Palacký University Olomouc

Faculty of Science

Department of Ecology and Environmental Sciences

Phylogenetic species delimitation in *Pinnularia subgibba* group and its congruence with morphology

Master Thesis

Bc. Jan Kollár

Supervisor: Prof. RNDr. Aloisie Poulíčková, CSc.

Co-supervisor: Dr. Pieter Vanormelingen

Co-supervisor: Prof. Wim Vyverman

Olomouc 2016

Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Katedra ekologie a životního prostředí

Fylogenetická delimitace druhů skupiny *Pinnularia subgibba* a její soulad s morfologií

Diplomová práce

Bc. Jan Kollár

Vedoucí: Prof. RNDr. Aloisie Poulíčková, CSc.

Spoluvedoucí: Dr. Pieter Vanormelingen

Spoluvedoucí: Prof. Wim Vyverman

Olomouc 2016

AFFIRMATION

I hereby declare that this master thesis has been written exclusively by myself and without using other resources than those listed in the “References” section.

In Olomouc on 10th May 2016 …….

ACKNOWLEDGEMENTS

I am most grateful to my supervisor prof. Aloisie Poulíčková for being excellent and versatile mentor. I am also most grateful to my co-supervisor Dr. Pieter Vanormelingen who initiated and coordinated the research on the genus *Pinnularia* and introduced me into molecular phylogenetics. I am also most grateful to my co-supervisor prof. Wim Vyverman for allowing me to cooperate with Laboratory of Protistology and Aquatic Ecology (University of Ghent) and for his overall support. I also thank prof. Vyverman, prof. Koen Sabbe and Eveline Pinseel for their cooperation, advice and constructive critique of the research. I thank Sofie D'Hondt who introduced me into the world of molecular biology. I thank Dr. Petr Dvořák for advice in the fields of molecular biology and molecular phylogenetics. I thank prof. Bart Van de Vijver who provided most of the morphology-based identification and Dr. Markéta Fránková and Dr. Petr Marvan who shared their vast experience in morphological study of diatoms. I thank Dr. Caroline Souffreau for micrographs of her strains. I am grateful to all other researchers who helped to expand Diatoms Collection of Belgian Co-ordinated Collections of Micro-organisms (BCCM/DCG) hosted by Laboratory of Protistology and Aquatic Ecology. Finally, I am most grateful to my parents and family for the overall support during my studies. The research was supported by Endowment Fund of Palacký University, Internal grant agency Faculty of Sciences Palacký University IGA Prf-2015-001 and Prf-2016-001 and Dr. Pieter Vanormelingen was financially supported by a postdoc grant from the Research Fund - Flanders (FWO).

BIBLIOGRAPHIC IDENTIFICATION

First name and surname of the author: Jan Kollár

Name of the thesis: Phylogenetic species delimitation in *Pinnularia subgibba* group and its congruence with morphology

Type of thesis: Master

Workplace: Department of Botany, Faculty of Science

Thesis supervisor: Prof. RNDr. Aloisie Poulíčková, CSc.

Year of defence: 2016

ABSTRACT

Pinnularia Ehrenberg (1843) represents one of the most species-rich genus of freshwater diatoms (Bacillariophyceae). In this study, evolutionary distances between *Pinnularia* sequences of five molecular markers were compared. Nuclear marker 28S rDNA and mitochondrial COX1 have been found to be the most variable and therefore possess the greatest potential to distinguish between closely related species of the genus. These two markers were used to infer phylogeny and to suggest primary hypothesis on species limits within *Pinnularia subgibba* group represented by 59 strains. Total of 15 monophyletic species was delimited using three methods (SPNA, GMYC, PTP). Moreover, congruence of primary hypothesis with morphology was evaluated and other available lines of evidence (e.g. biogeography, reproductive compatibility) were taken into consideration. To summarize, results suggest that several clades of *Pinnularia subgibba* group represent morphologically distinct species which might be unknown to science. Others can be relativelly clearly associated with decribed taxa. Most of the clades, however, seem to be part of large polyphyletic gibba-subgibba-parvulissima species supercomplex and it must be investigated in much more detail before coming to reliable conclusion on their true biological identity.

KEY WORDS

28S rDNA, COX1, Bacillariophyceae, diatoms, microevolution, molecular phylogenetics, polyphasic systematics

BIBLIOGRAFICKÁ IDENTIFIKACE

Jméno a příjmení autora: Jan Kollár

Název práce: Fylogenetická delimitace druhů skupiny Pinnularia subgibba a její soulad s morfologií

Typ práce: Magisterská

Pracoviště: Katedra botaniky, Přírodovědná fakulta

Vedoucí práce: Prof. RNDr. Aloisie Poulíčková, CSc.

Rok obhajoby: 2016

ABSTRAKT

Pinnularia Ehrenberg (1843) reprezentuje jeden z druhově nejbohatších rodů sladkovodních rozsivek (Bacillariophyceae). V této studii byly porovnány evoluční vzdálenosti mezi sekvencemi pěti různých genů. Bylo zjištěno, že jaderný gen 28S rDNA a mitochondriální COX1 jsou u rodu *Pinnularia* nejvariabilnější a mají tak největší potenciál pro delimitaci blízce příbuzných druhů tohoto rodu. Tyto dva geny byly použity pro odovození fylogeneze a navržení primární hypotézy o hranicích druhů uvnitř skupiny *Pinnularia subgibba*, která byla repreprezentována 59 kmeny. Třemi různými metodami (SPNA, GMYC, PTP) bylo delimitováno celkem 15 monofyletických druhů skupiny *P. subgibba*. Dále byla vyhodnocena shoda této na DNA založené primární hypotézy s morfologií a dalšími dostupnými typy důkazů (např. biogeografie či reprodukční kompatibilita). Výsledky naznačují, že některé klády mohou být poměrně jasně přiřazeny k již popsaným taxonům. Jiné představují morfologicky poměrně dobře rozlišitelné druhy, které však mohou být dosud nepopsané. Většina kládů je však součástí rozsáhlého polyfyletického superkomplexu gibba-subgibba-parvulissima a k odhalení jejich skutečné biologické identity je nezbytné důkladnější prošetření tohoto komplexu.

KLÍČOVÁ SLOVA

28S rDNA, COX1, Bacillariophyceae, mikroevoluce, molekulární fylogenetika, rozsivky, polyfazická systematika

CONTENTS

1 INTRODUCTION

1.1 Characteristics of diatoms and genus *Pinnularia*

Diatoms (Bacillariophyceae, Diatomea) are a monophyletic taxon (Cavalier-Smith & Chao 2006) of unicellular photoautotrophic eukaryots which is uniquely distinguished from other taxa by two features: (1) complex siliceous cell wall called the frustule (Round, Crawford & Mann 1990) and (2) gradual cell size reduction in a process of reproduction (Mann & Marchant 1989). The second feature is direct consequence of the first. By morphology, they can be divided into two distinct groups – centric and pennate diatoms. Centric diatoms (Centrales) are characterized by radial symetry of the cell, whereas pennate diatoms (Pennales) are organized bilaterally. However, this traditional morphology-based division is not supported by molecular evidence (Medlin et al. 1993; Alverson & Theriot 2005, Evans et al. 2007) and paraphyly of centric diatoms is suggested (Round, Crawford & Mann 1990; Medlin et al. 1996). On the other hand, pennate diatoms represent monophyletic taxon within centric diatoms (e.g. Simonsen 1979; Medlin et al. 1996; Theriot et al. 2009) and can be divided into two morphologically distinct groups – raphid and araphid diatoms. Raphid diatoms are distinguished from araphid by a presence of characteristic longitudal slits through the valve which are called the raphe system (Round, Crawford & Mann 1990). This system allows locomotion and adhesion of the cell to the substratum (Edgar & Pickett-Heaps 1983). Raphid diatoms appeared in fossil record in the Upper Cretaceous (around 70.6-55.8 Ma; Pantocsek 1889; Chacon-Baca et al. 2002; Singh et al. 2006) and since then they diversified enormously. Nowadays raphid diatoms represent vast majority of 200 000 extant diatom species estimated to exist (Mann & Droop 1996). The rate of diversification indicates evolutionary advantage bestowed by the raphe system (Sims et al. 2006).

Diatom genus *Pinnularia* Ehrenberg (1843) is classified as a member of order Naviculales and family Pinnulariaceae (Round, Crawford & Mann 1990) or Naviculaceae (Scott & Thomas 2005). The origin of the genus is not yet satisfactorily dated. The oldest known fossils of diatoms which are reliably assigned to the genus are 35-32 Ma old (Lohman & Andrews 1968). However, most recent estimation based on time-calibrated multi-gene phylogeny suggests the genus to be 50-75 Ma old (Souffreau et al. 2011). Since then it diversified greatly and became one of the most species-rich genus of raphid diatoms. To date, there is 2703 taxon names in Algaebase of which 693 are currently

taxonomically accepted (Guiry & Guiry 2016). Almost all of these taxa live solitary but few of them form chains (Krammer 2000) or band-shaped colonies (Round 1988). The genus is considered cosmopolitan and occurs (with the exeption of 10 marine forms) in freshwater habitats, preferably with a low electrolyte content (Round, Crawford & Mann 1990; Krammer 2000). Cells are very variable in size. Their naviculoid valves are linear, lanceolate or even elliptical. Poles are blunt, sometimes rostrate or capitate. Cell wall is complex and composed of inner and outer wall. This double-walled valve structure and multiseriate usually alveolar striae are characteristic for the genus. Raphe system is usually central with expanded central endings and long hooked terminal endings. Raphe may sometimes be complex, especially at large forms. Two large plate-like plastids are either separate and paralel or united by central isthmus, forming single H-shaped plastid (Round, Crawford & Mann 1990).

Phylogenetic relationships within the genus has been unknown until recently. The phylogeny of the genus has been inferred dividing it into three clades and several subclades (Fig. 1; Souffreau et al. 2011). Clade B sensu Souffreau et al. (2011) consists of three distinct subclades called grunowii, nodosa and subgibba. The subgibba subclade is in the focus of this thesis research. Both Clade B and subbgiba subclade are very well statistically supported (Fig. 1). Moreover, the specimens within these clades posses certain combination of morphological features which distinguish them from other clades of *Pinnularia* (Fig. 2). They contain small linear cells with drop-like central raphe endings (Fig. 2a) and large alveolar openings (Fig. 2b). In addition, specimens of subgibba subclade (henceforth referred to as *Pinnularia subgibba* group) are usually elongated in shape with two girdle-apressed paralel plastids (Fig. 2c) and broad non-porous central area called the fascia (Fig. 2d; Souffreau et al. 2011).

The study of diatoms is important for several reasons. Firstly, diatoms are enormously diverse group of microorganisms with more than 200 000 extant diatom species estimated to exist (Mann & Droop 1996). Vast majority of these species is unknown. Secondly, diatoms are ecologicaly extremely significant. They contribute cca. 20% to global primary production (Werner 1977; Mann 1999) and are important part of food webs in most aquatic ecosystems. Thirdly, they are being used for routine assessments of general water quality, eutrophication and acidification in many countries (Whitton & Rott 1996). Finally, they synthesize relatively high amount of energy-rich oils which gives them great biotechnological potential for production of biofuel (e.g. Chen 2012).

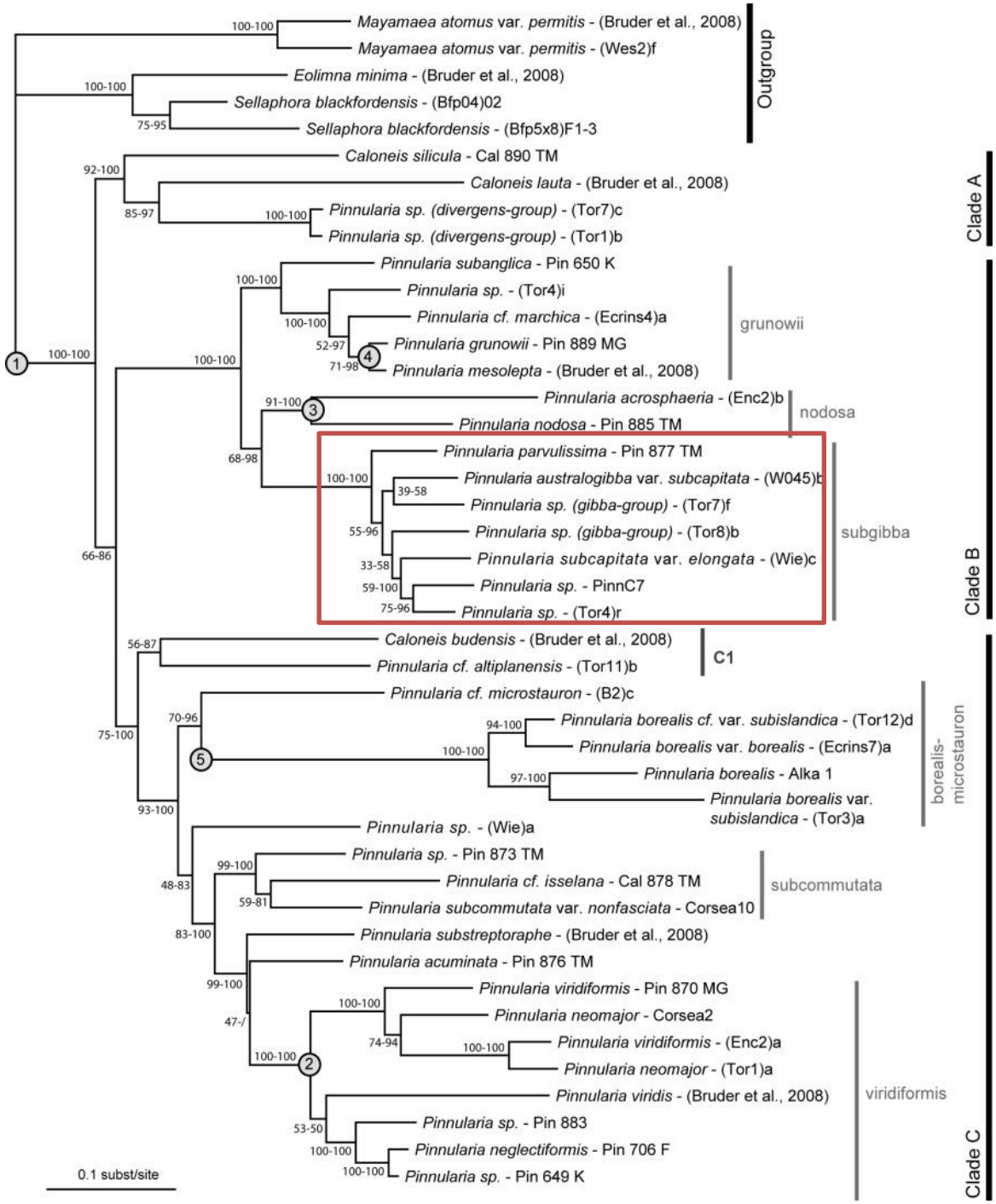

Figure 1. Adjusted from Souffreau et al. (2011) and edited. Phylogenetic relationships within genus *Pinnularia* inferred from five-locus DNA alignment using maximum likelihood under partitioned model. Numbers at nodes indicates statistical support, ML bootstrap proportions - BI posterior probabilities (both given as percentages). Encircled numbers represent nodes constrained in the relaxed molecular clock analysis. Subgibba subclade is indicated by red rectangle.

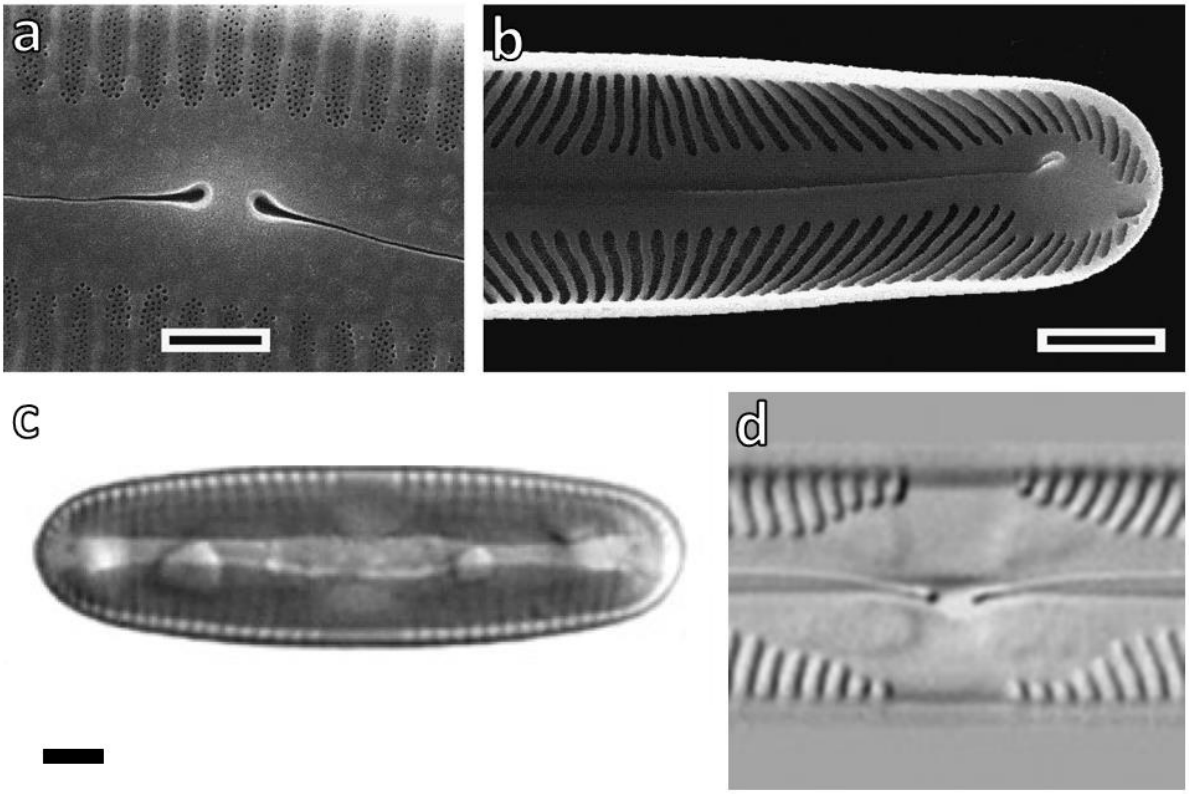

Figure 2. Adjusted from Souffreau et al. (2011) and edited. Supporting SEM (a and b) and LM (c and d) micrographs showing morphological characteristics of *Pinnularia subgibba* group: external drop-like raphe endings (a), large alveolar openings (b), paralel plastids (c) and fascia (d). Scale bars represent 2 μm (a and c) and 5 μm (b), respectively.

1.2 Identification of diatom species

In general, discovery of microorganisms is connected with the invention of light microscope in 17th century and microscopy remained one of the major tools for identification of microscopic species ever since. This tool was further enhanced by invention of electron microscope in 1930s. Microscopic observations also allowed to gather biogeographical and ecological data about microorganisms. Further developments in life sciences (esp. emersion of evolutionary biology, genetics, phylogenetics and molecular biology) led to introduction of molecular characters into systematics of life in 1980s and especially in 1990s. Moreover, development in cultivation technics allowed study of physiology (incl. reproduction) and life cycles of microorganisms. Also in diatoms, morphological, ecological, biogeographical, physiological and molecular characters can be, in theory, used for identification of species.

