Palacký University Olomouc

Faculty of Science

Department of Botany

Polyphasic approach to the study of diatom diversity

Doctoral Thesis

by

Mgr. Jan Kollár

Supervisor: prof. RNDr. Aloisie Poulíčková, CSc.

Olomouc 2020

"*The love for all living creatures is the most noble attribute of man.*"

- Charles Darwin -

AFFIRMATION

I hereby declare that this doctoral thesis has been written exclusively by myself and without using other resources than those listed in the "References" section. All published results included in this thesis have been approved by the co-authors.

In Olomouc on 30th June 2020 ……

ACKNOWLEDGEMENTS

Working on this research project was a long and exciting journey. Along the road, I had an honour of meeting, cooperating with and/or being advised by some of the brightest minds in the field. First and foremost, my eternal gratitude belongs to my supervisor, Dr Aloisie Poulíčková, who guided both my scientific and personal growth through the entirety of my university studies. Her unwavering patience, trust and dedication (to all her students) always represented a rare certainty in an uncertain sea of university life. I am most grateful to Dr Eveline Pinseel for her friendship, contagious enthusiasm, guidance and hard work on the co-authored manuscripts. I am grateful to Dr Petr Dvořák and Dr Petr Hašler for their overall support as well as for many hours of exciting scientific conversations. I am grateful to Dr Pieter Vanormelingen for guiding my first steps through the intricate world of the molecular phylogenetics. I thank Dr David M. Williams for his patient guidance through the worlds of cladistics and diatom systematics as well as for exciting scientific conversations. I thank Dr Wim Vyverman for overall support and cooperation. I thank Dr Markéta Fránková and Dr Petr Marvan for introducing me into the depths of diatom morphology and taxonomy. I thank Dr David G. Mann for cooperation as well as for his writings on diatoms which were one of the greatest sources of inspiration throughout my studies. My gratitude also belongs to all the members of the Department of Botany (UP), the research group of Protistology and Aquatic Ecology (UGENT) and the Division of Algae, Fungi and Plants of the Natural History Museum in London. I consider myself lucky for having an opportunity to work in such friendly, supportive and stimulating environments which all of them created. I am grateful to all the members of the UP Crowd whose friendship, unwavering enthusiasm and dedication to science created the perfect environment for the growth in the field of science popularization. I thank all the teachers who guided me through the various stages of my education. Last but not least, my eternal gratitude belongs to my family and especially to my parents, Jaroslava Veselá and Vladimír Veselý. To you, without whom I would never stand here, I dedicate this work.

ABSTRACT

Diatoms are one of the most species-rich groups of algae. As a group, they are cosmopolitan and adapted to almost all aquatic and some terrestrial habitats. Consequently, they are of great significance not only to fundamental research in the fields of biodiversity, evolutionary biology, ecology, and paleolimnology, but also for applied disciplines such as biomonitoring, biotechnology, and nanotechnology. For many of them, it is key to use correct taxonomic identifications which reflect an evolutionary history of the lineages in question. In this regard, one of the most serious challenges faced by contemporary diatomologists is the problem of cryptic and pseudocryptic species diversity. In the light shed by new technologies and approaches, many traditionally described diatom species appeared to be (pseudo)cryptic species complexes rather than single species. Several presumed species complexes are included in the genus *Pinnularia*, the *P. gibba* group among them. In this dissertation, I have gathered a data set of 105 *P. gibba* group strains with worldwide origin and used it to delimit species by means of a polyphasic approach. The primary (i.e., DNA-based) species delimitation based on two genetic markers and three automatized species delimitation methods revealed that the data set covered 15 species of the group. The secondary species delimitation rested in confrontation of this primary hypothesis with other available lines of evidence, namely with morphological, geographic, environmental and/or reproductive data. I conclude that the data set indeed covered 15 species many of which are (pseudo)cryptic with limited geographic distributions. An addition of more conserved genetic marker along with the fossil data allowed for an inference of time-calibrated phylogeny. The phylogeny further allowed for formulation and (in same cases) statistical testing of some evolutionary and biogeographic hypotheses. For example, significant difference in cell-size was detected in different parts of the tree and the difference is currently best explained by an evolutionary origin. Concerning the identification of the delimited species, proposed diatom DNA barcode markers in *rbcL* and SSU rDNA can unambiguously distinguish even between (pseudo)cryptic sister species of the group. Furthermore, the taxonomic review of the group was initiated by formal description of one of the delimited species as *P. lacustrigibba* sp. nov. Finally, the dissertation led to a development of a novel probabilistic model of speciation which, perhaps, may form a theoretical basis for future development of a new kind of probabilistic species delimitation methods applicable even to incipient species (many of which are presumably cryptic) and most of the asexual lineages.

KEY WORDS

biogeography, cryptic species diversity, LSU rDNA, morphology, *Pinnularia*, polyphasic systematics, *rbcL*, species, SSU rDNA, taxonomy

ABSTRAKT

Rozsivky jsou jednou z druhově nejbohatších skupin řas. Jako skupina jsou kosmopolitní a adaptovány na život téměř ve všech vodních a mnohých terestrických habitatech. Využití nacházejí v mnoha odvětvích základního (např. biodiverzita, evoluční biologie, ekologie, paleolimnologie) i aplikovaného výzkumu (nebo přímo jeho aplikacích; např. biomonitoring, biotechnologie, nanotechnologie). Pro mnohá z nich je klíčové pracovat se správným taxonomickým určením, které odráží evoluci daných evolučních linií. V této souvislosti je jedním z nejvážnějších problémů současných diatomologů tzv. kryptická a pseudokryptická druhová diverzita. V posledních desetiletích se ukazuje že mnohé tradičně popsané rozsivkové druhy jsou ve skutečnosti komplexy několika (pseudo)kryptických druhů. Několik z předpokládaných (pseudo)kryptických druhových komplexů obsahuje i rod *Pinnularia*. Jedním z nich je tzv. skupina *P. gibba*. V této disertační práci jsem shromáždil dataset 105 kmenů skupiny *P. gibba* a polyfázicky v něm druhy delimitoval. Primární (tj. na DNA založená) delimitace druhů s využitím dvou genetických markerů a tří auto-matizovaných molekulárně-delimitačních metod odhalila, že dataset pokryl 15 druhů skupiny *P. gibba*. Sekundární delimitace pak spočívala v konfrontaci této primární hypotézy s veškerými morfologickými, geografickými, environmentálními a reprodukčními daty dostupnými k daným kmenům. Finální hypotézou je, že dataset skutečně zahrnuje 15 druhů. Většina z nich je (pseudo)kryptická a vykazuje omezené geografické rozšíření. Po přidání dalšího genetického markeru a dat z fosilního záznamu jsem dále zkonstruoval časově kalibrovaný evoluční strom. Ten dále umožnil formulaci a (v některých případech i) statistické testování evolučních či biogografickcých hypotéz. Mimo jiné byl v různých částech stromu objeven signifikantní rozdíl v buněčné velikosti daných druhů, který je nejlépe vysvětlitelný evolučním původem. Bylo též zjištěno, že genetické markery navržené pro DNA barkóding rozsivek (*rbcL* a SSU rDNA) dokážou jednoznačně rozlišit i (pseudo)kryptické sesterské druhy studované skupiny. V této práci byla zahájena i taxonomická revize skupiny a to popisem jednoho z delimitovaných druhů jako *P. lacustrigibba* sp. nov. Tato disertační práce dále vedla k vývoji nového probabilistického modelu speciace, který by snad v budoucnu mohl posloužit jako teoretický základ k vývoji nových probabilistických delimitačních metod, které by teoreticky mohly být aplikovatelné nejen na mladé druhy (z nichž mnohé mohou být kryptické), ale dokonce i na většinu asexuálních linií.

KLÍČOVÁ SLOVA

biogeografie, *cox1*, druh, kryptická druhová diverzita, LSU rDNA, morfologie, *Pinnularia*, polyfazická systematika, *rbcL*, SSU rDNA, taxonomie

List of abbreviations

AB, AlgaeBase (www.algaebase.org)

BI, Bayesian inference

BIC, Bayesian information criterion

BoLD, Barcode of Life Database (www.boldsystems.org)

BS, bootstrap support

cox1, mitochondrial gene encoding cytochrome c oxidase subunit I

cox1-5P, 5' region of *cox1*

D1, D2 and D3 LSU rDNA, regions of LSU rDNA

DB, DiatomBase (www.diatombase.org)

EM, electron microscopy

GB, Genbank (www.ncbi.nlm.nih.gov)

LM, light microscopy

LSU rDNA, nuclear gene encoding large ribosomal subunit

MCMC, Markov chain Monte Carlo

ML, maximum likelihood

ORF, open reading frame

PP, posterior probability

psbA, plastid gene encoding photosystem II protein D1

PTP, Poissson tree processes

rbcL, plastid gene encoding RUBISCO

rbcL-3P, 3' region of *rbcL*

SEM, scanning electron microscopy

sGMYC, single threshold generalised mixed Yule-coalescent

SPNE, statistical parsimony network estimation

SSU rDNA, nuclear gene encoding small ribosomal subunit

UPCEL, the Universal Probabilistic Concept of Evolutionary Lineages

V4 SSU rDNA, region of SSU rDNA

List of papers

(the thesis is based on these four papers - see Appendices for full reprints/manuscripts)

Kollár J., Pinseel E., Vanormelingen P., Poulíčková A., Souffreau C., Dvořák P. & Vyverman W. 2019: A Polyphasic approach to the delimitation of diatom species: a case study for the genus *Pinnularia* (Bacillariophyta). *J. Phycol.* 55, 365–379. (IF_{2018} = 2,831; Q1)

Kollár J., Pinseel E., Vyverman W. & Poulíčková A.: A time-calibrated phylogeny provides an insight into the evolution, taxonomy and DNA barcoding of the *Pinnularia subgibba* group (Bacillariophyta). *Fottea*, (*accepted*). (IF_{2018} = 1,727; Q2)

Poulíčková A., **Kollár J.**, Hašler P., Dvořák P. & Mann D. G. 2018: A new species *Pinnularia lacustrigibba* sp. nov. within the *Pinnularia gibba* group (Bacillariophyceae). *Diatom Res.* 33, 273–282. (IF_{2018} = 1.169; Q3)

Kollár J., Poulíčková A. & Dvořák P.: On the relativity of species, or the probabilistic solution to the species problem. *Syst. Biol.*, (*submitted*). (IF_{2018} = 10.266; Q1)

CONTENTS

1. INTRODUCTION

1.1 Diatom species diversity

Diatoms (Bacillariophyta) are one of the most diverse groups of protists with an estimated number of extant species ranging between 30,000 and 200,000 (Andersen 1992, Mann and Vanormelingen 2013). The vastness of this number may be better apretiated in comparison to estimated species richness of other primary producers such as cyanobacteria (ca 6,000 spp.; Nabout et al. 2013), chrysophytes (ca 2,400 spp.; Andersen 1992), brown algae (ca 2,000 spp.; Andersen 1992), red algae (ca 5,500–20,000 spp.; Andersen 1992), green algae (ca 35,000–200,000 spp.; Andersen 1992, Adl et al. 2007), and streptophytes (over 300,000 extant spp.; Christenhusz and Byng 2016). Over 12,000 diatom species have been formally described (Guiry 2012) and 16 226 are marked as taxonomically accepted in the AlgaeBase (www.algaebase.org; Guiry and Guiry 2020). As a group, diatoms are cosmopolitan and adapted to almost all aquatic and some terrestrial habitats. Consequently, they are of great significance not only to fundamental research in the fields of biodiversity, evolutionary biology, ecology, and paleolimnology (e.g., Lewis et al. 2008), but also for applied disciplines such as biomonitoring (e.g., Smol and Stoermer 2010), biotechnology (e.g., Hu et al. 2008), and nanotechnology (e.g., Drum and Gordon 2003). However, an enormous diversity of diatoms leads not only to their significance but also to many difficulties with their systematics.

Traditionally, the descriptions of diatom species have been based on discontinuities in frustule morphology. Therefore, the majority of traditionally described and currently accepted diatom species may be considered morphospecies, i.e., groups of organisms distinguishable by morphology (Cronquist 1978), and their underlying biological identity remains questionable. After developement of the biological species concept (Mayr 1942, Mayr and Ashlock 1991), the differences in morphology were considered to reflect reproductive isolation (Mann 1999). However, even though this assumption may be correct for some species, it cannot be generalised as was proven by the widespread occurrence of cryptic and pseudocryptic species diversity in diatoms (e.g., Mann et al. 2004, Sarno et al. 2005, Evans et al. 2007, Vanormelingen et al. 2008a, Trobajo et al. 2009, Vanelslander et al. 2009, Poulíčková et al. 2010, Souffreau et al. 2013, Pinseel et al. 2017a, Pinseel et al. 2018, Amato et al. 2019, Pinseel et al. 2019). The cryptic and pseudocryptic species diversity means that species are often morphologically indistinguishable or distinguishable

only after employment of methods with higher resolution than light-microscopical observations alone (e.g., geometric morphometry, statistical analysis of morphological data, thorough examination in the electron microscope), respectively (Mann and Evans 2008). In addition, it was shown that (pseudo)cryptic species diversity in diatoms is widespread not only in taxonomical terms but also in terms of abundance. For example, an average of ca 65% of individuals in the studied epiphytic diatom communities (between 25% and 90% in dependance to the species of the macrophytic substrate) were members of presumed (pseudo)cryptic species complexes (Kollár et al. 2015). A disentanglement of diatom (pseudo)cryptic species complexes is a vital step towards their fine-grained taxonomy. In addition, correct taxonomical identifications reflecting the evolutionary history are vital for many fundamental and applied purposes. For example, the data from autecological studies of the (pseudo)cryptic species increase the resolution of the diatom-based bioindicative systems of water quality assessments such as the one implemented through the European Union's Water Framework Directive (e.g., Poulíčková et al. 2017).

Early in my university studies, I have developed an opinion that it is highly beneficial to study the selected subject broadly and to pay special attention to the integral fundamental phenomena and principles upon which it stands. In a study of species diversity of any group of organisms, only few questions may be considered as fundamental as 'What is a species?' Indeed, the identity of species is one of the most controversial topics of post-Darwinian biology (e.g., Dobzhansky 1937, Mayden 1997, Mann 1999, Wheeler and Meier 2000, De Queiroz 2007, John S. Wilkins 2011) and, therefore, I believe that at least relatively short introduction to this problematics will provide the reader with insight necessary for deeper understanding and, hopefully, appretiation of this thesis.

1.2 The species problem and a polyphasic approach

For centuries, 'species' has been one of the most widely used concepts in biology. At the same time, however, it is one of the most controversial concepts which potentially introduces confusion into evolutionary biology (including taxonomy), ecology and conservation biology. This so-called species problem is twofold. The problem of species category is that of species definition, while the problem of species taxon is that of species delimitation (e.g., Mayr 1982). Both are frequently interchanged and at least 27 more or less different species concepts were proposed (Wilkins 2011). In the studies focused on/working with species diversity, several issues related to the species problem

may arise. For example, researchers are rarely explicit about applied species concept and, therefore, it must be often inferred indirectly based on an evidence and approach used to formulate presented species hypotheses. Furthermore, even in cases when the same species concept is applied, the respective criterion may be studied with different methods and thus the species delimitations may result in recognition of different species boundaries. Further confusion is introduced by researchers who do not distinguish clearly between species category and species taxon thus confusing the problem of species definition with that of species delimitation. An exhaustive discussion on different species concepts may be found elsewhere (e.g., Mayr 1982, Wheeler and Meier 2000, Wilkins 2009). Here, I will briefly summarize only few widely used categories of species concepts which were applied in diatoms - phenetic, reproductive, evolutionary and phylogenetic.

The phenetic species concepts (e.g., Cronquist 1978) use discontinuities in morphological/anatomical variation to delimit a species. Thus, it is a species concept focused more on the delimitation of species taxa than the definition of species category. The criterion of morphological/anatomical similarity has been widely used especially before introduction of molecular techniques into the taxonomical practice. However, molecular data provided plenty of evidence that measurable morphological/anatomical diversity is often much lower than the diversity in nucleotide and amino acid sequences (e.g., Bickford et al. 2007). Extreme (but hardly rare; see above) cases of this fact led to a proposal of cryptic and pseudocryptic species. It is clear that (pseudo)cryptic species diversity introduces an important consideration for taxonomists, evolutionists and ecologists, especially those focused on the groups of organisms where measurable morphological/anatomical diversity is relatively low (e.g., microorganisms including diatoms; Mann 1999, Mann and Evans 2008).

The species concepts based on reproductive isolation were developed and advocated by many naturalists, the most prominent being Mayr (1942, 1963, 1982). Even though he had adjusted his famous ‘Biological Species Concept’ (BSC) over the years, its core remained unchanged - *species are groups of interbreeding populations which are reproductively isolated from other such groups*. For the sake of species delimitation, reproductive compatibility is operational criterion. In practice, it is often inferred indirectly based on other evidence at hand (e.g., morphology, ecology, behaviour; Mayr 1982). Nevertheless, for good reasons, the reproductive species concept has found many supporters. However, the direct testing of the reproductive compatibility in some groups

of sexual organisms is an intricate task. On the other hand, laborious crossing experiments allowed for an application of reproductive species concept even in several groups of eukaryotic sexual microorganisms including diatoms such as *Sellaphora* (Behnke et al. 2004, Mann et al. 2004), *Pseudo-nitzschia* (Quijano-Scheggia et al. 2009) and *Pinnularia* (Poulíčková et al. 2007). One of the most serious obstacles to the universal application of BSC is its definitional exclusion of asexual lineages. Mayr (1982, p. 273) tried to resolve the issue by adjusting the concept into "*a species is a reproductive community of populations (reproductively isolated from others) that occupies a specific niche in nature*" thus effectively merging it with those from the category of ecological species concepts (Van Valen 1976, Andersson 1990). Nevertheless, he always considered reproductive isolation primal for the definition of the species category (Mayr 1982).

One of the most important and often neglected properties of BSC is its 'nondimensionality'. By that Mayr (1982) meant that his concept is truly applicable exclusively if time dimension is excluded from considerations, i.e., on species distributed over space not through time (in my opinion, the term 'horizontal' is more appropriate adjective than Mayr's 'nondimensional' because the species as a natural unit always exists in some physical or other dimensions; see 3.4 On the relativity of species). The time dimension is incorporated into the evolutionary species concepts (Simpson 1951, 1961, Wiley 1978, Wiley and Mayden 2000). Let the definition of Wiley and Mayden (2000) serve as an example: "*An evolutionary species is an entity composed of organisms that maintains its identity from other such entities through time and over space and that has its own independent evolutionary fate and historical tendencies*". Obviously, the evolutionary species concepts are focused more on the problem of definition of species category than delimitation of species taxa. However, none of them specify (1) what exactly makes the lineages evolutionarily independent, (2) where in the time dimension is the borderline between the two subsequent species and (3) how species lineages differ from other lineages which reflect metapopulation structure rather than separate species.

Various versions of the phylogenetic species concepts (e.g., Hennig 1966, Nelson and Platnick 1981, Wheeler and Meier 2000) are based on the phylogeny of the groups of organisms in question. They are more focused on the delimitation of species taxa than on the definition of species category with monophyly and diagnosability being the key operational criteria. The criterion of diagnosability is altogether arbitrary (natural entities independent of human mind cannot be defined by human's ability to diagnose them), while

large parts of the genome violates the criterion of monophyly. Moreover, different phylogenetic methods/genetic markers often provide conflicting phylogenetic relationships (e.g., Rosenberg and Nordborg 2002, Leliaert et al. 2014). Therefore, it is unclear which of them reflect the species reality and thus should be used for reconstructing phylogeny. In any case, it is increasingly clear that units of genetic markers commonly used for molecular phylogenetic inferences are but a tip of an iceberg of actual genome diversity. Furthermore, ambiguous phylogenetic relationships were recently found in prokaryotes where horizontal gene transfer was pronounced to be important and frequent especially in closely related lineages (e.g., Dvořák et al. 2015). The same patterns were observed in eukaryotic genomes, although responsible processes differ between prokaryotes and eukaryotes (e.g., Martin et al. 2017).

De Queiroz (1998, 2007) synthesized proposed species concepts into the 'Unified Species Concept' (USC). Its major contribution was clear separation of the theoretical concept of what species is (i.e., the definition of species category) and the operational criteria used to delimit species taxa (e.g., morphological discontinuity, genetic discontinuity, reproductive discontinuity). The theoretical component of the unified species concept belongs to the category of evolutionary species concepts – species are segments of separately evolving metapopulation lineages. On the other hand, operational criteria emerge in time as a result of the evolutionary separation. The problem is that in different pairs of lineages different operational criteria emerge in different times, order or may not emerge at all (in case of morphological discontinuity leading to a cryptic species). Therefore, it should be borne in mind that these operational criteria may be considered just imperfect reflections of the underlying species reality which can merely help us formulate hypotheses on the numbers and limits of lineages, i.e., on their gene flow interconnection state. Since we cannot efficiently (i.e., in sufficient detail under reasonable time, effort, and financial expenses) study the gene flow directly in diatoms, I believe we should, in the spirit of science, at least consider as many operational criteria as possible before formulating a hypothesis on the species limits. This approach to systematics is sometimes called polyphasic (Colwell 1970, Vandamme et al. 1996, Komárek 2016). It spread from bacteriologists (e.g., Colwell 1970) and cyanobacteriologists (e.g., Komárek 2016) to protistologists (e.g., Škaloud 2008). In groups of macroorganisms this practice is called an integrative taxonomy (e.g., Will et al. 2005, Puillandre et al. 2012).

1.3 Genus *Pinnularia* and the *P. gibba* group

Genus *Pinnularia* Ehrenberg was selected as a focus of this research. Two of the main reasons follow. Firstly, it is among the most taxon-rich genera of raphid diatoms, with 723 currently accepted names in AlgaeBase (Guiry and Guiry 2018) and 1527 species names listed in DiatomBase (Kociolek et al. 2018), although, the taxonomic or nomenclatural status of the vast majority of the latter taxa is unassessed. In addition, the genus *Caloneis* Cleve was shown to be polyphyletic within *Pinnularia* (Bruder et al. 2008, Souffreau et al. 2011) raising the total number of accepted taxa in the *Pinnularia–Caloneis* complex (i.e., *Pinnularia* sensu lato) to 931 according to AlgaeBase and potentially making it, at the current state of knowledge, the second most taxon-rich diatom genus (outnumbered only by *Navicula* Bory with 1270 taxa, and followed by *Nitzschia* Hassall with 788 taxa; Guiry and Guiry 2018). And secondly, it contains several presumably (pseudo)cryptic species complexes. In recent years many new *Pinnularia* species have been described using the traditional morphological approach (e.g., Van de Vijver et al. 2011, Luís et al. 2012, Pereira et al. 2014) or using both molecular and morphological data (e.g., Pinseel et al. 2016). However, molecular studies suggest that many traditional *Pinnularia* morphospecies are species complexes with cryptic or pseudocryptic diversity (e.g., Souffreau et al. 2011, Pinseel et al. 2016).

The genus is cosmopolitan, occurring in a broad range of both terrestrial (e.g., Pinseel et al. 2020) and freshwater environments within assemblages such as the epipelon (e.g., Poulíčková et al. 2014), epibryon (e.g., Poulíčková et al. 2013) and epiphyton (e.g., Kollár et al. 2015, Poulíčková et al. 2017). Although their ecological demands vary, most species prefer low electrolyte content (Round et al. 1990, Krammer 2000). The genus is estimated to have originated in Paleocene some 60±15 Ma ago (Souffreau et al. 2011). To date, two DNA-based intrageneric phylogenetic hypotheses were proposed using concatenated alignments of three (SSU rDNA, LSU rDNA, and *rbcL*; Bruder et al. 2008) and five genetic markers (*cox1*, SSU rDNA, LSU rDNA, *rbcL* and *psbA*; Souffreau et al. 2011). Both revealed three major clades to be present within the genus. Clade B sensu Souffreau et al. (2011) consists of three distinct subclades called “grunowii”, “nodosa”, and “subgibba”. Because the whole genus would be too challenging for a single doctoral research project, subclade “subgibba” was selected as a model. Henceforth, it will be referred to as the “*Pinnularia subgibba* group” or “*Pinnularia gibba* group”.

The *P. gibba* group is estimated to originate in the upper Eocene around 35±10 Ma ago (Souffreau et al. 2011). Both the entire clade B and the *P. gibba* group are very well statistically supported (ML bootstrap proportions-BI posterior probabilities = 100-100) and can be distinguished by a certain combination of morphological features (Souffreau et al. 2011). Specimens of clade B are characterised by small linear cells with a drop-like central raphe endings and large alveolar openings. In addition, specimens of the *Pinnularia gibba* group are usually elongated in shape with two girdle-appressed parallel plastids and a broad non-porous central area called the fascia (Souffreau et al. 2011). Several species of the *Pinnularia gibba* group (e.g., *Pinnularia gibba* Ehrenberg, *Pinnularia parvulissima* Krammer, *Pinnularia subgibba* Krammer) are very uniform in morphology and are generally considered to be cosmopolitan (Krammer 2000). However, studies focusing on molecular species delimitations in the *P. gibba* group are entirely lacking. In the light shed by recent discoveries in the field of diatom biogeography, namely that many diatom species were shown to be geographicaly restricted or even endemic (e.g., Vyverman et al. 2007, Vanormelingen et al. 2008b, Verleyen et al. 2009, Pinseel et al. 2017b), this may suggest an underestimation of the species diversity within the group.

2. AIMS

The ultimate aim of this doctoral research project was investigating species diversity and evolution of the selected (pseudo)cryptic species complex of the genus *Pinnularia* (i.e., the *P. gibba* group). Firstly, species limits were explored by means of a polyphasic approach. This means that both primary (by using multiple genetic markers and automated species delimitation methods) and secondary (by considering all available morphological, environmental, geographic and reproductive data associated with the studied strains) species delimitations were performed. Secondly, the evolution of the group was investigated by inferring its time-calibrated multi-gene phylogeny. Finally, newly generated knowledge on species limits and phylogentic relationships was used to investigate whether DNA barcode markers proposed for routine identification of diatoms can unambiguously distinguish even between closely related (pseudo)cryptic species of the group.

To summarize, main aims of this doctoral research project were

(1) exploring species limits within the *Pinnularia gibba* group,
(2) inferring its time-calibrated multi-gene phylogeny and
(3) testing the taxonomic resolution of the proposed diatom DNA barcode markers on the species of the group.

In addition, newly generated knowledge along with the available morphological, environmental, geographic and reproductive data allowed for

(4) formulating and testing some evolutionary hypotheses and
(5) initiating the taxonomic review of the group.

Furthermore, an unexpected result of this work was a development of a novel view of the species (see 3.4 On the relativity of species).

3. RESULTS AND DISCUSSION

3.1 On the diversity of the *P. gibba* group

(this chapter is based on Kollár et al. 2019 - reprinted as Appendix 1)

The species limits within the *Pinnularia gibba* group were investigated using a data set of 105 strains with world-wide origin (see Figure 5 in Kollár et al. 2019; Appendix 1) by means of a polyphasic approach (Kollár et al. 2019). Firstly, a set of nuclear, chloroplast and mitochondrial markers was examined for their sequence variability. Secondly, two most variable markers (LSU rDNA and *cox1*) and three automated molecular species delimitation methods were used to formulate the primary (i.e., DNA-based) hypothesis on species limits within the group (see Figure 3 in Kollár et al. 2019). The most variable markers were used because they have the highest chance of covering intraspecific variability required by applied coalescence-based species delimitation methods. Finally, DNA-based species limits were confronted with other available lines of evidence such as morphology studied under both LM and SEM (e.g., Figure 4 in Kollár et al. 2019), geographic, environmental and/or reproductive data (all available evidence is summarized in Table S5 of Kollár et al. 2019).

Automated molecular species delimitations based on two most variable markers suggested the presence of 15 species in our data set and showed complete congruence between *cox1* and LSU rDNA (Figure 3 in Kollár et al. 2019). We then confronted our primary hypothesis with other lines of evidence. Whereas it is relatively easy to distinguish some of the potential species (clades A, B, C) based on morphological characteristics, it is not straightforward to distinguish others (clades D, E, F, G, H, I, J, and K, and singleton strains), especially the small specimens. It seems that different morphologies of large cells of different clades tend to converge to a very similar morphology when they become smaller and, therefore, may no longer be unambiguously identified (e.g., Fig.4g-i in Kollár et al. 2019). Nevertheless, the relative size of the fascia seems promising for the recognition of both higher natural species clusters (e.g., JK and GHI) and possibly even some species within these clusters (e.g., H) because it seems to be more stable for both small and large specimens compared to valve outline or length-to-width ratio. We found no clear utility of the indistinct markings accompanying the fascia (e.g., Fig 4j in Kollár et al. 2019) for species level identifications. Nevertheless, these markings may still be helpful for the correct assignment of specimens to the *P. gibba* group using LM as suggested by

Krammer (2000; although see nomenclatural note in 3.3 On the taxonomy of the *P. gibba* group). In any case, it seems that morphology alone lacks the resolution necessary to unambiguously delimit species within this group (e.g., G and I) and must be accompanied by genetic evidence. Moreover, evidence based on other diatom taxa (e.g., Sarno et al. 2005, Evans et al. 2007, Vanormelingen et al. 2008a, Trobajo et al. 2009, Vanelslander et al. 2009, Poulíčková et al. 2010, Souffreau et al. 2013, Pinseel et al. 2017a) confirms that (pseudo)cryptic diversity is rather widespread among diatoms. Therefore, we conclude that using morphology as the sole line of evidence for species delimitations and consequent taxonomic changes may lead to underestimation of diatom species diversity.

The delimited species have largely unknown biogeographies and seem mostly restricted to a single location. The only three species found in multiple regions were D (found on Amsterdam Island and the island of Réunion; several thousands of kilometres apart), G (found on Marion Island and Île de la Possession; thousand kilometres apart) and C (found throughout Europe; hundreds of kilometres apart), none of which was shared between hemispheres. Several of our species are geographically very distant from the sampling localities of the presumed corresponding type specimens (e.g., species A - *Pinnularia* cf. *microstauron*; Czech Republic vs. Brasil), suggesting that these taxa either have very wide geographic distributions, or that they do not correspond to single species, but rather could represent members of (pseudo)cryptic species complexes with restricted geographic distributions. It is recently becoming clear that many former cosmopolitan diatom taxa actually show hidden diversity with their members showing (relatively) restricted geographic distributions (Vanormelingen et al. 2008b), and/or niche-differentiation (Vanelslander et al. 2009, Souffreau et al. 2013b). However, significant increase of the sampling coverage is necessary to study the presence and nature of potential biogeographical patterns in the *P. gibba* group.

3.2 On the evolution of the *P. gibba* group

(this chapter is based on Kollár et al. *accepted* - reprinted as Appendix 2)

While the study design in Kollár et al. (2019) was optimized for species delimitation, namely that the data set was built in a way to maximize the probability of covering all the available intraspecific variability, the next step focused on phylogeny inference (Kollár et al. *accepted*). To this end, we added the more conserved plastid marker *rbcL* and information from the fossil record to infer a time-calibrated multi-gene phylogeny

of the group (see Figure 2 in Kollár et al. *accepted*). In addition, the phylogenetic results allowed for formulation and (in some cases) statistical testing of evolutionary hypotheses (see below or accepted manuscript in Appendix 2).

Our study (Kollár et al. *accepted*) represents the first attempt to infer the phylogenetic relationships within the *P. gibba* group. When considering only shared strains, the tree topologies of the *P. gibba* clade in our study (Figure 2 in Kollár et al. *accepted*) and those recovered by Souffreau et al. (2011), who covered the entire genus *Pinnularia*, are identical. Nevertheless, some topological differences were observed between the single-gene trees and the concatenated analysis in our study. However, none of these topological differences represented hard conflicts. Differences in tree topologies retrieved by analyzing different genes are generally thought to be the result of incomplete lineage sorting and are a common phenomenon across the tree of life (e.g., Leliaert et al. 2014). The results of the fossil time-calibrated molecular clock analysis suggest that the common ancestor of the *P. gibba* group probably diverged from the ancestor of the *P. grunowii* group in the Eocene, i.e., 36±9 Ma ago. The common ancestor of the terminal clade (i.e., F-G-H-I-S2-S3-S4) diverged 17±4 Ma ago. The Scottish sister species J (*P.* cf. *parvulissima*) and K (*P. lacustrigibba*) diverged 7±4 Ma ago, and species G (*P. acidicola* var. *acidicola*), H (*P. acidicola* var. *elongata*) and I (*P.* cf. *acidicola*), inhabiting volcanic islands of the Southern Hemisphere, diverged 6±3 Ma ago. In general, the overall results of our fossil-guided molecular time-calibration of the genus *Pinnularia*, were directly in line with earlier studies on/including the genus (Souffreau et al. 2011, Nakov et al. 2018).

