amplification and direct sequencing of fungal ribosomal RNA Genes for phylogenetics*. Jan. 1990.
- [29] *Staden Package*. Apr. 18, 2018. URL: <https://sourceforge.net/projects/staden/>.

- [30] Sudhir Kumar, Glen Stecher, and Koichiro Tamura. "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets". In: *Molecular Biology and Evolution* 33.7 (Mar. 2016), pp. 1870–1874. DOI: 10.1093/molbev/msw054.
- [31] Robert C. Edgar. "MUSCLE: multiple sequence alignment with high accuracy and high throughput". In: *Nucleic Acids Research* 32.5 (2004), pp. 1792–1797. DOI: 10.1093/nar/gkh340.
- [32] Jessica W. Leigh, David Bryant, and Shinichi Nakagawa. "popart: full-feature software for haplotype network construction". In: *Methods in Ecology and Evolution* 6.9 (), pp. 1110–1116. DOI: 10.1111/2041-210X.12410.
- [33] Tatiana Benaglia et al. "mixtools: An R Package for Analyzing Finite Mixture Models". In: *Journal of Statistical Software* 32.6 (2009), pp. 1–29. URL: <http://www.jstatsoft.org/v32/i06/>.
- [34] Douglas Bates et al. "Fitting Linear Mixed-Effects Models Using lme4". In: *Journal of Statistical Software* 67.1 (2015). DOI: 10.18637/jss.v067.i01.
- [35] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. "lmerTest Package: Tests in Linear Mixed Effects Models". In: *Journal of Statistical Software* 82.13 (2017). DOI: 10.18637/jss.v082.i13.
- [36] Andrew Gelman and Yu-Sung Su. *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models*. R package version 1.10-1. 2018. URL: <https://CRAN.R-project.org/package=arm>.
- [37] Alejandro H. Orsi, Thomas Whitworth, and Worth D. Nowlin. "On the meridional extent and fronts of the Antarctic Circumpolar Current". In: *Deep Sea Research Part I: Oceanographic Research Papers* 42.5 (1995), pp. 641–673. DOI: [https://doi.org/10.1016/0967-0637\(95\)00021-W](https://doi.org/10.1016/0967-0637(95)00021-W).
- [38] Giuseppe Cortese et al. "Glacial-interglacial size variability in the diatom *Fragilariopsis kerguelensis*: Possible iron/dust controls?" In: *Paleoceanography* 27.1 (Mar. 2012), pp. 1208–. DOI: 10.1029/2011pa002187.
- [39] Xavier Crosta. "Holocene size variations in two diatom species off East Antarctica: Productivity vs environmental conditions". In: *Deep Sea Research Part I: Oceanographic Research Papers* 56.11 (2009), pp. 1983–1993. DOI: <https://doi.org/10.1016/j.dsr.2009.06.009>.
- [40] Sunil Kumar Shukla and Xavier Crosta. "Fragilariopsis kerguelensis size variability from the Indian subtropical Southern Ocean over the last 42 000 years". In: *Antarctic Science* 29.2 (2017), pp. 139–146. DOI: 10.1017/S095410201600050X.
- [41] Pinkernell Stefan and Beszteri Bánk. "Potential effects of climate change on the distribution range of the main silicate sinker of the Southern Ocean". In: *Ecology and Evolution* 4.16 (), pp. 3147–3161. DOI: 10.1002/ece3.1138.
- [42] Lundholm Nina et al. "CRYPTIC AND PSEUDO-CRYPTIC DIVERSITY IN DIATOMS—WITH DESCRIPTIONS OF PSEUDO-NITZSCHIA HASLEANA SP. NOV. AND P. FRYXELIANA SP. NOV.1". In: *Journal of Phycology* 48.2 (), pp. 436–454. DOI: 10.1111/j.1529-8817.2012.01132.x.

Appendix

Appendix A

Additional results

A.1 Morphology

Figure A.1 shows the sample stations in the Antarctic Zone, which were left out in section 3.1.