It has allready been mentioned above that diatoms are characterized by siliceous cell wall or the frustule. This characteristic feature has several implications on both alpha taxonomy and systematics of diatoms in general. Firstly, frustule possess distinct morphological traits which are often considered to be specific and thus useful for identification of diatom species. These traits include stria density, raphe characteristics, detailed valve structure and overall shape, lenght and breadth of a valve. Secondly, solid character of frustule cause diatoms to leave relatively well preserved fossils. This rare trait among microorganisms is undoubtly great advantage for diatomologists (e.g. Theriot et al. 2006; Souffreau et al. 2011). Thirdly, two unequally sized halves of a frustule are passed on daughter cells during asexual reproduction which results in gradual cell size reduction in population (MacDonald 1869; Pfitzer 1869). However, size reduction has its limits (e.g. Werner 1971; Edlund & Stoermer 1991) and cell size restoration is vital for survival of the population. It is usually achieved via sexual reproduction (e.g. Geitler 1932; Round, Crawford & Mann 1990) and its necessity allows us to apply biological species concept (Mayr 1942; Mayr 1946) on diatoms. Finally, we can hypothesize that frustule synthesis is relatively expansive process resulting in relatively low rate of reproduction and consequently low rate of population growth. Moreover, considerable number of diatom species can survive and reproduce only under relatively specific environmental conditions (e.g. Round, Crawford & Mann 1990; Stenger-Kovacs et al. 2007; Toporowska et al. 2008).

To summarize, morphological features on a frustule, fossilization, obligate sexual reproduction and specific environmental requirements may be helpful in attempts to delimit and identify diatom species. On the other hand, low growth rate and also specific environmental requirements makes considerable number of diatom species relatively difficult to cultivate which limits the rate of investigation of their physiological (incl. reproduction), ecological and to some extent even molecular characters. It also limits the examination of phenotypic polymorphism, phenotypic convergence and consequently cryptic species diversity. These limitations and practical difficulties have serious consequences of which I will emphasize two: (1) progress in investigation of diatom species diversity is relativelly slow and (2) morphology of diatoms is still being used as the major or even exclusive line of evidence not only for routine identifications but also for description of new species, even though importance of polyphasic approach in systematics of microorganisms is emphasized. It is well known that morphology-based

identification always carries certain level of subjectivity and even experienced diatomologists frequently disagree about identifications (Mann et al. 2010). Possible solution could be DNA barcoding.

DNA barcoding is a diagnostic technique in which short DNA sequences are used for unique characterization of species. The main advantage of DNA barcodes is that it allows an objective and unambiguous identification. On top of that, the DNA barcode sequence can point at the possible presence of an additional unknown species diversity, but only if the chosen barcode marker has sufficient resolution. In recent years, the widespread applications of DNA barcoding have resulted in a fast accumulation of DNA barcode sequences. Also in diatoms, DNA barcoding can be very useful given the often difficult identifications and the presence of large and mostly unknown cryptic species diversity (Mann et al. 2010). However, to this end, we need to build a DNA barcode database for diatoms. To build up a functional and widely used DNA barcode database it is important to choose best fitting barcode markers. In general, there are several criteria for a barcode marker such as universality, resolution and practicality (based on Hamsher et al. 2011). Universality means potential of the marker to be used for amplifying sequences of as wide range of taxa as possible while keeping as good quality of identification as possible. Resolution means discriminative power of marker. It should be high enough to distinguish even between closely related species. Practicality is also criterion of high priority as barcoding is meant to be used routinely. It means ability to be used for as low time and financial expenses as possible. Both time and financial expenses arise from the number of primer pairs used in the process of obtaining sequences. Ideal barcode marker should have high universality, high resolution and low expenses. In reality it is important to find balanced compromise between all criteria mentioned above. To date, several barcode markers were evaluated and discussed, including 18S rDNA, 28S rDNA, ITS rDNA, UPA, rbcL, psbA and COX1 (for review see Mann et al. 2010).

2 AIMS

This master thesis research is part of superrior research project with the ultimate aim to suggest a well supported hypothesis about species limits within the genus *Pinnularia*, one of the most species-rich genus of raphid diatoms. Ideally, this hypothesis should be a result of polyphasic approach based on molecular, morphological, physiological (incl. reproductive), ecological and biogeographical characters. However, fulfillment of this ultimate aim would be too demanding (in terms of both time and finance) for master thesis project. Therefore, species limits within only one clade of the genus (*Pinnularia subgibba* group) was chosen to be investigated in detail.

In species delimitation process, primary and secondary hypothesis on species limits should be achieved (e.g. Goldstein & DeSalle 2011; Puillandre et al. 2012). Primary hypothesis is based exclusively on molecular data. In order to achieve it, decision about which molecular markers should be used have to be made first. To date, several molecular markers have been proposed for identification of diatom species (Evans et al. 2007; Jahn et al. 2007; Mann et al. 2010; Hamsher et al. 2011) and their potential to distinguish between closely related species of *Pinnularia* has been investigated and compared in this study. Primary hypothesis, thereafter, should be confronted with additional characters and criteria (polyphasic approach). This confrontation results in formulation of secondary hypothesis about species limits.

To sumarize, the aims of the master thesis are:

(1) to investigate and compare different molecular markers‘ potential to distinguish between closely related species of genus *Pinnularia*,

(2) to use markers with the greatest potential to suggest primary hypothesis about species limits within *Pinnularia subgibba* group and

(3) to evaluate a congruence of this hypothesis with morphology and, if possible, other available lines of evidence.

3 MATERIAL AND METHODS

In sake of clarity I found it usefull to include diagram of workflow of the thesis research (Fig. 3).

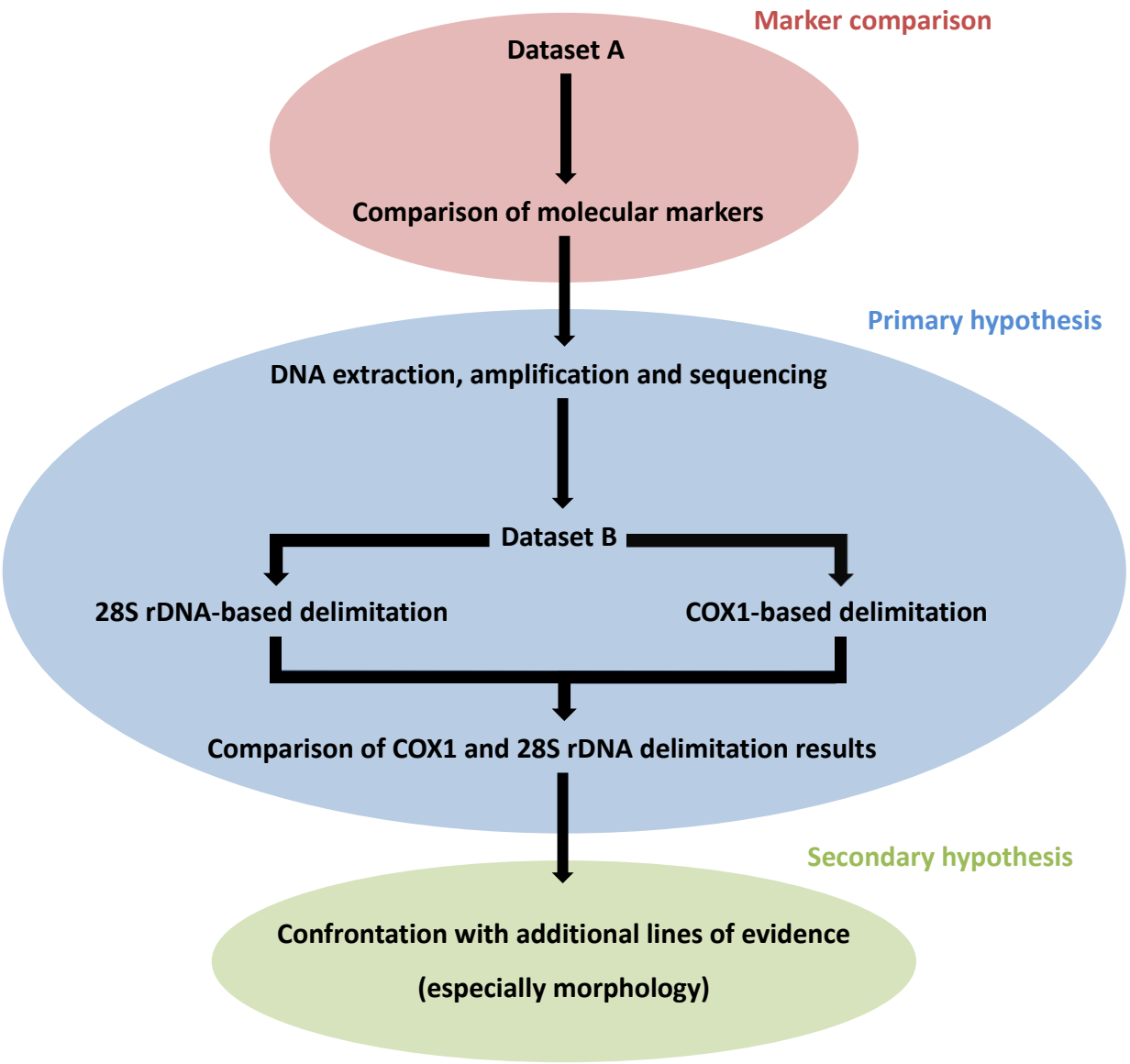

Figure 3. Simplyfied flowchart of the thesis research. Dataset A represents two subdatasets: A1 and A2. However, their divison is not important for subsequent work (primary and secondary hypotheses formulation) which is the reason why they are kept together in the diagram.

3.1 Comparison of molecular markers

Dataset A formation

One mitochondrial (COX1), two nuclear (28S rDNA, 18S rDNA) and two plastid (rbcL, psbA) markers were chosen for investigation and comparison of their potential for species delimitation and DNA barcoding of *Pinnularia*. Potential is based mainly on the criterion of resolution (i.e. discriminative power) of a marker which should be high enough to distinguish between closely related species (see 1.2 Identification of diatom species). Moreover, separate V4 region of nuclear 18S rDNA (V4 18S rDNA) and 748 bp at 3' end of plastid rbcL (rbcL-3P) were examined and compared with others for they are the most variable parts of 18S rDNA and rbcL, respectively (Hamsher et al. 2011; Zimmermann, Jahn & Gemeinholzer 2011). To distinguish V4 18S rDNA region in 18S rDNA sequences primer sequences D512-F and D978-R were used (Zimmermann, Jahn & Gemeinholzer 2011). To distinguish 748 bp of rbcL-3P CfD-F (Hamsher et al. 2011) and DPrbcL7-R (Levialdi-Ghiron 2006) primer sequences were used. In addition, we decided to compare D1, D2 and D3 regions of 28S rDNA with each other. This decision was based on the preliminary results of the comparison.

Sequences used for comparisons of COX1, 18S rDNA, V4 18S rDNA, 28S rDNA, rbcL, rbcL-3P and psbA were taken from published data (Souffreau et al. 2011) and formed Dataset A (Figure 3). It is divided into two subdatasets. Subdataset A1 was created by selection of 24 strains for which sequences of all markers were available (Appendix D - Supplementary Table I). Decision to use 100% complete dataset was made to avoid large inconsistency in output data which would make further work more difficult. Detailed characteristics of the alignments are given in Appendix D - Supplementary Table II. Comparison of D1, D2 and D3 regions of 28S rDNA is based on Subdataset A2. It consists of 27 strains for which full 28S rDNA sequences (i.e. all D1, D2 and D3 regions were included) were available (Appendix D - Supplementary Table III). Strains were selected in order to cover morphological variability within genus. To distinguish D2 region T16N-F (Hamsher et al. 2011) and D2C-R (Souffreau et al. 2011) primer sequences were used. Notice that problems occurred with distinguishing D1 and D3 regions by primer sequences and exact position ranges of the regions were not found. Therefore D1 and D3 regions as referred in this study may not be exactly identical with real D1 and D3 region borders. In this study D1 includes cca. 350 bp in front of the beginning of D2 region

(distinguished by T16N-F primer sequence as mentioned above) and D3 includes cca. 350 bp behind the end of D2 region (distinguished by D2C-R primer sequence).

Sequence alignment

Sequences of 24 strains for each marker were aligned automatically using MUSCLE algorithm as implemented in MEGA6 (Tamura et al. 2013). All alignments were then checked and corrected manually if needed. Separated alignments of D1, D2 and D3 regions were made following the same procedure.

Evolutionary distance based marker comparison

Comparison of molecular markers itself is based on evolutionary (or genetic) distances (p-distance and number of differences). Separate alignment of each marker was used to compute p-distances and numbers of differences from between the sequences using MEGA6 under default settings. Both kinds of evolutionary distances were used separately to compare each possible marker pair using visualization in MS Excel 2010.

3.2 Phylogenetic species delimitation

Taxon sampling and Dataset B formation

Firstly, all *Pinnularia* sensu lato (i.e. *Pinnularia* + *Caloneis*) strains were identified in Diatoms Collection of Belgian Co-ordinated Collections of Micro-organisms (BCCM/DCG) hosted by Laboratory of Protistology and Aquatic Ecology (PAE, Gent University, Belgium) and their 28S rDNA were sequenced as described below. In addition, all available *Pinnularia* sequences were downloaded from GenBank (www.ncbi.nlm.nih.gov/genbank). This process resulted in overall dataset of 246 *Pinnularia* strains. Records for these strains were created on Barcode of Life Data Systems (BOLD, www.boldsystems.org) and are continuously updated. However, they will be set publicly accessible after proper publication of the results. Secondly, preliminary maximum likelihood (ML) phylogeny based on 28S rDNA was produced in order to select clade for detailed examination of species limits. *Pinnularia subgibba* group (as distinguished by Souffreau et al. 2011) appeared to be the largest (in term of number of strains) and well supported (1000 generations bootstrap support = 0.99) clade of the tree. It consists of 105 of total 246 *Pinnularia* strains and therefore was selected to be examined in detail. Finally, 105 28S rDNA sequences of *Pinnularia subgibba* group strains were examined for presence of identical sequences. From each identical sequence three (if possible) were

selected (preferably from different samples) in order to include all species and intraspecific variability within the group. This process resulted in 59 strains of *Pinnularia subgibba* group (Appendix D - Supplementary Table IV) with no strains downloaded from GenBank. Information on the origin of these strains is given in Appendix D – Supplementary Table V. Decision to use 28S rDNA and COX1 for phylogenetic species delimitation was made according to the results of molecular markers comparison (4.1 Comparison of molecular markers). Attempts to obtain COX1 sequences for all 59 strains were made following the procedure described below. However, I repeatedly failed to obtain COX1 sequences for 12 of these 59 strains (indicated by asterisk in Appendix D – Supplementary Table IV). Possible causes of the failure are discussed below (5. Discussion). Whole process resulted in Dataset B which consists of 28S rDNA sequences for 59 strains and COX1 sequences for 47 strains.

DNA extraction, PCR amplification and sequencing

Diatoms were harvested and prepared for DNA extraction by centrifugation. DNA was extracted following Zwart et al. (1998) without the last purification step. 28S rDNA sequences were amplified in two separate reactions, one for D1-D2 region and the other for D2-D3 region. DNA sequences of D1-D2 region were amplified using PCR primers D1R-F (forward) and D2C-R (reverse). Sequences of D2-D3 region were amplified using PCR primers T16N (forward) and T24U (reverse). Details on PCR primers are given in Appendix D – Supplementary Table VI. PCR reaction mixtures contained: 5µl of 10X PCR buffer (Tris-HCl, $(NH_4)_2SO_4$, KCl, 15mM MgCl2, pH 8.7 at 20°C; ‘‘Buffer I’’, Applied Biosystems, Foster City, USA), 5µl of deoxynucleoside triphosphates at a concentration of 2mM each, 4µl of each primer at 5µM, 2µl of bovine serum albumin (BSA) at $0.4 mg.ml^{-1}$, 2µl of Taq polymerase at $1U.\mu l^{-1}$ (AmpliTaq, Perkin-Elmer, Wellesley, USA), 2µl of template DNA and 26µl of sterile water. Total reaction volume was 50µl. Following protocol for amplification of 28S rDNA was applied: initial denaturation for 5 min at 95°C, followed by 35 cycles of denaturation for 1 min at 94°C, annealing for 1 min at 55°C and elongation for 1 min at 74°C, and final elongtion for 10 min at 72°C. COX1 sequences were amplified using PCR primers GazF2 (forward) and KEdtmR (reverse). Details on PCR primers are given in Appendix D – Supplementary Table VI. PCR reaction mixtures contained same substances as described above. Total reaction volume was 50µl. Following protocol for amplification of COX1 was applied: initial denaturation for 3 min at 95°C, followed by 35 cycles of denaturation

for 0.5 min at 95°C, annealing for 1 min at 50°C and elongation for 1.5 min at 72°C, and final elongation for 5 min at 72°C. The success of the reaction was checked using gel electrophoresis on 1.5% agarose gel. PCR products were sent to be sequenced without any purification. Sequences were assembled using BioNumerics version 3.5.

Sequence alignment and model testing

28S rDNA sequences of 59 strains and COX1 sequences of 47 strains were aligned separately using automatic algorithm MUSCLE as implemented in MEGA6 (Tamura et al. 2013). All alignments were then checked and corrected manually if needed. Best fitting nucleotide substitutional models for each alignment were found using maximum likelihood method under default settings as implemented in MEGA6. Bayesian information criterion (BIC) was used to choose the best models (Appendix D – Supplementary Table VII).

Phylogeny inferrence

Phylogenies of *Pinnularia subgibba* group based on 28S rDNA and COX1 were inferred separately using two methods: (1) Maximum Likelihood phylogeny inference (ML) and (2) Bayesian Inferrence (BI). First method based on ML was conducted in MEGA6. GTR+G+I (for 28S rDNA; Nei & Kumar 2000) and TN93+G (for COX1; Tamura & Nei 1993) nucleotide substitution models were set following the results of model testing (Appendix D – Supplementary Table VII). Initial tree for the heuristic search was obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. All positions with less than 50% site coverage were eliminated. That is, fewer than 50% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 912 positions in the final dataset of 28S rDNA sequences and 671 positions in the final dataset of COX1 sequences. The percentage of trees in which the associated taxa clustered together was calculated using bootstrap with 1000 replications.

Second method based on BI was conducted in MrBayes 3.2.2 (Ronquist et al. 2012). GTR+G+I nucleotide substitution model was set for 28S rDNA. However, GTR+G substitution model was set for COX1 because TN93+G is not implemented in MrBayes. Two independent runs of Markov chain Monte Carlo (MCMC) analysis were performed for each marker separately for 2 000 000 generations. 1 cold and 3 heated chains were run for each MCMC analysis. Diagnostical frequency was set to 1000 generations. The Markov chain was sampled every 100 generations. 25% of samples were discarded.

Phylogenetic species delimitation

Phylogenetic species delimitation was performed for each marker separately using three methods: (1) Statistical Parsimony Network Analysis, (2) Generalized Mixed Yule Coalescent and (3) Poisson Tree Processes. Statistical Parsimony Network Analysis (SPNA; Hart & Sunday 2007) divides a haplotype network in different statistically independent networks which corresponds to putative species boundaries. Analysis was performed using TCS 1.21 (Clement, Posada & Crandall 2000). Gaps were treated as fifth state and connection limit was set to 95% which was calculated to 13 detectable mutations for 28S rDNA and 11 detectable mutations for COX1.

Generalized Mixed Yule Coalescent (GMYC, Pons et al. 2006) detects increase in branching rate at the transition between interspecific and intraspecific variation. Input ultrametric tree was inferred using MrBayes 3.2.2. Following the results of model testing (Appendix D – Supplementary Table VII) nucleotide substitution model was set to GTR+I+G for 28S rDNA and GTR+G for COX1. Strict molecular clock model was set. Two independent runs of MCMC analysis were performed for each marker separately for 2 000 000 generations. 1 cold and 3 heated chains were run for each MCMC analysis. Diagnostical frequency was set to 1000 generations. The Markov chain was sampled every 100 generations. 25% of samples were discarded. GMYC itself was performed in R 3.2.3 (R Core Team 2015) using packages splits (Ezard, Fujisawa & Barraclough 2014), ape (Paradis, Claude & Stimmer 2004), MASS (Venables & Ripley 2002) and paran (Dinno 2012).