Visual comparison of our phylogenetic results (Figure 2 in Kollár et al. *accepted*) with the valve morphology of the different species (e.g., Figure 4 in Kollár et al. 2019), suggested that the mean cell-sizes of the strains of the basal lineages (i.e., A-B, C, J-K, and S1-D-E) are larger than those of the strains of the lineages in the terminal clade (i.e., F-S2-S3-S4-G-H-I). We used the full data set of Kollár et al. (2019) and the randomization test to statistically test the null hypothesis (H_0: *there is no difference in the mean cell-size of the strains of the basal lineages and the mean cell-size of the strains of the lineages in the terminal clade*) and it was rejected. In other words, we found that the two groups in the *P. gibba* group (the terminal clade, and the basal grade) showed significant differences in cell-size. In microbes, cell-size plays a crucial role in their physiology and ecology (Marshall et al. 2012). However, an investigation of available environmental and geographic data associated with the *P. gibba* group strains (Kollár et al. 2019) revealed no connection between species within neither the terminal clade nor the basal grade.

In addition, the size of diatom cells usually reduce during the cell-cycle due to the mechanism of valve inheritance (MacDonald 1869, Pfitzer 1869). However, a hypothesis that strains of any of the two tested groups (i.e., terminal clade or basal grade) are more similar (relative to the strains of the second group) in cell-size because, by chance, they were measured in the similar phase of the cell-cycle, is refused by the statistical analysis itself. The observed pattern was significantly non-random. Therefore, we concluded that, at the current state of knowledge, observed cell-size difference is best explained by evolutionary origin, i.e., relatively diminutive appearance was probably inherited from the common ancestor of the species of the terminal clade. Moreover, in diatoms, patterns of cell-size evolution were observed through geological time (Finkel et al. 2005), and across different environments (Nakov et al. 2014). Furthermore, cell-size is often positively correlated with genome-size (e.g., Connolly et al. 2008, Koester et al. 2010, Mueller 2015). It is therefore not unlikely that the here observed reduction of cell-size may be correlated with the reduction of genome size. Nevertheless, this remains to be tested in future research.

We compared our phylogenies to biogeographical data available for the delimited species (Figure 5 of Kollár et al. 2019). Two species from the Northern Hemisphere (S2 from the Netherlands and S3 from the Czech Republic) appeared within the terminal clade, which is predominantly represented by species from the Southern Hemisphere. A similar pattern was observed in several other diatom complexes, such as the *P. borealis* complex (Pinseel et al. 2019), the *Hantzschia amphioxys* complex (Souffreau et al. 2013b), and the *Gomphonema parvulum* complex (Abarca et al. 2014). Along with the fact that many diatom species seem to be geographically restricted or even endemic (e.g., Van De Vijver et al. 2005, Vyverman et al. 2010, Pinseel et al. 2019), these results suggest that diatoms, similarly to some other protists (e.g., chrysophytes; Bock et al. 2017), are capable of effective dispersal between hemispheres. Nevertheless, the exact mechanisms of dispersal remain unknown. It is not unreasonable to suggest that dispersal by wind may be less effective between hemispheres than within hemispheres due to the equatorial atmospheric circulation: the Hadley cells (Hartmann 2015). The Hadley cells rise the air near the equator and turn it poleward in an altitude between 10 and 15 km above sea level. At such altitude, both atmospheric pressure (between 12 and 25 kPa) and temperature (ca -55°C) are very low (NASA 1976). To date, no experiments tested the tolerance of diatoms to low pressure, but we may assume that under such conditions they may suffer from desiccation which is lethal for many diatoms (Souffreau et al. 2010, 2013a).

Similarly, it seems that many diatoms are intolerant to freezing below -40°C, although there are some notable exceptions (Stock et al. 2018). In addition to passive dispersal by means of abiotic vectors, biotic vectors may also play a role. Previously, diatoms have been shown to colonize feathers of water birds (Croll and Holmes 1982), and they can even survive travel through the bird digestive system (e.g., Atkinson 1972, 1980, Stoyneva 2016). Twice a year, several bird species migrate between hemispheres in huge numbers. An extreme example is the Arctic Tern (*Sterna paradisaea*), which migrates between the Arctic and Antarctic region (Fijn et al. 2013). Its estimated population size is over 2,000,000 individuals and it fishes in both marine and freshwater habitats (BirdLife International 2018). It is thus not unlikely that migratory water birds may play a role in the (long-distance) dispersal of diatoms, in addition to abiotic vectors.

3.3 On the taxonomy of the *P. gibba* group

(this chapter is based on Kollár et al. *accepted* and Poulíčková et al. 2018 - reprinted as Appendices 2 and 3)

As mentioned above (see Introduction), unambiguous identification of diatom species is important not only for fundamental research but also for some applied disciplines. For example, more fine-grained species-level taxonomy increase the resolution of the diatom-based bioindicative systems of water quality assessments such as the one implemented through the European Union's Water Framework Directive (e.g., Poulíčková et al. 2017). To this end, however, diatom species must be (1) properly delimited and (2) unambiguously and routinely identifiable. In the *P. gibba* group, species limits were investigated by our polyphasic species delimitation (Kollár et al. 2019). However, most delimited species are members of large (pseudo)cryptic species complex *gibba-subgibba-parvulissima* and their morphology-based identification seems either challenging or impossible. Therefore, the diagnostic characters allowing for unambiguous and routine identification of species must be sought outside of morphology. For this reason, we tested whether the proposed diatom DNA barcode markers *rbcL* and SSU rDNA have sufficient resolution to distinguish even between closely related species in the *P. gibba* group.

The test revealed that both complete and partial *rbcL* (which is proposed for diatom-based biomonitoring in the United Kingdom and France; (Vasselon et al. 2017, Kelly et al. 2018) can unambiguously distinguish between different species, with a minimum divergence of 5 and 3 bp, respectively (Kollár et al. accepted). Similarly, Pinseel et al. (2019) found that

rbcL could distinguish even between the most closely related lineages of the cryptic species complex *P. borealis* (9-14 bp sequence difference on a total of 1,395 bp). Following the results of *rbcL*, complete and partial SSU rDNA was also able to distinguish between all *P. gibba* species (with a minimum divergence of 5 and 2 bp respectively), suggesting that both *rbcL* and SSU rDNA could be good barcode markers to distinguish even closely-related (pseudo)cryptic *P. gibba* group species. However, the sequence variability of SSU rDNA was generally lower than *rbcL*'s, and previous work indicated that different species can have identical V4 SSU rDNA sequences in the *P. borealis* and *A. minutissimum* species complexes (Pinseel et al. 2017b, Pinseel 2019). In addition, Mora et al. (2019) recently compared the resolutions of detailed morphology-based identifications with that of V4 SSU rDNA-based metabarcoding on epilithic diatom communities of streams in Central Mexico. They concluded that the most complete insight into the taxonomic structure of the communities is achieved by combining both approaches. However, while V4 SSU rDNA metabarcoding generally distinguished more taxa than morphological studies (266 vs. 216 taxa in Mora et al. 2019), it underestimated taxonomic diversity within 19 of the 49 genera (i.e., morphological investigations revealed more taxa within these genera than V4 SSU rDNA; see Table 2 in Mora et al. 2019) and, simultaneously, completely missed 14 of them. Concerning intrageneric taxonomic diversity in the genus *Pinnularia* and related genera (i.e., members of the same clade according to Nakov et al. 2018), V4 SSU rDNA revealed 7 taxa while morphology revealed 14 taxa in *Pinnularia*, 1 vs. 5 in *Caloneis*, 0 vs. 3 in *Eolimna*, and 7 vs. 19 in *Sellaphora*. The only genus closely related to *Pinnularia* in which V4 SSU rDNA outperformed morphological identifications (7 vs. 4 taxa) was *Mayamaea*. Therefore, despite our results suggest that V4 SSU rDNA can unambiguously distinguish closely related (pseudo)cryptic species of the *P. gibba* group, caution in using V4 SSU rDNA as a diatom DNA barcode marker still seems advisable.

To prevent possible future disputes regarding the taxonomy of the *P. gibba* group, I consider the following nomenclatural note important. From a morphological point of view, the *Pinnulariae* bearing the characteristic markings in the central area (so-called ghost striae, see Cox 1999) were sometimes called the "*gibba* group" (e.g., Krammer 2000). On the other hand, in the molecular phylogeny by Souffreau et al. (2011), the clade containing *gibba*-like taxa was called "the *subgibba* subclade" and, for sake of continuity, Kollár et al. (2019) followed this term, altering it to "the *Pinnularia subgibba* group"

or „*subgibba* group" in short. However, (Souffreau 2011) also observed ghost striae in one member of the "*grunowii*" subclade (strain PIN650K identified as *P. subanglica*) and one member of the "*nodosa*" subclade (PIN885TM identified as *P. nodosa*) and they may be observed also on LM photographs of *P. nodosa* (p. 308, Fig. 10-12) by Krammer (2000). Given that the character is probably homologous in all three subclades (together they form clade B sensu Souffrau et al. 2011), it cannot be considered apomorphic for the *P. gibba* group as implicitly suggested by Krammer's use of the term (e.g., Krammer 2000, p. 96). Nevertheless, while ghost striae cannot serve as an unambiguous diagnostic character of the *P. gibba* group on their own, for sake of identification, overall morphology is usually considered. Therefore, terms *subgibba* group sensu Souffreau and *gibba* group sensu Krammer may be considered synonymous.

The classifications within the genus *Pinnularia*, which developed during the last century (Hustedt 1930, Patrick and Reimer 1966, Krammer and Lange-Bertalot 1986, Krammer 1992), were either explicitly artificial (Krammer 2000, p. 17) or, to a great extent, reproductions of the classification by Per Theodor Cleve (1894, 1895). Krammer (2000, p. 17) stated that Cleve's classification was artificial. However, Cleve was one of the first Sweden followers of Darwinism and his intention to construct natural classification (although based on both common descent and degree of evolutionary change; Cleve 1894, p. 8-9) rather than an artificial one is, in my opinion, obvious from his statements:

"An ideal system should take in consideration the **evolution** of the different forms, but in the present fragmentary state of our knowledge, such consideration cannot be more than an approximation which may become closer to the truth as our knowledge of the forms becomes more extensive and perfect."

and

"As the known diatoms probably represent merely a fraction of those, which exist and have existed, it will be necessary in constructing a **natural** system to fill the gaps with conjectures." (Cleve 1894, p. 8)

For these reasons, I compared the *P. gibba* group primarily with Cleve's intendedly natural sections of the genus *Pinnularia*. The majority of species in our data set of the *P. gibba* group, would be placed in Cleve's section Tabellariae, namely species C (*P. macilenta*), D (*P.* cf. *australogibba*/*amsterdamensis*), E (*P.* cf. *vixconspicua*), F (*P.* sp.), G (*P. acidicola* var. *acidicola*), H (*P. acidicola* var. *elongata*), I (*P.* cf. *acidicola*), J (*P.* cf. *parvulissima*), K (*P. lacustrigibba*), S1 (*P.* sp.), S3 (*P.* sp.), and perhaps also B (*P.* cf. *lokana*). However, this section was later merged with section Divergentes (Patrick and Reimer 1966), formally preserving the name of the latter. The rest of species in our data set, namely A (*P.*

cf. *microstauron*), S2 (*P. subcapitata* var. *elongata*), and S4 (*P.* sp.) would be placed in the section Capitatae. Thus, I conclude that none of the Cleve's sections may be considered synonym of the *P. gibba* group.

One of the possible issues with morphology-based natural classifications may be that taxa are often defined with superficial non-structural characters (e.g., capitate valve ends in the section Capitatae) and that these characters are then used for phylogenetic inferences. However, such characters often do not reflect phylogenetic relationships but rather morphological convergence, ancestral state, or phenotypic plasticity (e.g., Cox 2009). Therefore, it is advisable to infer phylogenetic relationships and even define taxa based on structural apomorphic characters. On the other hand, as far as we know, there are truly cryptic species among diatoms (Pinseel et al. 2019, Poulíčková et al. 2018) and, thus, we cannot always follow this advice. Nevertheless, our study (Kollár et al. *accepted*) showed that deeper in evolutionary history patterns of morphological evolution (e.g., cell-size reduction in the common ancestor of the terminal clade) may be detected even in taxonomically complicated groups of diatoms with the widespread occurrence of (pseudo)cryptic species diversity. Although causes and circumstances of such patterns remain open for future research, they may still proof helpful in reviewing the higher classification of such groups.

The final contribution of this thesis to the taxonomy of the genus *Pinnularia* is an initiation of the taxonomical review of the *P. gibba* group by formal description of the species K sensu Kollár et al. (2019) as *Pinnularia lacustrigibba* sp. nov. (Poulíčková et al. 2018). The strains were earlier studied (including their reproductive compatibility) and identified as *Pinnularia* cf. *gibba* by Poulíčková et al. (2007). Comparison with morphologically similar taxa lead to the conlusion that, at current state of knowledge, it seems impossible to distinguish with reasonable certainty *P. lacustrigibba*, species J sensu Kollar et al. (2019; identified as *P.* cf. *parvulissima*), *P. parvulissima*, *P. subgibba* var. *subgibba*, *P. subgibba* var. *sublinearis* and *P. subgibba* var. *undulata* solely on the basis of LM (see Supplementary Table 1 of Poulíčková et al. 2018). The key morphological differences between these morphotaxa are often too subtle to allow for unambiguous identification in LM (e.g., valve sides 'very slightly convex' and 'parallel' in *P. subgibba* var. *subgibba* and *P. subgibba* var. sublinearis, respectively; Krammer 2000), especially considering the morphological variability we observed within polyphasically delimited species *P. lacustrigibba* (e.g., compare valve sides in Figures 1–7 of Poulíčková et al. 2018).

Nevertheless, analysis of species covered by data set of Kollár et al. (2019) revealed that they can be distinguished using genetic markers. For example, the genetic difference between sequences of species K (*P. lacustrigibba* sp. nov.) and J was found to be 2.3% (20 bp), 5.5% (36 bp) and 0.4% (6 bp) for LSU rDNA, *cox1* and *rbcL*, respectively. There was no (LSU rDNA) or low (*cox1*; strain PIN598A differed in two bp from the other four examined strains) intraspecific variability in the five available strains of *P. lacustrigibba*. More detailed investigations of valve ultrastructure under SEM, environmental requirements and geographical ranges of genetically distinct units within the morphotaxa are necessary to discover whether they can be distinguished without application of molecular techniques. Furthermore, species concepts and geographical distribution in relation to global diversity are among the most discussed problems in relation to the taxonomy, distribution and dispersal of microorganisms. In contrast to non-sexual microorganisms such as cyanobacteria (Dvořák et al. 2015), the biological species concept (Mayr 1942, 1963, Mann 1999) can potentially be applied to most diatoms, since most diatoms reproduce sexually and the main principles of their life cycle have been known for almost 150 years (Pfitzer 1869).

3.4 On the relativity of species

(this chapter is based on Kollár et al. *submitted* - reprinted as Appendix 4)

"*I have just been comparing definitions of species… It is really laughable to see what different ideas are prominent in various naturalists' minds, when they speak of 'species'… It all comes, I believe, from trying to define the undefinable.*" (C. Darwin's letter to J. D. Hooker from the 24th December 1856; Darwin 1887, p. 88)

The work on this thesis, coupled with thorough study of the species problem, led me to develop (at first glance perhaps surprising) opinion that Darwin was partly right in above metnioned letter from 1856. While the end of the terminal species (i.e., the end of the species lineage) is marked by an extinction event, its origin is hidden in the so-called grey zone of a speciation event where the evolutionary change of one species into the other is continual (with exception of saltational speciation). Due to a continual nature of evolution (regardless of its rate and constancy), species inevitably are undefinable as natural discontinuous units (with exception of presumably rare cases of saltational speciation, see below) whenever time dimension is taken into consideration. In other words, species cannot exist as natural units in time (with exception of those originating in saltational

speciation events) because an incorporation of time dimension inevitably transforms more or less discontinuous horizontal natural entities into continual evolutionary lineages (Figure 1 in Kollár et al. *submitted*). The existence or nonexistence of species as a natural unit is thus relative to its dimensionality.

The relativity of species' existence may be difficult to comprehend at first. However, Mayr (1982, p. 273) himself was aware that his 'biological species' are nonsense whenever time dimension is incorporated into considerations. In our work (Kollár et al. *submitted*), we simply extend this awareness by stating that the view of species as the natural discontinuous entities in time is nonsense concept in general (with exception of presumably rare saltational speciation). The meaningful concept in time is that of continual evolutionary lineages. Luckily, in most biological fields working with species (e.g., biodiversity, ecology, ethology, conservation biology, population genetics, medicine, some branches of evolutionary biology), the units of interest are horizontal species (e.g., Stamos 2004). Nevertheless, the study of species through time is of great importance for some branches of evolutionary biology (e.g., speciation, palaeontology, natural history).

Despite species-level taxonomists presumably deal mostly with horizontal species, they are often bitterly exposed to the species relativity whenever they deal with so-called incipient species. The incipient species are those which are at given time somewhere between the levels of subpopulation and independent species. They can either further diverge from the rest of the parental metapopulation or they can undergo backward merging thus never achieving relative independency of a fully developed species. However, it is usually impossible to predict their future with certainty. Therefore, they are virtually in a dual state of being and not being species at the same time. The most accurate way to reflect both the species relativity and the duality of incipient species is thus probabilistic. In other words, incipient species under certain conditions have certain probability that they will continue to diverge from the rest of the parental metapopulation and, eventually, achieve an independency of fully developed species (Figure 2 in Kollár et al. *submitted*). In theory, this probability is computable and thus have a potential to become an operational criterion even for the delimitation of incipient species.

In Kollár et al. (*submitted* - included as Appendix 4 of the thesis), we argue that such probabilistic model of divergence of evolutionary lineage, which we call the Universal Probabilistic Concept of Evolutionary Lineages (UPCEL), could potentialy form a theoretical basis for the development of probabilistic species delimitation methods based on the

operational criterion which is the fundamental cause of species' inner coherence, i.e., on gene flow. In theory, its advantages are that it should be universally applicable to all domains of life (with exception of strictly asexual lineages without any gene flow, if such exist) and also to incipient species. Nevertheless, it is clear that tremendously long and difficult journey stands between this first report of an interesting idea and its potential application.

4. CONCLUSIONS

The contributions to science made in this doctoral thesis may be summarized as follows:

(1) The species diversity covered by the data set of 105 clonal strains with worldwide origin (representing the selected *P. gibba* group) was studied by means of a poly-phasic approach. Primary (i.e., DNA-based) species delimitation based on LSU rDNA and *cox1* uncovered 15 species within the complex. This primary hypothesis was further confronted with other available lines of evidence (i.e., morphological, geographic, environmental and/or reproductive data) coming to the secondary hypothesis on species limits within the group, namely that our data set indeed covered 15 species of the group and that it is either challenging or imposible to distinguish majority of them solely on the basis of morphology.

(2) The fossil-guided time-calibrated multi-gene phylogeny of the *P. gibba* group was inferred using genetic markers from nuclear (LSU rDNA), plastid (*rbcL*) and mitochondrial (*cox1*) genomes.

(3) The knowledge on the limits of species combined with the knowledge on their phylogenetic relationships was used to design the test of taxonomic resolution of DNA barcode markers proposed for unambiguous and routine identification of diatoms. We have concluded that proposed DNA barcode markers of both *rbcL* and SSU rDNA are variable enough to unambiguously distinguish even cryptic sister species of the group.

(4) The phylogenetic results further allowed for a formulation and (in some cases) statistical testing of some evolutionary and biogeographic hypothesis. For example, we have found significant differences in cell-size in different parts of the tree which, at current state of knowledge, are best explained by an evolutionary origin.

(5) The taxonomic review of the group was initiated by describing one of the delimited species as *Pinnularia lacustrigibba* sp. nov. Several other uncovered species are new to science and should be formally decribed in the future.

(6) The work on this thesis, coupled with thorough study of the species problem, lead to a development of a novel probabilistic view of the species, the Universal Probabilistic Concept of Evolutionary Lineages (UPCEL). In the future, this view may potentially proof useful in developing probabilistic species delimitation methods based on gene flow interconnection state between evolutionary lineages.

5. REFERENCES

Abarca, N., Jahn, R., Zimmermann, J. & Enke, N. 2014. Does the cosmopolitan diatom *Gomphonema parvulum* (Kützing) Kützing have a biogeography? *PLoS One*. 9:1–18.

Adl, S.M., Leander, B.S., Simpson, A.G.B., Archibald, J.M., Anderson, O.R., Bass, D., Bowser, S.S. et al. 2007. Diversity, nomenclature, and taxonomy of protists. *Syst. Biol.* 56:684–9.

Andersen, R.A. 1992. Diversity of eukaryotic algae. *Biodivers. Conserv.* 1:267–92.

Andersson, L. 1990. The driving force: species concepts and ecology. *Taxon*. 39:375–82.

Atkinson, K.M. 1972. Birds as transporters of algae. *Br. Phycol. J.* 7:319–21.

Atkinson, K.M. 1980. Experiments in dispersal of phytoplankton by ducks. *Br. Phycol. J.* 15:49–58.

Behnke, A., Friedl, T., Chepurnov, V.A. & Mann, D.G. 2004. Reproductive compatibility and rDNA sequence analyses in the *Sellaphora pupula* species complex (Bacillariophyta). *J. Phycol.* 40:193–208.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K. et al. 2007. Cryptic species as a window on diversity and conservation. *Trends Ecol. Evol.* 22:148–55.

BirdLife, I. 2018. *Sterna paradisaea*.

Bock, C., Chatzinotas, A. & Boenigk, J. 2017. Genetic diversity in chrysophytes: comparison of different gene markers. *Fottea*. 17:209–21.

Bruder, K., Sato, S. & Medlin, L.K. 2008. Morphological and molecular investigations of naviculoid diatoms IV. *Pinnularia* vs. *Caloneis*. *Diatom*. 24:8–24.

Christenhusz, M.J.M. & Byng, J.W. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa*. 261:201–17.

Cleve, P.T. 1894. Synopsis of the naviculois diatoms, Part I. *Kongliga Sven. Akad. Handl.* 26:1–194.

Cleve, P.T. 1895. Synopsis of the naviculois diatoms, Part II. *Kongliga Sven. Akad. Handl.* 27:1–219.

Colwell, R.R. 1970. Polyphasic taxonomy of the genus *Vibrio*: numerical taxonomy of

Vibrio cholerae, *Vibrio parahaemolyticus*, and related *Vibrio* species. *J. Bacteriol.* 104:410–33.

Connolly, J.A., Oliver, M.J., Beaulieu, J.M., Knight, C.A., Tomanek, L. & Moline, M.A. 2008. Correlated evolution of genome size and cell volume in diatoms (Bacillariophyceae). *J. Phycol.* 44:124–31.

Cox, E.J. 1999. Variation in patterns of valve morphogenesis between representatives of six biraphid diatom genera (Bacillariophyceae). *J. Phycol.* 35:1297–312.

Cox, E.J. 2009. What's in a name? – Diatom classification should reflect systematic relationships. *Acta Bot.Croat.* 68:443–54.

Croll, D.A. & Holmes, R.W. 1982. A Note on the Occurrence of Diatoms on the Feathers of Diving Seabirds. *Auk*. 99:765–6.

Cronquist, A. 1978. Once again, what is a species? *In* Romberger, J. A. [Ed.] *Biosystematics in Agriculture*. Allanheld and Osmun, Montclair, pp. 3–20.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. *In* Howard, D. J. & Berlocher, S. H. [Eds.] *Endless Forms: Species and Speciation*. Oxford University Press, New York, pp. 57–75.

de Queiroz, K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56:879–86.

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia University Press, New York. 364 pp.

Drum, R.W. & Gordon, R. 2003. Star Trek replicators and diatom nanotechnology. *Trends Biotechnol.* 21:325–8.

Dvořák, P., Poulíčková, A., Hašler, P., Belli, M., Casamatta, D.A. & Papini, A. 2015. Species concepts and speciation factors in cyanobacteria, with connection to the problems of diversity and classification. *Biodivers. Conserv.* 24:739–57.

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*. 158:349–64.

Fijn, R.C., Hiemstra, D., Phillips, R.A. & van der Winden, J. 2013. Arctic Terns *Sterna paradisaea* from the Netherlands migrate record distances across three oceans to

Wilkes Land , East Antarctica. *Ardea*. 101:3–12.

Finkel, Z. V, Katz, M.E., Wright, J.D., Schofield, O.M.E. & Falkowski, P.G. 2005. Climatically driven macroevolutionary patterns in the size of marine diatoms over the Cenozoic. *Proc. Natl. Acad. Sci. U. S. A.* 102:8927–32.

Guiry, M.D. 2012. How many species of algae are there? *J. Phycol.* 48:1057–63.

Guiry, M.D. & Guiry, G.M. 2018. AlgaeBase. Available At: www.algaebase.org (last accessed November 13, 2018).

Guiry, M.D. & Guiry, G.M. 2020. AlgaeBase. Available At: www.algaebase.org (last accessed June 29, 2020).

Hartmann, D.L. 2015. Global physical climatology. 2nd ed. Elsevier Science. 498 pp.

Hennig, W. 1966. Phylogenetic systematics.

Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M. & Darzins, A. 2008. Microalgal triacylglycerols as feedstocks for biofuel production: Perspectives and advances. *Plant J.* 54:621–39.

Hustedt, F. 1930. Bacillariophyta (Diatomeae). Verlag von Gustav Fischer, Jena. 464 pp.

Kelly, M., Boonham, N., Juggins, S., Kille, P., Mann, D.G., Pass, D., Sapp, M. et al. 2018. A DNA based diatom metabarcoding approach for Water Framework Directive classification of rivers. *Environ. Agency*. Bristol.

Kociolek, J.P., Balasubramanian, K., Blanco, S., Coste, M., Ector, L., Liu, Y. & Kulikovskiy, M. 2018. *Pinnularia* C. G. Ehrneberg 1943. Available At: www.diatombase.org/aphia.php?p=taxdetails&id=149208 (last accessed May 25, 2018).

Koester, J.A., Swalwell, J.E., Von Dassow, P. & Armbrust, E.V. 2010. Genome size differentiates co-occurring populations of the planktonic diatom *Ditylum brightwellii* (Bacillariophyta). *BMC Evol. Biol.* 10:1–11.

Kollár, J., Fránková, M., Hašler, P., Letáková, M. & Poulíčková, A. 2015. Epiphytic diatoms in lotic and lentic waters - Diversity and representation of species complexes. *Fottea*. 15:259–71.

Kollár, J., Pinseel, E., Vanormelingen, P., Poulíčková, A., Souffreau, C., Dvořák, P. & Vyverman, W. 2019. A polyphasic approach to the delimitation of diatom species: A

case study for the genus *Pinnularia* (Bacillariophyta). *J. Phycol.* 55:365–79.

Kollár, J., Pinseel, E., Vyverman, W. & Poulĭcková, A. A time-calibrated multi-gene phylogeny provides insights into the evolution, taxonomy and DNA barcoding of the *Pinnularia gibba* group (Bacillariophyta). *Fottea* (*accepted*).

Kollár, J., Poulíčková, A. & Dvořák, P. On the Relativity of Species, or the Probabilistic Solution to the Species Problem. *Syst. Biol.* (*submitted*).

Komárek, J. 2016. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. *Eur. J. Phycol.* 51:346–53.

Krammer, K. 1992. *Pinnularia*, eine monographie de europäischen taxa. *Bibl. Diatomol.* 26:1–353.

Krammer, K. 2000. The genus *Pinnularia*. A.R.G. Ganter Verlag K.G., Ruggell. 703 pp.

Krammer, K. & Lange-Bertalot, H. 1986. 1. Teil: Naviculaceae. *In* Ettl, H., Gerloff, J., Heynig, H. & Mollenhauer, D. [Eds.] *Süsswasserflora von Mitteleuropa*. Gustav Fisher Verlag, Jena, p. 876.

Lewis, A.R., Marchant, D.R., Ashworth, A.C., Hedenäs, L., Hemming, S.R., Johnson, J. V, Leng, M.J. et al. 2008. Mid-Miocene cooling and the extinction of tundra in continental Antarctica. *Proc. Natl. Acad. Sci.* 105:10676–80.

Luís, A.T., Novais, M.H., van de Vijver, B., Almeida, S.F.P., da Silva, E.A.F., Hoffmann, L. & Ector, L. 2012. *Pinnularia aljustrelica* sp. nov. (Bacillariophyceae), a new diatom species found in acidic waters in the Aljustrel mining area (Portugal) and further observations on the taxonomy and ecology of *P. acidophila* Hofmann et Krammer and *P. acoricola* Hustedt. *Fottea*. 12:27–40.

MacDonald, J.D. 1869. On the structure of the diatomaceous frustule, and its genetic cycle. *Ann. Mag. Nat. Hist. Ser.* 4:1–8.

Mann, D.G. 1999. The species concept in diatoms. *Phycologia*. 38:437–95.

Mann, D.G. & Evans, K.M. 2008. The species concept and cryptic diversity. *In Proceedings of the 12th International Conference on Harmful Algae*. International Society for the Study of Harmful Algae and Intergovernmental Oceanographic Commission of UNESCO, Copenhagen, pp. 262–8.

Mann, D.G., McDonald, S.M., Bayer, M.M., Droop, S.J.M., Chepurnov, V.A., Loke, R.E.,

Ciobanu, A. et al. 2004. The *Sellaphora pupula* species complex (Bacillariophyceae): Morphometric analysis, ultrastructure and mating data provide evidence for five new species. *Phycologia*. 43:459–82.

Mann, D.G. & Vanormelingen, P. 2013. An inordinate fondness? The number, distributions, and origins of diatom species. *J. Eukaryot. Microbiol.* 60:414–20.

Martin, F.M., Uroz, S. & Barker, D.G. 2017. Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science (80).* 356.

Mayden, R.L. 1997. A hierarchy of species concepts: the denoument in the saga of the species problem. *In* Claridge, M. F., Dawah, H. A. & Wilson, M. R. [Eds.] *Species: The Units of Diversity*. Chapman and Hall, London, pp. 381–423.

Mayr, E. 1942. Systematics and the Origin of Species. Columbia University Press, New York. 334 pp.

Mayr, E. 1963. Animal Species and Evolution. Harvard University Press, Cambridge, MA. 797 pp.

Mayr, E. 1982. The Growth of Biological Thought. 11th ed. Harvard University Press, Cambridge, MA. 974 pp.

Mayr, E. & Ashlock, P.D. 1991. Principles of Systematic Zoology. McGraw-Hill Inc., New York. 475 pp.

Mora, D., Abarca, N., Proft, S., Grau, J.H., Enke, N. & Carmona, J. 2019. Morphology and metabarcoding : A test with stream diatoms from Mexico highlights the complementarity of identi fi cation methods. *Freshw. Sci.* 38:448–464.

Mueller, R.L. 2015. Genome biology and the evolution of cell-size diversity. *Cold Spring Harb. Perspect. Biol.* 7:1–9.

Nabout, J.C., da Silva Rocha, B., Carneiro, F.M. & Sant'Anna, C.L. 2013. How many species of Cyanobacteria are there? Using a discovery curve to predict the species number. *Biodivers. Conserv.* 22:2907–18.

Nakov, T., Theriot, E.C. & Alverson, A.J. 2014. Using phylogeny to model cell size evolution in marine and freshwater diatoms. *Limnol. Oceanogr.* 59:79–86.

NASA 1976. U.S. Standard Atmosphere. National Aeronautics and Space Administration. 241 pp.