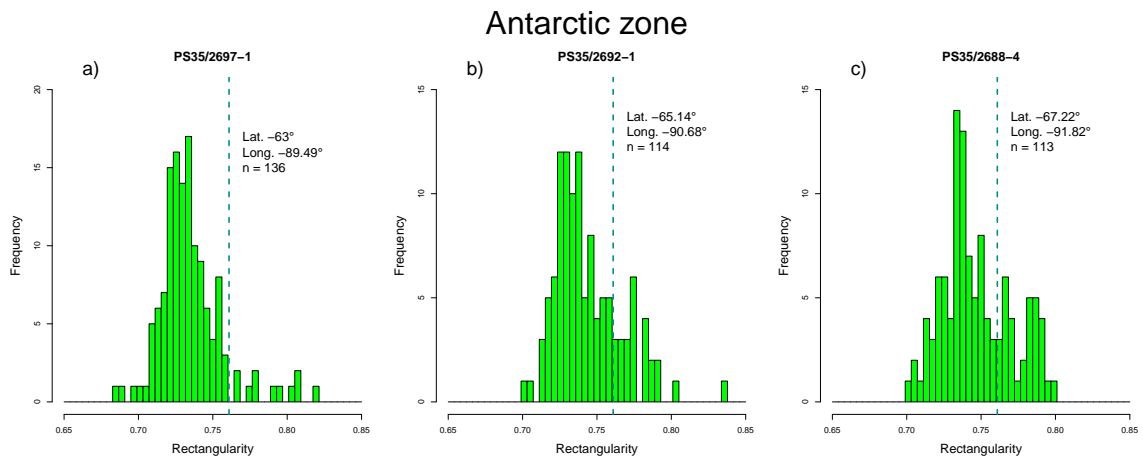


Fig. A.1: Additional histograms for the Antarctic Zone, all originating from transect PS35 and thus being sediment samples. The diagram titles display the station names. Coordinates, sample size and the threshold are also included.

Figure A.2 shows the additional sample stations in the Subpolar Zone which were not included in section 3.1.