Poisson Tree Processes (PTP; Zhang et al. 2013) is another coalescent-based delimitation method which infer putative species boundaries on a given phylogenetic input tree. PTP was performed using Species delimitation web server (www.species.h-its.org). Outgroup sequences were removed from the tree. Number of MCMC generations was set to 500 000 and 25% of samples were discarded.

3.3 Morphological study

Vouchers of cleaned material of the original natural samples were adjusted from BCCM/DCG and are held in PAE, Gent University, Belgium. Strains were studied using light microscopes Zeiss Axiophot 2 with objective Plan-Apochromat 100x/1.4 N.A., oil immersion, DIC (in Gent) and Zeiss Primo Star with objective Plan-Apochromat

100x/1.25 N.A., oil immersion (in Olomouc). Microphotographs were taken in various quality using camera AxioCam MRm. I focused especially on characters of diatom valves such as shape, length (L), breadth (B), stria density (S), character of central and axial area and raphe features. Identifications are based mostly on Krammer (2000).

4 RESULTS

4.1 Comparison of molecular markers

Comparison of five markers (COX1, 28S rDNA, 18S rDNA, rbcL and psbA) and five separate regions of these markers (D1, D2 and D3 regions of 28S rDNA, V4 region of 18S rDNA and 748 bp at 3' end of rbcL) is based on evolutionary distances (p-distances and numbers of differences) between sequences of 24 (Appendix D – Supplementary Table I) and 27 strains (Appendix D – Supplementary Table III), respectively. Six most representative comparisons based on p-distances are shown (Fig. 4). X-axis shows p-distances of one marker's sequence pair (e.g. proportional distance between two sequences of 28S rDNA in graph a) whereas y-axis shows p-distances of another marker's sequence pair (e.g. proportional distance between two sequences of COX1 in graph a). Each black dot represents correlation of p-distances between one strain pair's sequences of one marker (x-value) and another marker (y-value). Red line represents linear function x=y. If p-distance between two sequences have exactly the same value for both compared markers, black dots will cover the red line. If one marker's sequence pair have bigger p-distance value (i.e. sequences are more different) than another, black dot will be deflected in the direction of first marker's axis.

Following this assumption we can compare COX1 and 28S rDNA in graph a (Fig. 4). It is obvious that cloud of dots is slightly deflected in the direction of x-axis which represents p-distances of 28S rDNA. So p-distances between two sequences of 28S rDNA are slightly bigger than p-distances between two sequences of COX1 (i.e. two 28S rDNA sequences are more different from each other than same strains' COX1 sequences). Therefore, 28S rDNA is slightly less conservative than COX1 and have slightly bigger resolution. If we will apply same procedure on the rest of graphs shown in Figure 4 we will come to following conclusion. According to p-distance analysis 28S rDNA has the greatest resolution among compared markers and some of their most variable regions. COX1 has slightly less resolution and is followed by V4 region of 18S rDNA. 18S rDNA has slightly less resolution than V4 18S rDNA and is comparable with rbcL-3P. RbcL-3P has slightly more resolution than whole rbcL and psbA has the least resolution of compared markers and some of their regions.

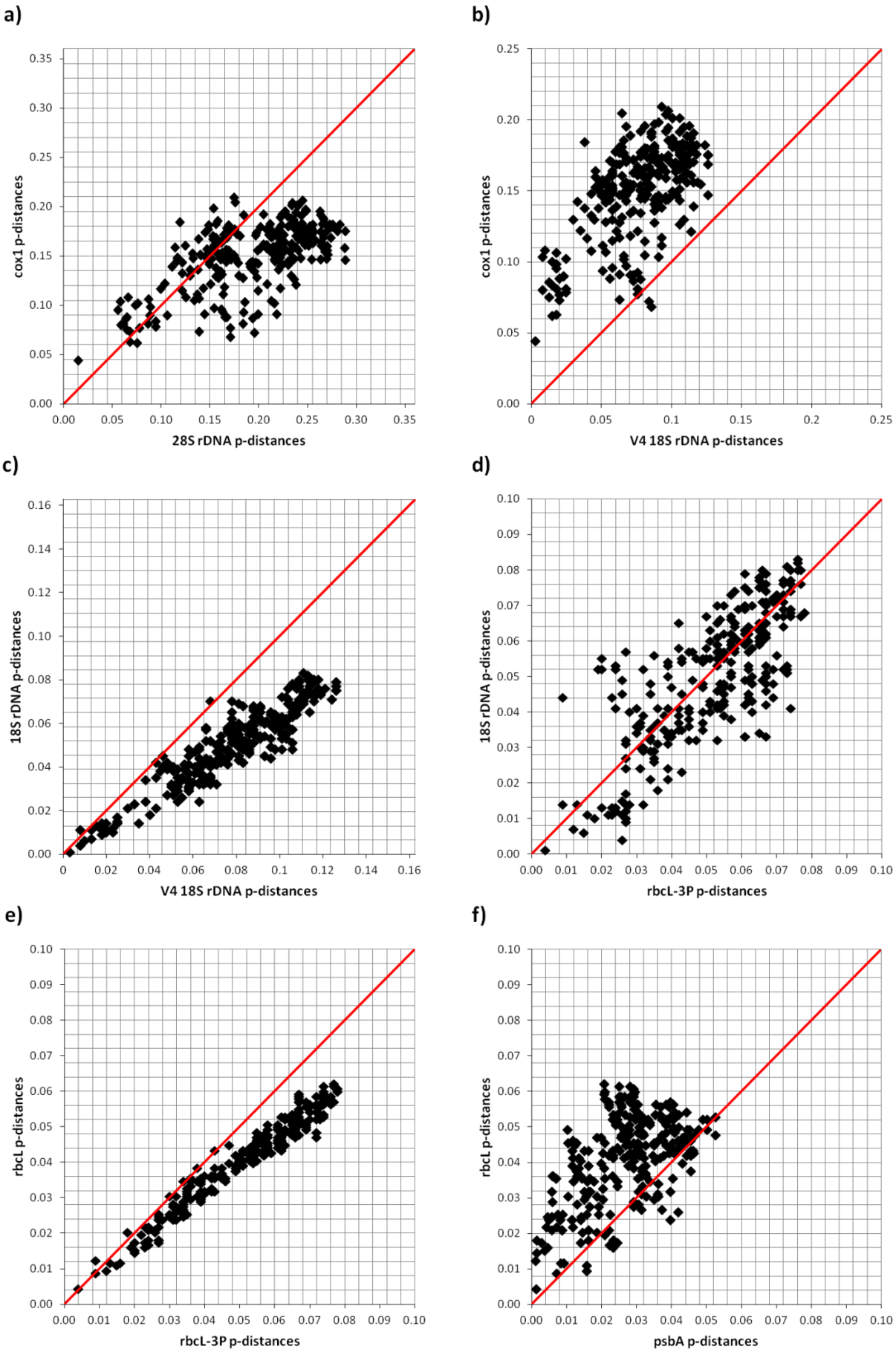

Figure 4. Comparison of p-distances from between nucleotide sequences computed for different marker pairs. P-distances were computed using MEGA6 (Tamura et al. 2013). The analysis involved 24 strains taken from published data (Souffreau et al. 2011). Six different marker pair comparisons are shown: (a) D1-D2 28S rDNA with COX1, (b) V4 18S rDNA with COX1, (c) V4 18S rDNA with 18S rDNA, (d) rbcL-3P with 18S rDNA, (e) rbcL-3P with rbcL and (f) psbA with rbcL.

Similarly, six most representative comparisons based on numbers of differences are shown (Fig. 5). If we will follow same procedure as in evaluation of the results based on p-distances (Fig. 4) we will come to following conclusion. According to numbers of differences 28S rDNA has the greatest resolution among compared markers and their most variable regions. Sequence of the rest of markers from the most variable (with the greatest resolution) to the most conservative is COX1, 18S rDNA, rbcL, rbcL-3P, V4 18S rDNA and psbA.

Results based on both p-distance and number of differences show 28S rDNA as most variable marker with the greatest discriminative power among compared markers. Therefore, we decided to compare three separate regions of this marker (i.e. D1, D2 and D3) one with each other. Three comparisons of D1, D2 and D3 regions based on p-distances (graphs a-c) and three comparisons based on numbers of differences (graphs d-f) are shown (Fig. 6). It is obvious that D2 region has the greatest resolution and is followed by D1 region. D3 region appeared to be the most conservative among the three regions of 28S rDNA.

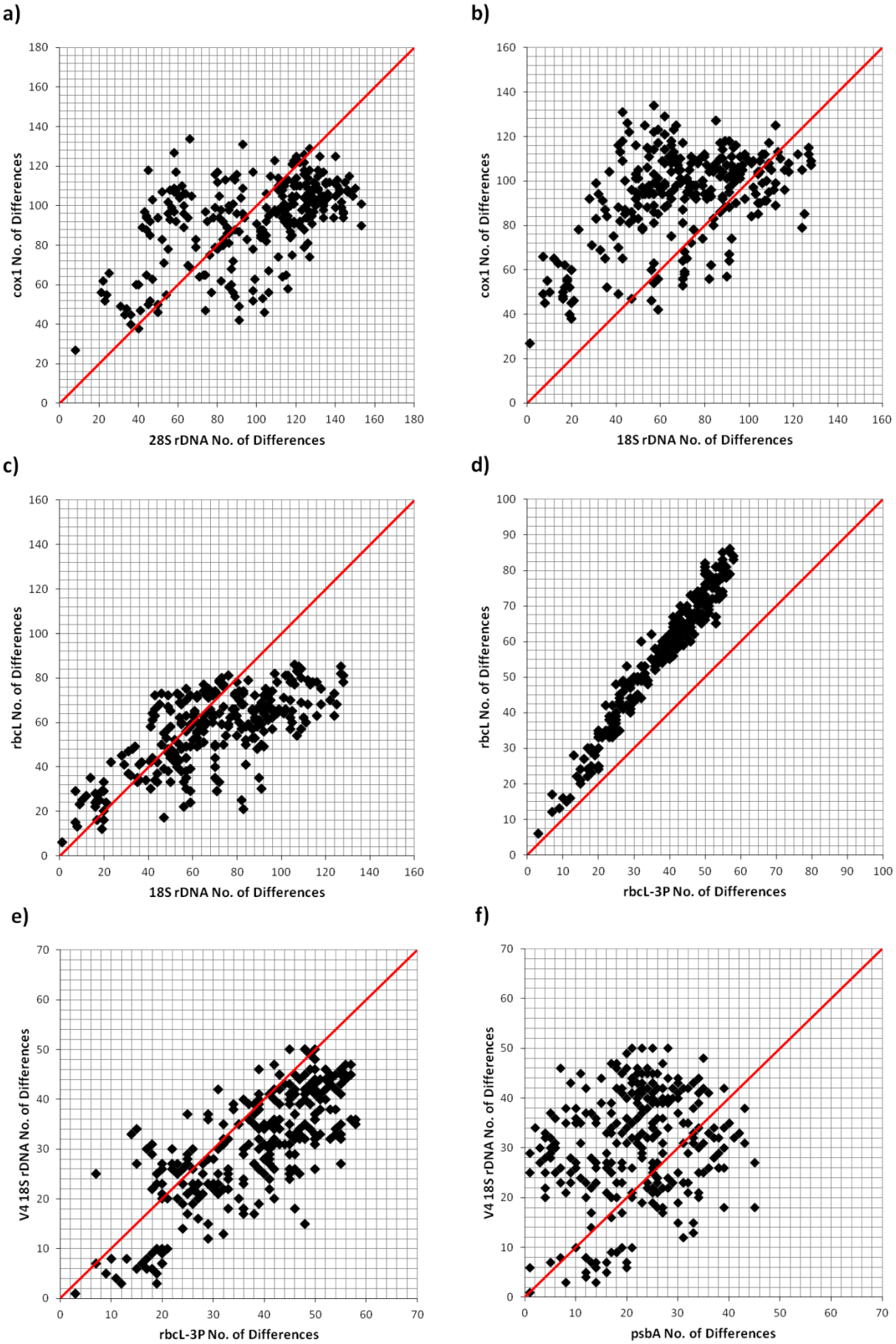

Figure 5. Comparison of numbers of differences between nucleotide sequences computed for different marker pairs. Numbers of differences were computed using MEGA6 (Tamura et al. 2013). The analysis involved 24 strains taken from published data (Souffreau et al. 2011). Six different marker pair comparisons are shown: (a) D1-D2 28S rDNA with COX1, (b) 18S rDNA with COX1, (c) 18S rDNA with rbcL, (d) rbcL-3P with rbcL, (e) rbcL-3P with V4 18S rDNA and (f) psbA with V4 18S rDNA.

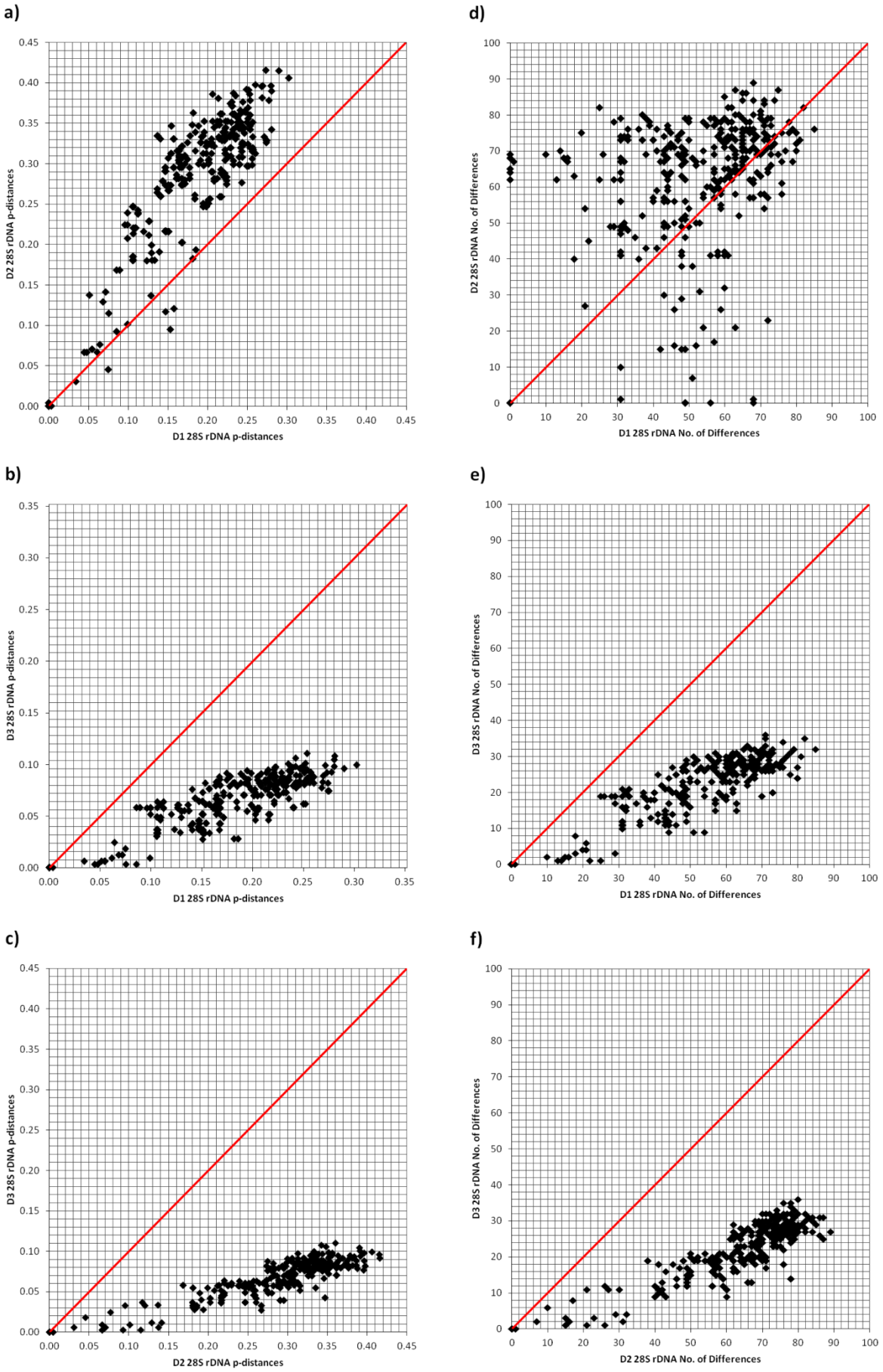

Figure 6. Comparison of p-distances and numbers of differences between nucleotide sequences computed for different 28S rDNA region pairs. Evolutionary distances were computed using MEGA6 (Tamura et al. 2013). The analysis involved 27 strains taken from BCCM/DCG. Three p-distance based (a,b,c) and three number of differences based (d, e, f) comparisons are shown: (a and d) D1 28S rDNA with D2 28S rDNA, (b and e) D1 28S rDNA with D3 28S rDNA, (c and f) D2 28S rDNA with D3 28S rDNA.

4.2 Phylogenetic species delimitation

Phylogeny inferrence and phylogenetic species delimitation results based on 28S rDNA (Fig. 7) and COX1 (Fig. 8) are shown. All three methods' results inferred from COX1 alignment are identical and reveals 15 species boundaries within *Pinnularia subgibba* group. Results based on 28S rDNA are identical with one exception. GMYC and PTP delimited cluster D (Fig. 7) as one species whereas SPNA divides it into two species.

Comparison of 28S rDNA with COX1 are shown (Fig. 9). In clades A, B, C, E, F, K and all singleton strains (i.e. Tor7f, CZECH_NOS-7 and Tor4r) there is perfect congruence in detected species limits across both genes and all delimitation methods. Clades G, I and L are congruent as well but one COX1 sequence per clade is missing due to amplification failure. No COX1 sequence of strains of clade H was amplified succesfully and thus congruence of clade H among the two genes can not be evaluated. The only incongruence among the two genes appeared in clade J. Species limits based on COX1 clearly distinguished strain MIC5_16 as a separate species, whereas 28S rDNA-based results suggest it to be part of the species boundary J.

To summarize, 15 (GMYC and PTP) or 16 (SPNA) species clusters were delimited using 28S rDNA sequences and 15 species clusters were delimited using COX1 sequences. In species limits based on 28S rDNA one incongruence in SPNA result in clade D appeared. On the other hand, there is perfect congruence in methods delimiting species limits based on COX1.