Nelson, G.J. & Platnick, N.I. 1981. Systematics and Biogeography: Cladistics and Vicariance. Columbia University Press, New York. 567 pp.

Patrick, R. & Reimer, C.W. 1966. The diatoms of the United States exclusive of Alaska and Hawaii. *Monogr. Acad. Nat. Sci. Philadelphia*. 1:688.

Pereira, A.C., Torgan, L.C. & Melo, S. 2014. Four new *Pinnularia* Ehrenberg (Bacillariophyta, Pinnulariaceae) species from Amazonian black water (Tupé Lake, Amazonas State, Brazil). *Phytotaxa*. 158:154–68.

Pfitzer, E. 1869. Über den Bau und die Zellteilung der Diatomeen. *Bot. Zeitung*. 27:774–6.

Pinseel, E. 2019. Diversity, biogeography and evolutionary history of diatoms: *Pinnularia borealis* as a case study. Ghent University Doctoral Thesis.

Pinseel, E., Hejduková, E., Vanormelingen, P., Kopalová, K., Vyverman, W. & de Vijver, B. Van 2016. *Pinnularia catenaborealis* sp. nov . (Bacillariophyceae), a unique chain-forming diatom species from James Ross Island and Vega Island (Maritime Antarctica). *Phycologia*. 56:94–107.

Pinseel, E., Janssens, S.B., Verleyen, E., Vanormelingen, P., Kohler, T.J., Biersma, E.M., Sabbe, K. et al. 2020. Global radiation in a rare biosphere soil diatom. *Nat. Commun.* 11:1–12.

Pinseel, E., Kulichová, J., Scharfen, V., Urbánková, P., Van de Vijver, B. & Vyverman, W. 2019. Extensive cryptic diversity in the terrestrial diatom *Pinnularia borealis* (Bacillariophyceae). *Protist*. 170:121–40.

Pinseel, E., Van de Vijver, B., Kavan, J., Verleyen, E. & Kopalová, K. 2017a. Diversity, ecology and community structure of the freshwater littoral diatom flora from Petuniabukta (Spitsbergen). *Polar Biol.* 40:533–51.

Pinseel, E., Vanormelingen, P., Hamilton, P.B., Vyverman, W., Van de Vijver, B. & Kopalová, K. 2017b. Molecular and morphological characterization of the *Achnanthidium minutissimum* complex (Bacillariophyta) in Petuniabukta (Spitsbergen, High Arctic) including the description of *A. digitatum* sp. nov. *Eur. J. Phycol.* 52:264–80.

Poulíčková, A., Dvořák, P., Mazalová, P. & Hašler, P. 2014. Epipelic microphototrophs: an overlooked assemblage in lake ecosystems. *Freshw. Sci.* 33:513–23.

Poulíčková, A., Hájková, P., Kintrová, K., Baťková, R., Czudková, M. & Hájek, M. 2013. Tracing decadal environmental change in ombrotrophic bogs using diatoms from herbarium collections and transfer functions. *Environ. Pollut.* 179:201–9.

Poulíčková, A., Kollár, J., Hašler, P., Dvořák, P. & Mann, D.G. 2018. A new species *Pinnularia lacustrigibba* sp. nov. within the *Pinnularia subgibba* group (Bacillariophyceae). *Diatom Res.* 33:273–82.

Poulíčková, A., Letáková, M., Hašler, P., Cox, E. & Duchoslav, M. 2017. Species complexes within epiphytic diatoms and their relevance for the bioindication of trophic status. *Sci. Total Environ.* 599–600:820–33.

Poulíčková, A., Mayama, S., Chepurnov, V.A. & Mann, D.G. 2007. Heterothallic auxosporulation, incunabula and perizonium in *Pinnularia* (Bacillariophyceae). *Eur. J. Phycol.* 42:367–90.

Poulíčková, A., Veselá, J., Neustupa, J. & Škaloud, P. 2010. Pseudocryptic diversity versus cosmopolitanism in diatoms: A case study on *Navicula cryptocephala* Kütz. (Bacillariophyceae) and morphologically similar taxa. *Protist*. 161:353–69.

Puillandre, N., Modica, M. V., Zhang, Y., Sirovich, L., Boisselier, M.C., Cruaud, C., Holford, M. et al. 2012. Large-scale species delimitation method for hyperdiverse groups. *Mol. Ecol.* 21:2671–91.

Quijano-Scheggia, S.I., Garcés, E., Lundholm, N., Moestrup, Ø., Andree, K. & Camp, J. 2009. Morphology, physiology, molecular phylogeny and sexual compatibility of the cryptic *Pseudo-nitzschia delicatissima* complex (Bacillariophyta), including the description of *P. arenysensis* sp. nov. *Phycologia*. 48:492–509.

Rosenberg, N.A. & Nordborg, M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.* 3:380–90.

Round, F.E., Crawford, R.M. & Mann, D.G. 1990. The Diatoms: Biology and Morphology of the Genera. Cambridge University Press, Cambridge, UK. 747 pp.

Sarno, D., Kooistra, W.H.C.F., Medlin, L.K., Percopo, I. & Zingone, A. 2005. Diversity in the genus *Skeletonema* (Bacillariophyceae). II. An assessment of the taxonomy of *S. costatum*-like species with the description of four new species. *J. Phycol.* 41:151–76.

Simpson, G.G. 1951. The species concept. *Evolution* 5(4):285–298.

Simpson, G.G. 1961. Principles of animal taxonomy. Columbia University Press, New York. 247 pp.

Škaloud, P. 2008. Polyphasic approaches in the taxonomy of green aerophytic algae. Charles University Doctoral Thesis, 111 pp.

Smol, J.P. & Stoermer, E.F. 2010. The Diatoms: Applications for the Environmental and Earth Sciences. 2nd ed. Cambridge University Press, Cambridge. 667 pp.

Souffreau, C. 2011. Biogeography, species diversity and stress tolerance of aquatic and terrestrial diatoms. Ghent University Doctoral Thesis, 285 pp.

Souffreau, C., Vanormelingen, P., Sabbe, K. & Vyverman, W. 2013a. Tolerance of resting cells of freshwater and terrestrial benthic diatoms to experimental desiccation and freezing is habitat-dependent. *Phycologia*. 52:246–55.

Souffreau, C., Vanormelingen, P., Van de Vijver, B., Isheva, T., Verleyen, E., Sabbe, K. & Vyverman, W. 2013b. Molecular evidence for distinct Antarctic lineages in the cosmopolitan terrestrial diatoms *Pinnularia borealis* and *Hantzschia amphioxys*. *Protist*. 164:101–15.

Souffreau, C., Vanormelingen, P., Verleyen, E., Sabbe, K. & Vyverman, W. 2010. Tolerance of benthic diatoms from temperate aquatic and terrestrial habitats to experimental dessication and temperature stress. *Phycologia*. 49:309–24.

Souffreau, C., Verbruggen, H., Wolfe, A.P., Vanormelingen, P., Siver, P.A., Cox, E.J., Mann, D.G. et al. 2011. A time-calibrated multi-gene phylogeny of the diatom genus *Pinnularia*. *Mol. Phylogenet. Evol.* 61:866–79.

Stamos, D.N. 2004. The Species Problem, Biological Species, Ontology and the Metaphysics of Biology. Lexington Books, Oxford. 390 pp.

Stock, W., Pinseel, E., De Decker, S., Sefbom, J., Blommaert, L., Chepurnova, O., Sabbe, K. et al. 2018. Expanding the toolbox for cryopreservation of marine and freshwater diatoms. *Sci. Rep.* 1–9.

Stoyneva, M.P. 2016. Allochthonous planktonic algae recorded during the last 25 years in Bulgaria and their possible dispersal agents. *Hydrobiologia*. 764:53–64.

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* (Bacillariophyceae). *Phycologia*. 48:443–59.

Van De Vijver, B., Gremmen, N.J.M. & Beyens, L. 2005. The genus *Stauroneis* (Bacillariophyceae) in the Antarctic region. *J. Biogeogr.* 32:1791–8.

Van de Vijver, B., Zidarova, R. & de Haan, M. 2011. Four new *Luticola* taxa (Bacillariophyta) from the South Shetland Islands and James Ross Island (Maritime Antarctic Region). *Nov. Hedwigia*. 92:137–58.

Van Valen, L. 1976. Ecological species, multispecies, and oaks. *Taxon*. 25:233–9.

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. & Swings, J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60:407–38.

Vanelslander, B., Créach, V., Vanormelingen, P., Ernst, A., Chepurnov, V.A., Sahan, E., Muyzer, G. et al. 2009. Ecological differentiation between sympatric pseudocryptic species in the estuarine benthic diatom *Navicula phyllepta* (Bacillariophyceae). *J. Phycol.* 45:1278–89.

Vanormelingen, P., Chepurnov, V.A., Mann, D.G., Sabbe, K. & Vyverman, W. 2008a. Genetic divergence and reproductive barriers among morphologically heterogeneous sympatric clones of *Eunotia bilunaris* sensu lato (Bacillariophyta). *Protist*. 159:73–90.

Vanormelingen, P., Verleyen, E. & Vyverman, W. 2008b. The diversity and distribution of diatoms: From cosmopolitanism to narrow endemism. *Biodivers. Conserv.* 17:393–405.

Vasselon, V., Rimet, F., Tapolczai, K. & Bouchez, A. 2017. Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island, France). *Ecol. Indic.* 82:1–12.

Verleyen, E., Vyverman, W., Sterken, M., Hodgson, D.A., De Wever, A., Juggins, S., Van De Vijver, B. et al. 2009. The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities. *Oikos*. 118:1239–49.

Vyverman, W., Verleyen, E., Sabbe, K., Vanhoutte, Koenradd, Sterken, M., Hodgson, D.A., Mann, D.G. et al. 2007. Historical processes constrain patterns in global diatom diversity. *Ecology*. 88:1924–31.

Vyverman, W., Verleyen, E., Wilmotte, A., Hodgson, D.A., Willems, A., Peeters, K., Van

de Vijver, B. et al. 2010. Evidence for widespread endemism among Antarctic micro-organisms. *Polar Sci.* 4:103–13.

Wheeler, Q.D. & Meier, R. 2000. Species concepts and phylogenetic theory: A debate. Columbia University Press, New York. 230 pp.

Wiley, E.O. 1978. The evolutionary species concept reconsidered. *Syst. Zool.* 27:17–26.

Wiley, E.O. & Mayden, R.L. 2000. The evolutionary species concept. *In* Wheeler, Q. D. & Meier, R. [Eds.] *Species Concepts and Phylogenetic Theory*. Columbia University Press, New York, pp. 70–89.

Wilkins, J.S. 2009. Species: A history of the idea. 1st ed. University of California Press, Berkeley. 320 pp.

Wilkins, J.S. 2011. Philosophically speaking, how many species concepts are there? *Zootaxa*. 2765:58-60.

Will, K.W., Mishler, B.D. & Wheeler, Q.D. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* 54:844–51.

6. APPENDICES

1. **Kollár J.**, Pinseel E., Vanormelingen P., Poulíčková A., Souffreau C., Dvořák P. & Vyverman W. 2019: A polyphasic approach to the delimitation of diatom species: a case study for the genus *Pinnularia* (Bacillariophyta). *J. Phycol.* 55, 365–379. (IF_{2018} = 2,831; Q1)

2. **Kollár J.**, Pinseel E., Vyverman W. & Poulíčková A.: A time-calibrated phylogeny provides an insight into the evolution, taxonomy and DNA barcoding of the *Pinnularia gibba* group (Bacillariophyta). *Fottea*, (*accepted*). (IF_{2018} = 1,727; Q2)

3. Poulíčková A., **Kollár J.**, Hašler P., Dvořák P. & Mann D. G. 2018: A new species *Pinnularia lacustrigibba* sp. nov. within the *Pinnularia subgibba* group (Bacillariophyceae). *Diatom Res.* 33, 273–282. (IF_{2018} = 1.169; Q3)

4. **Kollár J.**, Poulíčková A. & Dvořák P.: On the relativity of species, or the probabilistic solution to the species problem. *Syst. Biol.*, (*submitted*). (IF_{2018} = 10.266; Q1)

5. Author's curriculum vitae

Appendix 1

Paper I

J. Phycol. **55,** 365–379 (2019)

DOI: 10.1111/jpy.12825

A POLYPHASIC APPROACH TO THE DELIMITATION OF DIATOM SPECIES: A CASE STUDY FOR THE GENUS *PINNULARIA* (BACILLARIOPHYTA)[1]

Jan Kollár[2]

Department of Botany, Faculty of Science, Palacký University Olomouc, Šlechtitelů 27, CZ-78371 Olomouc, Czech Republic

Eveline Pinseel

Laboratory of Protistology and Aquatic Ecology, Department of Biology, Faculty of Science, Ghent University, Krijgslaan 281 - S8, 9000 Gent, Belgium

Department of Bryophyta and Thallophyta, Botanic Garden Meise, Nieuwelaan 38, 1860 Meise, Belgium

Ecosystem Management Research Group (ECOBE), Department of Biology, Faculty of Science, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium

Pieter Vanormelingen

Laboratory of Protistology and Aquatic Ecology, Department of Biology, Faculty of Science, Ghent University, Krijgslaan 281 - S8, 9000 Gent, Belgium

Aloisie Poulíčková

Department of Botany, Faculty of Science, Palacký University Olomouc, Šlechtitelů 27, CZ-78371 Olomouc, Czech Republic

Caroline Souffreau

Laboratory of Aquatic Ecology and Evolutionary Biology, University of Leuven, Charles Deberiotstraat 32, 3000 Leuven, Belgium

Petr Dvořák

Department of Botany, Faculty of Science, Palacký University Olomouc, Šlechtitelů 27, CZ-78371 Olomouc, Czech Republic

and Wim Vyverman

Laboratory of Protistology and Aquatic Ecology, Department of Biology, Faculty of Science, Ghent University, Krijgslaan 281 - S8, 9000 Gent, Belgium

Diatoms are one of the most abundant and arguably the most species-rich group of protists. Diatom species delimitation has often been based exclusively on the recognition of morphological discontinuities without investigation of other lines of evidence. Even though DNA sequences and reproductive experiments have revealed several examples of (pseudo)cryptic diversity, our understanding of diatom species boundaries and diversity remains limited. The cosmopolitan pennate raphid diatom genus *Pinnularia* represents one of the most taxon-rich diatom genera. In this study, we focused on the delimitation of species in one of the major clades of the genus, the *Pinnularia subgibba* group, based on 105 strains from a worldwide origin. We compared genetic distances between the sequences of seven molecular markers and selected the most variable pair, the mitochondrial *cox*1 and nuclear encoded LSU rDNA, to formulate a primary hypothesis on the species limits using three single-locus automated species delimitation methods. We compared the DNA-based primary hypotheses with morphology and with other available lines of evidence. The results indicate that our data set comprised 15 species of the *P. subgibba* group. The vast majority of these taxa have an uncertain taxonomic identity, suggesting that several may be unknown to science and/or members of (pseudo) cryptic species complexes within the *P. subgibba* group.

Key index words: ***cox*1; diatoms; LSU rDNA; molecular taxonomy; *Pinnularia subgibba*; *psb*A; *rbc*L; species diversity; SSU rDNA**

***Abbreviations*: BI, Bayesian inference; BIC, Bayesian information criterion; BoLD, Barcode of Life Database (www.boldsystems.org); BS, bootstrap support; *cox*1-5P, 5′ region of *cox*1; D1, D2, and D3 LSU rDNA, regions of LSU rDNA; GB, GenBank (www.ncbi.nlm.nih.gov); MCMC, Markov chain Monte Carlo; ML, maximum likelihood; PP, posterior**

[1]Received 21 March 2018. Accepted 7 November 2018. First Published Online 7 December 2018. Published Online 31 January 2019, Wiley Online Library (wileyonlinelibrary.com).
[2]Author for correspondence: e-mail jan.kollar@upol.cz.
Editorial Responsibility: O. De Clerck (Associate Editor)

probability; PTP, Poissson tree processes; *rbc*L-3P, 3′ region of *rbc*L; sGMYC, single threshold generalized mixed Yule-coalescent; SPNE, statistical parsimony network estimation; V4 SSU rDNA, region of SSU rDNA

Diatoms (Bacillariophyta) are the most species-rich group of algae with estimates of the total number of species ranging from 30,000 to 200,000 (Mann and Droop 1996, Mann and Vanormelingen 2013). Diatom species have traditionally been delimited using discontinuities in frustule morphology. However, several recent studies using DNA sequences, ecological data, and/or data on sexual reproduction have revealed an extensive (pseudo)cryptic species diversity in various diatom genera (Sarno et al. 2005, Evans et al. 2007, Vanormelingen et al. 2008a, Trobajo et al. 2009, Vanelslander et al. 2009, Poulíčková et al. 2010, Souffreau et al. 2013, Pinseel et al. 2017a, 2018). To date, over 12,000 diatom species have been formally described (Guiry 2012) many of which are adapted to specific environmental conditions. Consequently, they are frequently used as bioindicators for routine assessments of ecological status of aquatic ecosystems (Smol and Stoermer 2010) and in paleoecological studies focusing on past climate and environmental change (Lewis et al. 2008). With the increasing availability of molecular tools to identify diatom species in environmental analyses (Zimmermann et al. 2015), studies linking molecular species identifications with morphological data are needed. In addition, more fine-grained taxonomy will potentially for the more accurate definition of ecological niches of species and, consequently, increase the resolution of environmental analyses (Novais et al. 2015).

Recently, De Queiroz (2007) separated the theoretical concept of species from the operational criteria used for species delimitation and identification (e.g., genetics, morphology, reproduction, physiology, ecology, biogeography, and reciprocal monophyly). This unified species concept uses as many operational criteria as possible to delimit species. This holistic approach is often called integrative (Will et al. 2005, Puillandre et al. 2012) or polyphasic (Colwell 1970, Vandamme et al. 1996, Komárek 2016). In this regard, molecular data are often used to formulate a primary hypothesis to species limits and other lines of evidence can subsequently be used to test and, if necessary, adjust the primary hypothesis into a secondary hypothesis (Goldstein and DeSalle 2011, Puillandre et al. 2012). The main genes currently used for species-level molecular analyses of diatoms are the plastid *rbc*L, and the nuclear encoded LSU rDNA, SSU rDNA, and ITS rDNA (Sarno et al. 2005, Poulíčková et al. 2010, Kistenich et al. 2014). For molecular species delimitation, (i) a "barcode gap" should be present (i.e., the intraspecific variation must be distinctly lower than the interspecific variation) and (ii) the markers should be variable enough to distinguish between closely related species. In addition, DNA-based species limits, especially when inferred from a single marker, can be biased by incomplete lineage sorting (Funk and Omland 2003, Leliaert et al. 2014), or introgressive hybridization (Chase et al. 2005). Consequently, the use of multiple markers significantly increases the accuracy of species delimitations (Knowles and Carstens 2007). The reliability of the primary hypothesis may also be increased by the congruence of results obtained using multiple molecular species delimitation methods (Blair and Bryson 2017).

Pinnularia is among the most taxon-rich genera of raphid diatoms, with 723 currently accepted names in Algaebase (Guiry and Guiry 2018) and 1,527 species names listed in DiatomBase (Kociolek et al. 2018), although, the taxonomic or nomenclatural status of the vast majority of the latter taxa is unassessed. In addition, the genus *Caloneis* was shown to be polyphyletic within *Pinnularia* (Bruder et al. 2008, Souffreau et al. 2011) raising the total number of accepted taxa in the *Pinnularia–Caloneis* complex (i.e., *Pinnularia* sensu lato) to 931 according to Algaebase and making it, at the current state of knowledge, the second most taxon-rich diatom genus (outnumbered only by *Navicula* with 1,270 taxa, and followed by *Nitzschia* with 788 taxa; Guiry and Guiry 2018). The genus is estimated to have originated in the Paleocene, 60 ± 15 Mya (Souffreau et al. 2011). To date, two DNA-based intrageneric phylogenetic hypotheses were proposed using concatenated alignments of three (SSU rDNA, LSU rDNA, and *rbc*L; Bruder et al. 2008) and five (*cox*1, SSU rDNA, LSU rDNA, *rbc*L, and *psb*A; Souffreau et al. 2011) genes, respectively. Both revealed three major clades to be present within the genus. Clade B sensu Souffreau et al. (2011) consists of three distinct subclades named after their main representative taxa *P. grunowii*, *P. nodosa*, and *P. subgibba*. The last subclade is the focus of this study and, henceforth, will be referred to as the "*Pinnularia subgibba* group." It is estimated to have originated in the upper Eocene, 35 ± 10 Mya (Souffreau et al. 2011). Both the entire clade B and the *P. subgibba* group have maximal statistical support and can be distinguished by a combination of morphological features (Souffreau et al. 2011). Specimens of clade B are characterized by small linear cells with drop-like central raphe endings and large alveolar openings. In addition, members of the *P. subgibba* group are usually elongated in shape with two girdle-appressed parallel plastids and a broad nonporous central area called the fascia (Souffreau et al. 2011). The fascia is often accompanied by two to four indistinct markings (the pair on the primary side of the valve being considerably larger than the pair on the secondary side) but their utility for species-level taxonomy of the group is questionable.

Several taxa of the *P. subgibba* group (e.g., *Pinnularia gibba*, *Pinnularia parvulissima*, *P. subgibba*) are very uniform in morphology and are generally considered to be cosmopolitan (Krammer 2000).

In this study, species limits within the *Pinnularia subgibba* group were explored in a stepwise way (Fig. 1). First, using a data set of 24 *Pinnularia* strains, the sequence variability of seven molecular markers was assessed after which the two most variable markers were selected for species delimitation in *Pinnularia*. Second, using newly established diatom strains and GenBank sequences, a data set containing these two most variable genes of 58 strains belonging to the *P. subgibba* group was established (these were selected to represent original 105 *P. subgibba* group strains). Third, automated molecular species delimitation methods were used to delimit species-level clusters within the *P. subgibba* group. At last, other available lines of evidence (morphology, biogeography, ecology, and reproductive compatibility) were assessed in relation to the results of the molecular species delimitation.

MATERIALS AND METHODS

Data set A: variability of molecular markers. To assess the molecular markers most suitable for species delimitation in *Pinnularia*, a data set of 24 published *Pinnularia* strains (Data set A, Table S1 in the Supporting Information; Souffreau et al. 2011), including the *Pinnularia subgibba* group was analyzed for their sequence variability of seven molecular markers: *cox*1-5P (ca. 650 bp at the 5′ end of *cox*1), *rbc*L, *rbc*L-3P (748 bp at the 3′ end of *rbc*L; Hamsher et al. 2011), *psb*A, SSU rDNA, V4 SSU rDNA (ca. 400 bp of SSU rDNA including its V4 subregion; Zimmermann et al. 2011), and the D1-D2 subregion of LSU rDNA. The data set of Souffreau et al. (2011) included many genes and reasonably covered the morphological and geographic variability within *Pinnularia*, which made this data set the best available candidate to identify the most suitable markers for species delimitation. As Souffreau et al. (2011) amplified only the D1-D2 subregion of LSU rDNA, we additionally used a subset of 40 strains from Data set B (Table S2 in the Supporting Information) to analyze sequence variability of the separate subregions D1, D2, and D3.

Sequences of each marker were automatically aligned using the MUSCLE algorithm (Edgar 2004) as implemented in MEGA6 (Tamura et al. 2013) and checked visually. Alignments were trimmed so that at least 97% of the positions were available and alignments of protein-coding genes (*cox*1-5P, *rbc*L, and *psb*A) were trimmed according to their open reading frames. The most divergent regions in the alignments of ribosomal genes (SSU rDNA and LSU rDNA) were eliminated using the Gblocks Server 0.91b (Castresana 2000) to avoid possible nonhomologous positions of hypervariable loop regions. Gap positions were allowed in the final blocks along with the less strict flanking positions. The exact effect of the elimination on the alignments of the ribosomal genes can be explored in Table 1. All alignments built in this study (Table 1) are available in the JK's repository (https://www.researchgate.net/profile/Jan_Kollar5).

The genetic distances (p-distances and numbers of differences) between the aligned sequences were calculated in MEGA6. While numbers of differences are more informative for the species delimitation itself, p-distances may still be informative for other branches of molecular systematics such as DNA barcoding where efficiency (i.e., species resolution at low expanses) is a critical criterion for the selection of molecular markers. Therefore, we used both genetic distances for the comparison of molecular markers. Both transitions and transversions were included and all positions with less than 50% site coverage were eliminated. For each pair of strains,

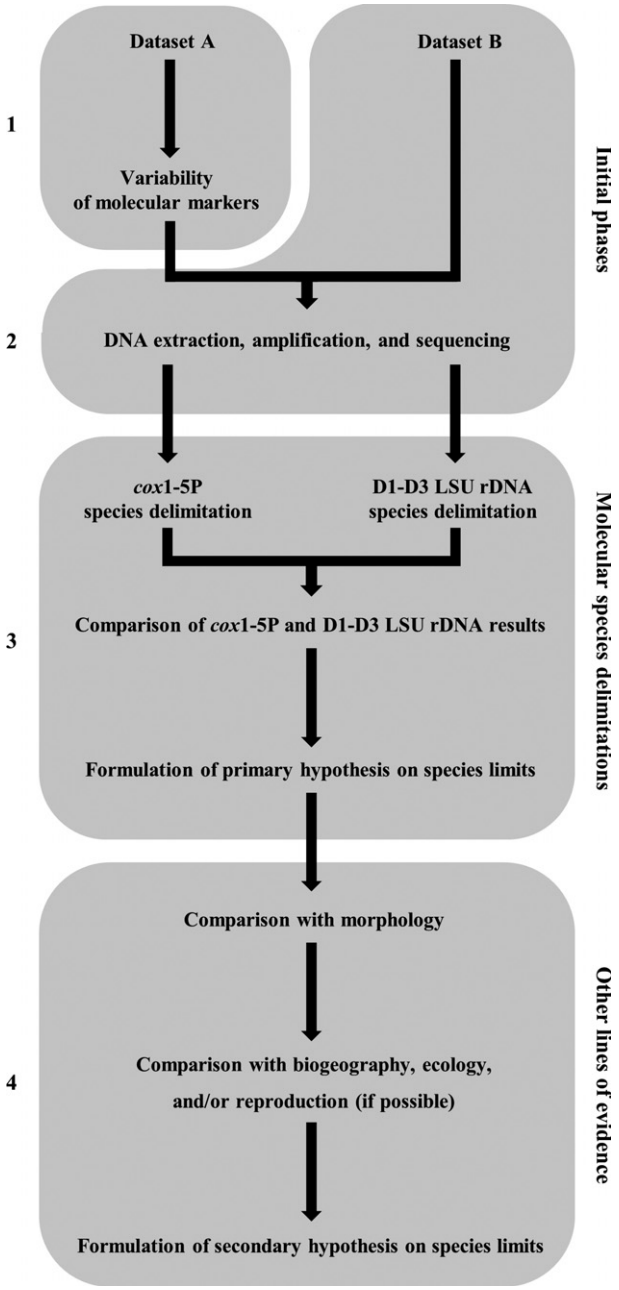


FIG. 1. The design of the study. The study included four major steps indicated in grey: (1) exploring variability of molecular markers, (2) building the data set for species delimitation, (3) automated molecular species delimitation resulting in the primary (i.e., DNA-based) hypothesis on species limits within the *Pinnularia subgibba* group, and (4) the comparison of the primary hypothesis with other lines of evidence resulting in the secondary hypothesis on species limits within the group. Data set A consisted of 24 strains of the genus *Pinnularia*. Data set B consisted of 58 strains of the *P. subgibba* group.

TABLE 1. Characteristics of alignments including total number of characters, number of variable characters (V), number of parsimony-informative characters (PI), and mean genetic distances between the sequences of molecular markers ± SD. Data set A consisted of 24 strains of the genus *Pinnularia*. Data set B consisted of 58 strains of the *P. subgibba* group. Original trimmed alignments of ribosomal markers are indicated by an asterisk.

Genetic marker	Total	V	PI	p-distances	Numbers of differences
Data set A					
*cox*1-5P	639	262	227	0.150 ± 0.033	93.0 ± 20.7
*rbc*L	1,386	225	134	0.040 ± 0.013	56.1 ± 17.4
*rbc*L-3P	748	149	92	0.052 ± 0.017	37.8 ± 12.4
*psb*A	888	99	58	0.026 ± 0.012	21.8 ± 9.9
D1-D2 LSU rDNA	519	264	210	0.187 ± 0.055	93.1 ± 31.1
D1-D2 LSU rDNA*	548	280	222	0.193 ± 0,056	98.6 ± 33.2
SSU rDNA	1,476	250	173	0.048 ± 0.018	64.5 ± 25.3
SSU rDNA*	1,755	299	198	0.049 ± 0.020	71.1 ± 29.3
V4 SSU rDNA	397	104	75	0.077 ± 0.028	30.4 ± 10.9
Data set B					
*cox*1-5P	660	203	169	0.082 ± 0.029	51.9 ± 18.2
D1-D3 LSU rDNA	866	167	143	0.062 ± 0.030	39.3 ± 17.4
D1-D3 LSU rDNA*	940	190	163	0.065 ± 0.030	44.5 ± 19.6

pair-wise genetic distances of one marker were plotted against pair-wise genetic distances of another marker. The most variable markers (D1-D3 LSU rDNA and *cox1*-5P) were selected for subsequent analyses as they have the highest probability of reflecting intraspecific variability which is necessary for the automated molecular species delimitation methods.

Data set B: the Pinnularia subgibba *group.* Samples for live diatom material originating from littoral zones of several water bodies from various locations worldwide (Greenland, Scotland, Belgium, the Netherlands, Czech Republic, Chile, Marion Island, Île de la Possession, Réunion, and Amsterdam Island) were collected between 2007 and 2013. Upon arrival in the laboratory of Protistology & Aquatic Ecology (PAE), small quantities of the natural material were incubated for a week in WC medium (Guillard and Lorenzen 1972), without pH adjustment or vitamin addition, at 4°C, 20–30 µmol photons $\cdot$ m^{-2} $\cdot$ s^{-1} and a 12:12 h light:dark cycle. Diatom cells were isolated from the natural samples under a stereomicroscope using a needle and a micropipette and grown at PAE in WC medium at standard culture conditions of 18°C, 20–30 µmol photons $\cdot$ m^{-2} $\cdot$ s^{-1}, and a 12:12 h light:dark cycle. The cultures were reinoculated every 2–3 weeks when reaching the late exponential phase. Since long-term preservation was not successful, all newly established cultures are no longer available.

Monoclonal diatom cultures were harvested in late exponential phase and centrifuged prior to DNA extraction. DNA was extracted following Zwart et al. (1998) without the last purification step. The D1-D2 and D2-D3 subregions of LSU rDNA were amplified separately using standard primers (Table S3 in the Supporting Information). For all genes, the PCR reaction mixtures contained: 5 µL of 10× PCR buffer (Tris-HCl, $(NH_4)_2SO_4$, KCl, 15 mM $MgCl_2$, pH 8.7 at 20°C; "Buffer I"; Applied Biosystems, Foster City, CA, USA), 5 µL of deoxynucleoside triphosphates at a concentration of 2 mM each, 4 µL of each primer at 5 µM, 2 µL of BSA at 0.4 mg $\cdot$ mL^{-1}, 2 µL of Taq polymerase at 1 U $\cdot$ μL^{-1} (AmpliTaq, Perkin-Elmer, Wellesley, MA, USA), 2 µL of template DNA, and 26 µL of sterile water. The total reaction volume was 50 µL. For amplification of LSU rDNA, the following protocol 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 a final elongation step for 10 min at 72°C. For amplification of *cox*1-5P, the following protocol 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 a final elongation step for 5 min at 72°C. PCR success was checked using gel electrophoresis on a 1.5% (w/v) agarose gel. Amplification of *cox*1-5P sequences of 13 strains repeatedly failed (indicated in Table S2). PCR products were purified and sequenced by MACROGEN, Inc. (http://www.dna.macrogen.com). Sequences were assembled using BioNumerics v.3.5 (Applied Maths, Kortrijk, Belgium). All extracted DNA is stored at PAE. All newly obtained sequences, along with additional data (including micrographs, geographical data, etc.), were uploaded to the Barcode of Life Database (BoLD; https://doi.org/10.5883/ds-pin2) and GenBank (Table S2). Next to the newly established cultures, all available LSU rDNA *Pinnularia* sequences were downloaded from GenBank resulting in an overall data set of 246 *Pinnularia* strains.