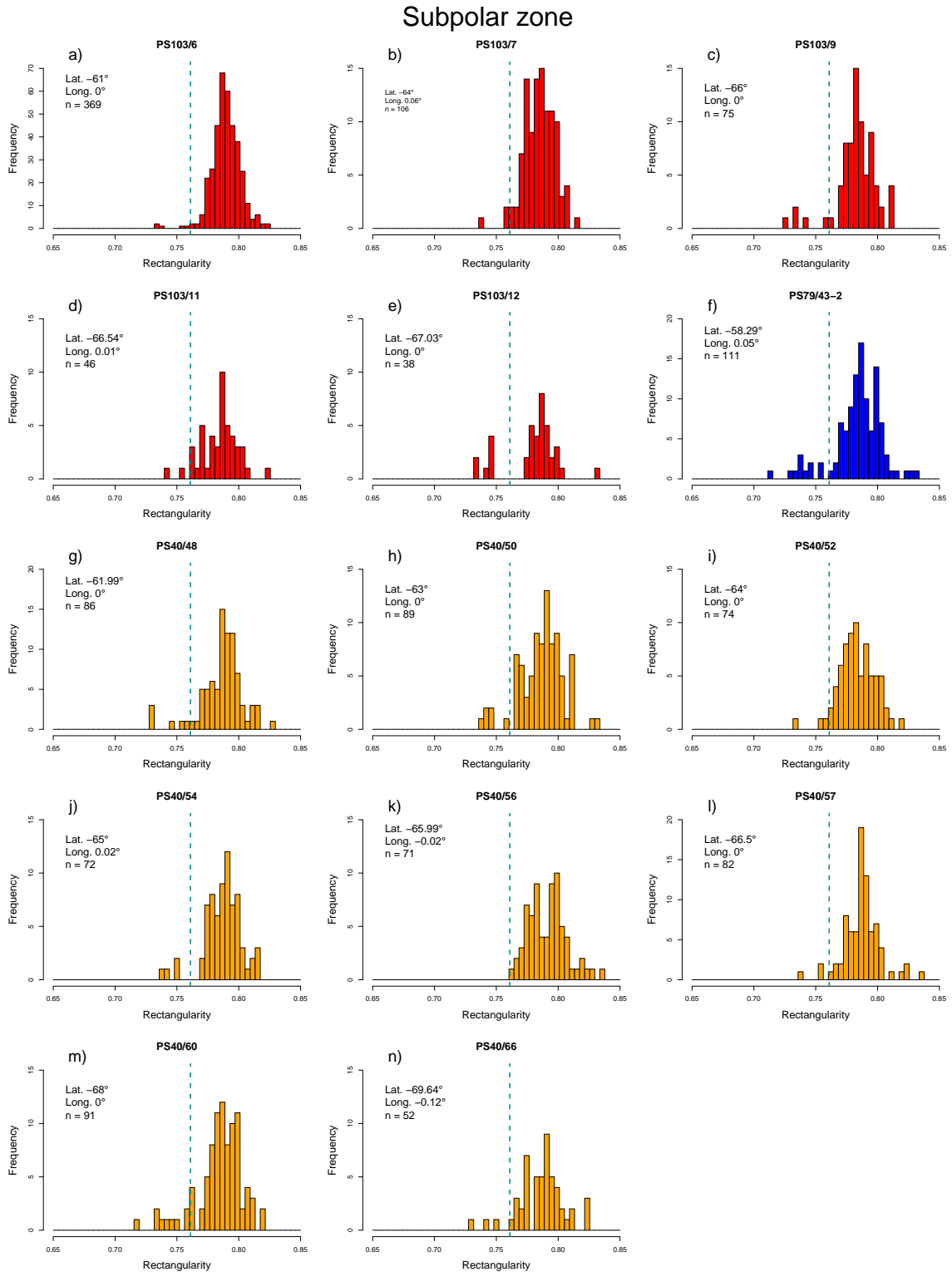


Fig. A.2: The figures shows the additional histograms from the sample stations in the Subpolar Zone. The stations names, coordinates, sample sizes and the threshold are included in the histograms

A.2 Genetic characterisation

Figure A.3 shows an exemplary result of the gel electrophoresis. On the left side is a QX DNA Size Marker 250bp–4kb v2.0 (Qiagen, Hilden, Germany) for comparison of fragment size, with the two fragments closest to the expected fragment size marked. The DNA fragments in the expected length are clearly visible making this sample suitable for sequencing.



Fig. A.3: Picture of fractionation of DNA samples done by gel electrophoresis. Fragments of ca. 700 bp are clearly visible and can be compared to the marker on the left side. The two marker fragments relevant for this analysis are marked red and green.

Table A.1 shows the variable positions in the alignment of all sequenced strains.

Tab. A.1: Table showing all variable positions in the alignment. W and K mark ambiguities, meaning that the alleles of the ITS regions differ at this position. Colours highlight different bases.