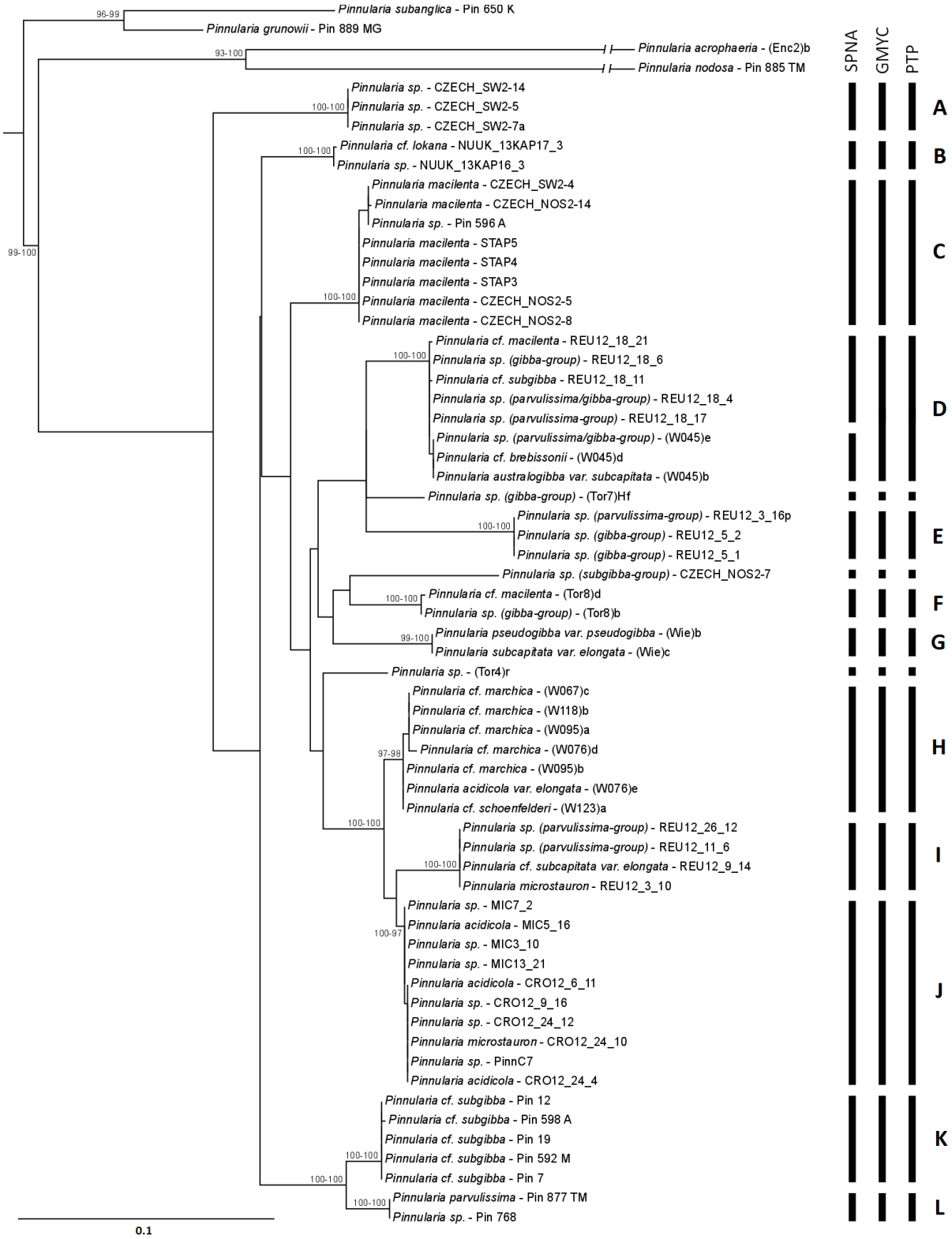

Figure 7. 28S rDNA phylogenetic species delimitation within *Pinnularia subgibba* group visualized on a phylogeny inferred from 28S rDNA alignment using maximum likelihood. Numbers at nodes indicates statistical support, ML bootstrap proportions - BI posterior probabilities (both given as percentages and only values above 90% are shown). Species boundaries inferred using three different methods (SPNA, GMYC, PTP) are indicated by black bars.

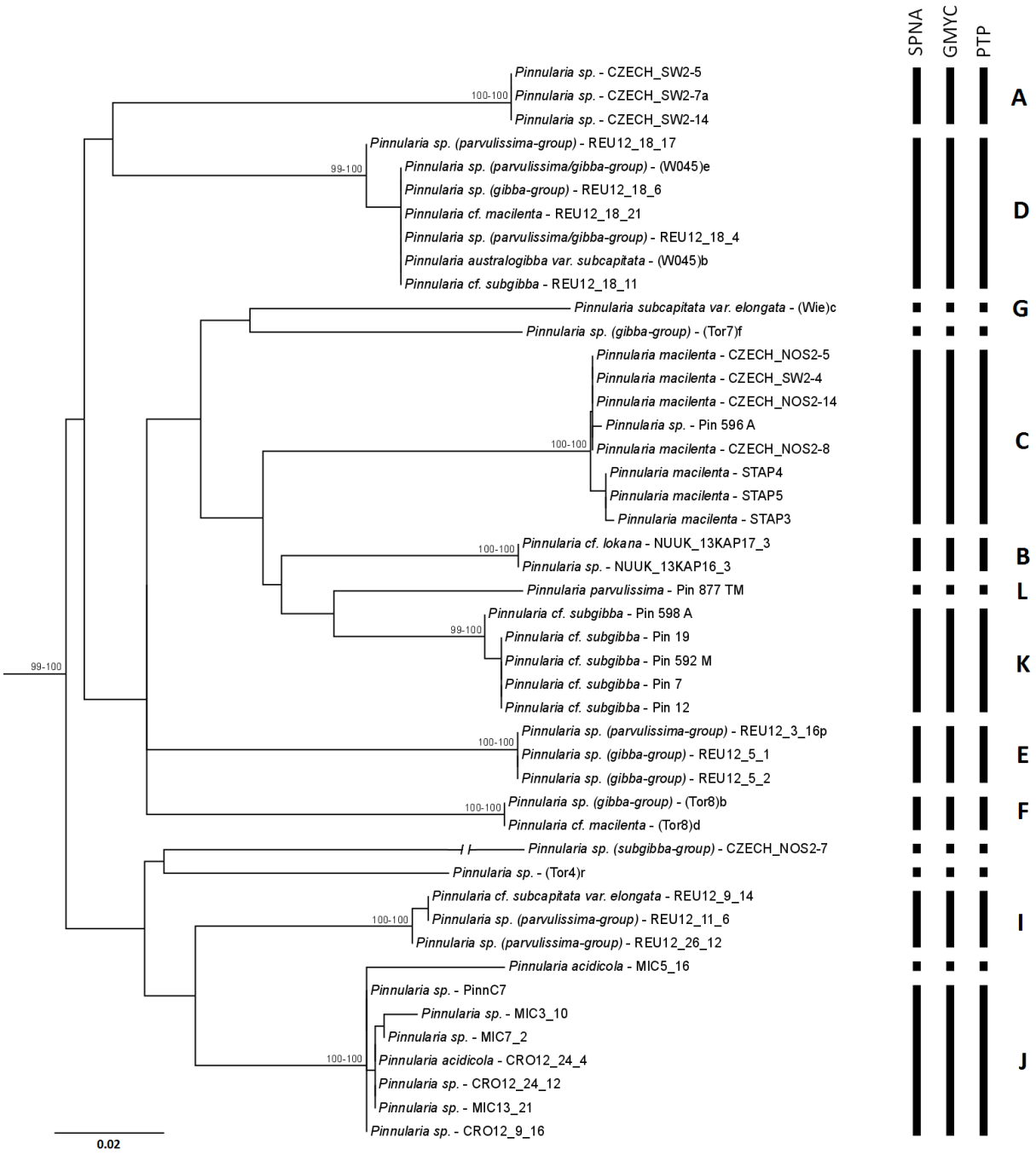

Figure 8. COX1 based phylogenetic species delimitation within *Pinnularia subgibba* group visualized on a phylogeny inferred from COX1 alignment using maximum likelihood. Numbers at nodes indicates statistical support, ML bootstrap proportions - BI posterior probabilities (both given as percentages and only values above 90% are shown). Species boundaries inferred using three different methods (SPNA, GMYC, PTP) are indicated by black bars. Outgroup sequences were deleted from the tree for sake of visualization.

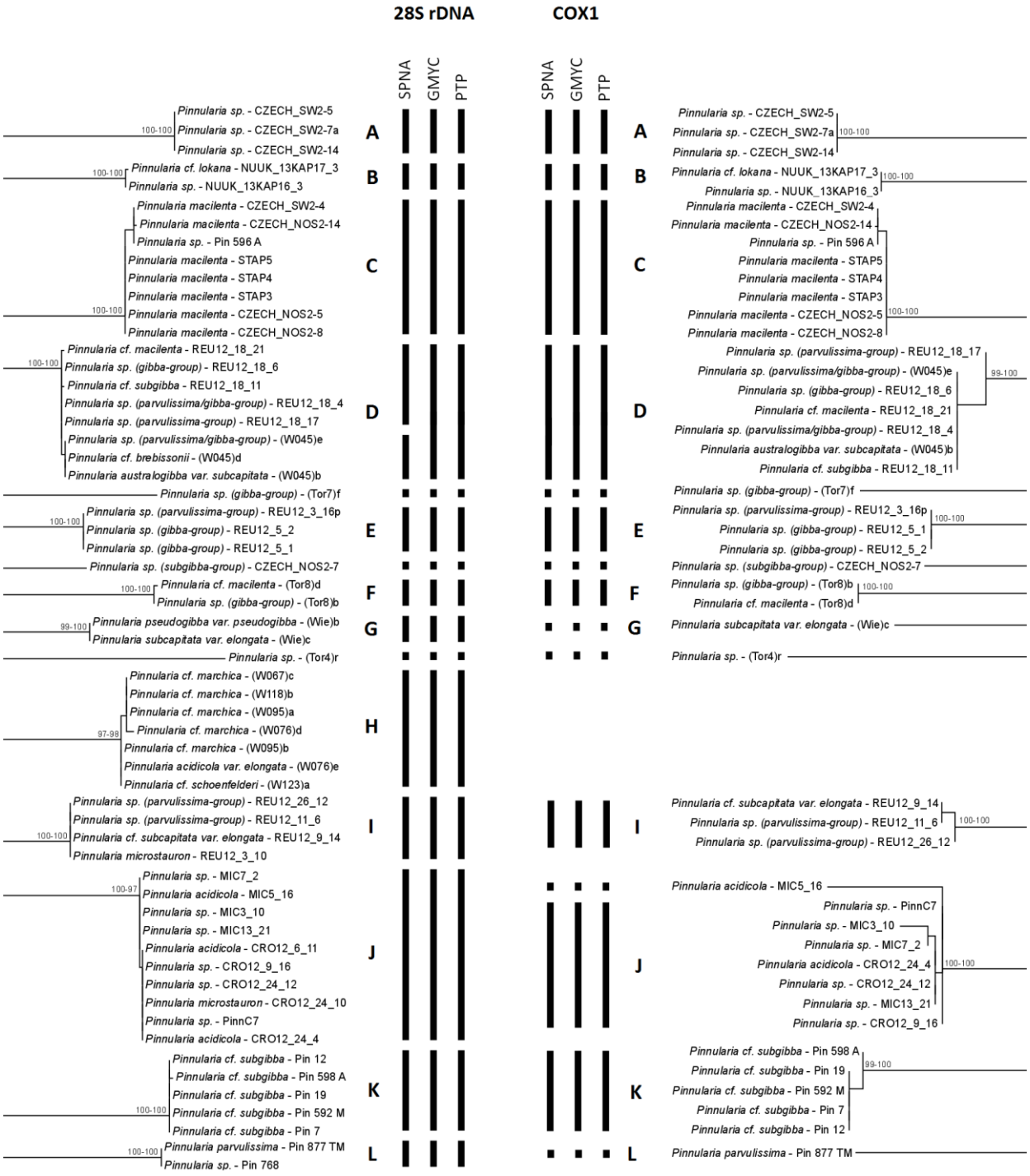

Figure 9. Comparison of 28S rDNA-based species limits with COX1-based species limits. Species boundaries inferred using three different methods (SPNA, GMYC, PTP) are indicated by black bars. Different size of black bars in G, I and L clades is caused by different number of 28S rDNA and COX1 sequences in these clades. 15 (GMYC and PTP) or 16 (SPNA) species clusters were delimited using 28S rDNA sequences. Similarly, 15 species clusters were delimited using COX1 sequences with perfect congruence across all delimitation method. However, none COX1 sequence of clade H was amplified sucessfully. Possible reasons of this failure are discussed below.

4.3 Morphological study

Morphological variability of strains were studied and documented. Representatives for strains are shown (Figures 10-17). Average length, breadth and stria density from 10 valves per strain along with standard deviations are given (Tables 1-8).

Clade A is relatively uniform in morphology (Fig. 10). Strains of this clade are characterized by linear valves with sides slightly concave in the center and convex in the ends. Maximal width of the valve is in $^1/_4$ the valve lenght. Ends are broadly subrostrate. Central area is narrow with asymetric fascia. Axial area is narrow and covers $^1/_4$ or less of the valve breadth. Transapical striae are slightly radiate in the middle and convergent near the poles. Average length, breadth and stria density from 10 valves per strain along with standard deviations are given (Tab. 1). I have identified this clade as *Pinnularia cf. microstauron*.

Strains of clade B (Fig. 10) are similar in morphology. They both have complex lateral raphe with large and strongly hooked terminal fissures. Axial area covers from $^1/_5$ to $^1/_3$ of the valve. It widens continuously from ends to the elongated rhombic central area. The fascia is small and asymetric. There are usually 2 or 4 indistinct markings accompanying the fascia. Valve end of both strains are broadly rounded. Shape of the valves, however, is different in both strains. Valves of NUUK13_KAP17_3 are linear with parallel sides, whereas valves of NUUK13_KAP16_3 are concave in the center. Morphometric measurements are given (Tab. 1). Strain NUUK_13KAP17_3 was identified as *Pinnularia cf. lokana*, while strain NUUK_13KAP16_3 was identified only to the genus level.

Clade C (Fig. 11) possess perhaps the most distinctive morphology among all clades of *Pinnularia subgibba* group. Most of the strains were identified as *Pinnularia macilenta*. The only exception is Pin 596 A which was identified only to the genus level. In general, the strains of this clade are characterized by linear valves with sides slightly convex in the center and slightly concave in the ends. Valve ends, however, are slightly subcapitate, sometimes broadly rounded. Central area is irregularly rhombic and covers from $^2/_3$ to $^3/_4$ of the valve breadth. Small asymetric transverse fascia is always present and is usually accompanied by large indistinct markings. Axial area covers from $^1/_4$ to $^1/_2$ of the valve breadth. Raphe is complex, broadly lateral and filiform on a short distance. Central pores are large, drop-like, close standing and laterally bent. Terminal fissures are large and hooked. Transapical striae are slightly radiate in the middle and convergent near

the poles. Average length, breadth and stria density from 10 valves per strain along with standard deviations are given (Tab. 2).

Clades D (Fig. 12; Tab. 3), E (Fig. 13e-g; Tab. 4), F (Fig. 13a-b; Tab. 4), I (Fig. 15a-c; Tab. 6), K (Fig. 17a-d; Tab. 8), L (Fig. 17e; Tab. 8) and singleton strain CZECH_NOS2-7 (Fig. 15e; Tab. 6) possess relatively uniform morphology. It can be characterized by lanceolate or rhombic-lanceolate outline, paralel or slightly convex sides and broadly rounded to broadly subcapitate ends. Raphe is lateral with large hooked terminal fissures. Central pores are relativelly large, drop-like, close standing and deflected to the edge of the valve. Central area is of various size and usually iregularly rhombic. Axial area often widens from the ends to the central area. Striae slightly radiate in the middle and slightly to strongly convergent in the ends. Strains of these clades were often identified as members of *Pinnularia gibba*-group, *P. subgibba*-group or *P. parvulissima*-group. Morphometric measurements are given in tabels mentioned above.

Clades H (Fig. 14) and J (Fig. 16) are very similar in morphology to clades mentioned in previous paragraph. However, it seems that their central area is larger and less rhombic. Also their axial area is always narrow. Average length, breadth and stria density from 10 valves per strain along with standard deviations are given for clade H (Tab. 5) and J (Tab. 7). Clade G (Fig. 13c-d) seems very disunited in morphology. Strain (Wie)b is generally braoder and it has broadly rounded ends of the valve. In comparison, valve of strain (Wie)c possess very distinct shape characterized by sides concave in the middle and convex in the ends. Ends themselves are subcapitate. Central area is also extremely different in two strains. Morphometric measurements are given (Tab. 4). Singleton strain (Tor4)r (Fig. 15d) is relativelly similar to the strains of clade A. However, it is more narrow and its central area is larger. On the other hand, singleton strain (Tor7)f (Fig. 15e) is more similar to the strains of clade B. This similarity is especially in shape of the valve but other characters (esp. central area, breadth) are quite different. Average length, breadth and stria density from 10 valves of singleton strains along with standard deviations are given (Tab. 6).

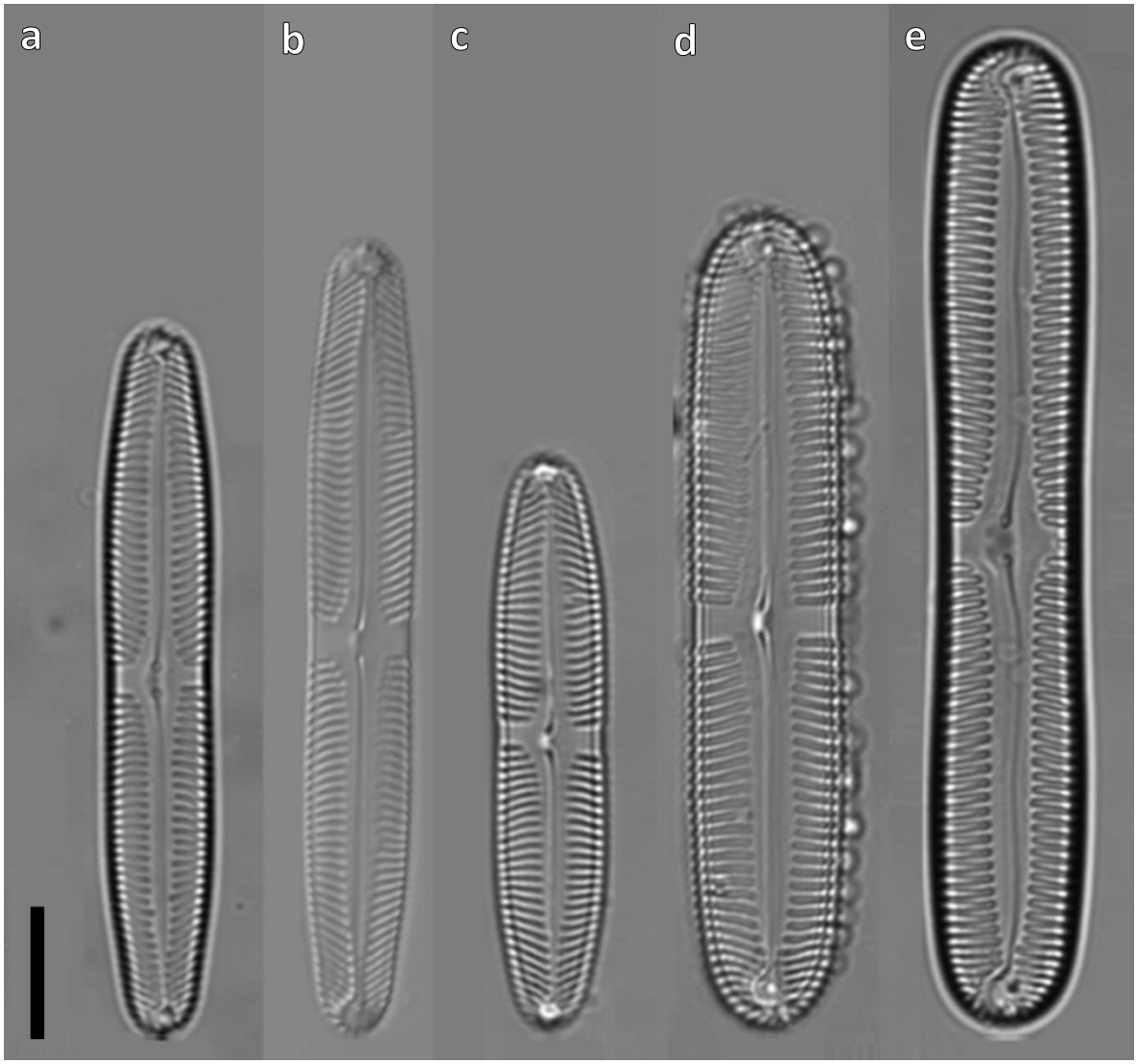

Figure 10. Supporting LM micrographs showing morphological variability of strains in clade A (a-c) and B (d-e). Micrographs are scaled. Scale bar represents 10 μm. Order of the strains is as follows: CZECH_SW2-5 in (a), CZECH_SW2-7a in (b), CZECH_SW2-14 in (c), NUUK_13KAP17_3 in (d) and NUUK_13KAP16_3 in (e).