All *Pinnularia* sequences were aligned according to the methods outlined above, using MEGA6 and the Gblocks Server 0.91b (see Table 1 for more details). In order to select all strains belonging to the *Pinnularia subgibba* group, a guide maximum likelihood (ML) phylogeny based on LSU rDNA was inferred using the GTR+G+I (Nei and Kumar 2000) substitution model as inferred by the Bayesian information criterion (BIC) in MEGA6. This revealed the 105 strains belonging to the *P. subgibba* group. All 105 LSU rDNA sequences belonging to the *P. subgibba* group were examined for the presence of identical sequences. In order to include potential intraspecific variability not reflected in LSU rDNA, we selected three specimens (if available) from each identical LSU rDNA sequence. Identical sequences from strains sampled at a different location and/or time were preferred. This resulted in Data set B, which consisted of 58 strains of the *P. subgibba* group (Table S2). The alignments contained a total of 866 positions for D1-D3 LSU rDNA and 660 positions for *cox*1-5P.

Automated molecular species delimitation. Best fit models of nucleotide substitution were selected for *cox*1-5P and D1-D3 LSU rDNA according to BIC as calculated in MEGA6. Single-gene phylogenies of the *Pinnularia subgibba* group based on D1-D3 LSU rDNA and *cox*1-5P were inferred using ML and Bayesian inference (BI). ML was conducted in MEGA6 using substitution models TN93+G (Tamura and Nei 1993) and GTR+G+I (Nei and Kumar 2000) for *cox*1-5P and D1-D3 LSU rDNA, respectively. An initial tree for the heuristic search was obtained by the Neighbor-Joining method to a matrix of pair-wise distances estimated using the maximum composite

likelihood approach. Less than 50% alignment gaps, missing data, and ambiguous characters were allowed at any position. The proportion of trees in which the associated taxa clustered together was calculated using bootstrap with 1,000 generations. Bayesian inference was conducted in MrBayes v.3.2.6 (Ronquist et al. 2012). Nucleotide substitution models HKY+G (Hasegawa et al. 1985) and GTR+G+I were defined for *cox*1-5P and D1-D3 LSU rDNA, respectively. Two independent runs of Markov chain Monte Carlo (MCMC) analysis were performed for each marker separately for 1,000,000 generations. One cold and three heated chains were run for each MCMC analysis. Diagnostic frequency was defined to 1,000 generations and the Markov chain was sampled every 100th generation. Convergence was checked in Tracer v.1.6.0 (Rambaut et al. 2014). Twenty-five percent of the samples were discarded as burn-in.

Molecular species delimitation was performed for the two markers separately using three methods: (i) statistical parsimony network estimation (SPNE), (ii) single threshold generalized mixed Yule-coalescent (sGMYC), and (iii) Maximum likelihood Poissson tree processes (PTP). SPNE (Clement et al. 2000, Hart and Sunday 2007) divides a haplotype network in different statistically independent networks which are considered to correspond with species boundaries. SPNE was conducted in TCS v.1.21 (Clement et al. 2000). Gaps were treated as a fifth state and connection limit (default 95%) corresponded with 11 and 12 detectable mutations for *cox*1-5P and D1-D3 LSU rDNA, respectively. The sGMYC (Pons et al. 2006, Monaghan et al. 2009, Fujisawa and Barraclough 2013) detects an increase in branching rate at the transition between interspecific and intraspecific variation, which is considered to correspond with species limits. The ultrametric input trees were inferred using MrBayes v.3.2.6 under the same settings as BI plus defining a strict molecular clock model. sGMYC was then performed in R v.3.4.3 (R Core Team 2016) using package splits (Ezard et al. 2017), ape (Paradis et al. 2004), MASS (Venables and Ripley 2002), and paran (Dinno 2012). The PTP models branching events are based on the number of substitutions represented by the branch length on an input phylogenetic tree (Zhang et al. 2013). It was performed using the species delimitation web server (species.h-its.org) under default settings.

Morphological observations. All 58 strains of Data set B were studied by LM under oil immersion in a Zeiss Axiophot 2 with objective Plan-Apochromat at 100×/1.4 N.A. using DIC, and a Zeiss Primo Star light microscope with objective Plan-Apochromat 100×/1.25 N.A. Micrographs were taken using an AxioCam MRm camera. Length, width, and stria density of 10 valves per strain were measured and mean ± standard deviations were calculated (Table S4 in the Supporting Information). Additionally, valve outline, orientation of striae, characteristics of the central and axial areas, fasciae, and raphe features were examined. The terminology used in this study is based on Round et al. (1990) and Krammer (2000). For some strains, data on ecology and reproductive experiments were available. All oxidized material and voucher slides are stored at PAE.

RESULTS

Variability of molecular markers. For a subset of 24 *Pinnularia* strains (Dataset A, Table S1) genetic distances (p-distance, Fig. 2, a–f and number of differences, Fig. 2, g–l) were used to compare the sequence variability of seven molecular markers (mitochondrial *cox*1-5P, plastid *rbc*L, *rbc*L-3P, and *psb*A, and nuclear encoded D1-D2 LSU rDNA, SSU rDNA, and V4 SSU rDNA; Table 1). According to the p-distances, D1-D2 LSU rDNA was the most variable marker, followed by *cox*1-5P, V4 SSU rDNA, *rbc*L-3P, SSU rDNA, and *rbc*L, with *psb*A being the most conserved. According to the number of differences between sequences, D1-D2 LSU rDNA was the most variable followed by *cox*1-5P, SSU rDNA, *rbc*L, *rbc*L-3P, V4 SSU rDNA, and *psb*A. Within the D1-D3 region of LSU rDNA, the D2 subregion was found to be the most variable, followed by D1, with D3 being the most conserved (Fig. S1 in the Supporting Information) based on a subset of 40 strains of Data set B (Table S2). Altogether, *cox*1-5P and D1-D3 LSU rDNA were selected as the two most variable markers, most suitable for species delimitation in the *P. subgibba* group.

Phylogenetic trees. Single-locus ML (Figs. S2 and S3 in the Supporting Information) and BI (not shown) phylogenies based on *cox*1-5P and D1-D3 LSU rDNA were not identical in topology. On the other hand, all four phylogenies included clades which were later named A-K in Figure 3. Bootstrap supports (BS) for these clades were 59%–100% and 99%–100% based on D1-D3 LSU rDNA and *cox*1-5P, respectively. Posterior probabilities (PP) for these clades were 92%–100% and 100% based on D1-D3 LSU rDNA and *cox*1-5P, respectively. Moreover, both ML and BI phylogenies based on D1-D3 LSU rDNA included congruent clusters of these clades, namely GHI and JK, both with maximal BS and PP. Similar (clade H was not present due to the sequence amplification failure) clusters were included also in ML phylogeny based on *cox*1-5P although BS were lower (49% and 38% for GI and JK, respectively). BI phylogeny based on *cox*1-5P included GI (PP 62%) but not JK due to a polytomy. The position of the rest of the clades varied between different phylogenies (for example compare Figs. S2 and S3).

Molecular species delimitation. Automated molecular species delimitation of the *Pinnularia subgibba* group was performed on 58 strains (Data set B, Table S2) using the two most variable molecular markers (*cox*1-5P and D1-D3 LSU rDNA) and three single-locus delimitation methods (sGMYC, PTP, and SPNE). Figure 3 compares the results based on *cox*1-5P and D1-D3 LSU rDNA. For *cox*1-5P, the results based on the three delimitation methods were congruent, delimiting a total of 14 potential species-level lineages. For D1-D3 LSU rDNA, all results obtained by the three methods were congruent with exception of clades G and H (Fig. 3) which were delimited as two lineages by sGMYC and PTP, and as a single lineage by SPNE. Due to failed PCR amplifications, one purported species (clade H) was not present in the *cox*1-5P data set. In all other clades and all singleton strains (i.e., Wiec, Tor7f, CZECH_NOS2_7, and Tor4r), there was a complete congruence in species limits across all methods and both molecular markers. Based on both genetic markers, the monophyly of all clades delimited as species was well supported (Fig. 3).

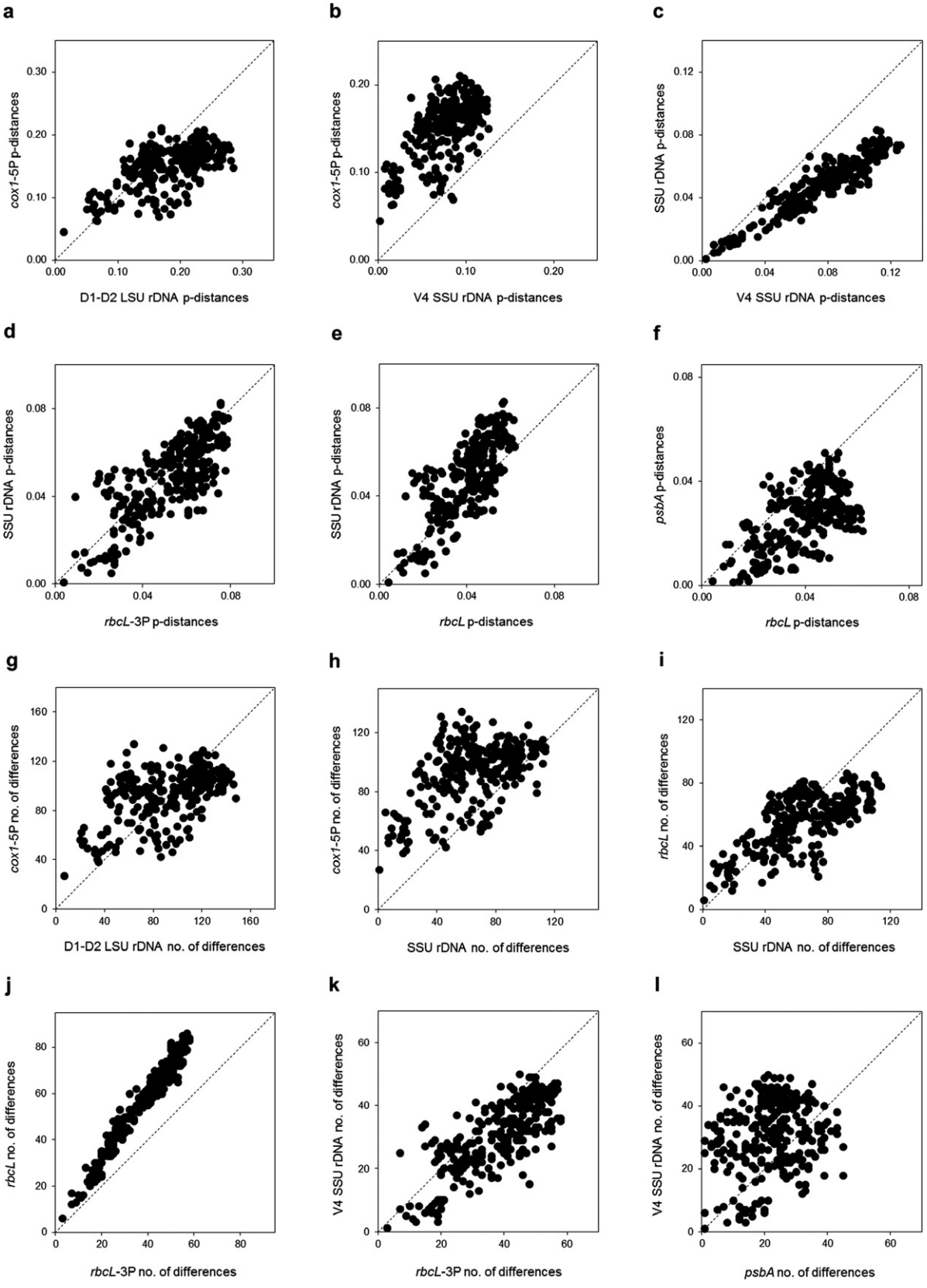


FIG. 2. Pair-wise comparison of genetic distances among all combinations of 24 *Pinnularia* strains (Data set A) for all molecular markers *cox*1-5P, *rbc*L, *rbc*L-3P, *psb*A, D1-D2 LSU rDNA, SSU rDNA, and V4 SSU rDNA. Six comparisons based on p-distances (a–f) and six based on numbers of differences are shown (g-l). The dashed line represents the linear function $x = y$.

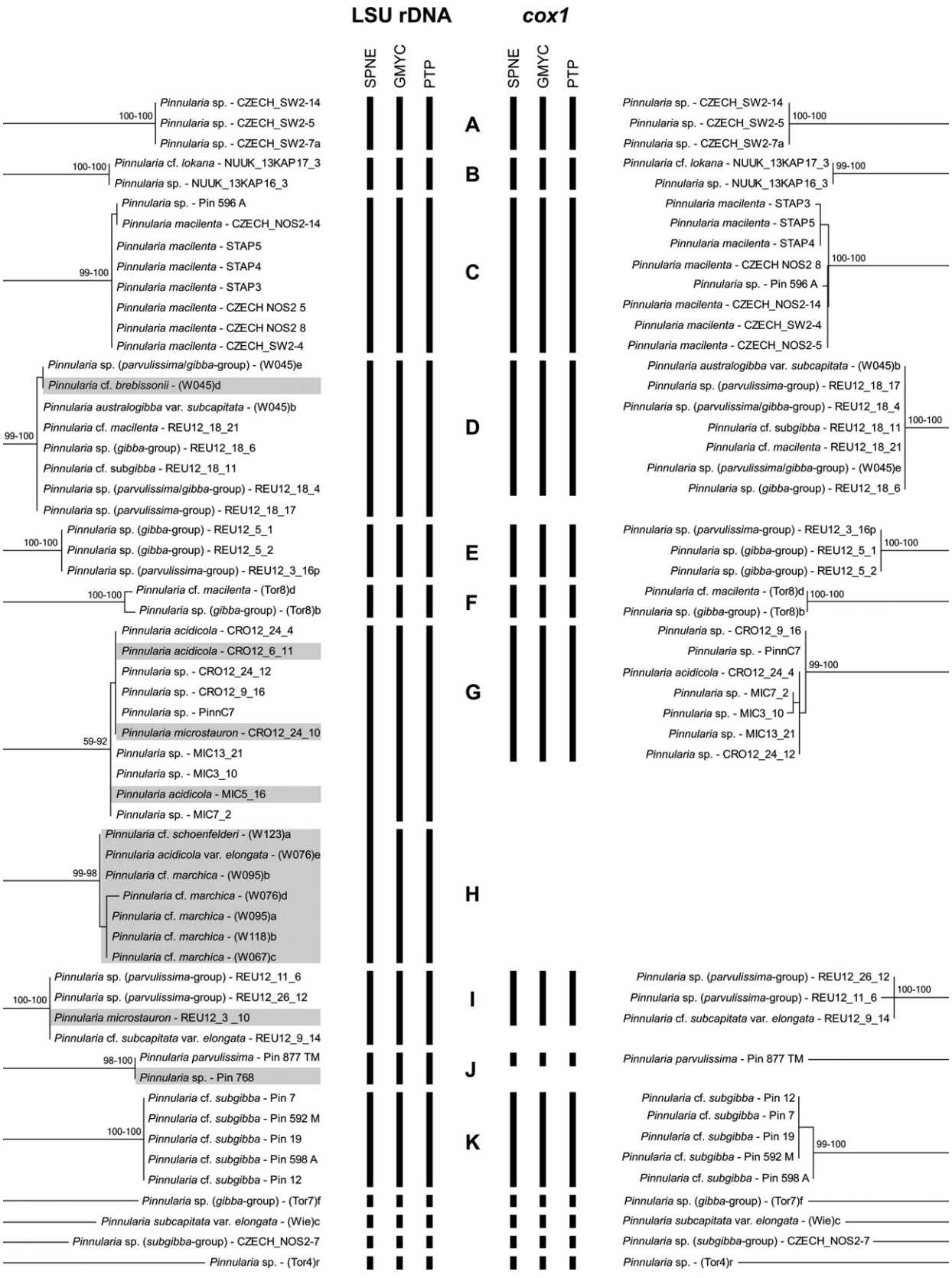


FIG. 3. Comparison of species limits based on *cox*1-5P and LSU rDNA. Species boundaries were delimited using three methods (SPNE, sGMYC, and PTP) as indicated with black bars. Strains for which *cox*1-5P could not be amplified, are indicated in grey in the LSU rDNA data set. The visualization is based on a ML phylogeny. The numbers at the nodes represent ML bootstrap supports – BI posterior probabilities.

Comparison with morphology and other available lines of evidence. The results of the DNA-based species delimitation were compared to other available lines of evidence (Table S5 in the Supporting Information): morphology (Fig. 4; Table S4), geographic origin (Fig. 5; Table S6 in the Supporting Information), and ecology and reproductive compatibility. Table 2 summarizes our final hypothesis on the taxonomical identity of the delimited species.

Clade A was delimited as a single species by both D1-D3 LSU rDNA and *cox*1-5P. All strains belonging to this clade were relatively uniform in morphology, being characterized by linear valves with slightly concave valve outlines in the middle and broadly rounded rostrate apices (Fig. 4a). This clade was identified as *Pinnularia* cf. *microstauron*. All three were isolated from the Swamp area of Lake Mácha (Czech Republic). The type population of *P. microstauron* was sampled from soil on roots of plants in South America and the morphospecies has been recorded worldwide (e.g., Hartley 1986, Day et al. 1995, Edlund et al. 2001, Kociolek 2005, Eskinazi-Leça et al. 2010, Smith et al. 2015).

Clade B was delimited as a single species by both markers. All strains were characterized by robust valves with parallel or almost parallel oriented striae and a relatively small fascia. However, the valve outline was different in the two strains (Fig. 4b). Strain NUUK_13KAP17_3 was identified as *Pinnularia* cf. *lokana*, whereas strain NUUK13_KAP16_3 was identified only to the genus level. Both strains were sampled near the Kapisillit settlement (Greenland). The morphospecies *Pinnularia lokana* was described from fossil sediments of Lokabad (Sweden; Krammer 2000). Since then it was found in the quaternary sediments from Velanská cesta, a former lake in the Czech Republic (Bešta et al. 2009), Great Xing'an Mountains in China (Liu et al. 2018) and possibly also in Russia (Medvedeva and Nikulina 2014).

Clade C was delimited as a single species by both markers. All strains were very distinct in their morphology, comprising of large cells with a distinctive valve outline (convex in the middle of the valve and slightly concave toward the apices) and a very small or absent fascia (Fig. 4c). The exception was small specimens of strain PIN596A (Fig. $4c_1$). The strains resembled *Pinnularia macilenta* sensu Ehrenberg (1843) or *P. gibba* sensu Hustedt (1930). They were sampled in Loch Achray in Scotland (PIN596A), Lake Mácha in the Czech Republic (CZECH_NOS2_5, 8, 14, and CZECH_SW2_4), and Stappersven in Belgium (STAP3-5). *Pinnularia macilenta* has been described from South America (Ehrenberg 1843) and observed in Europe (e.g., Hartley 1986, Veen et al. 2015), Asia (e.g., Hirano 1969, Medvedeva and Nikulina 2014), North America (e.g., Kociolek 2005, Johansen et al. 2007), South America (e.g., Montoya-Moreno et al. 2013), and New Zealand (e.g., Harper et al. 2012).

Clades D (Fig. 4d), E (Fig. 4e), F (Fig. 4f), G (Fig. 4g), H (Fig. 4h), I (Fig. 4i), J (Fig. 4j), and K (Fig. 4k), and singleton strains Tor7f (Fig. $4s_1$), CZECH_NOS2_7 (Fig. $4s_2$) and Tor4r (Fig. $4s_3$) were very uniform in their morphology (especially the small specimens), resembling *Pinnularia gibba* sensu Krammer and Lange-Bertalot, *P. subgibba*, and *P. parvulissima*. All three morphospecies were considered to be cosmopolitan (Foged 1981, Krammer 2000, Edlund et al. 2001, Kociolek 2005, Eskinazi-Leça et al. 2010, Montoya-Moreno et al. 2013, Medvedeva and Nikulina 2014, Smith et al. 2015, Veen et al. 2015). However, after detailed morphometric examination in LM, it seemed that at least three morphogroups could be distinguished within these clades using a combination of characters such as valve outline, fascia morphology, and/or length-to-width ratio. These three morphogroups did not correspond to the three morphospecies mentioned above and corresponded to clades J and K (morphogroup 1), G, H, and I (morphogroup 2), and D, E, and F (morphogroup 3). The latter was the most variable in its morphology (especially the fasciae). The first (JK) and the second morphogroup (GHI) were monophyletic and statistically well supported according to the ML and BI phylogenies based on LSU rDNA (BS-PP supports 100-100; Fig. S2) and they were monophyletic also in the phylogenies based on *cox*1-5P, even though statistical support was lower (Fig. S3). On the other hand, the last morphogroup (DEF) was not monophyletic in either of the phylogenetic trees.

The JK morphogroup (Fig. 4, j and k) could be distinguished from the other two by relatively small fascia covering between 7% and 8% of the valve length. The valves were relatively large with average lengths and widths equalling 51.2–61.4 μm, and 8.8–10.1 μm, respectively, with a length-to-width ratio of 5.8–6.8 μm. The valve apices were broadly rounded or almost indistinguishably rostrate (Fig. 4j). Based on morphology, all strains of clade K were identified as *Pinnularia* cf. *subgibba*, while the strains of clade J were identified as *P.* cf. *parvulissima* (PIN877TM) or *Pinnularia* sp. (PIN768). The strains were isolated from the Scottish lakes Achray (PIN598A), Craiglush (PIN7Cra and PIN19Cra), Menteith (PIN592M), and Threipmuir reservoir (PIN 877TM) and were all epipelic (living in fine bottom mud).

Members of the GHI morphogroup (Fig. 4, g–i) were characterized by slightly convex valves with slightly protracted ends (especially in large specimens) and a large fascia which was always present and covered between 9% and 24% of the valve length. Difference in fascia coverage could be found in each of the three clades: it covered between 9% and 12% of the valve length in clade H (11% on average), between 17% and 24% in clade I (20% on average), and between 16% and 21% in clade G (15% on average). This suggested that members of clade H could be distinguished from members of clades I and G by a considerably smaller fascia.

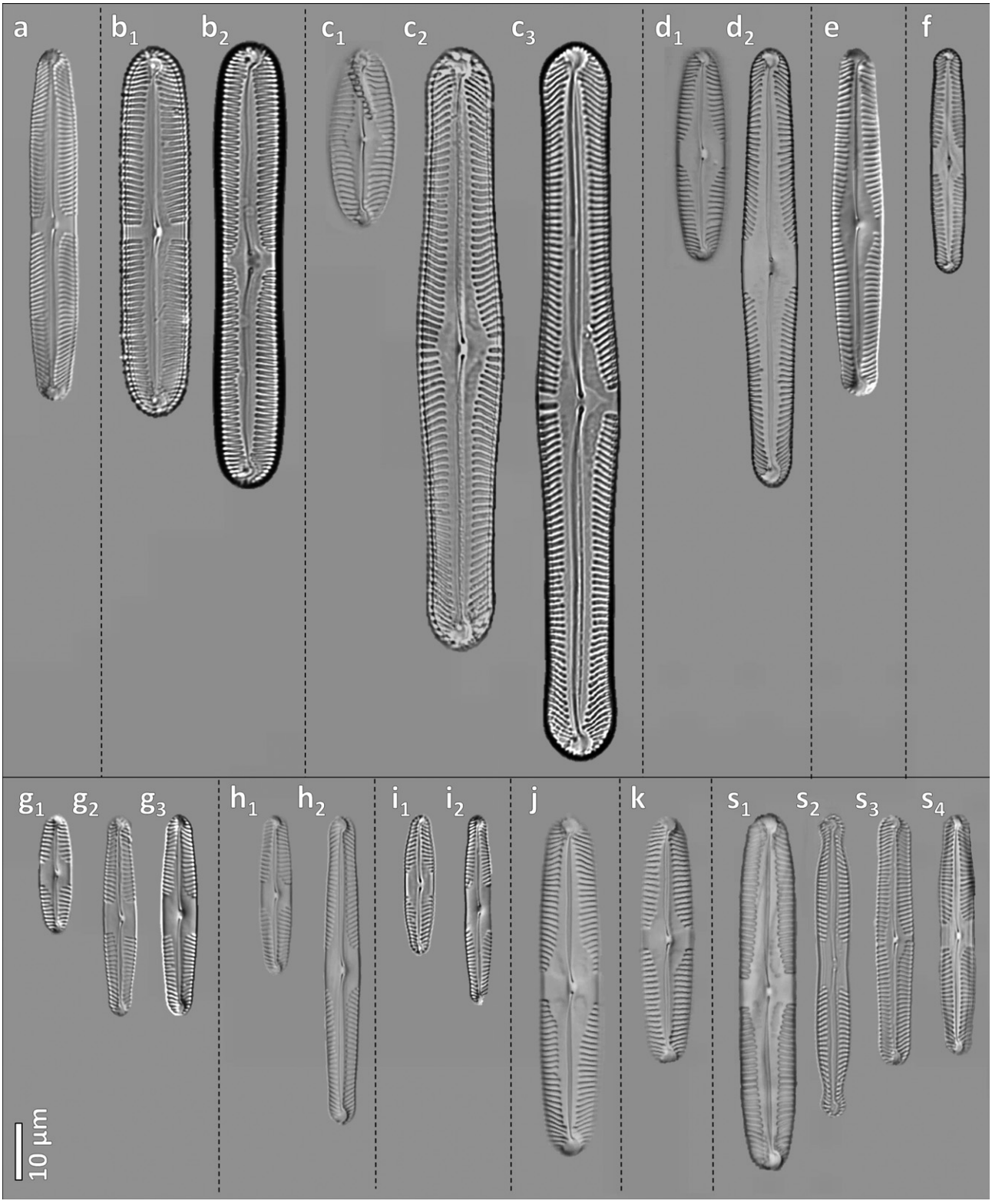


FIG. 4. Morphological variability of clades within the *Pinnularia subgibba* group as delimited by both *cox*1-5P and LSU rDNA. Representatives are shown for each clade (a–k) and singleton strain (s). Clade A is represented by strain CZECH_SW2_7a (a); clade B by NUUK_13-KAP17_3 (b_1), and NUUK_13KAP16_3 (b_2); clade C by PIN596A (c_1), CZECH_NOS2_5 (c_2), and STAP4 (c_3); clade D by W045d (d_1), and W045e (d_2); clade E by REU12_5_1 (e); clade F by Tor8b (f); clade G by MIC7_2 (g_1), CRO12_9_16 (g_2), and MIC5_16 (g_3); clade H by W076d (h_1), and W076e (h_2); clade I by REU12_3_10 (i_1), and REU12_9_14 (i_2); clade J by PIN877TM (j); clade K by PIN19Cra (k); singleton strains are Tor7f (s_1), Wiec (s_2), CZECH_NOS2_7 (s_3), and Tor4r (s_4). More microphotographs may be found in the BoLD.

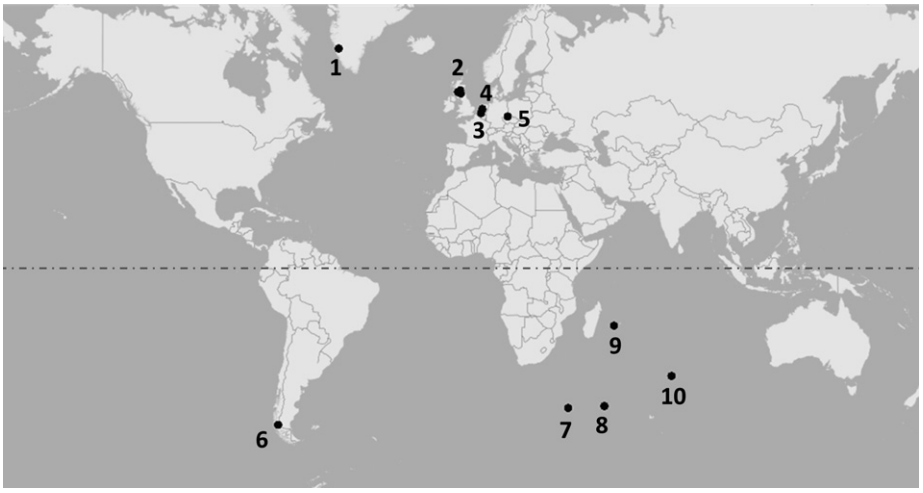


FIG. 5. Geographical origin of the 58 strains of the *Pinnularia subgibba* group. (1) Greenland – Kapisilit, (2) Scotland – Lake Menteith, Loch Achray, Loch of Craiglush, and Threipmuir reservoir, (3) Belgium – Stappersven, (4) the Netherlands – De Wieden, (5) Czech Republic – Lake Mácha, (6) Chile – Torres del Paine, (7) Prince Edward Islands – Marion Island, (8) Crozet Islands – Île de la Possession, (9) Réunion, (10) Amsterdam Island.

TABLE 2. A hypothesis on the taxonomic identity of each delimited species. Members of *Pinnularia gibba-subgibba-parvulissima* species complex are indicated by an asterisk.

Species	Taxon
Clade A	*P.* cf. *microstauron* (Ehrenberg) Cleve
Clade B	*P.* cf. *lokana* Krammer
Clade C	*P. macilenta* Ehrenberg
Clade D*	*P.* cf. *australogibba* Chattová, Van de Vijver & Metzeltin *P.* cf. *amsterdamensis* Chattová, Van de Vijver & Metzeltin
Clade E*	*P.* cf. *vixconspicua* Chattová, Van de Vijver & Metzeltin
Clade F*	*Pinnularia* sp.
Clade G*	*P. acidicola* var. *acidicola* B. van de Vijver & R. Le Cohu
Clade H*	*P. acidicola* var. *elongata* B. van de Vijver & R. Le Cohu
Clade I*	*P.* cf. *acidicola* B. van de Vijver & R. Le Cohu
Clade J*	*P.* cf. *parvulissima* Krammer
Clade K*	*P.* cf. *subgibba* Krammer
Strain Tor7f*	*Pinnularia* sp.
Strain Wiec	*P. subcapitata* var. *elongata* Krammer
Strain CZECH_NOS2_7*	*Pinnularia* sp.
Strain Tor4r*	*Pinnularia* sp.

Morphometric measurements (Table S4) indicated that none of the clades of GHI could be reliably distinguished by length, width, length-to-width ratio or stria density. Large specimens of clade I (e.g., strain REU12_9_14), however, might be distinguished by a combination of a concave valve outline in the center of the valve and a large fascia. All strains of this clade were isolated from the island of Réunion (Indian Ocean) and were delimited as a separate monophyletic species by both D1-D3 LSU rDNA and *cox*1-5P. Strains of clade H were isolated from the Amsterdam Island (Southern Indian Ocean) and represented a single species according to D1-D3 LSU rDNA (*cox*1-5P failed to be amplified). Most strains were identified as *Pinnularia* cf. *marchica*. However, one large specimen (W076e) was identified as *P. acidicola* var. *elongata*. The type localities of *P. marchica* and *P. acidicola* var. *elongata* were in Germany and Île de la Possession (Crozet Archipelago, Southern Indian Ocean), respectively. Morphospecies *P. marchica* was found throughout Europe (Souffreau et al. 2010, Veen et al. 2015) and North America (Bahls 2009), while *P. acidicola* var. *elongata* has so far only been reported from the Île de la Possession (Van de Vijver et al. 2002) and North America (Johansen et al. 2007). Strains of clade G were isolated from Marion Island (Southern Indian Ocean) and Île de la Possession and identified as *P. acidicola* var. *acidicola* (Van de Vijver et al. 2002).