species and strain id	Position in alignment																		
	25	36	163	165	168	200	224	252	268	474	478	482	520	568	572	704	723	732	734
<i>F. obliquecostata</i> _strain: 34-27	T	G	A	T	C	C	T	A	T	G	G	T	G	C	A	G	T	T	A
<i>F. ritscheri</i> _strain: 04-05	T	G	A	T	C	C	T	A	T	G	G	T	A	C	A	G	T	T	A
<i>F. ritscheri</i> _strain: 04-18	T	G	A	T	C	C	T	A	T	G	G	T	A	C	A	G	T	T	A
<i>F. ritscheri</i> _strain: 04-45	T	G	A	T	C	C	T	A	T	G	G	T	A	C	A	G	T	T	A
<i>F. ritscheri</i> _strain: 04-48	T	G	A	T	C	C	T	A	T	G	G	T	A	C	A	G	T	T	A
<i>F. sublinearis</i> _strain: 17-15	C	A	T	A	T	T	A	T	C	A	T	T	A	G	A	G	C	G	T
<i>F. sublinearis</i> _strain: 17-16	C	A	T	A	T	T	A	T	C	A	T	T	A	G	A	G	C	G	T
<i>F. kerguelensis</i> _strain: 01-07	T	G	A	T	C	C	A	A	T	A	G	T	A	C	T	G	T	T	A
<i>F. kerguelensis</i> _strain: 01-20	T	G	A	T	C	C	A	A	T	A	G	T	A	C	T	G	T	T	A
<i>F. kerguelensis</i> _strain: 01-59	T	G	A	T	C	C	A	A	T	A	G	T	A	C	T	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-14	T	G	A	T	C	C	A	A	T	A	G	T	A	C	T	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-18	T	G	A	T	C	C	A	A	T	A	G	T	A	C	T	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-19	T	G	A	T	C	C	A	A	T	A	G	T	A	C	T	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-24	T	G	A	T	C	C	A	A	T	A	G	T	A	C	T	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-27	T	G	A	T	C	C	A	A	T	A	G	T	A	C	T	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-23	T	G	A	T	C	C	A	A	T	A	G	T	A	C	W	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-01	T	G	A	T	C	C	A	A	T	A	G	T	A	C	A	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-02	T	G	A	T	C	C	A	A	T	A	G	T	A	C	A	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-04	T	G	A	T	C	C	A	A	T	A	G	T	A	C	A	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-09	T	G	A	T	C	C	A	A	T	A	G	T	A	C	A	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-17	T	G	A	T	C	C	A	A	T	A	G	T	A	C	A	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-25	T	G	A	T	C	C	A	A	T	A	G	T	A	C	A	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-26	T	G	A	T	C	C	A	A	T	A	G	T	A	C	A	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-29	T	G	A	T	C	C	A	A	T	A	G	T	A	C	A	G	T	T	A
<i>F. kerguelensis</i> _strain: 14-30	T	G	A	T	C	C	A	A	T	A	G	T	A	C	A	G	T	T	A
<i>F. kerguelensis</i> _strain: 01-51	T	G	A	T	C	C	A	A	T	A	G	K	A	C	A	C	T	T	A
<i>F. kerguelensis</i> _strain: 01-60	T	G	A	T	C	C	A	A	T	A	G	K	A	C	A	C	T	T	A
<i>F. kerguelensis</i> _strain: 01-52	T	G	A	T	C	C	A	A	T	A	G	K	A	C	A	C	T	T	A
<i>F. kerguelensis</i> _strain: 01-09	T	G	A	T	C	C	A	A	T	A	G	G	A	C	A	C	T	T	A
<i>F. kerguelensis</i> _strain: 01-14	T	G	A	T	C	C	A	A	T	A	G	G	A	C	A	C	T	T	A
<i>F. kerguelensis</i> _strain: 01-48	T	G	A	T	C	C	A	A	T	A	G	G	A	C	A	C	T	T	A
<i>F. kerguelensis</i> _strain: 01-54	T	G	A	T	C	C	A	A	T	A	G	G	A	C	A	C	T	T	A
<i>F. kerguelensis</i> _strain: 01-56	T	G	A	T	C	C	A	A	T	A	G	G	A	C	A	C	T	T	A
<i>F. kerguelensis</i> _strain: 01-58	T	G	A	T	C	C	A	A	T	A	G	G	A	C	A	C	T	T	A
<i>F. kerguelensis</i> _strain: 01-61	T	G	A	T	C	C	A	A	T	A	G	G	A	C	A	C	T	T	A

Appendix B

Acknowledgements

First I would like to thank Dr. Bánk Beszteri for supervising my thesis and for patiently answering all the questions I had during this project.

I would also like to thank Prof. Dr. Wittke for supervising my thesis.

A big thanks goes out to Lena, for keeping the cultures alive and supplying them with new growth medium and new homes every five weeks. Without you and all the diatoms none of this would have happened.

I thank Michael for the crash course in scientific writing and sharing your office with me and everybody else from Hustedt Diatom Study Center for their support during my thesis.

I would also like to thank Nancy Kuehne and Annegret Mueller from the Ecological Chemistry division for letting me use their laboratories and operating the sequencer for me.

Last but not least I would like to thank my friends and family who supported me during this journey and always encouraged me.

Appendix C

Declaration of Originality

I hereby declare that this thesis is my own work and was written by me without further assistance. I confirm that I have clearly referenced all sources used in this work, that I have not falsified the data and findings used in this work and that this work has not been published before.

Date

Signature

Appendix D

Data medium

Accompanying data medium for this thesis, containing:

- Digital version of the thesis (Thesis-Glemser.pdf).
- Sequencer trace files of all sequenced strains for forward and reverse primers and resulting consensus.
- SHERPA outputs for every analysed slide.
- R scripts for diagrams created with R and used .csv files.