Table 1. Morphometric measurements of strains in clades A and B (length, breadth and stria density from 10 valves per strain ± standard deviation).

Strain	Lenght (μm)	Breadth (μm)	No. of striae per 10 μm
CZECH_SW2-5	62.8 ± 0.8	8.0 ± 0.0	12.0 ± 0.0
CZECH_SW2-7a	59.2 ± 2.0	8.6 ± 0.5	11.6 ± 0.5
CZECH_SW2-14	42.8 ± 1.1	8.0 ± 0.0	12.2 ± 0.4
NUUK_13KAP17_3	65.8 ± 1.4	12.5 ± 0.5	9.8 ± 0.3
NUUK_13KAP16_3	77.4 ± 1.1	11.7 ± 0.5	9.5 ± 0.4

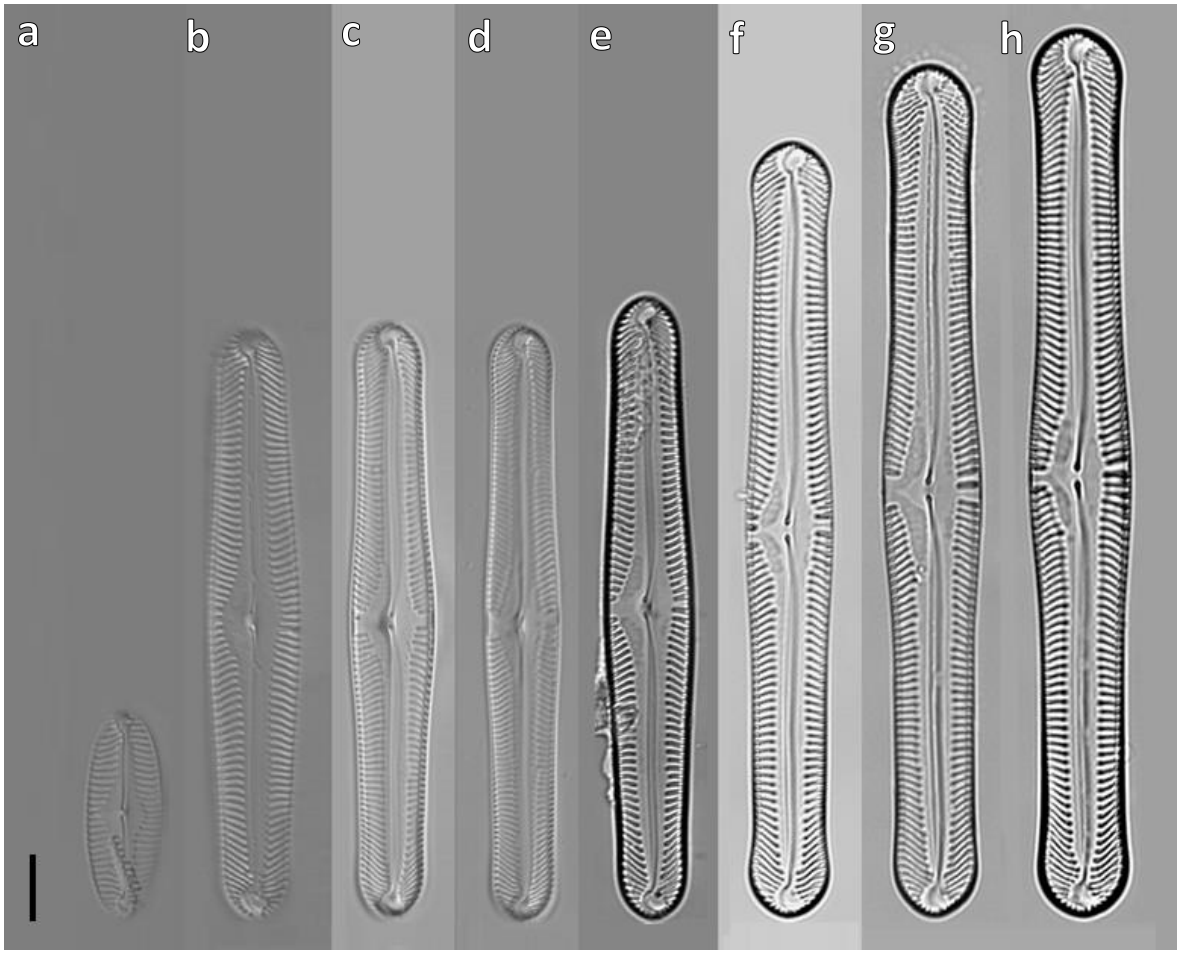

Figure 11. Supporting LM micrographs showing morphological variability of strains in clade C. Micrographs are scaled. Scale bar represents 10 μm. Order of the strains is as follows: Pin 596 A in (a), CZECH_NOS2-14 in (b), CZECH_NOS2-8 in (c), CZECH_SW2-4 in (d), CZECH_NOS2-5 in (e), STAP5 in (f), STAP4 in (g) and STAP3 in (h).

Table 2. Morphometric measurements of strains in clade C (length, breadth and stria density from 10 valves per strain ± standard deviation).

Strain	Lenght (μm)	Breadth (μm)	No. of stria per 10 μm
Pin 596 A	35.4 ± 0.8	11.9 ± 0.5	8.6 ± 0.5
CZECH_NOS2-14	89.0 ± 1.4	13.0 ± 0.0	9.2 ± 0.5
CZECH_NOS2-8	103.4 ± 1.3	14.5 ± 0.5	9.3 ± 0.5
CZECH_SW2-4	110.4 ± 1.1	14.0 ± 0.2	9.5 ± 0.4
CZECH_NOS2-5	93.2 ± 1.0	13.4 ± 0.5	10.3 ± 0.5
STAP5	116.0 ± 1.3	13.2 ± 0.4	9.2 ± 0.4
STAP4	125.2 ± 2.1	14.4 ± 0.5	9.0 ± 0.0
STAP3	128.6 ± 1.6	15.0 ± 0.0	8.7 ± 0.5

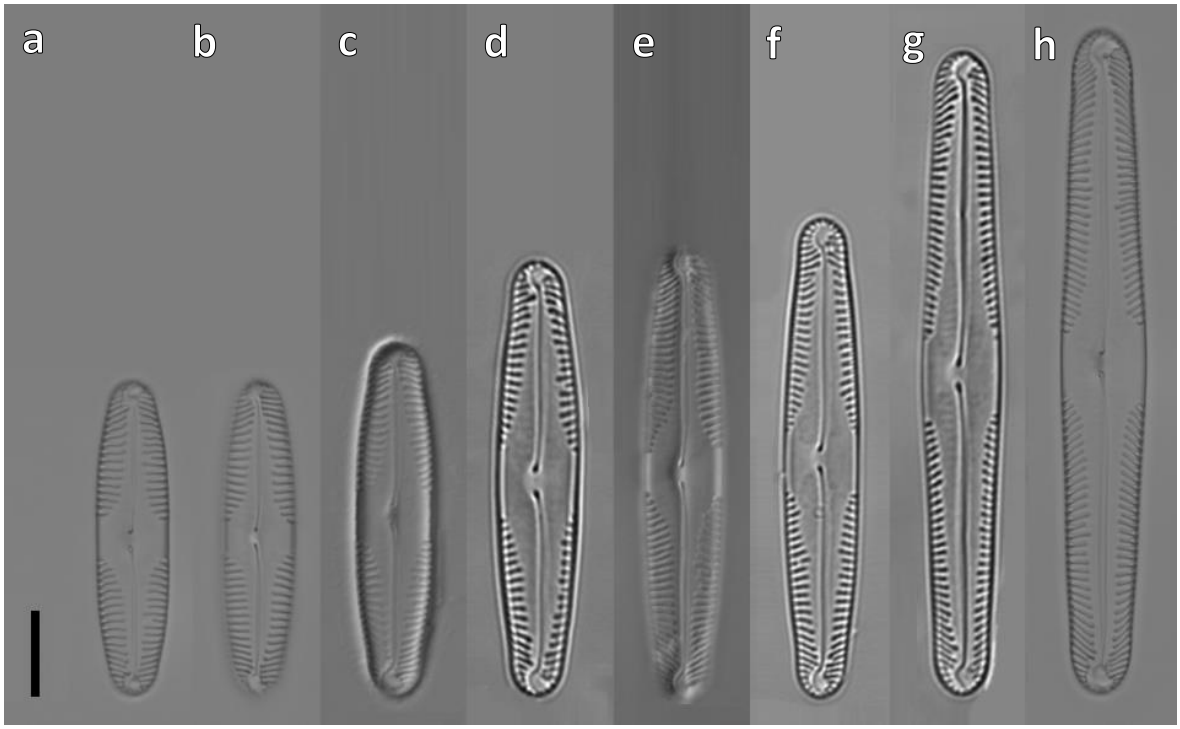

Figure 12. Supporting LM micrographs showing morphological variability of strains in clade D. Micrographs are scaled. Scale bar represents 10 μm. Order of the strains is as follows: (W045)d in (a), (W045)b in (b), REU12_18_4 in (c), REU12_18_6 in (d), REU12_18_17 in (e), REU12_18_11 in (f), REU12_18_21 in (g) and (W045)e in (h).

Table 3. Morphometric measurements of strains in clade D (length, breadth and stria density from 10 valves per strain ± standard deviation; n.m. = not measured).

Strain	Lenght (μm)	Breadth (μm)	No. of striae per 10 μm
(W045)d	n.m.	n.m.	n.m.
(W045)b	37.1 ± 0.8	8.7 ± 0.2	11.2 ± 0.6
REU12_18_4	41.2 ± 0.9	8.8 ± 0.4	12.0 ± 0.0
REU12_18_6	44.7 ± 0.8	8.7 ± 0.5	10.2 ± 0.4
REU12_18_17	53.4 ± 1.4	8.8 ± 0.4	10.6 ± 0.8
REU12_18_11	56.4 ± 0.5	9.0 ± 0.0	10.0 ± 0.0
REU12_18_21	81.0 ± 0.9	9.7 ± 0.7	9.8 ± 0.4
(W045)e	n.m.	n.m.	n.m.

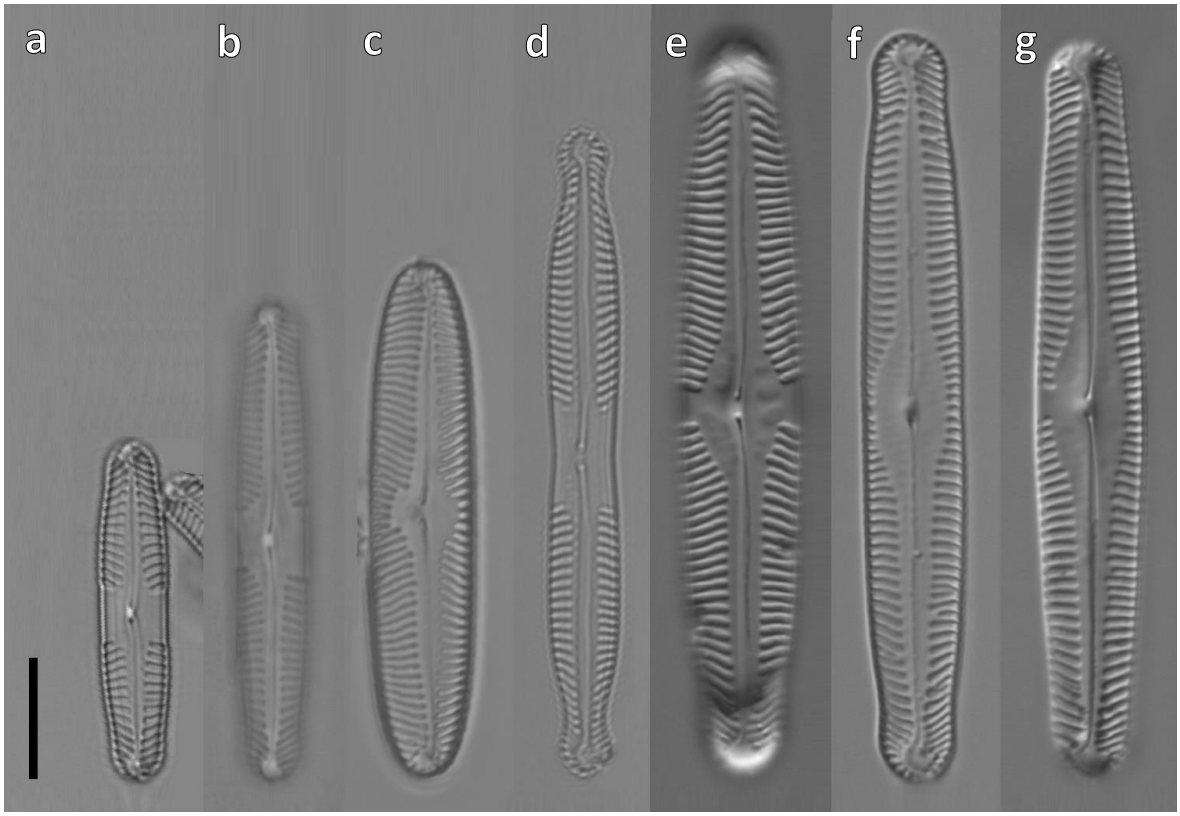

Figure 13. Supporting LM micrographs showing morphological variability of strains in clades E (e-g), F (a-b) and G (c-d). Micrographs are scaled. Scale bar represents 10 µm. Micrographs a-d are courtesy of Caroline Souffreau. Order of the strains is as follows: (Tor8)d in (a), (Tor8)b in (b), (Wie)b in (c), (Wie)c in (d), REU12_3_16p in (e), REU12_5_2 in (f) and REU12_5_1 in (g).

Table 4. Morphometric measurements of strains in clade E, F and G (length, breadth and stria density from 10 valves per strain ± standard deviation).

Strain	Lenght (µm)	Breadth (µm)	No. of striae per 10 µm
(Tor8)d	28.8 ± 1.2	5.6 ± 0.5	11.8 ± 0.4
(Tor8)b	40.0 ± 0.5	5.9 ± 0.3	11.8 ± 0.4
(Wie)b	61.5 ± 1.1	9.9 ± 0.3	11.7 ± 0.7
(Wie)c	53.3 ± 0.6	6.3 ± 0.3	11.9 ± 0.3
REU12_3_16p	61.5 ± 1.1	9.9 ± 0.3	11.7 ± 0.7
REU12_5_2	62.5 ± 0.8	8.8 ± 0.4	10.1 ± 0.4
REU12_5_1	61.7 ± 0.8	9.0 ± 0.2	10.8 ± 0.5

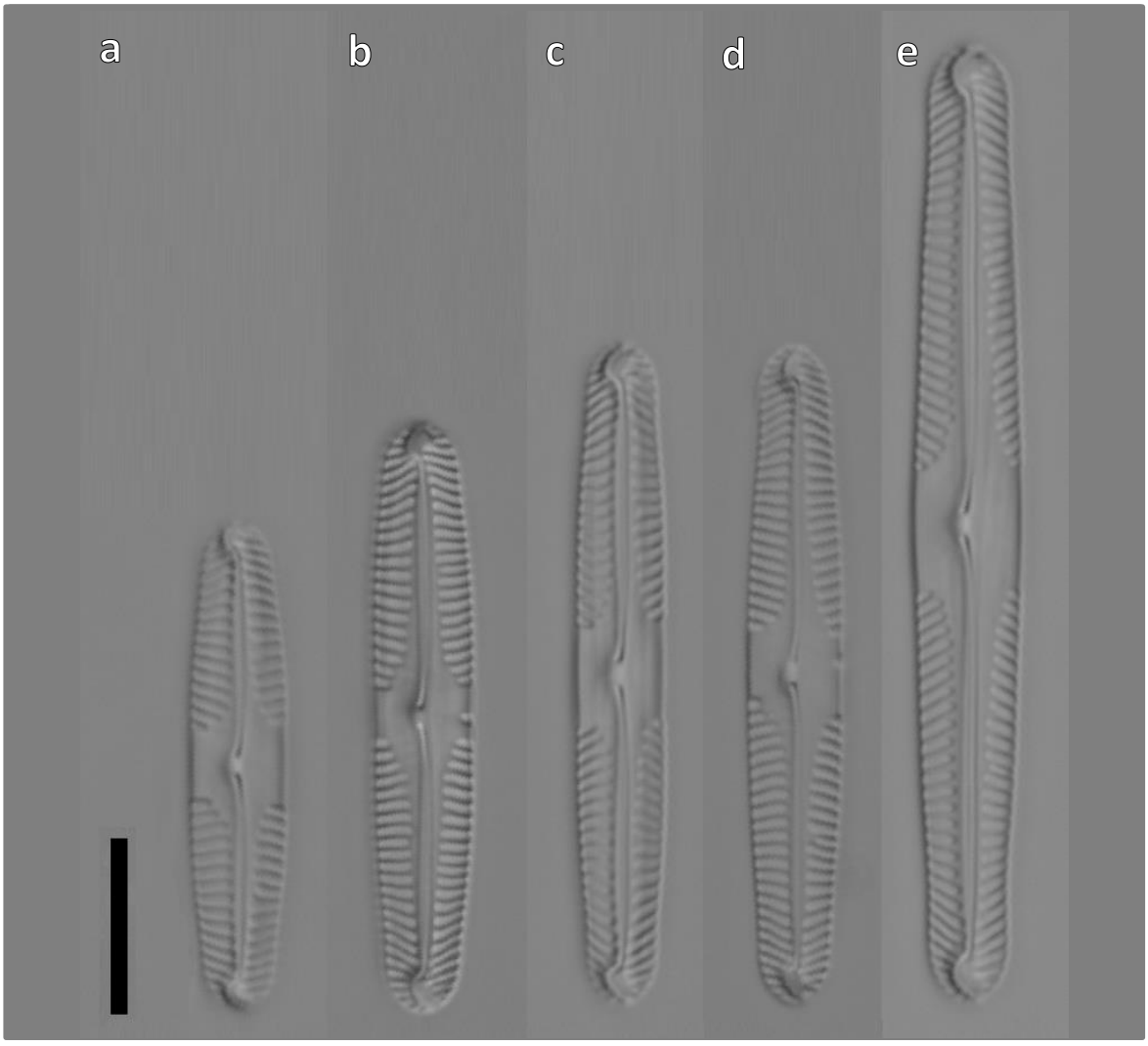

Figure 14. Supporting LM micrographs showing morphological variability of strains in clade H. Micrographs are scaled. Scale bar represents 10 µm. All micrographs are courtesy of Caroline Souffreau. Order of the strains is as follows: (W076)d in (a), (W067)c in (b), (W118)b in (c), (W095)b in (d) and (W076)e in (e). For strains (W095)a and (W123)a micrographs are not available.

Table 5. Morphometric measurements of strains in clade H (length, breadth and stria density from 10 valves per strain ± standard deviation; n.m. = not measured).