The last morphogroup (DEF) was the most variable in morphology and aphyletic (i.e., in need of further investigation; Williams and Ebach 2017) according to the ML and BI phylogenies of both D1-D3 LSU rDNA and *cox*1-5P. The valve structure resembled morphogroup GHI rather than morphogroup JK, while cell size was more similar to the latter. The fascia was absent or very small in clade E, covering on maximum 4% of the valve length. In clades D and F, on the other hand, the fascia was always relatively large, covering on average 12% of the valve length. In clade D, the axial area distinctly widened toward the large rhomboid fascia. Strains of clade D were isolated from Amsterdam Island and Réunion. Although the axial area of the small specimens suggested that strains of clade D might be closely related to *Pinnularia australogibba* described from Amsterdam Island, valve dimensions and outlines of the majority of the strains differed.

On the other hand, the large specimens strongly resembled *Pinnularia amsterdamensis*, which has also been described from the Amsterdam Island. Members of clade E might be distinguished from the members of clade D by slightly more parallel valve outlines, a very small or absent fascia and a considerably smaller axial area. There was some resemblance to *Pinnularia vixconspicua*. All strains of clade E were isolated from Réunion, whereas the type locality of *P. vixconspicua* was on the Amsterdam Island. Clade F possessed the typical morphology of the *P. gibba-subgibba-parvulissima* complex. Strains of clade F were isolated from Chile.

Singleton strain Wiec (Fig. $4s_4$) could be clearly identified as *Pinnularia subcapitata* var. *elongata* and originated from De Wieden (the Netherlands). The type locality was in Belgium (Krammer 1992) and the taxon has been observed throughout Europe (Veen et al. 2015), South America (Eskinazi-Leça et al. 2010), Asia (Medvedeva and Nikulina 2014), and New Zealand (Harper et al. 2012).

DISCUSSION

In this study, the species limits within the *Pinnularia subgibba* group were investigated using molecular and morphological data. In order to select molecular markers appropriate for species delimitation in the genus *Pinnularia*, and the *P. subgibba* group in particular, a set of nuclear, chloroplast, and mitochondrial markers were examined for their sequence variability. This analysis revealed D1-D2 LSU rDNA to be the most variable genetic marker in this study followed by *cox*1-5P. To date, the complete or partial LSU rDNA have been used in investigations of various taxonomic groups including algae (Daugbjerg et al. 2000, Saunders and Lehmkuhl 2005, Linton et al. 2010), plants (Qiu et al. 2007), fungi (James et al. 2006, Mapook et al. 2016), and animals (Subbotin et al. 2005, Uribe et al. 2016). Also in diatoms, the gene has been widely used for both phylogeny inference (Alverson et al. 2007, Trobajo et al. 2009, Tan et al. 2015), and investigation of species diversity (Sarno et al. 2005, Poulíčková et al. 2010, Kistenich et al. 2014). Specifically in *Pinnularia*, it was used for phylogenetic inferences of the entire genus, including *Caloneis* (Bruder et al. 2008, Souffreau et al. 2011), and phylogenetics and species delimitation of the *Pinnularia borealis* species complex (Souffreau et al. 2013, Pinseel et al. 2017b). Similarly, *cox*1-5P has been used for both phylogeny inference of both *Pinnularia* sensu lato (Souffreau et al. 2011) and *Skeletonema* (Yamada et al. 2017) and the investigation of diatom species diversity in relation to DNA barcoding (Trobajo et al. 2010, Hamsher et al. 2011). Trobajo et al. (2010) compared *cox*1-5P, D1-D2 LSU rDNA, and *rbc*L for the *Nitzschia palea* species complex and found that *cox*1-5P is the most variable marker, followed by D1-D2 LSU rDNA and *rbc*L. Similarly, Hamsher et al. (2011), who compared *cox*1-5P, D2-D3 LSU rDNA, *rbc*L, *rbc*L-3P, and UPA for six diatom genera found *cox*1-5P to be the most variable, followed by *rbc*L, *rbc*L-3P, and D2-D3 LSU rDNA. Surprisingly, in our study both the complete and partial *rbc*L was found to be relatively conserved in *Pinnularia*, which is in contradiction with studies based on other diatom genera (Trobajo et al. 2010, Hamsher et al. 2011, Urbánková and Veselá 2013) and possibly even with the subsequent proposal of *rbc*L as a primary barcode marker for diatoms (e.g., Mann et al. 2010). Whether the taxonomical resolution of *rbc*L is sufficiently high to reflect species level (for DNA Barcoding) or even intraspecific level (for species delimitation) structure within *Pinnularia* remains unknown and should be further investigated. The same may be applicable in the case of SSU rDNA, which has also been suggested as one of the barcode markers for diatoms (Zimmermann et al. 2011). Previously, Pinseel et al. (2017b) already questioned the usefulness of SSU rDNA for the delimitation of diatom species based on the observation of low interspecific molecular variation in the *Achnanthidium minutissimum* complex. Another molecular marker with intraspecific resolution in diatoms is the Internal Transcribed Spacer (ITS rDNA). In the genus *Pseudo-nitzschia* the variation in ITS1-5.8S-ITS2 rDNA was correlated with the variation in morphology (Casteleyn et al. 2008) and ITS2 rDNA even discriminated reproductively isolated groups (Amato et al. 2007). However, its amplification in the *Pinnularia* is difficult due to the presence of extensive intragenomic variation preventing routine PCR amplifications without the use of clone libraries (E. Pinseel and C. Souffreau, unpub. data).

Automated molecular species delimitations based on the two most variable markers in this study, LSU rDNA and *cox*1-5P, suggested the presence of 15 species belonging to the *Pinnularia subgibba* group in our data set and showed complete congruence between *cox*1-5P and LSU rDNA. We confronted our primary hypothesis with other lines of evidence. Whereas, it is relatively easy to distinguish some of the potential species (clades A, B, C) based on morphological characteristics, it is not straightforward to distinguish others (clades D, E, F, G, H, I, J, and K, and singleton strains), especially the small specimens. It seems that different morphologies of large cells of different clades tend to converge to a very similar morphology when they become smaller and, therefore, may no longer be unambiguously identified (e.g., Fig. 4, g–i). Nevertheless, the relative size of the fascia seems to be promising for the recognition of both higher natural species clusters (e.g., JK and GHI) and possibly even some species within these clusters (e.g., H) because it seems to be more stable for both small and large specimens compared to valve outline or length-to-width ratio. We found no clear utility of the indistinct markings accompanying

the fascia (e.g., Fig. 4j) for species-level identifications. Nevertheless, these markings may still be helpful for the correct assignment of specimens to the *Pinnularia subgibba* group using LM as suggested by Krammer (2000). In any case, it seems that morphology alone lacks the resolution necessary to unambiguously delimit species within this group (e.g., G and I) and must be accompanied by genetic evidence. Moreover, evidence based on other diatom taxa (Sarno et al. 2005, Evans et al. 2007, Vanormelingen et al. 2008a, Trobajo et al. 2009, Vanelslander et al. 2009, Poulíčková et al. 2010, Souffreau et al. 2013, Pinseel et al. 2017a) confirms that (pseudo)cryptic diversity is rather widespread among diatoms. Therefore, using morphology as the sole line of evidence for species delimitations and consequent taxonomic changes may lead to underestimation of diatom species numbers.

The delimited species have largely unknown biogeographies and seem mostly restricted to a single location. The only three species found in multiple regions were D (found on Amsterdam Island and the island of Réunion; several thousands of kilometers apart), G (found on Marion Island and Île de la Possession; thousand kilometers apart), and C (found throughout Europe; hundreds of kilometers apart), none of which was shared between hemispheres. Several of our species are geographically very distant from the sampling localities of the presumed corresponding type specimens (e.g., species A – *Pinnularia* cf. *microstauron*; Czech Republic vs. Brazil), suggesting that these taxa either have very wide geographic distributions, or that they do not correspond to single species, but rather could represent members of (pseudo)cryptic species complexes with restricted geographic distributions. It is recently becoming clear that many former cosmopolitan diatom taxa actually show hidden diversity with their members showing (relatively) restricted geographic distributions (Vanormelingen et al. 2008b), and/or niche-differentiation (Vanelslander et al. 2009, Souffreau et al. 2013). However, it is clear that it will be necessary to significantly increase the sample coverage in order to study the presence and nature of potential biogeographical patterns in the *P. subgibba* group.

Although our polyphasic framework of species delimitation proved to be feasible in *Pinnularia subgibba* group, further sampling, combining morphological and molecular investigations, preferably in the presence of mating experiments and ecophysiological experiments for niche-differentiation, are needed to review the identity of the specimens in this study and to reveal whether they truly represent cosmopolitan taxa, endemics, or something in-between. Such an in-depth study will be invaluable to a taxonomic revision of the *P. subgibba* group. In this light, inclusion of molecular data, preferably from fast evolving markers such as LSU rDNA and *cox*1 is invaluable.

This research was financially supported by the Endowment Fund of Palacký University in Olomouc (JK), the Internal grant agency of the Faculty of Science at Palacký University in Olomouc IGA Prf-2018-001 and Prf-2019-004 (JK, AP and PD), and the Fund for Scientific Research FWO-Flanders (funding of EP and CS as PhD students and PV as a postdoctoral fellow). Sampling on Crozet (Île de la Possession) has been made possible thanks to the logistic and financial support of the French Polar Institute-Paul-Emile Victor in the framework of the terrestrial program 136 (Ir. Marc Lebouvier & Dr Yves Frenot) and the Belspo CCAMBIO project (SD/BA/03). Sampling on Marion Island was funded by the Belspo CCAMBIO project (SD/BA03) and FWO-Flanders, upon invitation of Dr. Steven Chown. Sampling in Greenland was made possible thanks to the EU-INTERACT project Mibipol. We thank Dr. Elie Verleyen, Dr. Wim Van Nieuwenhuyze, Dr. Bart Van de Vijver, Dr. Koen Sabbe, Dr. Jana Kulichová, Pavla Urbánková, and Dagmar Obbels for sampling. We thank Dr. David G. Mann for providing isolates of Scottish strains. Part of the culture work was performed under facilities of the BCCM/DCG diatom culture collection (Belspo). We thank Dr. Bart Van de Vijver for the morphological identifications of many strains, Dr. Koen Sabbe and Dr. Elie Verleyen for advice and constructive critique of the research, and Sofie D'hondt for advice on molecular techniques. The authors declare no competing financial interests.

Alverson, A. J., Jansen, R. K. & Theriot, E. C. 2007. Bridging the Rubicon: phylogenetic analysis reveals repeated colonizations of marine and fresh waters by thalassiosiroid diatoms. *Mol. Phylogenet. Evol.* 45:193–210.

Amato, A., Kooistra, W. H. C. F., Levialdi Ghiron, J. H., Mann, D. G., Pröschold, T. & Montresor, M. 2007. Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist* 158:193–207.

Bahls, L. L. 2009. A checklist of diatoms from inland waters of the northwestern United States. *P. Acad. Nat. Sci. Phila.* 158:1–35.

Bešta, T., Šafránková, J., Pouzar, M., Novák, J. & Nováková, K. 2009. Late Pleistocene-early Holocene transition recorded in the sediments of a former shallow lake in the Czech Republic. *Hydrobiologia* 631:107–20.

Blair, C. & Bryson, R. W. 2017. Cryptic diversity and discordance in single-locus species delimitation methods within horned lizards (Phrynosomatidae: Phrynosoma). *Mol. Ecol. Resour.* 17:1168–82.

Bruder, K., Sato, S. & Medlin, L. K. 2008. Morphological and molecular investigations of naviculoid diatoms IV. *Pinnularia* vs. *Caloneis*. *Diatom* 24:8–24.

Casteleyn, G., Chepurnov, V. A., Leliaert, F., Mann, D. G., Bates, S. S., Lundholm, N., Rhodes, L., Sabbe, K. & Vyverman, W. 2008. *Pseudo-nitzschia pungens* (Bacillariophyceae): a cosmopolitan diatom species? *Harmful Algae* 7:241–57.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17:540–52.

Chase, M. W., Salamin, N., Wilkinson, M., Dunwell, J. M., Kesanakurthi, R. P., Haidar, N. & Savolainen, V. 2005. Land plants and DNA barcodes: short-term and long-term goals. *Philos. Trans. R. Soc. B Biol. Sci.* 360:1889–95.

Clement, M., Posada, D. & Crandall, K. A. 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9:1657–9.

Colwell, R. R. 1970. Polyphasic taxonomy of the genus *Vibrio*: numerical taxonomy of *Vibrio cholerae, Vibrio parahaemolyticus*, and related *Vibrio* species. *J. Bacteriol.* 104:410–33.

Daugbjerg, N., Hansen, G., Larsen, J. & Moestrup, O. 2000. Phylogeny of some of the major genera of dinoflagellates based on ultrastructure and partial LSU rDNA sequence data, including the erection of three new genera of unarmoured dinoflagellates. *Phycologia* 39:302–17.

Day, S. A., Wiskham, R. P., Entwisle, T. J. & Tyler, P. A. 1995. *Bibliographic Checklist of Non-marine Algae in Australia*. Australian Biological Resources Study, Canberra, Australia, 276 pp.

De Queiroz, K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56:879–86.

Dinno, A. 2012. paran: Horn's Test of Principal Components/ Factors. Available at: https://cran.r-project.org/package=paran.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with improved accuracy and speed. *Nucleic Acids Res.* 32:1792–7.

Edlund, M. B., Soninkhishig, N., Williams, R. M. & Stoermer, E. F. 2001. Biodiversity of Mongolia: checklist of diatoms, including new distributional reports of 31 taxae. *Nov. Hedwigia* 72:59–90.

Ehrenberg, C. G. 1843. Verbreitung und Einfluss des mikroskopischen Lebens in Süd-und Nord-Amerika. Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin 1841, 466 pp.

Eskinazi-Leça, E., Cunha, M. G. G. S., Santiago, M. F., Borges, G. C. P., Lima, J. C., Silva, M. H., Lima, E. P. & Menezes, M. 2010. Bacillariophyceae. *In* Forzza, R. C. [Ed.] *Catálogo de Plantas E Fungos Do Brasil*. Andrea Jakobsson Estúdio, Rio de Janeiro, pp. 262–309.

Evans, K. M., Wortley, A. H. & Mann, D. G. 2007. An assessment of potential diatom "barcode" genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). *Protist* 158:349–64.

Ezard, T., Fujisawa, T. & Barraclough, T. G. 2017. splits: Species' LImits by Threshold Statistics. Available at: https://r-forge.r-project.org/projects/splits/.

Foged, N. 1981. Diatoms in Alaska. *Bibl. Phycol.* 53:1–317.

Fujisawa, T. & Barraclough, T. G. 2013. Delimiting species using single-locus data and the generalized mixed yule coalescent approach: a revised method and evaluation on simulated data sets. *Syst. Biol.* 62:707–24.

Funk, D. J. & Omland, K. E. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst.* 34:397–423.

Goldstein, P. Z. & DeSalle, R. 2011. Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. *BioEssays* 33:135–47.

Guillard, R. R. L. & Lorenzen, C. J. 1972. Yellow-green algae with chlorophyllide c. *J. Phycol.* 8:10–4.

Guiry, M. D. 2012. How many species of algae are there? *J. Phycol.* 48:1057–63.

Guiry, M. D. & Guiry, G. M. 2018. AlgaeBase. Available at: http://www.algaebase.org (last accessed November 13, 2018).

Hamsher, S. E., Evans, K. M., Mann, D. G., Poulíčková, A. & Saunders, G. W. 2011. Barcoding diatoms: exploring alternatives to COI-5P. *Protist* 162:405–22.

Harper, M. A., Cassie Cooper, V., Chang, F. H., Nelson, W. A. & Broady, P. A. 2012. Phylum Ochrophyta: brown and golden-brown algae, diatoms, silicioflagellates, and kin. *In* Gordon, D. P. [Ed.] *New Zealand Inventory of Biodiversity*. Canterbury University Press, Christchurch, New Zealand, pp. 114–63.

Hart, M. W. & Sunday, J. 2007. Things fall apart: biological species form unconnected parsimony networks. *Biol. Lett.* 3:509–12.

Hartley, B. 1986. A check-list of the freshwater, brackish and marine diatoms of the British Isles and adjoining coastal waters. *J. Mar. Biol. Assoc. U.K.* 66:531–610.

Hasegawa, M., Kishino, K. & Yano, T. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–74.

Hirano, M. 1969. Freshwater algae from Rangtang Himal, Nepal Himalaya. *Contrib. Biol. Lab. Kyoto Univ.* 22:1–42.

Hustedt, F. 1930. *Bacillariophyta (Diatomeae)*. Verlag von Gustav Fischer, Jena, 464 pp.

James, T. Y., Kauff, F., Schoch, C. L., Matheny, P. B., Hoffstetter, V., Cox, C., Celio, G. et al. 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443:818–22.

Johansen, J. R., Lowe, R. L., Carty, S., Fučiková, K., Olsen, C. E., Fitzpatrick, M. H., Ress, J. & Furey, P. C. 2007. New algal species records for Great Smoky Mountains National Park, with an annotated checklist of all reported algal taxa for the park. *Southeast. Nat.* 6:99–134.

Kistenich, S., Dressler, M., Zimmermann, J., Hubener, T., Bastrop, R. & Jahn, R. 2014. An investigation into the morphology and genetics of *Cyclotella comensis* and closely related taxa. *Diatom Res.* 29:423–40.

Knowles, L. L. & Carstens, B. 2007. Delimiting species without monophyletic gene trees. *Syst. Biol.* 56:887–95.

Kociolek, J. P. 2005. A checklist and preliminary bibliography of the recent, freshwater diatoms of inland environments of the continental United States. *Sci. York.* 56:395–525.

Kociolek, J. P., Balasubramanian, K., Blanco, S., Coste, M., Ector, L., Liu, Y., Kulikovskiy, M. et al. 2018. DiatomBase. Pinnularia C.G. Ehrenberg, 1843. Available at: http://www.diatombase.org/aphia.php?p=taxdetails&id=149208 (last accessed May 25, 2018).

Komárek, J. 2016. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. *Eur. J. Phycol.* 51:346–53.

Krammer, K. 1992. *Pinnularia* eine Monographie de europäischen Taxa. *Bibl. Diatomol.* 26:1–353.

Krammer, K. 2000. *The genus Pinnularia*. A.R.G. Ganter Verlag K.G, Ruggell, 703 pp.

Leliaert, F., Verbruggen, H., Vanormelingen, P., Steen, F., López-Bautista, J. M., Zuccarello, G. C. & De Clerck, O. 2014. DNA-based species delimitation in algae. *Eur. J. Phycol.* 49:179–96.

Lewis, A. R., Marchant, D. R., Ashworth, A. C., Hedenäs, L., Hemming, S. R., Johnson, J. V., Leng, M. J. et al. 2008. Mid-Miocene cooling and the extinction of tundra in continental Antarctica. *Proc. Natl. Acad. Sci. USA* 105:10676–80.

Linton, E. W., Karnkowska-ishikawa, A., Kim, J. I. & Shin, W. 2010. Reconstructing Euglenoid evolutionary relationships using three genes: nuclear SSU and LSU, and chloroplast SSU rDNA sequences and the description of *Euglenaria* gen. nov. (Euglenophyta). *Protist* 161:603–19.

Liu, Y., Kociolek, J. P., Wang, Q. X. & Fan, Y. W. 2018. The diatom genus *Pinnularia* from Great Xing'an Mountains, China. *Bibl. Diatomol.* 65:1–298.

Mann, D. G. & Droop, S. J. M. 1996. Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia* 336:19–32.

Mann, D. G., Sato, S., Trobajo, R., Vanormelingen, P. & Souffreau, C. 2010. DNA barcoding for species identification and discovery in diatoms. *Cryptogamie* 31:557–77.

Mann, D. G. & Vanormelingen, P. 2013. An inordinate fondness? The number, distributions, and origins of diatom species. *J. Eukaryot. Microbiol.* 60:414–20.

Mapook, A., Boonmee, S., Liu, J. K., Jones, E. B. G., Bahkali, A. H. & Hyde, K. D. 2016. Taxonomic and phylogenetic placement of *Phaeodimeriella* (Pseudoperisporiaceae, Pleosporales). *Cryptogam., Mycol.* 37:157–76.

Medvedeva, L. A. & Nikulina, T. V. 2014. *Catalogue of Freshwater Algae of the Southern Part of the Russian Far East*. Dalnauka, Vladivostok, 271 pp.

Monaghan, M. T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D. J. G., Lees, D. C., Ranaivosolo, R., Eggleton, P., Barraclough, T. G. & Vogler, A. P. 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Syst. Biol.* 58:298–311.

Montoya-Moreno, Y., Sala, S., Vouilloud, A., Aguirre, N. & Plata-Díaz, Y. 2013. Lista de las diatomeas de ambientes continentales de Colombia. *Biota Colomb.* 14:13–78.

Nei, M. & Kumar, S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford, UK, 333 pp.

Novais, M. H., Jüttner, I., Van de Vijver, B., Morais, M. M., Hoffmann, L. & Ector, L. 2015. Morphological variability within the *Achnanthidium minutissimum* species complex (Bacillariophyta): comparison between the type material of *Achnanthes minutissima* and related taxa, and new freshwater *Achnanthidium* species from Portugal. *Phytotaxa* 224:101–39.

Paradis, E., Claude, J. & Strimmer, K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–90.

Pinseel, E., Hejduková, E., Vanormelingen, P., Kopalová, K., Vyverman, W. & Van de Vijver, B. 2017b. *Pinnularia catenaborealis* sp. nov. (Bacillariophyceae), a unique chain-forming diatom species from James Ross Island and Vega Island (Maritime Antarctica). *Phycologia* 56:94–107.

Pinseel, E., Kulichová, J., Scharfen, V., Urbánková, P., Van de Vijver, B. & Vyverman, W. 2018. Extensive cryptic diversity in the terrestrial diatom *Pinnularia borealis* (Bacillariophyceae). *Protist* (https://doi.org/10.1016/j.protis.2018.10.001).

Pinseel, E., Vanormelingen, P., Hamilton, P. B., Vyverman, W., Van de Vijver, B. & Kopalova, K. 2017a. Molecular and morphological characterization of the *Achnanthidium minutissimum* complex (Bacillariophyta) in Petuniabukta (Spitsbergen, High Arctic) including the description of *A. digitatum* sp. nov. *Eur. J. Phycol.* 52:264–80.

Pons, J., Barraclough, T., Gomez-Zurita, J., Cardoso, A., Duran, D., 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., Veselá, J., Neustupa, J. & Škaloud, P. 2010. Pseudocryptic diversity versus cosmopolitanism in diatoms: a case study on *Navicula cryptocephala* Kütz. (Bacillariophyceae) and morphologically similar taxa. *Protist* 161:353–69.

Puillandre, N., Modica, M. V., Zhang, Y., Sirovich, L., Boisselier, M. C., Cruaud, C., Holford, M. & Samadi, S. 2012. Large-scale species delimitation method for hyperdiverse groups. *Mol. Ecol.* 21:2671–91.

Qiu, Y. L., Li, L., Wang, B., Chen, Z., Dombrovska, O., Lee, J., Kent, L. et al. 2007. A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. *Int. J. Plant Sci.* 168:691–708.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut, A., Suchard, M. A., Xie, D. & Drummond, A. J. 2014. Tracer v1.6. Available at: http://tree.bio.ed.ac.uk/software/tracer.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, 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 model space. *Syst. Biol.* 61:539–42.

Round, F. E., Crawford, R. M. & Mann, D. G. 1990. *The Diatoms: Biology and Morphology of the Genera*. Cambridge University Press, Cambridge, UK, 747 pp.

Sarno, D., Kooistra, W. H. C. F., Medlin, L. K., Percopo, I. & Zingone, A. 2005. Diversity in the genus *Skeletonema* (Bacillariophyceae). II. An assessment of the taxonomy of *S. costatum*-like species with the description of four new species. *J. Phycol.* 41:151–76.

Saunders, G. W. & Lehmkuhl, K. V. 2005. Molecular divergence and morphological diversity among four cryptic species of *Plocamium* (Plocamiales, Florideophyceae) in northern Europe. *Eur. J. Phycol.* 40:293–312.

Smith, T. E., Smith, C. J. & Nii Yemoh Annang, T. 2015. *Taoxnomic Catalogue of Algae from Ghana (Africa) and New Additions*. Algae Press, Ave Maria, 101 pp.

Smol, J. P. & Stoermer, E. F. 2010. *The Diatoms: Applications for the Environmental and Earth Sciences*, 2nd edn. Cambridge University Press, Cambridge, UK, 667 pp.

Souffreau, C., Vanormelingen, P., Van de Vijver, B., Isheva, T., Verleyen, E., Sabbe, K. & Vyverman, W. 2013. Molecular evidence for distinct Antarctic lineages in the cosmopolitan terrestrial diatoms *Pinnularia borealis* and *Hantzschia amphioxys*. *Protist* 164:101–15.

Souffreau, C., Vanormelingen, P., Verleyen, E., Sabbe, K. & Vyverman, W. 2010. Tolerance of benthic diatoms from temperate aquatic and terrestrial habitats to experimental dessication and temperature stress. *Phycologia* 49:309–24.

Souffreau, C., Verbruggen, H., Wolfe, A. P., Vanormelingen, P., Siver, P. A., Cox, E. J., Mann, D. G., Van de Vijver, B. & Vyverman, W. 2011. A time-calibrated multi-gene phylogeny of the diatom genus *Pinnularia*. *Mol. Phylogenet. Evol.* 61:866–79.

Subbotin, S. A., Vovlas, N., Crozzoli, R., Sturhan, D., Lamberti, F., Moens, M. & Baldwin, J. G. 2005. Phylogeny of Criconematina Siddiqi, 1980 (Nematoda: Tylenchida) based on morphology and D2–D3 expansion segments of the 28S-rRNA gene sequences with application of a secondary structure model. *Nematology* 7:927–44.

Tamura, K. & Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10:512–26.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30:2725–9.

Tan, S. N., Lim, H. C., Teng, S. T., Lim, P. T. & Leaw, C. P. 2015. *Pseudo-nitzschia* species (Bacillariophyceae) identification and delineation using mitochondrial *cox*1 gene sequences as compared to LSU rDNA. *Fish. Sci.* 81:831–8.

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* (Bacillariophyceae). *Phycologia* 48:443–59.

Trobajo, R., Mann, D. G., Clavero, E., Evans, K. M., Vanormelingen, P. & McGregor, R. C. 2010. The use of partial *cox*1, *rbc*L and LSU rDNA sequences for phylogenetics and species identification within the *Nitzschia palea* species complex (Bacillariophyceae). *Eur. J. Phycol.* 45:413–25.

Urbánková, P. & Veselá, J. 2013. DNA-barcoding: a case study in the diatom genus *Frustulia* (Bacillariophyceae). *Nov. Hedwigia* 142:147–62.

Uribe, J. E., Colgan, D., Castro, L. R., Kano, Y. & Zardoya, R. 2016. Phylogenetic relationships among superfamilies of Neritimorpha (Mollusca: Gastropoda). *Mol. Phylogenet. Evol.* 104:21–31.

Van de Vijver, B., Frenot, Y. & Beyens, L. 2002. Freshwater diatoms from Île de la Possession (Crozet Archipelago, sub-Antarctica). *Bibl. Diatomol.* 46:1–412.

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. & Swings, J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60:407–38.

Vanelslander, B., Créach, V., Vanormelingen, P., Ernst, A., Chepurnov, V. A., Sahan, E., Muyzer, G., Stal, L. J., Vyverman, W. & Sabbe, K. 2009. Ecological differentiation between sympatric pseudocryptic species in the estuarine benthic diatom *Navicula phyllepta* (Bacillariophyceae). *J. Phycol.* 45:1278–89.

Vanormelingen, P., Chepurnov, V. A., Mann, D. G., Sabbe, K. & Vyverman, W. 2008a. Genetic divergence and reproductive barriers among morphologically heterogeneous sympatric clones of *Eunotia bilunaris* sensu lato (Bacillariophyta). *Protist* 159:73–90.

Vanormelingen, P., Verleyen, E. & Vyverman, W. 2008b. The diversity and distribution of diatoms: from cosmopolitanism to narrow endemism. *Biodivers. Conserv.* 17:393–405.

Veen, A., Hof, C. H. J., Kouwets, F. A. C. & Berkhout, T. 2015. Rijkswaterstaat Waterdienst, Informatiehuis Water. Available at: http://ipt.nlbif.nl/ipt/resource?r=checklist-twn (last accessed May 18, 2017).

Venables, W. N. & Ripley, B. D. 2002. *Modern Applied Statistics with S*, 4th edn. Springer, New York, NY, 498 pp.

Will, K., Mishler, B. & Wheeler, Q. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* 54:844–51.

Williams, D. M. & Ebach, M. C. 2017. Aphyly: identifying the flotsam and jetsam of systematics. *Cladistics* 34:459–66.

Yamada, M., Otsubo, M., Tsutsumi, Y., Mizota, C., Nakamura, Y., Takahashi, K. & Iwataki, M. 2017. Utility of mitochondrial-encoded cytochrome c oxidase I gene for phylogenetic analysis and species identification of the planktonic diatom genus *Skeletonema*. *Phycol. Res.* 65:217–25.

Zhang, J., Kapli, P., Pavlidis, P. & Stamatakis, A. 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29:2869–76.

Zimmermann, J., Glöckner, G., Jahn, R., Enke, N. & Gemeinholzer, B. 2015. Metabarcoding vs. morphological

identification to assess diatom diversity in environmental studies. *Mol. Ecol. Resour.* 15:526–42.

Zimmermann, J., Jahn, R. & Gemeinholzer, B. 2011. Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Org. Divers. Evol.* 11:173–92.

Zwart, G., Huismams, R., Agterveld, V. P. M., Van Peer, Y., De Rijk, P., Eenhoorm, H., Muyzer, G., Hannen, E. J., Gons, J. H. & Laandroek, J. H. 1998. Divergent members of the bacterial divison Verrucomicrobiales in a temperate fresh water lake. *Microbiol. Ecol.* 25:159–69.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Figure S1. Pair-wise comparison of genetic distances among all combinations of 40 *Pinnularia subgibba* group strains (Data set B) for all subregions of LSU rDNA D1, D2, and D3. Two comparisons based on p-distances (a, b) and two based on numbers of differences are shown (c, d). The dashed line represents the linear function $x = y$.

Figure S2. Maximum likelihood phylogeny of the *Pinnularia subgibba* group based on D1-D3 LSU rDNA sequences of 58 strains. Substitution model GTR+G+I was specified according to the BIC. There were a total of 866 characters in the alignment. The numbers at nodes represent ML bootstrap supports – BI posterior probabilities (only supports above 50% shown). The tree is drawn to scale with branch lengths measured in the number of substitutions per site.

Figure S3. Maximum likelihood phylogeny of the *Pinnularia subgibba* group based on cox1-5P sequences of 45 strains. Substitution model TN93+G was specified according to the BIC. There were a total of 660 characters in the alignment. The numbers at nodes represent ML bootstrap supports – BI posterior probabilities (only supports above 50% shown). The tree is drawn to scale with branch lengths measured in the number of substitutions per site.

Table S1. List of 24 strains of Data set A used for comparison of cox1-5P, rbcL, rbcL-3P, psbA, D1-D2 LSU rDNA, SSU rDNA, and V4 SSU rDNA, with taxon names and Barcode of Life Database identifiers. DOI of Data set A on BoLD is dx.doi.org/10.5883/DS-PIN1. GenBank identifiers can be found in Souffreau et al. (2011). Additional information can be found on BoLD.

Table S2. List of 58 strains of Data set B used for primary species delimitation based on cox1-5P and D1-D3 LSU rDNA with initial morphological identifications, Barcode of Life Database and GenBank identifiers. Strains used to compare the variability of D1, D2, and D3 subregions of LSU rDNA are indicated by an asterisk (n.a. = not available).

Table S3. The primer sequences used in this study.

Table S4. Morphometric measurements of 58 strains of Data set B (mean length, width, and stria density from 10 valves per strain ± SD; n.a. = not available).

Table S5. Summary of available evidence for each operational taxonomic unit (OTU), i.e., clade, singleton strain, or morphogroup (1 = available for all strains; 0 = unavailable for all strains; 7/8 = available for 7 of the total 8 strains). Evidence include primary species delimitation results based on LSU rDNA and cox1-5P, monophyly based on both molecular markers, morphological analysis, coherent biogeography (geographical origin of strains), and ecology (including living mode and in some cases also environmental factors such as pH, conductivity, altitude, water temperature, and salinity) and reproductive compatibility. Colors indicate falsifying (dark grey) and unevaluable or ambiguous evidence (light grey) to the hypothesis that the OTU is a single distinct species or, in the case of the three morphogroups (i.e., DEF, GHI, and JK), that the group is natural.