Strain	Lenght (µm)	Breadth (µm)	No. of striae per 10 µm
(W076)d	n.m.	n.m.	n.m.
(W067)c	n.m.	n.m.	n.m.
(W118)b	n.m.	n.m.	n.m.
(W095)b	30.6 ± 0.7	5.9 ± 0.2	13.0 ± 0.7
(W076)e	n.m.	n.m.	n.m.
(W095)a	32.8 ± 1.7	5.7 ± 0.5	13.8 ± 0.6
(W123)a	16.4 ± 0.5	6.0 ± 0.0	13.6 ± 0.8

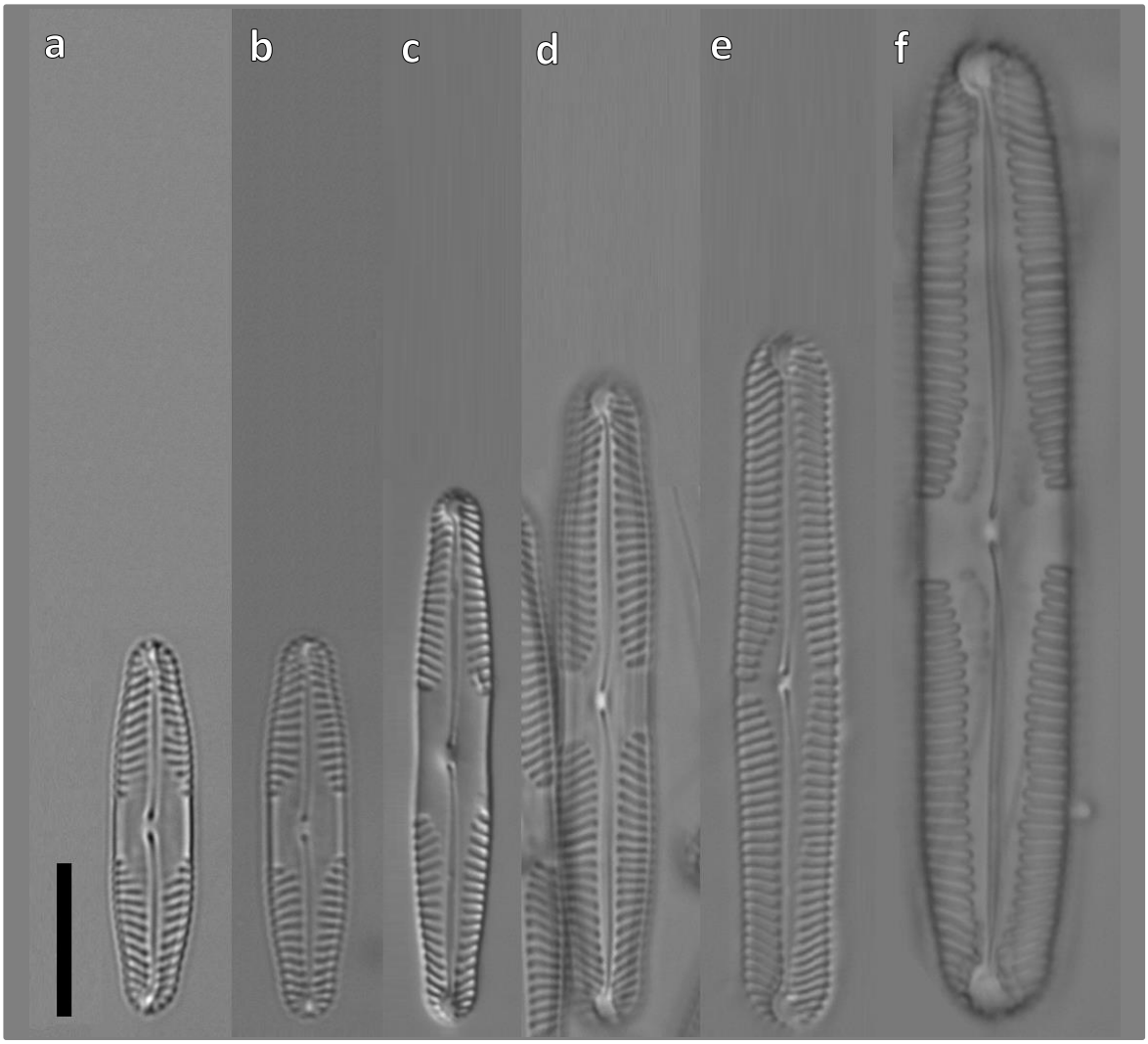

Figure 15. Supporting LM micrographs showing morphological variability of strains in clade I (a-c) and singleton strains (d-f). Micrographs are scaled. Scale bar represents 10 µm. Micrographs d and f are courtesy of Caroline Souffreau. Order of the strains is as follows: REU12_3_10 in (a), REU12_26_12 in (b), REU12_9_14 in (c), (Tor4)r in (d), CZECH_NOS2-7 in (e) and (Tor7)f in (f). For strain REU12_11_6 of clade I no micrographs are available.

Table 6. Morphometric measurements of strains in clade I and singleton strains (length, breadth and stria density from 10 valves per strain ± standard deviation; n.m. = not measured).

Strain	Lenght (µm)	Breadth (µm)	No. of striae per 10 µm
REU12_3_10	23.4 ± 1.6	5.2 ± 0.3	14.0 ± 0.0
REU12_26_12	n.m.	n.m.	n.m.
REU12_9_14	35.5 ± 0.5	5.1 ± 0.2	13.0 ± 1.1
(Tor4)r	42.6 ± 0.4	5.8 ± 0.3	12.4 ± 0.5
CZECH_NOS-7	47.1 ± 1.3	7.0 ± 0.0	11.0 ± 0.8
(Tor7)f	61.2 ± 1.6	8.9 ± 0.2	10.1 ± 0.5
REU12_11_6	23.4 ± 0.8	5.4 ± 0.5	13.2 ± 1.0

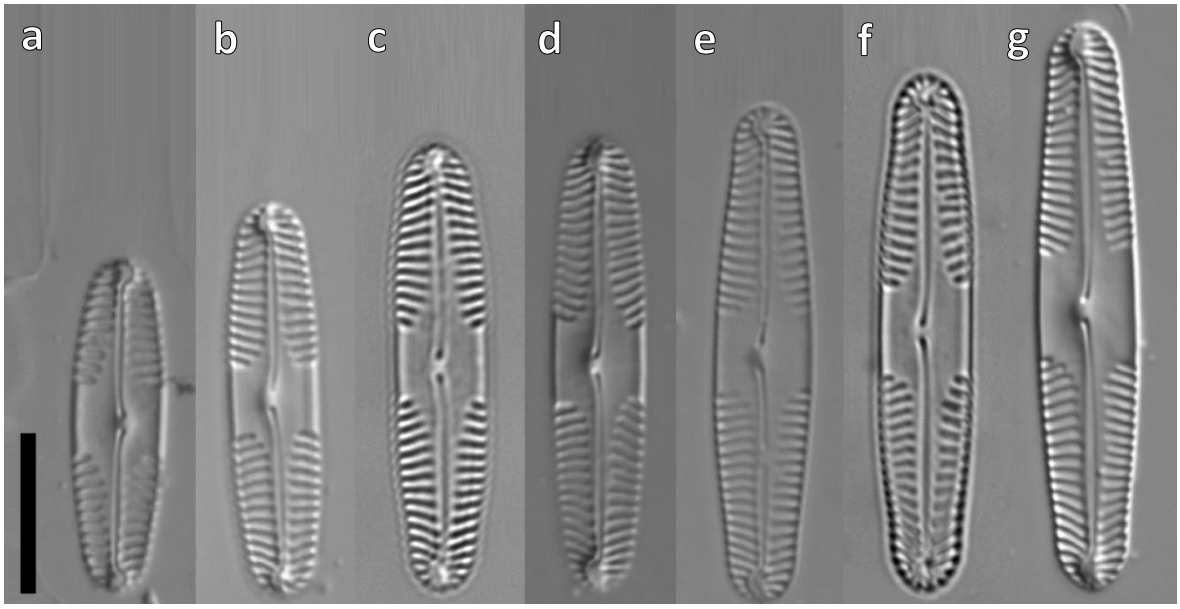

Figure 16. Supporting LM micrographs showing morphological variability of strains in clade J. Micrographs are scaled. Scale bar represents 10 µm. Order of the strains is as follows: MIC7_2 in (a), MIC5_16 in (b), MIC3_10 in (c), MIC13_21 in (d), CRO12_6_11 in (e), CRO12_9_16 in (f) and CRO12_24_4 (g). For strains CRO12_24_12, CRO12_24_10 and PinnC7 no micrographs are available.

Table 7. Morphometric measurements of strains in clade J (length, breadth and stria density from 10 valves per strain ± standard deviation).

Strain	Lenght (µm)	Breadth (µm)	No. of striae per 10 µm
MIC7_2	21.4 ± 0.5	5.9 ± 0.4	12.4 ± 0.8
MIC5_16	35.0 ± 0.4	5.9 ± 0.2	12.4 ± 0.8
MIC3_10	24.2 ± 0.8	5.9 ± 0.2	13.2 ± 1.1
MIC13_21	31.2 ± 0.8	6.0 ± 0.0	12.4 ± 1.3
CRO12_6_11	32.3 ± 0.5	6.0 ± 0.2	13.0 ± 1.1
CRO12_9_16	28.8 ± 0.6	6.0 ± 0.2	12.4 ± 0.8
CRO12_24_4	28.2 ± 0.4	6.1 ± 0.2	13.9 ± 0.3
CRO12_24_12	29.2 ± 0.9	5.9 ± 0.3	12.2 ± 0.8
CRO12_24_10	24.9 ± 2.5	6.2 ± 0.2	12.7 ± 1.2
PinnC7	11.6 ± 0.7	6.3 ± 0.5	12.8 ± 1.0

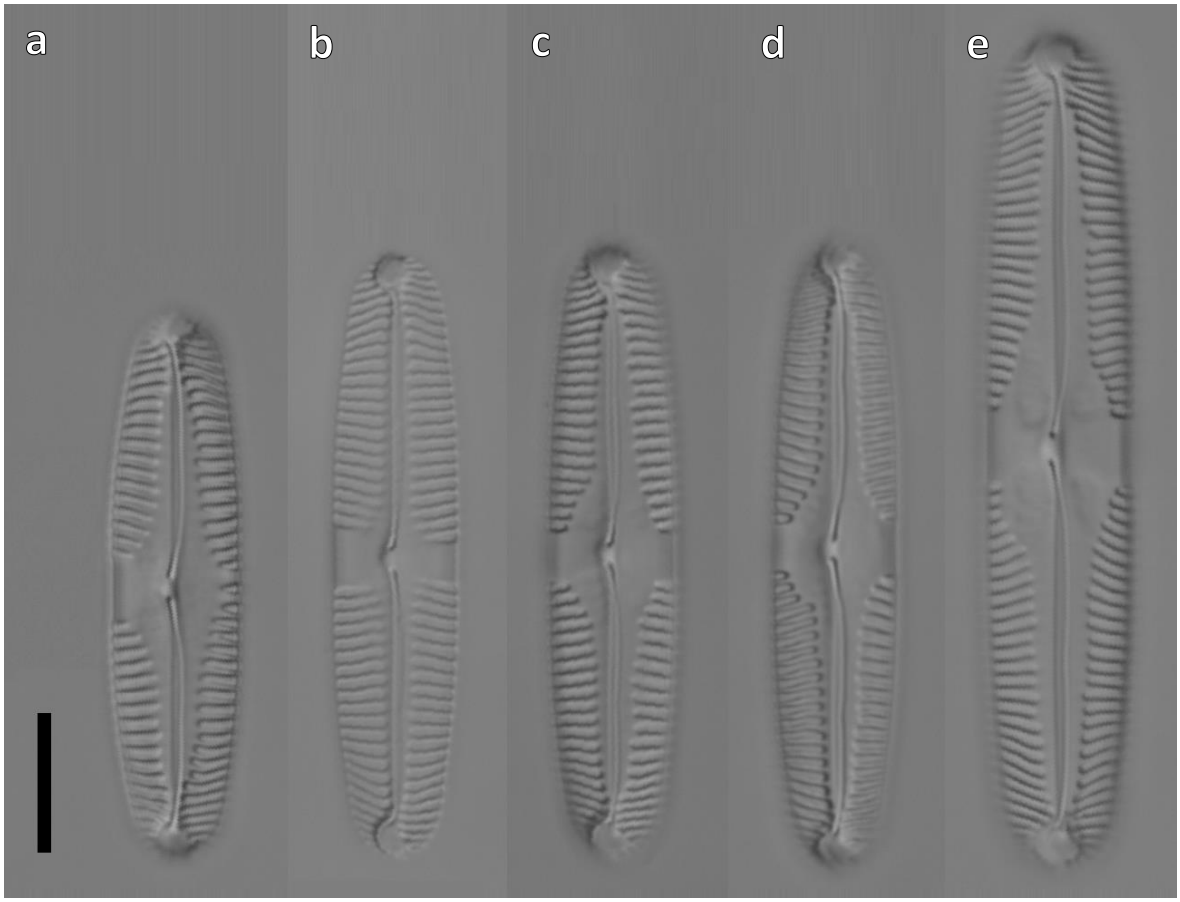

Figure 17. Supporting LM micrographs showing morphological variability of strains in clades K (a-d) and L (e). Micrographs are scaled. Scale bar represents 10 µm. All micrographs are courtesy of Caroline Souffreau. Order of the strains is as follows: Pin 592 M in (a), Pin 598 A in (b), Pin 19 in (c), Pin 7 in (d) and Pin 877 TM in (e). For strains Pin 12 and Pin 768 micrographs are not available.

Table 8. Morphometric measurements of strains in clades K and L (length, breadth and stria density from 10 valves per strain ± standard deviation; n.m. = not measured).

Strain	Lenght (µm)	Breadth (µm)	No. of striae per 10 µm
Pin 592 M	51.5 ± 0.9	8.8 ± 0.5	10.6 ± 0.5
Pin 598 A	59.9 ± 0.9	8.8 ± 0.4	10.5 ± 0.6
Pin 19	59.6 ± 1.5	9.0 ± 0.3	10.5 ± 0.5
Pin 7	61.4 ± 1.2	9.2 ± 0.3	10.9 ± 0.7
Pin 12	51.2 ± 1.1	9.0 ± 0.2	10.5 ± 0.7
Pin 877 TM	59.8 ± 2.3	10.1 ± 0.3	10.6 ± 0.5
Pin 768	n.m.	n.m.	n.m.

5 DISCUSSION

This study presents hypothesis about species limits within *Pinnularia subgibba* group based on genetic, morphological and, if available, other lines of evidence. This group has never been studied as a whole in such detail. I have found (using evolutionary distance analysis) that genes 28S rDNA and COX1 have the greatest power among five compared genes (18S rDNA, 28S rDNA, COX1, psbA and rbcL) to distinguish *Pinnularia* species. I have used these two genes to infer phylogeny (using ML and BI) and to delimit 15 or 16 phylogenetic species of *Pinnularia subgibba* group (using SPNA, GMYC and PTP). Finally, I have evaluated the congruence of the phylogenetic species delimitation results with morphology and other available lines of evidence. This evaluation is disscused below.

Five genes (18S rDNA, 28S rDNA, COX1, psbA and rbcL) and two separate regions of these genes (V4 18S rDNA and rbcL-3P) have been compared and nuclear 28S rDNA and mitochondrial COX1 have been identified as the most powerful in discrimination of *Pinnularia* species. Indeed, partial or complete 28S rDNA has been found to have relatively high discriminative power before (Trobajo et al. 2010; Hamsher et al. 2011) and it is being used in phylogeny inferrence of diatoms (e.g. Souffreau et al. 2011; Souffreau et al. 2013). However, it is recommended to use it for species level systematics along with rbcL (Trobajo et al. 2010; Hamsher et al. 2011). This recommendation, however, does not seem to be substantial for genus *Pinnularia*. According to the results of this study, 28S rDNA is the most powerful marker among the compared markers and, more importantly, rbcL appeared to be relatively conservative in *Pinnularia*. Therefore, the power of rbcL to distinguish closely related species of the genus, which is vital for both species delimitation and identification, is rather questionable. In addition, I have compared separate regions D1, D2 and D3 of 28S rDNA one with each other and found that D2 and D1 are the most variable. Therefore D1/D2 should be used prior to D2/D3 which is in contradiction with recommendations based on studies of different diatom genera (Mann et al. 2010; Hamsher et al. 2011). COX1 is generaly considered to be appropriate for species delimitation and identification and it is often used for barcoding of animals (e.g. Hebert, Ratnasingham & deWaard 2003). It also performed well in red algae (e.g. Le Gall & Saunders 2010) and even more closer relatives of diatoms such as brown algae (e.g. Kucera & Saunders 2008) and dictyochophytes (Riisberg & Edvardsen 2008). Its dicriminative power was tested on a model group of diatom genus *Sellaphora* (Evans et al. 2007) delimiting all putative species and surpassing other examined genes

(18S rDNA, ITS rDNA and rbcL). High discriminative power of COX1 was also confirmed in *Nitzschia palea* species complex (Trobajo et al. 2010). In some of the *Nitzschia* strains, however, COX1 failed to be amplified. This is in congruence with the result of this study in which COX1 was the second most powerful marker but it failed to be amplified in some strains. It seems reasonable to hypothesize that variability of COX1 in diatoms is so high that mutations occur in conventional primers' binding sites which prevents amplification. On the other hand, its discriminative power is so high that, in my opinion, it should be further investigated and new primers should be designed in order to use it for delimitation and identification of diatom species. To summarize, 28S rDNA and COX1 should be used for *Pinnularia* species delimitation and identification, although it would probably be necessary in the future to develop alternative primers for COX1 in order to cover its variability in both *Pinnularia* and diatoms in general.

Phylogenetic species of *Pinnularia subgibba* group were delimited independently using two most powerful genes (28S rDNA and COX1) and morphology of these putative species was examined. The congruence between genetic and morphological evidence is evaluated below. Moreover, all other available lines of evidence (e.g. biogeography, reproduction) were taken in consideration. Clade A was delimited as a separate species by all methods and by both genes. Moreover, it possess relatively distinct morphology characterized by linear valves with sides slightly concave in the center and convex in the ends. All three strains of this clade are originated in the lake Máchovo jezero, the Czech Republic. Considering morphological traits, I have identified this clade as *Pinnularia cf. microstauron*. However, morphological variability and relatively wide ranges of reference morphometric measures (i.e. L = 20-100 μm, B = 7-15 μm, S = 9-14/10 μm; Krammer 2000) indicates that *P. microstauron* is probably species complex rather than single species. This suggestion is supported by Krammer (2000) himself and Zidarova, Kopalová & Van de Vijver (2012) even decribed species of similar morphology from Livingston Island (Antarctica) as *P. australomicrostauron*. Moreover, my results confirm presumed polyphyletic character of this species complex. The confirmation is based on a fact that I have found this *P. microstauron* clade within *P. subgibba* group, while Souffreau et al. (2011) places *P. microstauron* into different group of *Pinnularia* as a sister lineage to *P. borealis*. On top of that, one strain identified as *P. microstauron* (CRO12_24_10) appeared in the clade J as delimited in this study. It is possible, and in my

opinion very probable, that other lineages of *P. microstauron* complex are spread across the tree of *Pinnularia*. Therefore, genetic variability of this species complex should be investigated. In terms of biogeography, it was shown that diatoms are regulated by the same processes that influence distribution of macroscopic organisms (Vanormelingen, Verleyen & Vyverman 2008). Thus, considering type locality of the taxon *Pinnularia microstauron* (Ehrenberg) Cleve 1891 (i.e. Rio de Janeiro, Brasil; Krammer 2000), it is very probable that clade A as delimited in this study represents species unknown to science. This hypothesis still remains to be tested. If it is correct, however, then taxon name *Pinnularia bohemica* would be appropriate given the finding locality in the Czech Republic.

Clade B was also delimited as a single species with perfect congruence across all methods and genes. Morphology of its two strains is relatively distinctive (especially in broad valves with broadly rounded ends) from other strains of *Pinnularia subgibba* group. However, there are differences among the two strains (especially the overal shape of the valve) which suggests either relatively high intraspecific morphological variability or interspecific genetic uniformity in 28S rDNA and COX1. Strain NUUK13_KAP17_3 have been identified as *Pinnularia cf. lokana*. I have searched reference literature and indeed I did not found any described species to be more similar to clade B (especially strain NUUK13_KAP17_3) than *Pinnularia lokana* Krammer 2000. There are distinct differences, however, of which the most notable is breadth of axial area. Moreover, both strains of the clade B are from Kapisillit, Greenland, while type locality of *P. lokana* is in Lokabad, Sweden. It is quite possible, that clade B represents another species unknown to science. In this case, however, I would recommend searching for additional strains and extensive investigation of its genetic and morphological variability. Ideally, reproductive compatibility of the strains should be tested as well before coming to conclusion on the identity of this species.