Table S6. List of 58 strains of Data set B, with source locality and date of collection and identity of collector. GPS coordinates may be found on BoLD.

Appendix 2

Paper II

A time-calibrated multi-gene phylogeny provides insights into the evolution, taxonomy and DNA barcoding of the *Pinnularia gibba* group (Bacillariophyta)

Jan Kollár[1]*, Eveline Pinseel[2,3 +], Wim Vyverman[2] and Aloisie Poulíčková[1]

[1]Department of Botany, Faculty of Science, Palacký University Olomouc, Šlechtitelů 27, CZ-78371 Olomouc, CZ

[2]Laboratory of Protistology and Aquatic Ecology, Department of Biology, Faculty of Science, Ghent University, Krijgslaan 281 - S8, 9000 Gent, BE

[3]Research Department, Meise Botanic Garden, Nieuwelaan 38, 1860 Meise, BE

[+]Current working address: Department of Biological Sciences, University of Arkansas, 850 W. Dickson St., Fayetteville AR 72701-1201, USA

*Corresponding author: jan.kollar@upol.cz

Running title: Phylogeny of *Pinnularia gibba* group

Abstract

Diatoms are one of the most species-rich groups of protists. However, despite increasing evidence for widespread (pseudo)cryptic species diversity, diatom species descriptions are still often based solely on morphological features which potentially confuses taxonomy in many diatom groups. One of the groups with a particularly confused taxonomy is the *Pinnularia gibba* group. Recently, Kollár et al. (2019) used a global data set of 105 strains of the *P. gibba* group for species delimitation by means of a polyphasic approach, providing an evidence for the existence of fifteen species. In order to further guide the systematic revision of the group, the present study focuses on the evolution and morphology of the *P. gibba* group. Using representatives of the fifteen species we used fossils to constrain and calibrate a multi-gene species-level phylogeny, suggesting that the common ancestor of the *P. gibba* group diverged in the Eocene some 36±9 Ma ago. Several (pseudo)cryptic species that were found to inhabit the same geographic regions, e.g. Scotland and volcanic islands in the Indian Ocean, were found to have diverged several millions of years ago. This further highlights the fact that they represent distinct evolutionary lineages, despite the fact that they are often difficult to distinguish using morphological characters only. Although many species are morphologically highly similar, significant differences in cell size were detected in different sections of the tree, suggestive of ongoing morphological differentiation between different *P. gibba* clades. We further used the phylogenetic analysis to assess marker resolution for DNA barcoding, showing that the proposed diatom barcode markers *rbcL* and V4 SSU rDNA can distinguish closely related (pseudo)cryptic species of the group. Given the recent surge in metabarcoding studies focusing on diatoms, this offers a promising avenue to gain a deeper understanding of species distributions and ecologies in the *P. gibba* group.

Keywords:

cryptic species, diatoms, *subgibba* group, molecular phylogenetics, molecular clocks, species complex

Introduction

Diatoms (Bacillariophyta) are one of the most diverse groups of protists with an estimated number of extant species ranging between 30,000 and 200,000 (Andersen 1992, Mann & Vanormelingen 2013). Among primary producers, such species richness is matched only by green algae (ca. 35,000 –200,000 spp.; Andersen 1992, Adl et al. 2007), and streptophytes (over 300,000 extant spp. although it is argued that species concepts applied in vascular plants are relatively narrow in comparison to those applied in microalgae; Christenhusz & Byng 2016). As a group, diatoms are cosmopolitan and adapted to almost all aquatic and some terrestrial habitats. Consequently, they are of great significance not only to fundamental research in the fields of biodiversity, evolutionary biology, ecology, and paleolimnology (e.g., Lewis et al. 2008), but also for applied disciplines such as biomonitoring (e.g., Smol & Stoermer 2010), biotechnology (e.g., Hu et al. 2008), and nanotechnology (e.g., Drum & Gordon 2003). However, the fact that many groups remain understudied and that taxonomy remains largely based on morphology, complicates progress in the systematics of this group. One of the major issues is the widespread occurrence of cryptic and pseudocryptic species diversity (e.g., Mann et al. 2004, Vanormelingen et al. 2008, Trobajo et al. 2009, Poulíčková et al. 2010, Amato et al. 2019, Pinseel et al. 2019) meaning that species are often morphologically indistinguishable or distinguishable only after employment of methods with higher resolution than LM (e.g., statistical analysis of morphological data, thorough examination in the electron microscope), respectively (Mann & Evans 2008). Therefore, molecular methods may be cost-efficient for (1) disentangling diatom species complexes and primary species delimitation, and (2) species identifications (e.g., DNA barcoding; Hebert et al. 2003) for both fundamental and applied purposes. In this context, studying the phylogenetic relationships between the delimited species provides a framework for taxonomic revisions (Cox 2009) as well as for testing the resolution of the genetic markers proposed for species identifications (i.e., barcode markers; e.g., Mann et al. 2010).

One of the taxa with a particularly complex taxonomy (reviewed in detail in, e.g., Krammer 1992, 2000) is *Pinnularia gibba* described by C.G. Ehrenberg (1843a). In the last three decades, detailed morphological studies have led to the description of several *P. gibba*-like taxa: e.g., *P. subgibba* (Krammer 1992), *P. gibbiformis* (Krammer 1992), *P. acidicola* (Van de Vijver et al. 2002), *P. australogibba* (Van de Vijver et al. 2012), *P. amsterdamensis* (Van de Vijver et al. 2012), and *P. vixconspicua* (Van de Vijver et al. 2012). In addition, several species have been described that resemble *P. gibba* more distantly: e.g., *P. macilenta*

(Ehrenberg 1843b), *P. pseudogibba* (Krammer 1992) and *P. lokana* (Krammer 2000). Recently, Kollár et al. (2019) studied the phenotypic (morphology) and genotypic (nuclear LSU rDNA and mitochondrial *cox1*) diversity and biogeographical distribution patterns within the *P. gibba* group (also called the *P. subgibba* group; see the nomenclatural note in the Discussion). As a result, fifteen putative species were uncovered, the majority of which were members of the (pseudo)cryptic species complex *gibba-subgibba-parvulissima*. Furthermore, Poulíčková et al. (2007) investigated sexual compatibility of one of these cryptic species (species K sensu Kollár et al. 2019). Following from the new insights obtained by molecular data, this species was recently described as *Pinnularia lacustrigibba* sp. nov. (Poulíčková et al. 2018), representing the latest contribution to the taxonomy of the group.

While the study design in Kollár et al. (2019) was optimized for species delimitation, namely that the dataset was built in a way to maximize the probability of covering all the available intraspecific variability, the present study focuses on phylogeny inference. To this end, we added the more conserved plastid marker *cox1* and information from the fossil record to infer a time-calibrated multi-gene phylogeny of the group. We used the phylogenetic results to test (1) whether there are significant differences in cell morphology in different parts of the *P. gibba* group tree, and (2) whether the proposed diatom barcode markers *rbcL* and SSU rDNA have sufficient resolution to distinguish between closely related species in the *P. gibba* group. Finally, the taxonomy of the *P. gibba* group is discussed.

Methods

Data sampling: For the sake of DNA-based phylogenetic analyses, one representative strain was selected for each of the fifteen species delimited by Kollár et al. (2019; Table 1). Sampling, culturing and DNA extraction protocols are described in Kollár et al. (2019). In addition to the already available nuclear-encoded D1-D3 LSU rDNA (28S rDNA) and the mitochondrial *cox1*-5P, the plastid marker *rbcL* was sequenced to extend the data set for the species-level phylogenetic analyses. As Kollár et al. (2019) repeatedly failed to amplify *cox1* sequences of all strains of species H, we used the *cox1* primers PBORcox2F and PBORcox2R (Table S1), designed by Pinseel (2019) to amplify the *cox1* sequence of the selected representative of species H (i.e., strain W095b). Furthermore, in order to test the resolution of the different diatom barcode markers, the missing SSU rDNA sequences of species K, G, H, and I were amplified. For *cox1*, the PCR mixtures contained: 2.5µl of 10x PCR buffer (Tris-HCl, $(NH_4)_2SO_4$, KCl, 15mM $MgCl_2$, pH 8.7 at 20°C; "Buffer I", Applied Biosystems, Foster City, USA), 200µM of

deoxynucleoside triphosphates at a concentration of 2mM each, 0.5µM of each primer (Table S1 - primers), 0.4µg.ml^{-1} of bovine serum albumin (BSA), 1.25U of Taq polymerase (AmpliTaq, Perkin-Elmer, Wellesley, USA) and 1µl of template DNA. The total reaction volume was adjusted to 25µl by adding high performance liquid chromatography water (Sigma, St. Louis, Missouri, USA). For *rbcL*, 0.4 µM of each primer was used. For amplification of *cox1*, the following protocol was applied: 35 cycles (0.5 min at 95°C, 1 min at 55°C and 1.5 min at 72°C), with an initial denaturation step of 3 min at 95°C, and a final elongation step of 5 min at 72°C. For amplification of *rbcL*, 40 cycles (1 min at 94°C, 1 min at 55°C and 1.5 min at 72°C), with an initial denaturation step of 3 min at 94°C, and a final elongation step of 5 min at 72°C. PCR success was checked using gel electrophoresis on a 1.5% (w/v) agarose gel. PCR products were purified and sequenced by MACROGEN, Inc. (https://www.dna.macrogen.com). Sequences were assembled using BioNumerics v.3.5 (Applied Maths, Kortrijk, Belgium).

Sequence alignment: Sequences of each marker were automatically aligned using the MUSCLE algorithm (Edgar 2004) as implemented in MEGAX (Kumar et al. 2018), and checked visually. Alignments of protein-coding genes (*cox1* and *rbcL*) were trimmed according to their open reading frames (Table 2). The most divergent regions in the alignments of LSU rDNA were automatically eliminated using the Gblocks Server 0.91b (Castresana 2000) to avoid possible non-homologous positions of hypervariable loop regions. Gap positions were allowed in the final blocks along with the less strict flanking positions.

Substitution saturation: If sites are saturated by substitutions, the phylogenetic signal is lost. As a consequence, such sites are no longer informative about underlying evolutionary processes (e.g., Verbruggen & Theriot 2008, Lemey et al. 2009). In protein-coding genes, the 3rd codon position usually saturates first, because its mutations are often synonymous (e.g., Bofkin & Goldman 2007) and hence beyond the reach of purifying selection. Therefore, the substitution saturation of all genetic markers used in this study was examined and, for the protein-coding genes, the 1st/2nd codon positions were analysed separately from the 3rd. Substitution saturation was analysed by two methods implemented in DAMBE 6.4.29 (Xia 2013). Firstly, proportions of invariant sites were computed from a neighbour-joining tree and specified for each alignment, following Xia et al. (2003). Only fully resolved sites were analysed. Secondly, the number of transitions and transversions were plotted against genetic distance (GTR-distance), and visually examined.

Phylogenetic analyses: A Maximum Likelihood (ML) phylogeny was computed in IQ-TREE v1.6.3 (Trifinopoulos et al. 2016). The best-fit substitution models TIM3+F+I+G4 (Posada 2003), K3Pu+F+I+G4 (Kimura 1981) and GTR+F+I+G4 (Tavaré 1986) were automatically detected by IQ-TREE according to the Bayesian Information Criterion (BIC) for LSU rDNA, *rbcL* and *cox1*, respectively. Branch support was computed using the UltraFast bootstrap approximation (UBS; Minh et al. 2013) with 1,000 replicates. Bayesian inference (BI) was conducted in MrBayes v3.2.6 (Ronquist et al. 2012). Best-fit substitution models GTR+I+G, HKY+I+G (Hasegawa et al. 1985) and GTR+I+G were selected according to BIC computed in MEGAX for LSU rDNA, *rbcL* and *cox1*, respectively. Two independent runs of Markov chain Monte Carlo (MCMC) analysis were performed for 1,000,000 generations. One cold and three heated chains were run for each MCMC analysis. The diagnostic frequency was defined to 1,000 generations and the Markov chain was sampled every 100th generation. The convergence of independent runs was checked in Tracer v1.7.1 (Rambaut et al. 2018). Twenty-five per cent of the samples were discarded as burn-in. To root the trees, two outgroup strains were included in the analyses, PIN885TM and PIN889MG representing the *P. nodosa* and *P. grunowii* groups sensu Souffreau et al. (2011), respectively.

Molecular clock analysis: A time-calibrated phylogeny was computed in BEAST v2.6.1 (Bouckaert et al. 2019), using a relaxed lognormal clock model and a Yule tree prior. The same substitution models as in the bayesian analysis outlined above were selected for each partition. To obtain four fossil-based calibration points, we added 36 outgroup strains to the data set (Table S2). The calibration points were (1) the emergence of the *Pinnularia-Caloneis* complex (based on Souffreau et al. 2011, Nakov et al. 2018), (2) the emergence of *P. borealis* (based on Pinseel 2019), (3) the split between *P. acrosphaeria* and *P. nodosa* (based on Souffreau et al. 2011), and (4) the radiation of *P. viridiformis* subclade sensu Souffreau et al. (2011; based on Saint Martin and Saint Martin 2005). All calibration points are indicated in Figure 2. A uniform probability distribution was used for all calibration points, constrained the ages to to 64.9-76.8 Ma, 14.07-60 Ma, 14.5-44 Ma and 11.7-44 Ma, respectively. The fossil record of *Pinnularia* and the rationale behind the calibration strategy was discussed in detail by Souffreau et al. (2011) and Pinseel (2019). It is important to notice that, in general, the fossil record in microbes is scarce which may negatively affect the accuracy of the time-calibration. Five independent markov chains were run for 500 million generations and sampled every 50,000 generations. The results of the five runs were combined using LogCombiner v2.6.1 (Bouckaert et al. 2019)

and the convergence of Markov chains was checked in Tracer v1.6. Ten per cent of the samples were discarded as burn-in.

Statistical testing of cell size: Visual comparison of our phylogenetic results (Figure 1) with the valve morphology of the different species (see Fig. 4 in Kollár et al. 2019), suggested that the mean cell-sizes of *P. gibba* strains of the basal lineages (i.e., C, A-B, J-K, and D-E-S1) are larger than those of the strains of the lineages in the terminal clade (i.e., F-S2-S3-S4-G-H-I). To statistically test this hypothesis, we used the full data set of Kollár et al. 2019 (i.e., 50 strains of the *P. gibba* group for which morphometric data were available; see Table S4 in Kollár et al. 2019). The cell-size was expressed using four variables: valve length, valve width, length:width ratio and length*width. We calculated the mean values of these variables for each strain using ten valves per strain. The standard deviations reached maxima of ± 2.5 μm and ± 0.7 μm for length and width, respectively. The null hypothesis (H_0: there is no difference in the mean cell-size of the strains of the basal lineages and the mean cell-size of the strains of the lineages in the terminal clade) was statistically tested for each of the four variables separately in R v3.5.1. (R Core Team 2016) by randomization test with 10,000 replications (using only core R functions without any packages). In other words, the values of the given variable were 10,000 times randomly divided into two groups (of the same sizes as the original basal and terminal groups) and a P-value was calculated.

Resolution of DNA barcode markers: To assess the resolution of standard DNA barcode markers applied in environmental metabarcoding of diatom diversity, we compared their sequence divergence between the most closely related species in our study. The evaluated DNA barcode markers were V4 SSU rDNA (i.e., ca 400 bp of SSU rDNA including its V4 subregion; Zimmermann et al. 2011), *rbcL*-3P (i.e., 748 bp at the 3' end of *rbcL*; Hamsher et al. 2011), 312 bp fragment of *rbcL* used in France (Vasselon et al. 2017), and 331 bp fragment of *rbcL* used in the UK (Kelly et al. 2018). In addition, complete SSU rDNA and *rbcL* were evaluated. Since SSU rDNA (18S rDNA) sequences were not available for several of these species, they were obtained when necessary using standard PCR protocols. The PCR mixtures contained the same mixture as described above, with exception of the primer concentration which equalled 1μM for each primer (Table S1 - primers). For PCR amplification, the same protocol as *rbcL* was used, but with an initial denaturation step of 7 min at 94°C and a final elongation step of 10 min at 72°C. SSU rDNA was amplified in two separate PCRs, using primer sets P2/P12 and P4/P14 (Elwood et al. 1985, Gunderson et al. 1986, Guillou et al. 1999, Van Hannen et al.

1999). Numbers of bp differences in both complete and partial *rbcL* and SSU rDNA between closely related species were computed in MEGAX.

Data availability: Extracted DNA, voucher slides and oxidised frustules are stored at PAE (Laboratory of Protistology and Aquatic Ecology, Ghent University, Belgium), and are available upon request. All newly obtained sequences were uploaded to the Barcode of Life Database (BoLD; DOI: XXX) and GenBank (GenBank accession numbers can be found in BoLD). Alignments and R scripts used in this study are available in JK's repository (https://www.researchgate.net/profile/Jan_Kollar5).

Results

Data set properties: A full set of LSU rDNA, *rbcL* and *cox1* sequences was sampled for the fifteen species of the *P. gibba* group (Table 1). LSU rDNA and *cox1* were the most variable markers in the data set, whereas the number of parsimony informative positions was distinctly lower for *rbcL* (Table 2). The concatenated alignment contained 2,883 characters, 9% of which were parsimony informative.

Substitution saturation: In the protein-coding genes (*rbcL* and *cox1*), the $1^{st}/2^{nd}$ codon sites, and the most variable 3^{rd} codon sites were examined separately. The index of substitution saturation (Iss) was significantly lower (at a significance level $\alpha=0.05$) than the critical Iss value (Iss.c) for all genes and codon positions. The Iss values equalled 0.113 for LSU rDNA, 0.027 for the $1^{st}/2^{nd}$ codon sites of *rbcL*, 0.072 for the 3^{rd} codon sites of *rbcL*, 0.061 for the $1^{st}/2^{nd}$ codon sites of *cox1*, and 0.349 for the 3^{rd} codon sites of *cox1*. The Iss.c values (for symmetrical and asymmetrical trees) equalled 0.753 and 0.559 for LSU rDNA, 0.756 and 0.565 for the $1^{st}/2^{nd}$ codon sites of *rbcL*, 0.704 and 0.506 for the 3^{rd} codon sites of *rbcL*, 0.665 and 0.474 for the $1^{st}/2^{nd}$ codon sites of *cox1*, and 0.629 and 0.480 for the 3^{rd} codon sites of *cox1*. According to these results, none of these genes were saturated. Similarly, following a visual examination of the number of transitions and transversions plotted against GTR-distance (Figure S1), none of the genes or their codon sites experienced critical saturation. Together, this indicates the presence of a phylogenetic signal in the data. Therefore, all genetic markers and all codon sites were used for the downstream phylogenetic analyses.

Phylogenetic analyses: Figure 1 shows the ML phylogeny based on the concatenated alignment of LSU rDNA, *rbcL* and *cox1*. Single-locus phylogenies can be found in the supplementary material (Figures S2-S4). Even though the tree-topologies based on the different

genetic markers/alignment strategies/methods of phylogenetic inference were not identical, there were common patterns. First, the ingroup always remained monophyletic. Second, species J (*P.* cf. *parvulissima*) and K (*P. lacustrigibba*) were always sister species and, similarly, species G (*P. acidicola* var. *acidicola*), H (*P. acidicola* var. *elongata*) and I (*P.* cf. *acidicola*) always formed a well-supported clade, although its internal topology varied. Both the ML and BI topologies based on the concatenated alignment recovered the topology (G,(H,I)), whereas the topology based on LSU rDNA (Figure S2) equalled (H,(G,I)) and the topologies based on *rbcL* (Figure S3) and *cox1* (Figure S4) were (I,(G,H)). Third, in both the ML and BI phylogenies based on the concatenated alignment (Figure 1) and in the ML and BI phylogenies based on LSU rDNA (Figure S2), the lineages from the Northern Hemisphere (C, A-B and J-K) appeared on the basis of the tree while the remaining clade consisted of the lineages from the Southern Hemisphere (D-E-S1, F, S4, G-H-I) except for one species from the Netherlands (S2) and one from the Czech Republic (S3). Finally, in the same set of phylogenies as in the third point, there was a terminal clade consisting of species F, G, H, I, S2, S3 and S4 while remaining species (i.e., A, B, C, J-K and D-E-S1) formed a basal grade.

Molecular clock analysis: The results of the fossil time-calibrated molecular clock analysis suggest that the common ancestor of the *P. gibba* group probably diverged from the ancestor of the *P. grunowii* group in the Eocene, i.e., 36±9 Ma ago. The common ancestor of the terminal clade (i.e., F-G-H-I-S2-S3-S4) diverged 17±4 Ma ago. The Scottish sister species J (*P.* cf. *parvulissima*) and K (*P. lacustrigibba*) diverged 7±4 Ma ago, and species G (*P. acidicola* var. *acidicola*), H (*P. acidicola* var. *elongata*) and I (*P.* cf. *acidicola*), inhabiting volcanic islands of the Southern Hemisphere, diverged 6±3 Ma ago.

Statistical testing of cell size: The observed difference between the mean cell-sizes of the strains of the basal lineages (i.e., C, A-B, J-K, and D-E-S1) and mean cell-sizes of the strains of the lineages in the terminal clade (consisting of F, S2, S3, S4, G, H, and I) could not be explained by chance using none of the four variables chosen as the representation of the cell-size (i.e., valve length, P = 0.000; valve width, P = 0.000; length:width, P = 0.001; length*width, P = 0.000). The average valve lengths were 70.2 μm in the basal lineages and 30.4 μm in the terminal clade. The average valve widths were 10.6 μm in the basal lineages and 5.9 μm in the terminal clade. The average length to width ratios were 6.5 in the basal lineages and 5.1 in the terminal clade. The average products of lengths and widths were 791.9 μm^2 in the basal lineages and 180.4 μm^2 in the terminal clade. Therefore, we rejected the null hypothesis and concluded that the observed difference in the cell-size is real. Despite that morphological variability in the

P. gibba group is generally limited, this observation is suggestive of ongoing morphological differentiation between different *P. gibba* lineages.

Resolution of DNA barcode markers: The universal phylogenetic patterns (i.e., the relationship between J-K and G-H-I) were used to investigate the resolution of the proposed barcode markers *rbcL* and SSU rDNA. The complete *rbcL* sequences used in this study (1,365 bp long) differed 6 bp between the two sister species J and K (five differences were in the third codon position and one in the first codon position). Differences in G-H-I equalled 5 bp, 9 bp and 10 bp between G and H, H and I, and G and I, respectively (all differences were in the third codon position except for one between G and H, and one between H and I, both in the first codon position). The fragment of 748 bp at the 3' end of *rbcL* (*rbcL*-3P; Hamsher et al. 2011) differed between J and K in 3 bp, whereas differences in G-H-I were 4 bp, 6 bp, and 8 bp between G and H, H and I, and G and I, respectively. Both French (312 bp; Vasellon et al. 2017) and British (331 bp; Kelly et al. 2018) barcode fragments of *rbcL* differed between J and K in 2 bp, whereas differences in G-H-I were 2 bp, 3 bp, and 5 bp between G and H, H and I, and G and I, respectively. The SSU rDNA sequences used in this study (1,575 bp for G-H-I; 1,190 bp for J and K) differed between two sister species J and K in 5 bp, whereas differences in G-H-I were 7 bp, 7 bp and 6 bp between G and H, H and I, and G and I, respectively. The V4 region (397 bp) of SSU rDNA (V4 SSU rDNA; Zimmermann et al. 2011) differed between J and K in 2 bp, whereas differences in G-H-I were 2 bp, 2 bp, and 4 bp between G and H, H and I, and G and I, respectively. According to these results, both proposed barcode regions *rbcL*-3P and V4 SSU rDNA can unambiguously distinguish closely related species of the group.

Discussion

This study represents the first attempt to infer the phylogenetic relationships within the *P. gibba* group. When only considering shared strains, the tree topologies of the *P. gibba* clade in this study (Figure 1) and those recovered by Souffreau et al. (2011), who covered the entire genus *Pinnularia*, are identical. In addition, the overall results of our fossil-guided molecular time-calibration of the genus *Pinnularia* were directly in line with earlier studies (Souffreau et al. 2011, Nakov et al. 2018). Nevertheless, some topological differences were observed between the single-gene trees and the concatenated analysis in our study. None of these topological differences represented hard conflicts. Differences in tree topologies retrieved by analyzing different genes are generally thought to be the result of incomplete lineage sorting, and are a common phenomenon across the tree of life (e.g., Leliaert et al. 2014). The three

(pseudo)cryptic species-level lineages G (*P. acidicola* var. *acidicola*), H (*P. acidicola* var. *elongata*) and I (*P.* cf. *acidicola*), inhabiting volcanic islands of the Southern Hemisphere, diverged 6±3 Ma ago. However, those volcanic islands are far younger (e.g, Hall et al. 2011) suggesting that the species must have evolved elsewhere and colonized respective islands somewhen after their emergence.

We found that two groups in the *P. gibba* group (the terminal clade, and the basal grade) showed significant differences in cell-size. This suggests that the diminutive appearance of species in the terminal clade, relative to the basal grade, was inherited from the common ancestor of the species in this clade. In microbes, cell-size plays a crucial role in their physiology (e.g., influencing flux across cytoplasmatic membranes) and ecology (e.g., influencing ability to attach effectively to external surfaces; Marshall et al. 2012). Moreover, in diatoms, patterns of cell-size evolution were observed through geological time (Finkel et al. 2005), and across different environments (Nakov et al. 2014). Furthermore, cell-size is often positively correlated with genome-size (e.g., Connolly et al. 2008, Koester et al. 2010, Mueller 2015) and evolutionary patterns of genome duplications were found in diatoms (Parks et al. 2018). On the other hand, diatom cell-size can vary without genome duplication (Mann and Poulíčková 2010). Nevertheless, the reduction of cell-size observed here may be correlated with the reduction of genome size. This remains to be tested in future research.

We compared our phylogenies to biogeographical data available for the delimited species (see Figure 5 in Kollár et al. 2019). Two species from the Northern Hemisphere (S2 from the Netherlands and S3 from the Czech Republic) appeared within the terminal clade, which is predominantly represented by species from the Southern Hemisphere. According to our chronogram (Figure 2), these two species diverged from the ancestors of other terminal clade representatives (namely S4, G, H and I) in the Miocene and had ca 14 Ma to disperse into the Northern Hemisphere. Nevertheless, a similar pattern was observed in several other diatom complexes, such as the *P. borealis* complex (Pinseel et al. 2019), the *Hantzschia amphioxys* complex (Souffreau et al. 2013b), and the *Gomphonema parvulum* complex (Abarca et al. 2014). These results suggest that diatoms, similarly to some other protists (e.g., chrysophytes; Bock et al. 2017), are capable of effective long-distance dispersal between hemispheres. On the other hand, many diatom species seem to be geographically restricted or even endemic (e.g., Van de Vijver et al. 2005, Vyverman et al. 2010, Pinseel et al. 2019). It is probable that increased sampling accompanied by genetic analyses will shed more light on the puzzle of diatom biogeography in the future. Nevertheless, for now, the exact mechanisms of dispersal

remain unknown. It could be suggested that dispersal by wind may be less effective between hemispheres than within hemispheres due to the equatorial atmospheric circulation: the Hadley cells (Hartmann 2015). Hadley cells rise the air near the equator and turn it poleward in an altitude between 10 and 15 km above sea level. At such altitude, both atmospheric pressure (between 12 and 25 kPa) and temperature (ca. -55°C) are very low (NASA 1976). To date, no experiments tested the tolerance of diatoms to low pressure, but we may assume that under such conditions they may suffer from desiccation which is lethal for many diatoms (Souffreau et al. 2010, 2013a). Similarly, it seems that many diatoms are intolerant to freezing below -40°C, although there are some notable exceptions (Stock et al. 2018). In addition to passive dispersal by means of abiotic vectors, biotic vectors might also play a role. Previously, diatoms have been shown to colonize feathers of water birds (Croll & Holmes 1982), and they can even survive travel through the bird digestive system (e.g., Atkinson 1972, 1980, Stoyneva 2016). Twice a year, several bird species migrate between hemispheres in huge numbers. An extreme example is the Arctic Tern (*Sterna paradisaea*), which migrates between the Arctic and Antarctic region (e.g., Fijn et al. 2013). It is thus not unlikely that migratory water birds may play a role in the (long-distance) dispersal of diatoms, in addition to abiotic vectors.

Recently, Kollár et al. (2019) found that, contrary to other diatom genera (e.g., Mann et al. 2010, Zimmermann et al. 2011), the proposed diatom barcode genes *rbcL* and SSU rDNA were relatively conserved in the genus *Pinnularia* (see Figure 2 in Kollár et al. 2019). Nevertheless, the taxonomic resolution of both barcode markers was not tested in their study. Here, we found that both complete and partial (i.e., 748 bp fragment proposed by Hamsher et al. 2011, 312 bp fragment proposed by Vasselon et al. 2017, and ca 340 bp fragment proposed by Kelly et al. 2018) *rbcL* can unambiguously distinguish between different species, with a minimum divergence of 5 and 2 bp, respectively. Similarly, Pinseel et al. (2019) found that *rbcL* could distinguish even between the most closely related lineages of the cryptic species complex *P. borealis* (9-14 bp sequence difference on a total of 1,395 bp). Complete and partial SSU rDNA were also able to distinguish between all *P. gibba* species here uncovered (with a minimum divergence of 5 and 2 bp respectively), suggesting that both *rbcL* and SSU rDNA could be good barcode markers to distinguish closely-related *P. gibba* species in environmental surveys. However, its sequence diversity was generally lower than *rbcL*, and previous work indicated that different species can have identical V4 SSU rDNA sequences in the *P. borealis* and *A. minutissimum* species complexes (Pinseel et al. 2017, Pinseel 2019). In addition, a recent study by Mora et al. (2019) suggested that SSU rDNA metabarcoding consistently

underestimates diversity in the genus *Pinnularia*, relatively to morphology-based diversity assessments. Finally, it has to be noted that for both *rbcL* and SSU rDNA the sequence divergences between closely related species are generally low, and likely to be overlooked when using the standard 97% similarity cut-off OTU-based pipelines in metabarcoding studies. In order to distinguish closely-related diatom species using metabarcoding, more sensitive protocols that can uncover haplotype diversity by means of Amplicon Sequence Variants should be adopted (e.g., DADA2 or UNOISE; Callahan et al. 2016, Edgar 2016).

To prevent possible future disputes regarding the taxonomy of the *P. gibba* group, we consider the following nomenclatural note important. From a morphological point of view, those *Pinnularia* species that bear 'ghost striae' in the central area (slight thinnings of the valve that correspond in size and spacing to the normal striae; see also Cox 1999) have sometimes been called the "*gibba* group" (e.g., Krammer 2000). On the other hand, in the molecular phylogeny by Souffreau et al. (2011), the clade containing *gibba*-like taxa was called "the *subgibba* subclade" and, for sake of continuity, Kollár et al. (2019) followed this term, altering it to "the *Pinnularia subgibba* group" or „*subgibba* group" in short. However, Souffreau (2011b) also observed ghost striae in one member of the "*grunowii*" subclade (strain PIN650K identified as *P. subanglica*) and one member of the "*nodosa*" subclade (PIN885TM identified as *P. nodosa*) and they may be observed also on LM photographs of *P. nodosa* (p. 308, Fig. 10-12) by Krammer (2000). Given that the character is probably homologous in all three subclades (together they form clade B sensu Souffrau et al. 2011), it cannot be considered apomorphic for the *P. gibba* group as implicitly suggested by Krammer's use of the term (e.g., Krammer 2000, p. 96). Nevertheless, while ghost striae cannot serve as an unambiguous diagnostic character of the *P. gibba* group on their own, for sake of identification, overall morphology is usually considered. Therefore, terms *subgibba* group sensu Souffreau and *gibba* group sensu Krammer may be considered synonymous. Further taxonomy-related considerations may be found in Kollár (2020).

Among other things, our study showed that deeper in evolutionary history patterns of morphological evolution may be detected even in taxonomically complicated groups of diatoms with the widespread occurrence of (pseudo)cryptic species diversity. Although causes and circumstances of such patterns remain open for future research, they may proof helpful in reviewing the higher classification of such groups while species delimitations and identifications may be achieved efficiently using mere units of genetic markers.

Acknowledgements

JK is most grateful to Dr David M. Williams for sharing his vast knowledge and ideas, as well as for many insightful and excited discussions. JK is further grateful to Dr David M. Williams and the Natural History Museum in London for an unlimited access to its vast collection of diatom material and literature. EP is a postdoctoral fellow from the Belgian American Education Foundation and Fulbright Belgium, and was supported as a PhD student by the Fund for Scientific Research Flanders (FWO), through grant numbers 1104315N and 1104317N.

References

Abarca, N., Jahn, R., Zimmermann, J. & Enke, N. 2014. Does the cosmopolitan diatom *Gomphonema parvulum* (Kützing) Kützing have a biogeography? *PLoS One*. 9:1–18.

Adl, S.M., Leander, B.S., Simpson, A.G.B., Archibald, J.M., Anderson, O.R., Bass, D., Bowser, S.S. et al. 2007. Diversity, nomenclature, and taxonomy of protists. *Syst. Biol.* 56:684–9.

Amato, A., Kooistra, W.H.C.F. & Montresor, M. 2019. Cryptic Diversity: a Long-lasting Issue for Diatomologists. *Protist*. 170:1–7.

Andersen, R.A. 1992. Diversity of eukaryotic algae. *Biodivers. Conserv.* 1:267–92.

Atkinson, K.M. 1972. Birds as transporters of algae. *Br. Phycol. J.* 7:319–21.

Atkinson, K.M. 1980. Experiments in dispersal of phytoplankton by ducks. *Br. Phycol. J.* 15:49–58.

Bock, C., Chatzinotas, A. & Boenigk, J. 2017. Genetic diversity in chrysophytes: Comparison of different gene markers. *Fottea*. 17:209–21.

Bofkin, L. & Goldman, N. 2007. Variation in evolutionary processes at different codon positions. *Mol. Biol. Evol.* 24:513–21.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J. et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary

analysis. *PLoS Comput. Biol.* 15:e1006650.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. & Holmes, S.P. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods*. 13:581–3.

Castresana, J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* 17:240–552.

Christenhusz, M.J.M. & Byng, J.W. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa*. 261:201–17.

Connolly, J.A., Oliver, M.J., Beaulieu, J.M., Knight, C.A., Tomanek, L. & Moline, M.A. 2008. Correlated evolution of genome size and cell volume in diatoms (Bacillariophyceae). *J. Phycol.* 44:124–31.

Cox, E.J. 2009. What's in a name? – Diatom classification should reflect systematic relationships . *Acta Bot.Croat.* 68:443–54.

Croll, D.A. & Holmes, R.W. 1982. A Note on the Occurrence of Diatoms on the Feathers of Diving Seabirds. *Auk*. 99:765–6.

Drum, R.W. & Gordon, R. 2003. Star Trek replicators and diatom nanotechnology. *Trends Biotechnol.* 21:325–8.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with improved accuracy and speed. *Proceedings. 2004 IEEE Comput. Syst. Bioinforma. Conf. 2004. CSB 2004.* 689–90.

Edgar, R.C. 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv*. 81257.

Ehrenberg, C.G. 1843a. Mittheilungen über 2 neue asiatische Lager fossiler Infusorien-Erden aus dem russischen Trans-Kaukasien (Grusien) und Sibirien. *Bericht über die zur Bekanntmachung geeigneten Verhandlungen der Königlich-Preussischen Akad. der Wissenschaften zu Berlin*. 43–

9.

Ehrenberg, C.G. 1843b. Verbreitung und Einfluss des mikroskopischen Lebens in Süd-und Nord-Amerika. Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin 1841. 466 pp.

Elwood, H.J., Olsen, G.J. & Sogin, M.L. 1985. The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytricha nova* and *Stylonychia pustulata*. *Mol. Biol. Evol.* 2:399–410.

Fijn, R.C., Hiemstra, D., Phillips, R.A. & van der Winden, J. 2013. Arctic Terns *Sterna paradisaea* from the Netherlands Migrate Record Distances Across Three Oceans to Wilkes Land, East Antarctica. *Ardea*. 101:3–12.

Finkel, Z. V, Katz, M.E., Wright, J.D., Schofield, O.M.E. & Falkowski, P.G. 2005. Climatically driven macroevolutionary patterns in the size of marine diatoms over the Cenozoic. *Proc. Natl. Acad. Sci. U. S. A.* 102:8927–32.

Guillou, L., Medlin, L.K., Oce, O. & Vaulot, D. 1999. *Bolidomonas*: A new genus with two species belonging to a new algal class, the Bolidophyceae (Heterokonta). *J. Phycol.* 35:368–81.

Gunderson, J.H., McCutchan, T.F. & Sogin, M.L. 1986. Sequence of the Small Subunit Ribosomal RNA Gene Expressed in the Bloodstream Stages of *Plasmodium berghei*: Evolutionary Implications. *J. Protozool.* 33:525–9.

Hall, K., Meiklejohn, I. & Bumby, A. 2011. Marion Island volcanism and glaciation. *Antarct. Sci.* 23:155–63.

Hamsher, S.E., Evans, K.M., Mann, D.G., Poulíčková, A. & Saunders, G.W. 2011. Barcoding diatoms: Exploring alternatives to COI-5P. *Protist*. 162:405–22.

Hartmann, D.L. 2015. Global physical climatology. 2nd ed. Elsevier Science. 498 pp.

Hasegawa, M., Kishino, H. & Yano, T. aki 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–74.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R. 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* 270:313–21.

Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M. & Darzins, A. 2008. Microalgal triacylglycerols as feedstocks for biofuel production: Perspectives and advances. *Plant J.* 54:621–39.

Kelly, M., Boonham, N., Juggins, S., Kille, P., Mann, D.G., Pass, D., Sapp, M. et al. 2018. A DNA based diatom metabarcoding approach for Water Framework Directive classification of rivers. *Environ. Agency*. Bristol.

Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences b. *Proc. Natl. Acad. Sci. U. S. A.* 78:454–8.

Koester, J.A., Swalwell, J.E., Von Dassow, P. & Armbrust, E.V. 2010. Genome size differentiates co-occurring populations of the planktonic diatom *Ditylum brightwellii* (Bacillariophyta). *BMC Evol. Biol.* 10:1–11.

Kollár, J. 2020. Polyphasic approach to the study of diatom diversity (Doctoral Thesis). Palacký University Olomouc.

Kollár, J., Pinseel, E., Vanormelingen, Pieter Poulíčková, A., Souffreau, C., Dvořák, P. & Vyverman, W. 2019. A Polyphasic approach to the delimitation of diatom species: a case study for the genus *Pinnularia* (Bacillariophyta). *J. Phycol.* 55: 365–379.

Krammer, K. 1992. Pinnularia eine Monographie de europäischen Taxa. *Bibl. Diatomol.* 26:1–353.

Krammer, K. 2000. The genus *Pinnularia*. A.R.G. Ganter Verlag K.G., Ruggell. 703 pp.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35:1547–9.

Leliaert, F., Verbruggen, H., Vanormelingen, P., Steen, F., López-Bautista, J.M., Zuccarello, G.C. & De Clerck, O. 2014. DNA-based species delimitation in algae. *Eur. J. Phycol.* 49:179–96.

Lemey, P., Salemi, M. & Vandamme, A.-M. (Eds.) 2009. The Phylogenetic Handbook. 2nd ed. Cambridge University Press, New York.

Lewis, A.R., Marchant, D.R., Ashworth, A.C., Hedenäs, L., Hemming, S.R., Johnson, J. V, Leng, M.J. et al. 2008. Mid-Miocene cooling and the extinction of tundra in continental Antarctica. *Proc. Natl. Acad. Sci.* 105:10676–80.

Mann, D.G. & Evans, K.M. 2008. The species concept and cryptic diversity. *In Proceedings of the 12th International Conference on Harmful Algae*. International Society for the Study of Harmful Algae and Intergovernmental Oceanographic Commission of UNESCO, Copenhagen, pp. 262–8.

Mann, D.G., McDonald, S.M., Bayer, M.M., Droop, S.J.M., Chepurnov, V.A., Loke, R.E., Ciobanu, A. et al. 2004. The *Sellaphora pupula* species complex (Bacillariophyceae): Morphometric analysis, ultrastructure and mating data provide evidence for five new species. *Phycologia*. 43:459–82.

Mann, D.G. & Poulíčková, A. 2010. Mating system, auxosporulation, species taxonomy and evidence for homoploid evolution in *Amphora* (Bacillariophyta). *Phycologia*. 49:183–201.

Mann, D.G., Sato, S., Trobajo, R., Vanormelingen, P. & Souffreau, C. 2010. DNA barcoding for species identification and discovery in diatoms. *Cryptogam. Algol.* 31:557–77.

Mann, D.G. & Vanormelingen, P. 2013. An inordinate fondness? the number, distributions, and origins of diatom species. *J. Eukaryot. Microbiol.* 60:414–20.

Marshall, W.F., Young, K.D., Swaffer, M., Wood, E., Nurse, P. & Kimura, A. 2012. What determines cell size? *BMC Biol.* 10:1–22.

Minh, B.Q., Nguyen, M.A.T. & Von Haeseler, A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30:1188–95.

Mora, D., Abarca, N., Proft, S., Grau, J.H., Enke, N. & Carmona, J. 2019. Morphology and metabarcoding: a test with stream diatoms from Mexico highlights the complementarity of

identification methods. *Freshw. Sci.* 38:448–464.

Mueller, R.L. 2015. Genome biology and the evolution of cell-size diversity. *Cold Spring Harb. Perspect. Biol.* 7:1–9.

Nakov, T., Beaulieu, J.M. & Alverson, A.J. 2018. Accelerated diversification is related to life history and locomotion in a hyperdiverse lineage of microbial eukaryotes (Diatoms, Bacillariophyta). *New Phytol.* 219:462–73.

Nakov, T., Theriot, E.C. & Alverson, A.J. 2014. Using phylogeny to model cell size evolution in marine and freshwater diatoms Using phylogeny to model cell size evolution in marine and freshwater diatoms. *Limnol. Oceanogr.* 59:79–86.

NASA 1976. U.S. Standard Atmosphere. National Aeronautics and Space Administration. 241 pp.

Parks, M.B., Nakov, T., Ruck, E.C., Wickett, N.J. & Alverson, A.J. 2018. Phylogenomics reveals an extensive history of genome duplication in diatoms (Bacillariophyta). *Am. J. Bot.* 105:330–47.

Pinseel, E. 2019. Diversity, biogeography and evolutionary history of diatoms: *Pinnularia borealis* as a case study (Doctoral Thesis). Ghent University.

Pinseel, E., Kulichová, J., Scharfen, V., Urbánková, P., Van de Vijver, B. & Vyverman, W. 2019. Extensive Cryptic Diversity in the Terrestrial Diatom *Pinnularia borealis* (Bacillariophyceae). *Protist*. 170:121–40.

Pinseel, E., Vanormelingen, P., Hamilton, P.B., Vyverman, W., Van de Vijver, B. & Kopalova, K. 2017. Molecular and morphological characterization of the *Achnanthidium minutissimum* complex (Bacillariophyta) in Petuniabukta (Spitsbergen, High Arctic) including the description of *A. digitatum* sp. nov. *Eur. J. Phycol.* 52:264–80.

Posada, D. 2003. Using Modeltest and PAUP* to Select a Model of Nucleotide Substitution. *In Current Protocols in Bioinformatics*. John Wiley & Sons, Inc., New York, pp. 6.5.1-6.5.14.

Poulíčková, A., Kollár, J., Hašler, P., Dvořák, P. & Mann, D.G. 2018. A new species *Pinnularia*

lacustrigibba sp. nov. within the *Pinnularia subgibba* group (Bacillariophyceae). *Diatom Res.* 33:273–82.

Poulíčková, A., Mayama, S., Chepurnov, V.A. & Mann, D.G. 2007. Heterothallic auxosporulation, incunabula and perizonium in *Pinnularia* (Bacillariophyceae). *Eur. J. Phycol.* 42:367–90.

Poulíčková, A., Veselá, J., Neustupa, J. & Škaloud, P. 2010. Pseudocryptic Diversity versus cosmopolitanism in diatoms: A case study on *Navicula cryptocephala* Kütz. (Bacillariophyceae) and morphologically similar taxa. *Protist*. 161:353–69.

R Core Team 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G. & Suchard, M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67:901–4.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B. et al. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61:539–42.

Saint Martin, S. & Saint Martin, J.P. 2005. The diatom assemblages as environmental evolution recording of the Paratethys area during Sarmatian times. *Comptes Rendus Palevol*. 4:191–201.

Smol, J.P. & Stoermer, E.F. 2010. The Diatoms: Applications for the Environmental and Earth Sciences. 2nd ed. Cambridge University Press, Cambridge. 667 pp.

Souffreau, C. 2011. Biogeography, species diversity and stress tolerance of aquatic and terrestrial diatoms (Doctoral Thesis). Ghent University, 285 pp.

Souffreau, C., Vanormelingen, P., Sabbe, K. & Vyverman, W. 2013a. Tolerance of resting cells of freshwater and terrestrial benthic diatoms to experimental desiccation and freezing is habitat-dependent. *Phycologia*. 52:246–55.

Souffreau, C., Vanormelingen, P., Van de Vijver, B., Isheva, T., Verleyen, E., Sabbe, K. & Vyverman,

W. 2013b. Molecular Evidence for Distinct Antarctic Lineages in the Cosmopolitan Terrestrial Diatoms *Pinnularia borealis* and *Hantzschia amphioxys*. *Protist*. 164:101–15.

Souffreau, C., Vanormelingen, P., Verleyen, E., Sabbe, K. & Vyverman, W. 2010. Tolerance of benthic diatoms from temperate aquatic and terrestrial habitats to experimental dessication and temperature stress. *Phycologia*. 49:309–24.

Souffreau, C., Verbruggen, H., Wolfe, A.P., Vanormelingen, P., Siver, P.A., Cox, E.J., Mann, D.G. et al. 2011. A time-calibrated multi-gene phylogeny of the diatom genus *Pinnularia*. *Mol. Phylogenet. Evol.* 61:866–79.

Stock, W., Pinseel, E., De Decker, S., Sefbom, J., Blommaert, L., Chepurnova, O., Sabbe, K. et al. 2018. Expanding the toolbox for cryopreservation of marine and freshwater diatoms. *Sci. Rep.* 1–9.

Stoyneva, M.P. 2016. Allochthonous planktonic algae recorded during the last 25 years in Bulgaria and their possible dispersal agents. *Hydrobiologia*. 764:53–64.

Tavaré, S. 1986. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *Lect. Math. Life Sci.* 17:57–86.

Trifinopoulos, J., Nguyen, L.T., von Haeseler, A. & Minh, B.Q. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44:W232–5.

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* (Bacillariophyceae). *Phycologia*. 48:443–59.

Van de Vijver, B., Chattová, B., Metzeltin, D. & Lebouvier, M. 2012. The genus *Pinnularia* (Bacillariophyta) on Ile Amsterdam (TAAF, Southern Indian Ocean). *Nov. Hedwigia*. 141:201–36.

Van de Vijver, B., Frenot, Y. & Beyens, L. 2002. Freshwater diatoms from Île de la Possession

(Crozet Archipelago, sub-Antarctica). *Bibl. Diatomol.* 46:1–412.

Van De Vijver, B., Gremmen, N.J.M. & Beyens, L. 2005. The genus *Stauroneis* (Bacillariophyceae) in the Antarctic region. *J. Biogeogr.* 32:1791–8.

Van Hannen, E.J., Mooij, W., Van Agterveld, M.P., Gons, H.J. & Laanbroek, H.J. 1999. Detritus-dependent development of the microbial community in an experimental system: Qualitative analysis by denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* 65:2478–84.

Vanormelingen, P., Chepurnov, V.A., Mann, D.G., Sabbe, K. & Vyverman, W. 2008. Genetic Divergence and Reproductive Barriers among Morphologically Heterogeneous Sympatric Clones of *Eunotia bilunaris* Sensu Lato (Bacillariophyta). *Protist*. 159:73–90.

Vasselon, V., Rimet, F., Tapolczai, K. & Bouchez, A. 2017. Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island, France). *Ecol. Indic.* 82:1–12.

Verbruggen, H. & Theriot, E.C. 2008. Building trees of algae: Some advances in phylogenetic and evolutionary analysis. *Eur. J. Phycol.* 43:229–52.

Vyverman, W., Verleyen, E., Wilmotte, A., Hodgson, D.A., Willems, A., Peeters, K., Van de Vijver, B. et al. 2010. Evidence for widespread endemism among Antarctic micro-organisms. *Polar Sci.* 4:103–13.

Xia, X. 2013. DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* 30:1720–8.

Xia, X., Xie, Z., Salemi, M., Chen, L. & Wang, Y. 2003. An index of substitution saturation and its application. *Mol. Phylogenet. Evol.* 26:1–7.

Zimmermann, J., Jahn, R. & Gemeinholzer, B. 2011. Barcoding diatoms: Evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Org. Divers. Evol.* 11:173–92.

Table 1. List of strains selected as representatives of the fifteen species of the *Pinnularia gibba* group delimited by Kollár et al. (2019).

Species	Strain code	Taxon	BoLD identifier
A	CZECH_SW2_7a	*P.* cf. *microstauron*	PINN213-14
B	NUUK_13KAP17_3	*P.* cf. *lokana*	PINN146-14
C	STAP3	*P. macilenta*	PINN219-15
D	W045b	*P.* cf. *australogibba*/*amsterdamensis*	PINN172-14
E	REU12_5_1	*P.* cf. *vixconspicua*	PINN103-14
F	Tor8b	*P.* sp.	PINN171-14
G	MIC5_16	*P. acidicola* var. *acidicola*	PINN003-14
H	W095b	*P. acidicola* var. *elongata*	PINN209-14
I	REU12_9_14	*P.* cf. *acidicola*	PINN105-14
J	PIN877TM	*P.* cf. *parvulissima*	PINN162-14
K	PIN19Cra	*P. lacustrigibba*	PINN192-14
S1	Tor7f	*P.* sp.	PINN170-14
S2	Wiec	*P. subcapitata* var. *elongata*	PINN176-14
S3	CZECH_NOS2_7	*P.* sp.	PINN039-14
S4	Tor4r	*P.* sp.	PINN164-14

Table 2. Characteristics of alignments including total number of characters, number of variable characters (V), number of parsimony-informative characters (PI), and mean genetic distances between the sequences of molecular markers ± standard deviation.

Genetic marker	Total	V	PI	p-distance	number of differences
D1-D3 LSU rDNA	858	178	108	0.063 ± 0.015	53.4 ± 13.2
rbcL	1365	104	48	0.019 ± 0.010	25.5 ± 8.0
cox1-5P	660	200	115	0.093 ± 0.010	55.5 ± 11.0

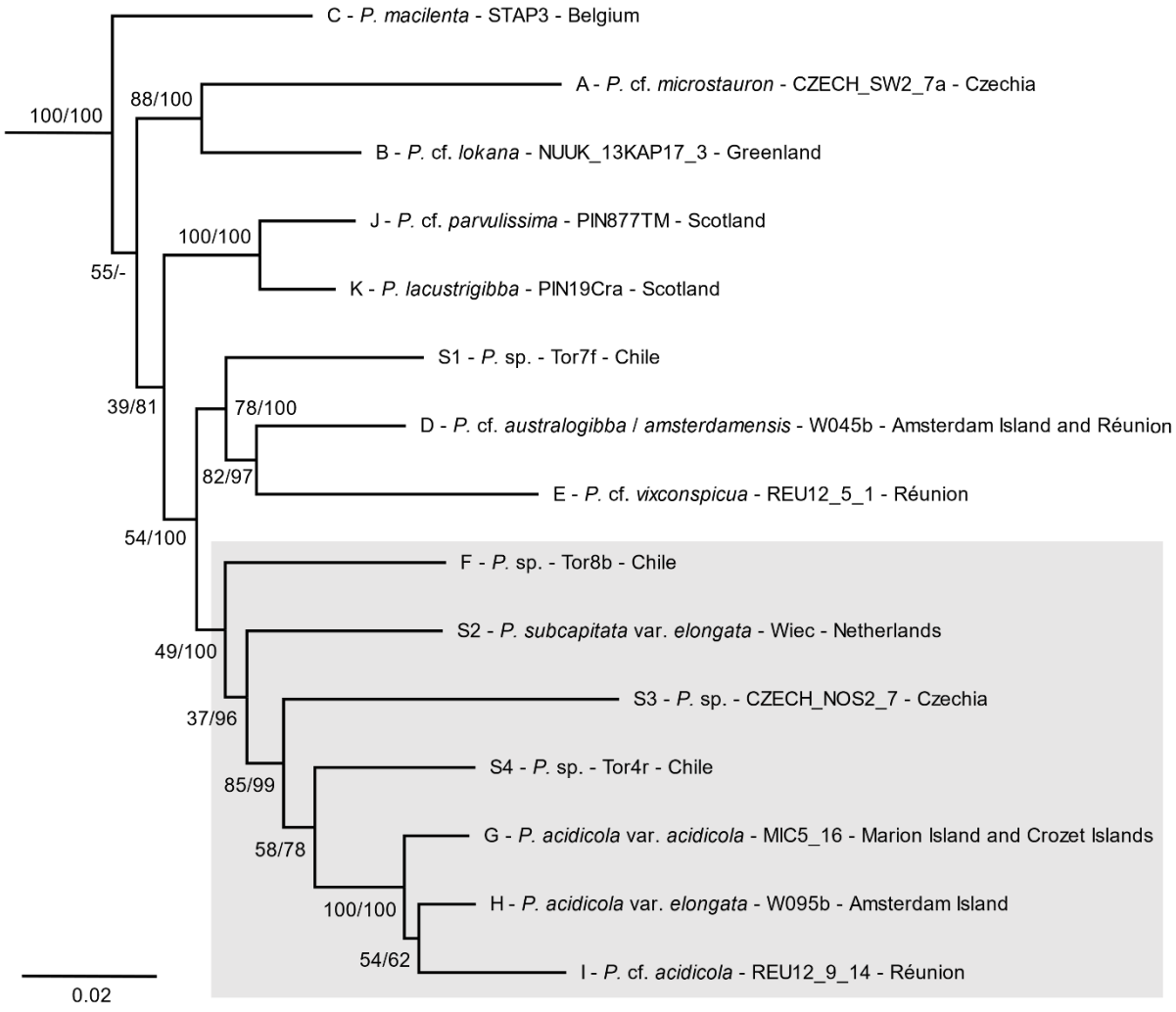


Figure 1. Rooted phylogeny of the *Pinnularia gibba* group inferred from a three-locus DNA alignment using ML under a partitioned model. Values at nodes indicate statistical supports given as percentages, ML ultrafast bootstrap approximation / BI posterior probabilities. Whenever the ML and BI topologies disagreed, a '-' is given for the BI posterior probability support. The tip labels correspond to the delimited species sensu Kollár et al. (2019): species code - taxon - strain - known locations. The terminal clade used in statistical hypothesis testing is indicated in grey. For the sake of presentation, the two outgroup strains are not shown.

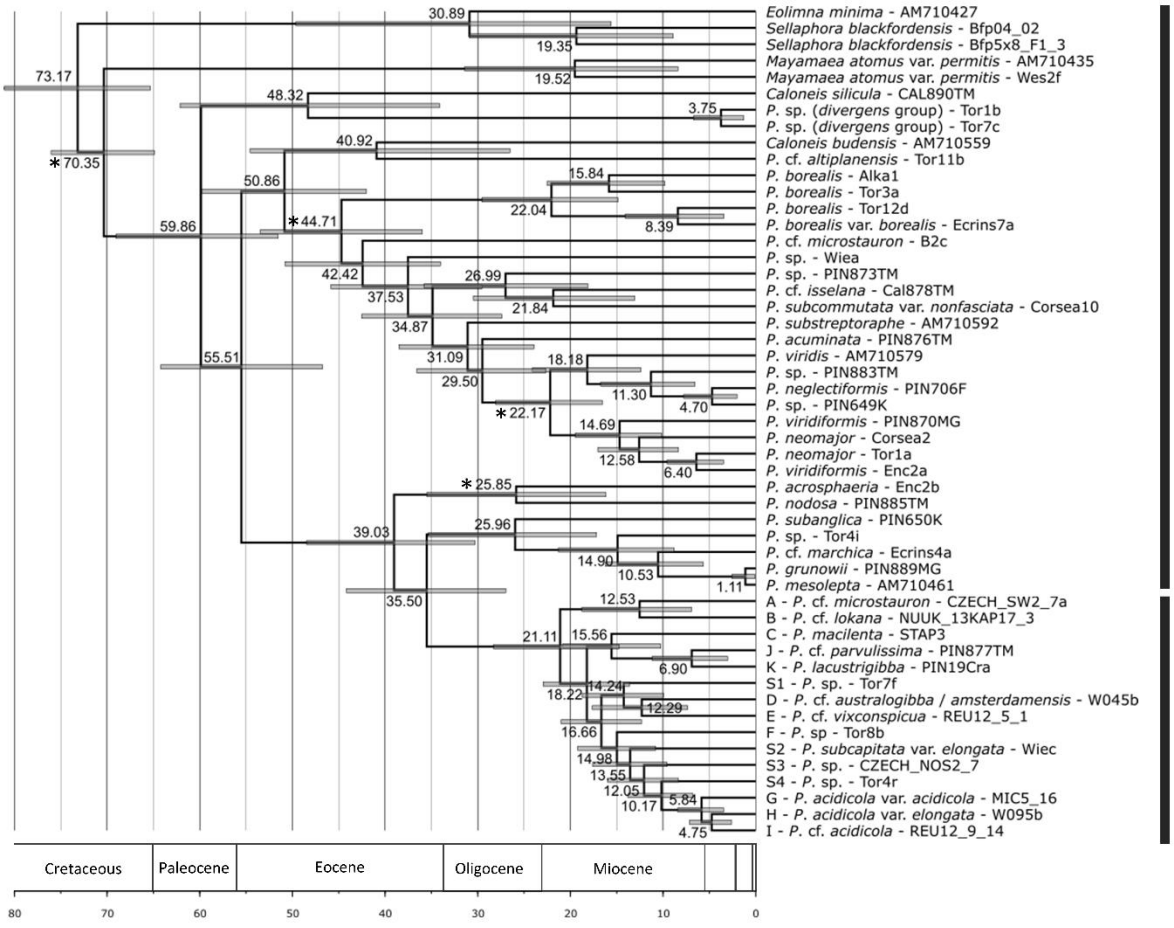


Figure 2. A time-calibrated phylogeny of the *Pinnularia gibba* group inferred from a three-locus DNA alignment using Bayesian relaxed molecular clock analysis with four fossil-based time constraints (indicated by asterisks). Values at nodes are mean node ages. Grey bars are 95% HPD (highest posterior density) intervals. Ingroup tip labels correspond to the delimited species sensu Kollár et al. (2019): species code - taxon - strain.

Appendix 3

Paper III

Diatom Research, 2018
Vol. 33, No. 3, 273–282, https://doi.org/10.1080/0269249X.2018.1513869

A new species *Pinnularia lacustrigibba* sp. nov. within the *Pinnularia subgibba* group (Bacillariophyceae)

ALOISIE POULÍČKOVÁ[1]*, JAN KOLLÁR[1], PETR HAŠLER[1], PETR DVOŘÁK[1] & DAVID G. MANN[2,3]

[1]*Department of Botany, Faculty of Sciences, Palacký University in Olomouc, Olomouc, Czech Republic*
[2]*Royal Botanic Garden Edinburgh, Edinburgh, UK*
[3]*Institute for Food and Agricultural Research and Technology (IRTA), Catalonia, Spain*

Some traditional diatom species have long been considered 'cosmopolitan', i.e., they are ubiquitous in suitable habitats. However molecular methods have revealed that some of these diatoms are complexes consisting of a few or many species, whose identification by eye with LM is difficult or impossible. This is true of several groups within *Pinnularia*, including the *P. subgibba* group, in which molecular species delimitation based on cox1-5P and D1–D3 LSU rDNA markers revealed 15 species within a dataset of 58 unique genotypes, representing 105 strains of the group. One clade comprises reproductively compatible clones from Scottish lakes and its morphology does not correspond exactly to any of the known *Pinnularia* species; thus the description of a new species '*P. lacustrigibba*' sp. nov. is proposed.

Keywords: *cryptic diversity, cytology, epipelon, genetics, morphology, molecular systematics, new species,* Pinnularia lacustrigibba *sp. nov*

Introduction

Pinnularia Ehrenberg is one of the most species-rich raphid diatom genera, with 704 accepted taxon names according to AlgaeBase (Guiry & Guiry 2018) and 1527 species names according to DiatomBase (Kociolek et al. 2018). In recent years many new *Pinnularia* species have been described using the traditional morphological approach (Van de Vijver & Zidarova 2011, Luís et al. 2012, Pereira et al. 2014) or using both molecular and morphological data (Pinseel et al. 2016). Molecular studies suggest that many traditional *Pinnularia* morphospecies are species complexes with cryptic or pseudocryptic diversity (Souffreau et al. 2011, Pinseel et al. 2016). The genus is cosmopolitan, occurring in a broad range of freshwater and terrestrial environments within assemblages such as the epipelon (Poulíčková et al. 2014), epibryon and epiphyton (Poulíčková et al. 2013, 2017, Kollár et al. 2015). Although their ecological demands vary, most species prefer low electrolyte content (Round et al. 1990). To date, only one molecular phylogeny (Souffreau et al. 2011) has focused exclusively on *Pinnularia* and this distinguished several morphologically distinct clades and subclades within the genus. The *subgibba* subclade is often characterized by relatively small linear cells with drop-like external central raphe endings, ghost striae at the valve centre, and a fascia (Souffreau et al. 2011). The *subgibba* subclade sensu Souffreau et al. (2011), equivalent to the *Pinnularia subgibba* group sensu Kollár et al. (in press), may be considered inclusive of, or even synonymous with, the *gibba* group sensu Krammer (2000).

Pinnularia clones isolated in 2004 and 2005 from three lakes in Scotland reproduced vigorously via a heterothallic sexual process and auxosporulation was observed not only within the population of each lake ('sympatric' experimental crosses), but also between populations from different lakes ('allopatric' crosses; Poulíčková et al. 2007). Although the species was studied morphologically, there was insufficient evidence to make a confident identification or to describe it as a new taxon. It was therefore identified and reported as *Pinnularia* cf. *gibba* (Poulíčková et al. 2007). Later, it became possible to obtain molecular data, which confirmed that all these clones belong to one monophyletic clade within the *P. subgibba* group (Kollár et al. in press). The valve morphology of the Scottish clones slightly resembles *P. gibba* sensu Hustedt (1930) or Krammer & Lange-Bertalot (1986) but does not correspond precisely to any named *Pinnularia* species and thus they represent a new species for science. This study therefore focuses on the description of a new species, *Pinnularia lacustrigibba* sp. nov.

*Corresponding author. E-mail: aloisie.poulickova@upol.cz

(Received 9 May 2018; accepted 18 July 2018)

Published online 09 Jan 2019

Materials and methods

Samples were collected from lakes in Scotland, UK, at altitudes of 100 m or less: the Loch of Craiglush (29.5 ha; sampled at UK National Grid Reference NO 041444, ± equivalent to 56° 34′ 54″N, 3° 33′ 46″W) on 8 December 2004, and the Lake of Menteith (212.4 ha; sampled at NN 567009, ± equivalent to 56° 10′ 43.5″N, 4° 18′ 34″W) and Loch Achray (77.1 ha; sampled at NN 506068, ± equivalent to 56° 13′ 49″N, 4° 24′ 40″W;) on 29 September 2005. Localities represent relatively unimpacted oligotrophic natural lakes in glacial hollows (kettle holes), with pH ranging from 6.4 to 7.5 and conductivity ranging from 38 to 127 µS cm^{-1} (Poulíčková et al. 2007).