Another putative species which limits were congruent using all methods and genes belong to Clade C. Strains of this clade possess probably the most distinctive morphology among all examined clades of *P. subgibba* group. Most of them are relatively large (up to average length 128.6 ± 1.6 μm in strain STAP3) and valves have characteristic shape with sides slightly convex in the centre and slightly concave in the ends. Ends of the valve themselves, however, are usually slightly capitate. Seven of total eight strains were identified as *Pinnularia macilenta*. Three strains of this clade are from Belgium

(STAP strains), four from the Czech Republic (CZECH strains) and one from Scotland (Pin 596 A). The Scotish strain is an exception to morphological uniformity of the clade and it was identified only to genus level. It is very different in size (average length only 35.4 ± 0.8 μm, reference length for *P. macilenta* is 60-134 μm; Krammer 2000). However, its average breadth (B = 11.9 ± 0.5 μm) and stria density (S = 8.9 ± 0.5/10 μm) are more or less within the range of reference measures for *P. macilenta* (B = 12-15.5 μm, S = 8-10/10 μm; Krammer 2000). Also character of its central and axial areas is similar to those of other strains from the clade. So except for lenght it seems to be part of the *P. macilenta* clade. If it truly is *P. macilenta*, what is the cause of its extremely small lenght and loss of characteristic features of the valve then? The most probable explanation seems to be valve size reduction caused by asexual reproduction. It is known that diatoms reduce their size (especially length) and generally become more simple in morphology (e.g. Geitler 1932; Hustedt 1937). In *Stephanodiscus niagarae* the cells before sexual reproduction and consequent size restoration can be even 43% of average cell size in the population (Edlund & Stoermer 1991). Another explanation would be phenotypic plasticity which is documented to be present in diatoms (Jahn et al. 1986; Trobajo 2009). However, phenotypic plasticity is generally not considered to be serious problem for diatom systematics (e.g. Mann 1999). Unfortunatelly, I do not have access to the culture of the strain so I am not able to check the two possibilities mentioned above at the moment. There are also another possible explanation, of course, but there is no way I can check them either at the moment and thus disscusing them would be pure speculation. In my opinion, cell size reduction caused by asexual reproduction seems to be most probable. However, this should be checked before clade C can be unreservedly associated with taxon *P. macilenta* Ehrenberg 1843. Strains REU12_18_21 from clade D and (Tor8)d from clade F were identified by morphology as *Pinnularia cf. macilenta*. However, if examined in detail they are very different and definitely are not *P. macilenta*.

In comparison to clades A, B and C, other delimited clades of the *Pinnularia subgibba* group seem to be more uniform in morphology. Clades D, E, F, I, K, L and singleton strain CZECH_NOS2-7 seem to be parts of large polyphyletic gibba-subgibba-parvulissima supercomplex. In order to unweave and revise the taxonomy of this supercomplex, clades must be investigated in more detail than I am able to do in the range of this master thesis. However, it seems that at least some of these clades will be distinguishable by morphological features even though overlaps in lenght, breadth, stria density

and lenght-to-breadth ratio are quite often. For example, all strains of clade D possess the largest rhombic central area among all the clades of the group, which, along with relatively broad axial area widening from the ends to the centre of the valve, might serve as disriminative trait for this clade. Species delimitation method SPNA based on 28S rDNA split the clade into two separate haplotype networks. One unites the strains from Amsterdam Island, while the other unites the strains from Réunion. Both Islands can be found in the Indian Ocean. However, I did not found any difference in morphology between strains from Amsterdam Island and strains from Réunion. Moreover, other two methods based on 28S rDNA and all three methods based on COX1 delimited clade D as a single species. Therefore, I think it is more probable that clade D still represents a single species and SPNA based on 28S rDNA detected genetical variability caused by limited gene flow between the two geographicaly isolated populations of the species. If gene flow is or will be reduced completely, however, than we might be witnesses of the speciation in its early stage. At the moment, however, evidence speaks in favor of the hypothesis that clade D represents single species and that morphological features distinguishing this clade may be defined upon more detailed examination of the clade. On the other hand, strains of clade K are very similar one with each other but they does not seem to possess some distinctive feature which would distinguish them from other clades of the supercomplex. However, all of these strains of clade K have their origin in Scotland and it was even found that they interbreed one with each other (Poulíčková et al. 2007). Therefore, clade K fulfills even the biological definition of a species which suggests correctness of the hypothesis on the species limits within *Pinnularia subgibba* group delimited in this study. However, whole gibba-subgibba-parvulissima supercomplex have to be investigated in much more detail as mentioned above.

Clade H was delimited by all three methods based on 28S rDNA. COX1-based delimitation results, however, are not available because it failed to be amplified. The most probable reason of the failure is genetic variability in primer binding site. Morphology of this clade is similar to the morphology of gibba-subgibba-parvulissima supercomplex although it seems from the micrographs that they might be distinguishable by lenght-to-breadth ratio. Strains of the clade seem to be generally more narrow than gibba-subgibba-parvulissima strains. However, I can not test it by morphometrical measurements because I do not have voucher slides of most of the strains at disposal

for the moment. Most of the strains were identified as *Pinnularia cf. marchica* and all of them are from Amsterdam Island.

Clade J was delimited by all three methods based on both 28S rDNA and COX1. Morphology of the clade is characteristic by relatively small cell size (average lenght is between 11.6 ± 0.7 μm and 35.0 ± 0.4 μm). Overall morphology is very similar to the morphology of gibba-subgibba-parvulissima species supercomplex. All 10 strains are from sub-antarctic islands Crozet and Marion. Three strains were identified as *P. acidicola*, one as *P. microstauron* and the others were identified only to genus level. I think this clade can be associated with taxon *Pinnularia acidicola* B. van de Vijver & R. le Cohu 2002 which was described from Crozet. Its morphometric measures more or less fits into the range of reference morphometric measures of this taxon (L = 24-40 μm, B = 4.5-7 μm, S = 11-12/10 μm; Van de Vijver, Frenot & Beyens 2002) although some minor differences can be found (especially in stria density). Therefore, widening of the taxon description should be considered.

Evaluation of clade G is challenging. The clade was clearly delimited by all methods and both genes. However, two strains of this clade (Wie)b and (Wie)c are very different one from each other. They seems to be nearly as different as two strains of the group can be. It seems probable to me that some mistake was made in labeling slides, cultures or DNA samples and that morphology and DNA sequence of one of the stains does not belong to the same biological entity. This must be investigated before the congruence of genetic and morphological evidence can be properly discussed. Although the morphology of both strains is so different, both morphologies seem to be quite distinctive from morphologies of the other clades of the group. Therefore, after proper examination, clade G might represent another morphologically distinc species of *Pinnularia subgibba* group.

6 CONCLUSION

In this master thesis I have:

(1) found that nuclear gene 28S rDNA and mitochondrial gene COX1 have the greatest potential to distinguish between species of genus *Pinnularia*,
(2) used this two genes and three different methods to suggest primary hypothesis on species limits within *Pinnularia subgibba* group and
(3) evaluated congruence of this primary hypothesis with morphology and other available lines of evidence such as biogeography and reproductive compatibility.

Hereby, I evaluate the aims of this master thesis fulfilled. To summarize, results suggest that several clades of *Pinnularia subgibba* group represent morphologically distinct species which might be unknown to science (e.g. A, B). Others (e.g. C, J) can be relativelly clearly associated with decribed taxa. Most of the clades, however, seem to be part of large polyphyletic gibba-subgibba-parvulissima species supercomplex and it must be investigated in much more detail before coming to reliable conclusion on their true biological identity.

In conclusion, this master thesis gave some answers about systematics of *Pinnularia* which is one of the most species-rich genus of raphid diatoms. On the other hand, it arised vast number of new questions and it is obvious that tremendous amount of work remains to be done in the field. In the light of genetic evidence, it seems that after more than 200 years of diatom systematics we are still at the begining.

7 REFERENCES

Alverson A. J. & Theriot E. C. (2005): Comments on recent progress toward reconstructing the diatom phylogeny. – Journal of Nanoscience and Nanotechnology 5: 57-62.

Cavalier-Smith T. & Chao E. E. (2006): Phylogeny and Megasystematics of Phagotrophic Heterokonts (Kingdom Chromista). – Journal of Molecular Evolution 62: 388–420.

Clement M., Posada D. & Crandall K. A. (2000): TCS: a computer program to estimate gene genealogies. Molecular Ecology 9 (10): 1657-1660.

Chacon-Baca E., Beraldi-Campesi H., Cevallos-Ferriz S. R. S., Knoll A. H., Golubic S. (2002): 70 Ma nonmarine diatoms from northern Mexico. - Geology 30, 279-281.

Chen Y. (2012): The biomass and total lipid content and composition of twelve species of marine diatoms cultured under various environments. – Food Chemistry 131: 211-219.

Dinno A. (2012): paran: Horn's Test of Principal Components/Factors. R package version 1.5.1. – https://CRAN.R-project.org/package=paran.

Edgar L. A. & Picket-Heaps J. D. (1983): The Mechanism of Diatom Locomotion. 1. An Ultrastructural-Study of the Motility Apparatus. – Proceedings of the Royal Society of London. Series B-Biological Sciences 218: 331-343.

Edlund M. B. & Stoermer E. F. (1991): Sexual Reproduction in *Stephanodiscus niagarae* (Bacillariophyta). – Journal of Phycology 27: 780-793.

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

Ezard T., Fujisawa T. & Barraclough T. (2014): splits: Species’ LImits by Threshold Statistics. R package version 1.0-19/r51. – https://R-Forge.R-project.org/projects/splits/.

Geitler L. (1932): Der Fotmwechsel der pennaten Diatomeen (Kieselalgen). – Arch. Protistenk 78: 1-226.

Goldstein P. Z. & DeSalle R. (2011): Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. – BioEssays 33: 135-147.

Guiry M. D. & Guiry G. M. (2016): AlgaeBase. - World-wide electronic publication, National University of Ireland, Galway. http://www.algaebase.org; searched on 28 March 2016.

Hamsher S. E., Evans K. M., Mann D. G., Poulíčková A. & Saunders G. W. (2011): Barcoding Diatoms: Exploring Alternatives to COI-5P. – Protist 162: 405-422.

Hart M. W. & Sunday J. (2007): Things fall apart: biological species form unconnected parsimony networks. - Biol. Lett. 3: 509-512.

Hebert P. D., Ratnasingham S. & deWaard J. R. (2003): Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. – Proc. Biol. Sci. Lond. B (Suppl.) 270: S96-S99.

Hustedt F. (1937): Zur Systematik der Diatomeen. II. Der Begriff des "Typus" bei den Diatomeen und der Umfang der Diagnosen. III Konvergenzerscheinungen und Kummerformen. – Berichte der Deutschen botanischen Gesellschaft 55: 465-472.

Jahn R. (1986): A study of Gomphonema augur (Ehrenberg): the structure of the frustule and its variability in clones and populations. In: Ricard M. (ed.) Proceedings of the 8th International Diatom Symposium, Koeltz Scientific Books, Königstein: pp. 191-204.

Jahn R., Zetzsche H., Reinhardt R. & Gemeinholzer B. (2007): Diatoms and DNA barcoding: a pilot study on an environmental sample. In: Kusber W. H. & Jahn R. (eds), Proceedings of the 1st Central European Diatom Meeting 2007. Berlin, Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin: pp. 63-68.

Krammer K. (2000): Diatoms of Europe (Lange-Bertalot H., ed.) Volume 1: The Genus *Pinnularia*. – A. R. G. Gantner Verlag K. G., Ruggel, 703 pp.

Kucera H. & Saunders G. W. (2008): Assigning morphological variants of Fucus (Fucales, Phaeophyceae) in Canadian waters to recognized species using DNA barcoding. – Botany 86: 1065-1079.

Le Gall L. & Saunders G. W. (2010): DNA barcoding is powerful tool to uncover algal diversity: a case study of the Phyllophoraceae (Gigartinales, Rhodophyta) in the Canadian flora. - J. Phycol. 46: 374-389.

Levialdi-Ghiron J. H. (2006): Plastid phylogeny and chloroplast inheritance in the planktonic pennate diatom Pseudo-nitzschia (Bacillariophyceae). - Doctoral thesis, Universita Degli Studi Di Messina.

Lohman K. E. & Andrews G. W. (1968): Late Eocene Nonmarine Diatoms from the Beaver Divide Area, Fremont County, Wyoming. – United States Geological Survey Professional Paper 593-E.

MacDonald J. D. (1869): On the structure of the diatomaceous frustule, and its genetic cycle. – Annals and Magazine of Natural History Series 4(3): 1-8.

Mann D. G. & Droop S. J. M. (1996): Biodiversity, biogeography and conservation of diatoms. – Hydrobiologia 336: 19-132.

Mann D. G. & Marchant H. J. (1989): The Origins of the Diatom and Its Life-Cycle. - Systematics Association Special Volumes 38: 307-323.

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

Mann D. G., Sato S., Trobajo R., Vanormelingen P. & Souffreau C. (2010): DNA barcoding for species identification and discovery in diatoms. – Cryptogamie Algologie 31 (4): 557-577.

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

Mayr E. (1946): What is a species, and what is not? Philos Sci 63: 262-267.

Medlin L. K., Kooistra W. H. C. F., Gersonde R. & Wellbrock U. (1996): Evolution of the Diatoms (Bacillariophyta). II. Nuclear-Encoded Small-Subunit rRNA Sequence Comparisons Confirm a Paraphyletic Origin for the Centric Diatoms. – Molecular Biology and Evolution 13(1): 67-75.

Medlin L. K., Williams D. M. & Sims P. A. (1993): The evolution of the diatoms (Bacillariophyta). I. Origin of the group and assessment of the monophyly. – European Journal of Phycology 28: 261-275.

Nei M. & Kumar S. (2000): Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Pantocsek J. (1889): Beiträge zur Kenntniss der fossilen Bacillarien Ungarns. Teil II Brackwasser Bacillarien. Julíus Platzsko, Nagy-Tapolcsány, Hungary.

Paradis E., Claude J. & Strimmer K. (2004): APE: analyses of phylogenetics and evolution in R language. – Bioinformatics 20: 289-290.

Pfitzer E. (1869): Über den Bau und die Zellteilung der Diatomeen. – Botanische Zeitung 27: 774-776.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., & Vogler, A.P. (2006): Sequence-based species delimitation for the DNA taxonomy of undescribed insects. - Syst. Biol. 55: 595-609.

Poulíčková A., Mayama S., Chepurnov V. A. & Mann D. G. (2007): Heterothallic auxosporulation, incunabula and perizonium in *Pinnularia* (Bacillariophyceae). – Eur. J. Phycol 42(4): 367-390.

Puillandre N., Lambert A., Brouillet S. & Achay G. (2012): ABGD, Automatic Barcode Gap Discovery for primary species delimitation. – Molecular Ecology 21: 1864-1877.

R Core Team (2013): R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria. URL http://www.R-project.org/.

Riisberg I. & Edvardsen B. (2008): Genetic variation in bloom-forming ichthyotoxic *Pseudochattonella* species (Dictyochophyceae, Heterokonta) using nuclear, mitochondrial and plastid DNA sequence data. – Eur. J. Phycol. 43: 413-422.

Ronquist F., Teslenko M., van der Mark P., Ayres D. L., Darling A., Hohna S., Larget B., Liu L., Suchard M. A. & Huelsenbeck J. P. (2012): MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large ModelSpace. - Systematic Biology 61: 539-542.

Round F. E. (1988): A re-investigation of the diatom *Pinnularia cardinaliculus*. In Algae and the Aquatic Environment (Round F. E., ed), Biopress, Bristol, pp. 434-445.

Round F. E., Crawford R. M. & Mann D. G. (1990): The diatoms. Biology and morphology of the genera. - Cambridge University Press, Cambridge, 747 pp.

Saunders G. W. (2005): Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. - Phil. Trans. R. Soc. B 360: 1879–1888.

Scholin C. A., Herzog M., Sogin M. & Anderson D. M. (1994): Identification of group- and strain-specific genetic markers for globally distributed *Alexandrium* (Dinophyceae). II. Sequence analysis of a fragment of the LSU rDNA Gene. – J. Phycol. 30: 999-1011.

Scott F. J. & Thomas D. P. (2005): Diatoms. In: Antarctic marine protists (Scott F. J. & Marchant H. J., eds). Cranberra & Hobart: Australian Biological Resources Study, Australian Antarctic Divison, pp. 13-201.

Simonsen R. (1979): The diatom system: ideas on phylogeny. - Bacillaria 2: 9-71.

Sims P. A., Mann D. G. Medlin L. K. (2006): Evolution of the diatoms: insights from fossil, biological and molecular data. – Phycologia 45: 364-402.

Singh R. S., Stoermer E. F. & Kar R. (2006): Earliest freshwater diatoms from the deccan intertrappean (Maastrichtian) sediments of India. - Micropaleontology 52: 545-551.

Souffreau C., Verbruggen H., Wolfe A. P., Vanormelingen P., Siver P. A., Cox E. J., Mann D. G., Van de Vijver B., Sabbe K. & Vyverman W. (2011): A time-calibrated multi-gene phylogeny of the diatom genus *Pinnularia*. - Molecular Phylogenetics and Evolution 61: 866-879.

Souffreau C., Vanormelingen P., Van de Vijver B., Isheva T., Verleyen E., Sabbe K. & Vyverman W. (2013): Molecular evidence for distinct antarctic lineagues in the cosmopolitan terrestrial diatoms *Pinnularia borealis* and *Hantzschia amphioxys*. - Protist 164: 101-115.

Stenger-Kovács C., Buczkó K., Hajnal É. & Padisák J. (2007): Epiphytic, littoral diatoms as bioindicators of shallow lake trophic status: Trophic Diatom Index for Lakes (TDIL) developer in Hungary. – Hydrobiologia 589: 141-154.

Tamura K. & Nei M. (1993): Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. – Molecular Biology and Evolution 10: 512-526.

Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. (2013): MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. - Molecular Biology and Evolution 30: 2725-2729.

Theriot E. C., Cannone J. J., Gutell R. R. & Alverson A. J. (2009): The limits of nuclear-encoded SSU rDNA for resolving the diatom phylogeny. – European Journal of Phycology 44(3): 277-290.

Theriot E. C., Fritz S. C., Whitlock C. & Conley D. J. (2006): Late Quaternary rapid morphological evolution of an endemic diatom in Yellowstone Lake, Wyoming. – Paleobiology 32(1): 38-54.

Toporowska M., Pawlik-Skowrońska B. & Wojtal A. Z. (2008): Epiphytic algae on *Stratiotes aloides* L., *Potamogeton lucens* L., *Ceratophyllum demersum* L. and *Chara* spp. in a macrophyte-dominated lake. - International Journal of Oceanography and Hydrobiology 37: 51-63.

Trobajo R., Clavero E., Chepurnov V. A., Sabbe K., Mann D. G., Ishihara S. & Cox E. J. (2009): Morphological, genetic, and mating diversity within the widespread bioindicator *Nitzschia palea* (Bacillariophyta). – Phycologia 48: 443-459.

Trobajo R., Clavero E., Evans K. M., Vanormelingen P., McGregor R. C. & Mann D. G. (2010): The use of partial cox1, rbcL and LSU rDNA sequences for phylogenetics and species identification within Nitzschia palea complex (Bacillariophyceae). – European journal of phycology 45 (4): 413-425.

Van de Vijver B., Frenot Y. & Beyens L. (2002): Freshwater diatoms from Ile de la Possession (Crozet Archipelago, sub-Antarctaca). – Bibliotheca Diatomologica 46: 1-412.

Vanormelingen P., Verleyen E. & Vyverman W. (2008): The diversity and distribution of diatoms: from cosmopolitanism to narrow endemism. – Biodiversity Conservation 17: 393-405.

Venables W. N. & Ripley B. D. (2002): Modern Applied Statistics with S. - Fourth Edition, Springer, New York.

Werner D. (1971): The life cycle with sexual phase in marine diatom *Coscinodiscus asteromphalus*. 3. Differentiation and spermatogenesis. – Archiv für Mikrobiologie 80(2): 134-146.

Werner D. (1977): The Biology of Diatoms. – University of California Press, Berkeley, 498 pp.

Whitton B. A. & Rott E, eds. (1996): Use of algae for monitoring rivers II. – Proceedings of International Symposium, Innsbruck, Austria, 17-19 September, 1995, Institut für Botanik, University of Innsbruck; Innsbruck, Austria, pp. 196.

Zidarova R., Kopalová K. & Van de Vijver B. (2012): The genus *Pinnularia* (Bacillariophyta) excluding the section Distantes on Livingston Island (South Shetland Islands) with the description of twelve new taxa. – Phytotaxa 44: 11-37.

Zimmermann J., Jahn R. & Gemeinholzer B. (2011): Barcoding diatoms: evaluation of the V4 subregion on the 18S rDNA gene, including new primers and protocols. – Organisms Diversity and Evolution: published online.

Zwart G., Huismans R., van Agterveld M. P., Van de Per Y., De Rijk P., Eenhoorn H., Muyzer G., van Hannen E. J., Gons H. J. & Laanbroek H. J. (1998): Divergent members of the bacterial division *Verrucomicrobiales* in a temperate freshwater lake. – FEMS Microbiology Ecology 25: 159-169.

8 APPENDICES

Appendix A - List of tables

Appendix B - List of figures

Appendix C - List of abbreviations

18S rDNA	ribosomal gene encoding small subunit of eukaryotic cytoplasmic ribosomes
28S rDNA	ribosomal gene encoding large subunit of eukaryotic cytoplasmic ribosomes
AIC	Akaike information criterion
BCCM	Belgian Co-ordinated Collections of Micro-organisms
BI	Bayesian inference
BIC	Bayesian information criterion
BOLD	Barcode of Life Data Systems (www.boldsystems.org)
BS	bootstrap support

COX1	mitochondrial gene encoding Cytochrome c oxidase I protein
DCG	Diatoms Collection (part of BCCM)
DNA	deoxyribonucleic acid
GTR	General Time Reversible nucleotide substitutional model
HKY	Hasegawa-Kishino-Yano nucleotide substitutional model
ITS rDNA	ribosomal gene called Internal Transcribed Spacer
JC	Jukes-Cantor nucleotide substitutional model
K2	Kimura 2-parameter nucleotide substitutional model
lnL	maximum likelihood value
LSU	large subunit of cytoplasmic ribosomes
Ma	megaannus (10^6 years)
MCL	maximum composite likelihood
MCMC	Markov chain Monte Carlo
ML	maximum likelihood
PAE	Laboratory of Protistology and Aquatic Ecology, Gent University, Belgium
psbA	plastid gene encoding photosystem II protein D1
rbcL	plastid gene encoding large subunit of RuBisCO
RuBisCO	ribulose-1,5-bisphosphate carboxylase/oxygenase
SSU	small subunit of cytoplasmic ribosomes
TN93	Tamura-Nei nucleotide substitutional model
T92	Tamura 3-parameter nucleotide substitutional model
UPA	Universal Plastid Amplicon

Appendix D – Supplementary material

Supplementary Table I. 24 strains of subdataset A1 used for comparison of molecular markers.

Strain	Taxon
(B2)c	*Pinnularia cf. microstauron*
Pin 876	*Pinnularia acuminate*
(Tor11)b	*Pinnularia cf. altiplanensis*
Cal 878	*Pinnularia cf. isselana*
(Ecrins4)a	*Pinnularia cf. marchica*
Pin 889	*Pinnularia grunowii*
Pin 706	*Pinnularia neglectiformis*
(Tor1)a	*Pinnularia neomajor*
Corsea 2	*Pinnularia neomajor*
Pin 885	*Pinnularia nodosa*
Pin 877	*Pinnularia parvulissima*
(Tor4)i	*Pinnularia sp.*
(Tor4)r	*Pinnularia sp.*
Pin 873	*Pinnularia sp.*
PinnC7	*Pinnularia sp.*
(Tor7)f	*Pinnularia sp.* (gibba group)
(Tor8)b	*Pinnularia sp.* (gibba group)
Pin 649	*Pinnularia sp.* (subcommutata group)
Pin 883	*Pinnularia sp.* (subcommutata group)
Pin 650	*Pinnularia subanglica*
(Wie)c	*Pinnularia subcapitata var. elongata*
Corsea 10	*Pinnularia subcommutata var. nonfasciata*
(Enc2)a	*Pinnularia acrosphaeria*
Pin 870	*Pinnularia viridiformis*

Supplementary Table II. Characteristics of alignments used for distance analysis.

	Total No. of characters	No. of variable characters	No. of parsimony-informative variables
COX1	673	255	220
28S rDNA	619	401	340
psbA	909	115	69
rbcL	1542	293	187
18S rDNA	1816	1631	1027
V4 18S rDNA	399	129	99
rbcL-3P	748	173	121

Supplementary Table III. 27 strains of subdataset A2 used for comparison of D1, D2 and D3 regions of 28r DNA.

Strain	Taxon
MIC9_10	*Pinnularia cf. microstauron*
MIC3_13	*Pinnularia cf. silvatica*
MIC5_15	*Pinnularia acidicola*
CZECH_SITE3	*Pinnularia stomatophora var. irregularis*
CZECH_SWK9	*Pinnularia biceps*
CZECH_SWK13b	*Pinnularia brebissonii*
CZECH_SWK14a	*Pinnularia grunowii*
KATE_AZUL6	*Pinnularia australomicrostauron*
CZECHNOS2 12b	*Pinnularia viridiformis*
SPITS_1_2	*Pinnularia cf. frequentis*
SPITS_S5_2	*Pinnularia krammeri*
CZECH_OUCH15	*Pinnularia cf. rupestris*
CZECH_SWK2a	*Pinnularia microstauron*
CZECH_SW1_2a	*Pinnularia rhombarea*
CZECH_NOS2_8	*Pinnularia macilenta*
CRO12_18_12	*Pinnularia subcommutata*
CRO12_23_4	*Pinnularia aff. silvatica*
REU12_3_3	*Pinnularia cf. subcapitata*
VEGA_L35C_6	*Pinnularia rabenhorstii*
(Ban1)a	*Pinnularia borealis var. borealis*
VEGA_L2C_19	*Pinnularia borealis var. pseudolanceolata*
(Mo1)c	*Pinnularia borealis var. islandica*
VEGA_Mul40_8bis	*Pinnularia borealis var. unknown*
13Z43_5	*Pinnularia viridiformis*
REU12_18_13	*Pinnularia parvulissima*
13Z6_3bis	*Pinnularia lunata*
VEGA_L30C_2	*Pinnularia australomicrostauron*

Supplementary Table IV. 59 strains of subdataset B used in species delimitation. Strains which COX1 sequences failed to be amplified are indicated by *.

Strain	Taxon	Identification
MIC5_16	*Pinnularia acidicola*	B. Van de Vijver
MIC7_2	*Pinnularia sp.*	B. Van de Vijver
MIC3_10	*Pinnularia sp.*	B. Van de Vijver
MIC13_21	*Pinnularia sp.*	B. Van de Vijver
CZECH_NOS2-7	*Pinnularia sp.* (subgibba-group)	B. Van de Vijver
CZECH_NOS2-8	*Pinnularia macilenta*	B. Van de Vijver
CZECH_NOS2-14	*Pinnularia macilenta*	B. Van de Vijver
CZECH_SW2-4	*Pinnularia macilenta*	B. Van de Vijver
CZECH_SW2-5	*Pinnularia cf. microstauron*	J. Kollár
CZECH_SW2-7a	*Pinnularia cf. microstauron*	J. Kollár
CZECH_SW2-14	*Pinnularia cf. microstauron*	J. Kollár
CZECH_NOS2-5	*Pinnularia macilenta*	B. Van de Vijver
CRO12_6_11*	*Pinnularia acidicola*	B. Van de Vijver
CRO12_9_16	*Pinnularia sp.*	J. Kollár
CRO12_24_12*	*Pinnularia sp.*	J. Kollár
REU12_3_10*	*Pinnularia microstauron*	B. Van de Vijver
REU12_18_6	*Pinnularia sp.* (gibba-group)	B. Van de Vijver
REU12_18_11	*Pinnularia cf. subgibba*	B. Van de Vijver
REU12_18_21	*Pinnularia cf. macilenta*	B. Van de Vijver
CRO12_24_4	*Pinnularia acidicola*	B. Van de Vijver
REU12_3_16p	*Pinnularia sp.* (parvulissima-group)	B. Van de Vijver
REU12_5_1	*Pinnularia sp.* (gibba-group)	B. Van de Vijver
REU12_5_2	*Pinnularia sp.* (gibba-group)	B. Van de Vijver
REU12_9_14	*Pinnularia cf. subcapitata var. elongata*	B. Van de Vijver
REU12_11_6	*Pinnularia sp.* (parvulissima-group)	B. Van de Vijver
REU12_18_4	*Pinnularia sp.* (parvulissima/gibba-group)	B. Van de Vijver
REU12_18_17	*Pinnularia sp.* (parvulissima-group)	B. Van de Vijver
NUUK_13KAP16_3	*Pinnularia sp.*	B. Van de Vijver
NUUK_13KAP17_3	*Pinnularia cf. lokana*	B. Van de Vijver
REU12_26_12	*Pinnularia sp.* (parvulissima-group)	B. Van de Vijver
PinnC7	*Pinnularia sp.*	Souffreau et al. 2011
(W076)d*	*Pinnularia cf. marchica*	B. Van de Vijver?
(W123)a*	*Pinnularia cf. schoenfelderi*	B. Van de Vijver?
(W095)a*	*Pinnularia cf. marchica*	B. Van de Vijver?
(W118)b*	*Pinnularia cf. marchica*	B. Van de Vijver?
(W067)c*	*Pinnularia cf. marchica*	B. Van de Vijver?
(W095)b*	*Pinnularia cf. marchica*	B. Van de Vijver?
(W076)e*	*Pinnularia acidicola var. elongata*	B. Van de Vijver?
(Wie)b	*Pinnularia pseudogibba var. pseudogibba*	B. Van de Vijver?
(Tor8)d	*Pinnularia cf. marchica*	B. Van de Vijver?
PIN 596 A	*Pinnularia sp.*	Poulíčková, Mann
(Tor7)f	*Pinnularia sp.* (gibba-group)	Souffreau et al. 2011

Supplementary Table IV. – continued.

Strain	Taxon	Identification
(W045)e	*Pinnularia sp.* (parvulissima/gibba-group)	B. Van de Vijver?
(W045)d*	*Pinnularia cf. brebissonii*	B. Van de Vijver
(W045)b	*Pinnularia australogibba var. subcapitata*	Souffreau et al. 2011
PIN 768*	*Pinnularia sp.*	Poulíčková, Mann
Pin 877 TM	*Pinnularia parvulissima*	Souffreau et al. 2011
PIN 598 A	*Pinnularia cf. subgibba*	Poulíčková, Mann
PIN 19	*Pinnularia cf. subgibba*	Poulíčková, Mann
PIN 7	*Pinnularia cf. subgibba*	Poulíčková, Mann
PIN 592 M	*Pinnularia cf. subgibba*	Poulíčková, Mann
(Tor8)b	*Pinnularia sp.* (gibba-group)	Souffreau et al. 2011
(Wie)c	*Pinnularia subcapitata var. Elongata*	Souffreau et al. 2011
(Tor4)r	*Pinnularia sp.*	Souffreau et al. 2011
STAP3	*Pinnularia macilenta*	P. Vanormelingen
STAP4	*Pinnularia macilenta*	P. Vanormelingen
STAP5	*Pinnularia macilenta*	P. Vanormelingen
PIN 12	*Pinnularia cf. subgibba*	Poulíčková, Mann
CRO12_24_10	*Pinnularia microstauron*	B. Van de Vijver?

Supplementary Table V. Origin of 59 strains of subdataset B used in species delimitation.

Strain	Locality and date of colection	Collector
MIC5_16	Marion Island (Sub-Antarctic) - IV. 2011	W. Van Nieuwenhuyze and E. Verleyen
MIC7_2	Marion Island (Sub-Antarctic) - IV. 2011	W. Van Nieuwenhuyze and E. Verleyen
MIC3_10	Marion Island (Sub-Antarctic) - IV. 2011	W. Van Nieuwenhuyze and E. Verleyen
MIC13_21	Marion Island (Sub-Antarctic) - IV. 2011	W. Van Nieuwenhuyze and E. Verleyen
CZECH_NOS2-7	Lake Mácha (Czech Republic) - X.2011	J. Veselá, P. Urbánková and P. Vanormelingen
CZECH_NOS2-8	Lake Mácha (Czech Republic) - X.2011	J. Veselá, P. Urbánková and P. Vanormelingen
CZECH_NOS2-14	Lake Mácha (Czech Republic) - X.2011	J. Veselá, P. Urbánková and P. Vanormelingen
CZECH_SW2-4	Lake Mácha (Czech Republic) - X.2011	J. Veselá, P. Urbánková and P. Vanormelingen
CZECH_SW2-5	Lake Mácha (Czech Republic) - X.2011	J. Veselá, P. Urbánková and P. Vanormelingen
CZECH_SW2-7a	Lake Mácha (Czech Republic) - X.2011	J. Veselá, P. Urbánková and P. Vanormelingen
CZECH_SW2-14	Lake Mácha (Czech Republic) - X.2011	J. Veselá, P. Urbánková and P. Vanormelingen
CZECH_NOS2-5	Lake Mácha (Czech Republic) - X.2011	J. Veselá, P. Urbánková and P. Vanormelingen
CRO12_6_11	Iles Crozet (Sub-Antarctica) - XII.2012	B. Van de Vijver
CRO12_9_16	Iles Crozet (Sub-Antarctica) - XII.2012	B. Van de Vijver
CRO12_24_12	Iles Crozet (Sub-Antarctica) - XII.2012	B. Van de Vijver
CRO12_24_10	Iles Crozet (Sub-Antarctica) - XII.2012	B. Van de Vijver
CRO12_24_4	Iles Crozet (Sub-Antarctica) - XII.2012	B. Van de Vijver
REU12_3_10	Réunion (Sub-Antarctica) - XII.2012	B. Van de Vijver
REU12_18_6	Réunion (Sub-Antarctica) - XII.2012	B. Van de Vijver
REU12_18_11	Réunion (Sub-Antarctica) - XII.2012	B. Van de Vijver
REU12_18_21	Réunion (Sub-Antarctica) - XII.2012	B. Van de Vijver
REU12_3_16p	Réunion (Sub-Antarctica) - XII.2012	B. Van de Vijver

Supplementary Table V. – continued.

Strain	Locality and date of colection	Collector
REU12_5_1	Réunion (Sub-Antarctica) - XII.2012	B. Van de Vijver
REU12_5_2	Réunion (Sub-Antarctica) - XII.2012	B. Van de Vijver
REU12_9_14	Réunion (Sub-Antarctica) - XII.2012	B. Van de Vijver
REU12_11_6	Réunion (Sub-Antarctica) - XII.2012	B. Van de Vijver
REU12_18_4	Réunion (Sub-Antarctica) - XII.2012	B. Van de Vijver
REU12_18_17	Réunion (Sub-Antarctica) - XII.2012	B. Van de Vijver
REU12_26_12	Réunion (Sub-Antarctica) - XII.2012	B. Van de Vijver
NUUK_13KAP16_3	Kapisillit (Greenland) - VIII.2013	D. Obbels and K. Sabbe
NUUK_13KAP17_3	Kapisillit (Greenland) - VIII.2013	D. Obbels and K. Sabbe
PinnC7	Iles Crozet (Sub-Antarctica) - XII.2004	B. Van de Vijver
(W076)d	Amsterdam Island - XII.2007	B. Van de Vijver
(W123)a	Amsterdam Island - XII.2007	B. Van de Vijver
(W095)a	Amsterdam Island - XII.2007	B. Van de Vijver
(W118)b	Amsterdam Island - XII.2007	B. Van de Vijver
(W067)c	Amsterdam Island - XII.2007	B. Van de Vijver
(W095)b	Amsterdam Island - XII.2007	B. Van de Vijver
(W076)e	Amsterdam Island - XII.2007	B. Van de Vijver
(Wie)b	De Wieden (The Netherlands) - II.2007	P. Vanormelingen
(Tor8)d	Torres del Paine (Chile) - I.2007	C. Souffreau
PIN 596 A	Scotland - III.2006	A. Poulíčková and D. G. Mann
(Tor7)f	Torres del Paine (Chile) - I.2007	C. Souffreau
(W045)e	Amsterdam Island - XII.2007	B. Van de Vijver
(W045)d	Amsterdam Island - XII.2007	B. Van de Vijver
(W045)b	Amsterdam Island - XII.2007	B. Van de Vijver
PIN 768	Scotland - III.2006	A. Poulíčková and D. G. Mann
Pin 877 TM	Scotland - III.2006	A. Poulíčková and D. G. Mann
PIN 598 A	Scotland - III.2006	A. Poulíčková and D. G. Mann
PIN 19	Scotland - III.2006	A. Poulíčková and D. G. Mann
PIN 7	Scotland - III.2006	A. Poulíčková and D. G. Mann
PIN 592 M	Scotland - III.2006	A. Poulíčková and D. G. Mann
(Tor8)b	Torres del Paine (Chile) - I.2007	C. Souffreau
(Wie)c	De Wieden (The Netherlands) - II.2007	P. Vanormelingen
(Tor4)r	Torres del Paine (Chile) - I.2007	C. Souffreau
STAP3	Stappersven (Belgium) - 2014	P. Vanormelingen
STAP4	Stappersven (Belgium) - 2014	P. Vanormelingen
STAP5	Stappersven (Belgium) - 2014	P. Vanormelingen
PIN 12	Scotland - III.2006	A. Poulíčková and D. G. Mann

Supplementary Table VI. PCR primers details.

Marker	PCR primer sequence (5’ - 3’)	Reference
28S rDNA	D1R-F: ACCCGCTGAATTTAAGCATA	Scholin et al. 1994
	D2C-R: CCTTGGTCCGTGTTTCAAGA	Scholin et al. 1994
	T16N: AMAAGTACCRYGAGGGAAAG	Saunders unpublished
	T24U: SCWCTAATCATTCGCTTTACC	Hamsher & Saunders unpublished
COX1	GazF2: CAACCAYAAAGATATWGGTAC	Saunders 2005
	KEdtmR: AAACTTCWGGRTGACCAAAAA	Evans et al. 2007

Supplementary Table VII. Five best fitting nucleotide substitution models for two markers.

Marker	Model	Parameters	BIC	AICc	lnL
28S rDNA	GTR+G+I	133	9965.863	8803.702	-4268.465
	TN93+G+I	130	9985.386	8849.423	-4294.343
	K2+G+I	126	9994.399	8893.366	-4320.337
	HKY+G+I	129	9999.173	8871.943	-4306.609
	T92+G+I	127	10005.583	8895.818	-4320.557
COX1	TN93+G	105	9724.142	8842.003	-4315.664
	GTR+G	108	9726.844	8819.521	-4301.404
	TN93+G+I	106	9734.549	8844.016	-4315.664
	GTR+G+I	109	9737.251	8821.534	-4301.404
	HKY+G	104	9738.743	8865.000	-4328.169