To obtain epipelon including *Pinnularia* spp, fine bottom sediments were collected using a glass tube (Round 1953), transported to the laboratory in polyethylene bottles and left to sediment in plastic boxes overnight. After sedimentation in the dark, the supernatant was removed by suction, the sediment was covered with lens tissue and cover-slips were placed on the top. Under light conditions (c. 5 µmol photons m^{-2} s^{-1}), epipelic algae moved up onto cover-slips, which were collected and observed under a light microscope or used for isolating clones. Clones were isolated by streaking onto 2% agar plates and cultured in WC medium at 16–18°C, with 5–20 μmol photons m^{-2} s^{-1} from cool white tubes; the photoperiod was 12:12 h L:D (Poulíčková et al. 2007).

Cleaned valves were prepared by boiling in a mixture of nitric and sulphuric acids, washed with deionized water and mounted in Naphrax as described in detail elsewhere (Poulíčková et al. 2007). Bright field and differential interference contrast light microscopy were carried out using a Zeiss Axioimager (Oberkochen, Germany) with Zeiss Axiocam MRc5 digital camera and Zeiss Axiovision imaging software.

For scanning electron microscopy, cleaned valves were dried onto cover-slips attached by carbon pads to aluminium stubs, coated with platinum for 2 min in an Emitech K575X sputter coater, and examined using an LEO Supra 55VP Field Emission SEM operated at 5 kV. Images were captured as 3 Mb TIFF files. Voucher material and slides of clones and source populations are kept in the diatom herbarium of the Royal Botanic Garden Edinburgh (acronym E). The following material was examined:

(1) E3668. –sample of natural epipelon from the Loch of Craiglush, 8 December 2004.
(2) E3585 – sample of natural epipelon from Loch Achray, 29 September 2005.
(3) E3588 – sample of natural epipelon from the Lake of Menteith, 29 September 2005.
(4) E3602 – clone PIN2Cra isolated from the Loch of Craiglush, sampled 8 December 2004.
(5) E3603 – clone PIN12Cra isolated from the Loch of Craiglush, sampled 8 December 2004.
(6) E2604 – clone PIN13Cra isolated from the Loch of Craiglush, sampled 8 December 2004.
(7) E3605 – clone PIN19Cra isolated from the Loch of Craiglush, sampled 8 December 2004.
(8) E3680 – clone PIN592M isolated from the Lake of Menteith, sampled 29 September 2005.
(9) E3682 – clone PIN598A isolated from the Loch Achray, sampled 29 September 2005.

Terminology follows the conventions of Ross et al. (1979), Mann (1981), Barber & Haworth (1981), Round et al. (1990) and Kaczmarska et al. (2013).

Results

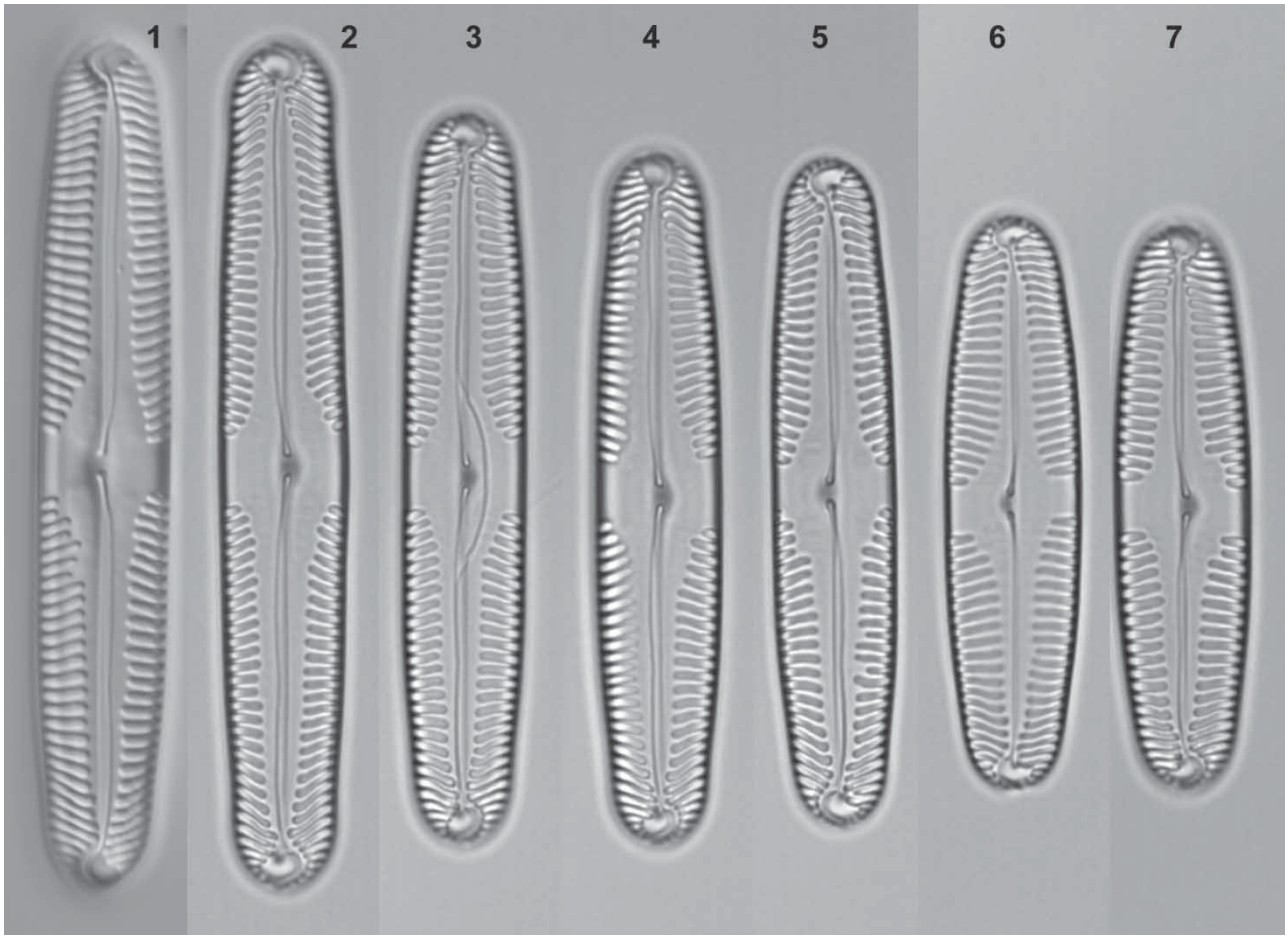
Pinnularia lacustrigibba Poulíčková, D.G. Mann & Kollár, *sp. nov.* (Figs 1–22)

LM observations

Longer valves are linear with rounded subcapitate ends (Figs 2, 8, 10, 11), whereas smaller valves have broadly rostrate ends (Figs 3–5, 12) and are almost linear-lanceolate (Figs 6, 7) as a consequence. All valves have a wide lanceolate axial area and almost always also a large transverse, unthickened area without striae (fascia) at the centre (Figs 1–13); the fascia covers approximately 7–8% of the valve length. The striae are radial near the centre and convergent at the poles (Figs 1–13). As usual in *Pinnularia*, the striae appear structureless in LM. Most valves of the clones grown in culture were 39–70 µm long, 8.5–11 µm wide (mostly < 10 µm), with 9–12 striae in 10 µm (Poulíčková et al. 2007, $n = 10$, table 1). However, following auxosporulation, cells attained lengths of up to 130 μm, with valves up to 12 μm wide (Poulíčková et al. 2007). The initial valves are sometimes slightly swollen at the centre (Poulíčková et al. 2007, fig. 54). In the source populations from which clones were obtained, the ranges of dimensions of putative *P. lacustrigibba* (e.g., Supplementary Fig. 1, from slide E3668) were similar to those of cultivated vegetative cells.

SEM observations

SEM reveals the valves to be shallow, especially around the poles, since there is a subapical 'step' in the height of the mantle (Fig. 14, arrows). The alveolate striae are almost fully open to the inside of the valve (i.e., the inner foramen of each alveolus is almost as large as the stria itself: Figs 15, 17, 20). Externally, each stria opens via 4–5 rows of small round poroids (Fig. 16), which are often more obvious internally than externally (Figs 15, 21, 22, contrast Fig. 18), since each poroid is occluded near or at its outer aperture by a hymen (sieve plate) with pores in a regular scatter (Fig. 19). The hymenes and the outer surfaces of the transapical ribs (virgae) combine to produce a smooth and, at first sight, structureless outer 'skin' to the valve (Fig. 18).



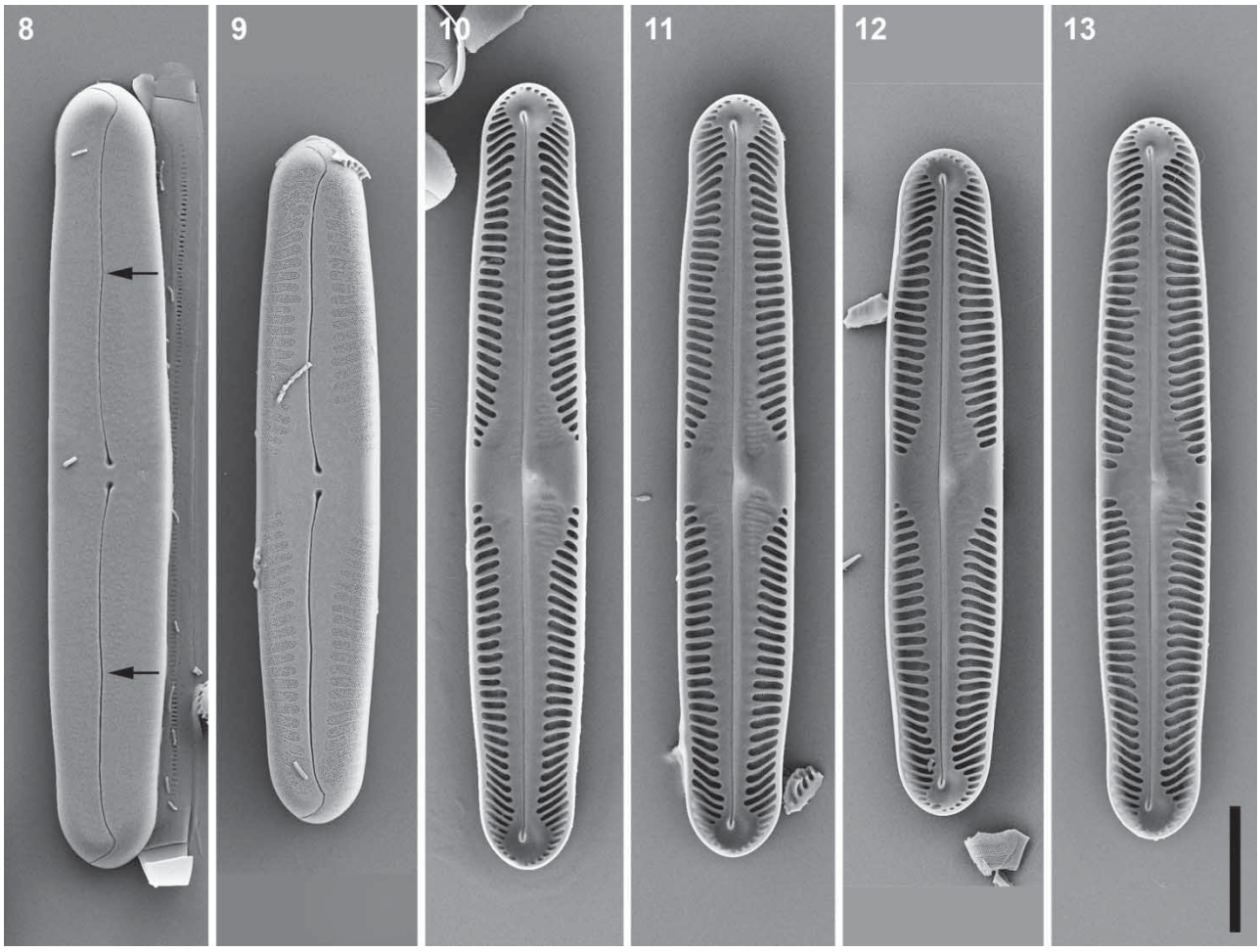
Figs 1–7. *Pinnularia lacustrigibba*, LM, Fig. 1. Natural sample from the Loch of Craiglush, 8 December 2004 (E3668), Fig. 2. Clone PIN19Cra (E2605), Figs 3, 4. Clone PIN12Cra (E3603), Fig. 5. Clone PIN2Cra (E3602), Figs 6, 7. Clone PIN13Cra (E3604). Scale bar = 10 µm.

The wide axial area bears many flat, roundish, blister-like accretions of silica externally (Figs 16, 18). Internally near the centre, the axial area often bears a series of transapically elongate markings (Figs 17, 20, contrast Fig. 21), which resemble infilled striae (cf. Cox 1999, who illustrated stages in the development of a similar *Pinnularia* species, though we have no evidence that this is how the fascia is formed in *P. lacustrigibba*). These markings are more pronounced and extensive on the primary side of the valve (Figs 10–13, 17, 20; in these photographs the primary side is on the right) and they are sometimes detectable in LM. The fascia is not thickened externally or internally and is simply a lateral expansion of the axial area, from which it is poorly differentiated. Occasionally the fascia is not developed on one or both sides of the valve, so that the striae continue around the centre of the valve (Fig. 21, contrast Fig. 20).

The raphe is 'oblique' (sensu Barber & Haworth 1981, fig. 12), i.e., the outer and inner raphe fissures are displaced from each other so that the raphe appears as two parallel lines in LM (Figs 1–7). Close inspection reveals, however, that whereas the inner fissure is almost exactly straight (Figs 10–13), the outer fissure not only bends towards the primary side at the centre (Figs 8, 9, 16, 18), but is also very slightly biarcuate, with an inflexion midway between centre and poles (e.g., Fig. 8, arrows). The external central raphe endings are expanded and pore-like and lie close to each other (Figs 8, 9, 18). The terminal fissures are hooked towards the secondary side (Figs 8, 9, 16) and continue onto the mantle. The internal raphe fissures are often continuous across the centre, the central endings being overlaid by a flange of silica extending beneath them from the primary side (Fig. 17). However, this feature is variable. Sometimes the internal fissure is clearly kinked at the centre (Fig. 21) and sometimes the fissures curve off and terminate (Fig. 20), though the interruption to the raphe is small. The valvocopula is a wide band bearing a single row of elongate poroids (Fig. 8, at right; Fig. 22).

Live cells and life cycle

Live cells and life cycle characteristics have been described and illustrated by Poulíčková et al. (2007). Cells



Figs 8–13. *Pinnularia lacustrigibba*, whole valves, SEM, zero tilt. Figs 8, 9. External views. Note the slightly biarcuate course of the external raphe fissures, with an inflection at the arrows (Fig. 8). Figs 10–13. Internal views. All are presented with the primary side of the valve to the right. Comparison of Figs 8, 9 with Figs 10–13 shows that the striae have the same transapical extent both externally and internally, reflecting their simple structure, with only very slight development of an internal wall in each alveolus; hence the internal foramen of each alveolus is almost as large as the stria. Figs 8, 10 and 11 show clone PIN19Cra (from unmounted material corresponding to the holotype E3605), Figs 9 and 12 show clone PIN12Cra (E3603 material), and Fig. 13 shows clone PIN598A (E3682 material). Scale bar = 10 µm.

have two chloroplasts, one appressed to each side of the girdle; there is no connection between them at the centre. Each chloroplast contains a small cushion-like pyrenoid at the centre, which is penetrated by a branching system of channels. The nucleus is apically elongate and lies in a bridge of cytoplasm between two large vacuoles, which occupy most of the cell lumen. The nucleus is slightly constricted by the pyrenoids and by the inward thickening of the valves around the central raphe endings. Vegetative cells move actively and produce extracellular polymeric substances (mucilage).

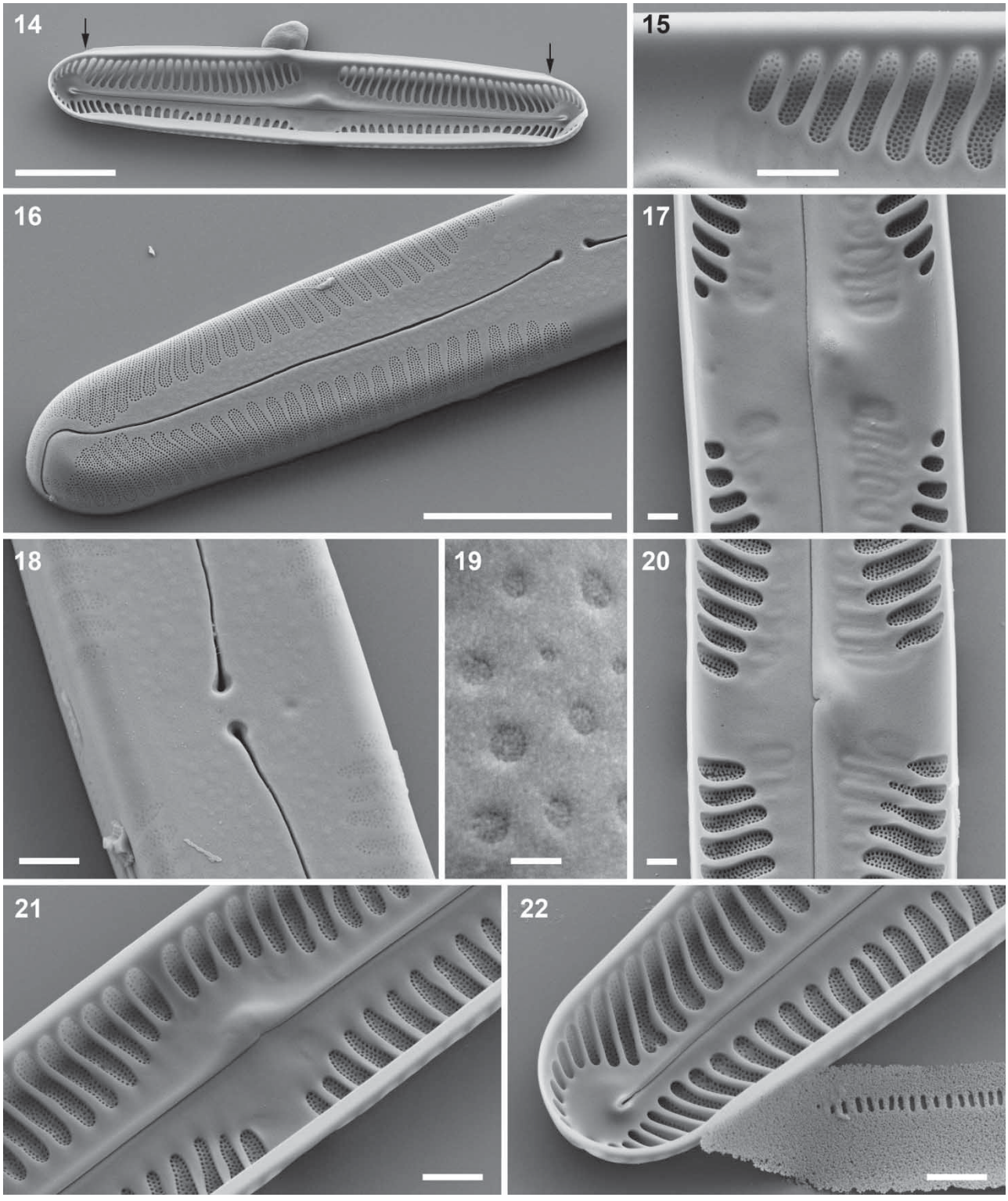
Clones with cells less than 60 µm long can reproduce sexually and exhibit heterothallic auxosporulation (Poulíčková et al. 2007), which takes place within a diffuse mucilage capsule and follows the pattern classified by Geitler (1973) as type IC (i.e., auxospores are formed via morphological and behavioural isogamy, in which each gametangium produces two gametes; the orientations of the expanding auxospores are not constant in relation to the gametangia).

Holotype: Voucher slide E3605, clone PIN19Cra isolated from the Loch of Craiglush, 8 December 2004. The slide is deposited in the diatom herbarium of the Royal Botanic Garden Edinburgh, UK.

Isotype: Voucher slide BR-4522, Botanic Garden Meise, Belgium

*Paratypes***:** Voucher slide E3603, clone PIN12Cra isolated from the Loch of Craiglush; slide E3668, a sample of the natural epipelon of the Loch of Craiglush (NO 041444), Scotland, UK., collected by D.G. Mann & A. Poulíčková, 8 December 2004. Voucher slides are deposited in the diatom herbarium of the Royal Botanic Garden Edinburgh, UK.

Genotype: Sequences (D1–D3 LSU rDNA, *cox*1-5P, *rbc*L-3P) of the holotype clone PINN19Cra (Process ID



Figs 14–22. *Pinnularia lacustrigibba*, SEM, specimens tilted 25° to the horizontal. Fig. 14. Whole valve, interior, showing the shallowness of the valve diminution in mantle height near the poles (at arrows); natural population, Loch Achray (E3605 material). Fig. 15. Internal detail of the alveolate multiseriate striae; clone PIN592M (E3680 material). Fig. 16. Slightly eroded half-valve, external view, showing the slightly biarcuate external raphe fissure, hooked terminal fissure, expanded central raphe endings deflected towards the opposite side from the terminal fissures (e.g., towards the primary side of the valve), and multiseriate striae of small round poroids; natural population, Loch Achray (E3605 material). Fig. 17. Valve centre, internal view, showing apparently continuous internal raphe fissure, fascia, and 'infilled' striae, especially on the primary side (right); natural population, Loch Achray (E3605 material). Fig. 18. Valve centre, outside; note the smooth external surface of this ± pristine valve, with hymenes still in place (contrast Fig. 16); clone PIN12Cra (E3603 material). Fig. 19. Detail of hymenes; clone PIN598A (E3682 material). Fig. 20. Internal view, valve centre, showing an intermissio separating two raphe branches; clone PIN598A (E3682 material). Fig. 21. Valve centre, internal view, showing an apparently continuous but slightly inflected internal raphe fissure and the absence of a fascia on one side of the valve; clone PIN592M (E3680 material). Fig. 22. Valve pole, internal view, showing helictoglossa; clone PIN592M (E3680 material). Scale bars = 10 μm (14, 16), 2 μm (15, 18, 21, 22), 1 μm (17, 20), or 100 nm (19).

PGIB003-14) were deposited in the public dataset DS-PIN2 (DOI: dx.doi.org/10.5883/DS-PIN2) in the Barcode of Life Database (BOLD; https://www.boldsystems.org). Sequences have also been deposited in GenBank (https://www.ncbi.nlm.nih.gov) with accessions MH670 426 (D1–D3 LSU rDNA), MH681074 (*cox*1-5P), and MH670406 (*rbc*L-3P; Kollár et al. in press).

Etymology: *Pinnularia lacustrigibba* means a species similar to *P. gibba*, living in lakes.

Habitat: Lake epipelon.

Ecology and distribution: All our strains were isolated from the epipelon of Scottish lakes. Six strains were isolated from the Loch of Craiglush (NO 041444; 29.5 ha), two from the Lake of Menteith (NN 567009; 212.4 ha) and one from Loch Achray (NN 506068; 77.1 ha). All three lochs are classified as 'oligotrophic' in the Scottish Natural Heritage lochs database (data reproduced by Palmer & Roy 2001) and have circumneutral pH (6.4–7.5) and quite low conductivity (38–127 µS cm^{-1}). Further details are given by Poulíčková et al. (2007), from long-term data on Achray and Menteith (personal communication of Dr. Laurence Carvalho to Poulíčková et al. 2007) and other data (from Wingfield et al. 2004).

Discussion

This species was earlier identified and reported as *Pinnularia* cf. *gibba* in Poulíčková et al. (2007). The identification was explained in a supplementary file giving a detailed account of the taxonomy of the species and comparisons with similar taxa. Unfortunately, for reasons that are unclear to us, the file was never made available to readers. The following discussion is based largely on the account we wrote in 2007, updated to take account of the conclusions drawn by Kollár et al. (in press).

According to two of the most influential freshwater diatom floras of the twentieth century – namely Hustedt (1930) and Krammer & Lange-Bertalot (1986) – the clones isolated from the Loch of Craiglush, Lake of Menteith and Loch Achray, and the corresponding natural material, could all have been classified as *P. gibba.* However, the high degree of variation documented for *P. gibba* by Hustedt and Krammer & Lange-Bertalot contrasts strongly with the morphological consistency within the Craiglush clones, and the close resemblance between them and the Menteith and Achray clones (illustrated and quantified by Poulíčková et al. 2007, figs 1–9 and table 1, and shown here in Figs 1–13). Other authors have also included a wide variety of different morphologies within *P. gibba* (e.g., Tynni 1976, Germain 1981, Hartley et al. 1996, pls 200, 201. Hartley et al. (1996) use a different name, *P. abaujensis* (Pantocsek) R. Ross, which was used instead of *P. gibba* following the advice of Ross [1947]; see also Hartley 1986, p. 597, note 18), again contrasting with the uniformity of the morphology in our clones and natural material.

Our results are therefore consistent with Krammer's (1992, 2000) conclusion that *P. gibba* sensu Hustedt (1930) or Krammer & Lange-Bertalot (1986) was heterogeneous and needed revision, which Krammer tried to achieve by separating several new species, morphotypes and varieties from *P. gibba*. One of the first species he separated from *P. gibba* was *P. subgibba* Krammer and according to Krammer's first monograph of European *Pinnularia* (Krammer, 1992), all our clones and natural material would fall within this species. However, in its 1992 circumscription, *P. subgibba* was still more heterogeneous than in our material. For example, it contained forms varying from 5.5 to 15 µm in width (in contrast to 8–12 µm in our material: this paper and Poulíčková et al. 2007, table 1), and forms both with and without a fascia (Krammer 1992, pl. 46, figs 5–7, pl. 47, figs 5, 6), even among specimens from the type variety and type locality (Rhiconich, near Durness, NW Scotland). By contrast, absence of a fascia was very rare in our material, which included the full range of sizes exhibited during the life cycle, from gametangia to post-auxospore valves. Almost all our specimens resemble only some of Krammer's (1992) valves of *P. subgibba*, especially one illustration of *P. subgibba* var. *subgibba* (his pl. 46, fig. 6) and one of *P. subgibba* var. *undulata* Krammer (1992, pl. 46, fig. 5).

In Krammer's later treatment of European *Pinnularia* (Krammer, 2000), the *P. subgibba* group was revised further. *Pinnularia subgibba* var. *subgibba* was again regarded as a taxon that varies with respect to the presence of a fascia (with a fascia in pl. 64, fig. 9, without in other figures), but *P. subgibba* var. *undulata* was illustrated as possessing a large fascia (Krammer 2000, pls 64, 66), as in our material. In this second monograph, however, a further species that closely resembles our material was described from Dresden, Germany, namely *P. parvulissima* Krammer (2000, p. 95, pl. 65, figs 9, 10; pl. 69, figs 7–11). The text gives no differential diagnosis to clarify the distinction between *P. parvulissima* and *P. subgibba* or other members of the *P. gibba* group ('the broad axial area and the markings combine the taxon with the *gibba* group') and the only species with which *P. parvulissima* is specifically compared is '*Navicula parva* sensu Grunow', which is not illustrated, does not seem to be referred to elsewhere by Krammer, either in 1992 or 2000, and seems to have no formal status. The illustrations and data of *P. parvulissima* and *P. subgibba* var. *undulata* provided by Krammer (2000) indicate great similarity in all respects (e.g., compare Krammer's illustrations of long specimens of *P. parvulissima* in pl. 69, figs 10, 11 with the short valves of var. *undulata* in pl. 64, figs 5 and 7), but Krammer's table of metric characters (Krammer 2000, table 8) indicates a slight difference in valve width, *parvulissima* having widths above 10 µm, whereas valves measuring less than 10 µm would belong instead to *P. subgibba* var.

undulata, even though valves of var. *undulata* are generally longer than those of *P. parvulissima* (52–84 µm and 34–70 µm respectively). However, one of the valves illustrated by Krammer (2000) for *P. parvulissima* (pl. 69, fig. 10) measures less than 10 µm in width, while at least two of the valves of var. *undulata* shown by Krammer (2000, pl. 66, figs 3, 4) are more than 10 µm wide. Valves identified by Bey & Ector (2013) as *P. subgibba* var. *undulata* have widths of c. 10 µm: the eight valves they show with lengths of 58–82 µm (i.e., within the range specified for var. *undulata* by Krammer 2000) measure 9.6–10.4 µm in width, with an average of 10.05 µm. They are therefore on the borderline between Krammer's two taxa. These inconsistencies and apparent intergradations throw some doubt on Krammer's separation of *P. parvulissima* and *P. subgibba* var. *undulata* and in 2007 we decided (Poulíčková et al. 2007) to avoid using either name for the species we studied. Instead, we referred to our clones as 'cf. *gibba*', because '*gibba*' would be more familiar than any others currently available and, as applied by Hustedt (1930) and Krammer & Lange-Bertalot (1986), also appropriate for our material.

In summary, given the current state of knowledge, it seems impossible to distinguish with reasonable certainty *P. lacustigibba*, *P. parvulissima*, clade J sensu Kollár et al. (in press; identified as *P.* cf. *parvulissima*), *P. subgibba* var. *subgibba*, *P. subgibba* var. *sublinearis* and *P. subgibba* var. *undulata* solely on the basis of LM (Supplementary Table 1). The key differences between described morphotaxa are often too subtle to allow for unambiguous identification in LM (e.g., valve sides 'very slightly convex' and 'parallel' in *P. subgibba* var. *subgibba* and *P. subgibba* var. *sublinearis*, respectively; Krammer 2000), especially considering the morphological variability we observed within polyphasically delimited species *P. lacustrigibba* (compare valve sides in Figs 1–7). More detailed investigations of valve ultrastructure, environmental requirements and geographical ranges of genetically distinct units within the morphotaxa are necessary to discover whether they can be distinguished without application of molecular techniques. Another similar taxon is *P. pseudogibba* var. *pseudogibba*, which may perhaps be distinguishable from *P. lacustrigibba* by the fascia, which is usually absent or strongly asymmetrical in *P. pseudogibba* (Krammer 2000).

Several pieces of information now persuade us to erect a new species for the clones referred to earlier as 'cf. *gibba*' by Poulíčková et al. (2007). These are (1) the availability of DNA sequence data that allow cells to be characterized and identified independently of morphology; (2) evidence from these DNA data of the existence of at least two separate species among diatoms assignable morphologically to *P. parvulissima–P. subgibba* var. *undulata*; (3) a difference in valve width between these species, one having valves almost always < 10 µm wide, which is supposed to be the lower limit for *P. parvulissima* according to Krammer (2000); and (4) the fact that this narrower species seems to have no name at the species level. The DNA data referred to above [in (1)] have been obtained as a result of progress in revising *Pinnularia* made by Kollár et al. (in press) using a polyphasic approach to delimit species within the *Pinnularia subgibba* group (sensu Souffreau et al. 2011). Kollár et al. used highly variable molecular markers (D1–D3 LSU rDNA and *cox*1-5P) to examine relationships and speciation in a dataset of 105 *Pinnularia* clones and showed the existence of 15 separate lineages, each worthy of species status. One of these clades (designated 'K') comprises the reproductively compatible clones from Scottish lakes studied by Poulíčková et al. (2007), while its sister clade (designated 'J') comprises clones with wider valves (c. 13 µm). Pending publication of the full study by Kollár et al. (in press), representative sequences of clades J and K have been deposited in GenBank and BOLD (for the type strain of *P. lacustrigibba,* PIN19Cra, and for a wider-valved clone of clade J, PIN877TM, respectively) to illustrate their differentiation. The genetic difference between sequences of PIN19Cra and PIN877TM was found to be 2.3% (20 bp), 5.5% (36 bp) and 0.4% (6 bp) for LSU rDNA, *cox*1 and *rbc*L, respectively. There was no (LSU rDNA) or low (*cox*1; PIN598A differed in two bp from the other four examined strains) intraspecific variability in the five strains of *P. lacustrigibba* examined by Kollár et al. (in press). Moreover, there are 2 bp differences between *P. lacustrigibba* and its closest relative (clade J) within the 300 + bp *rbc*L barcode regions developed for biomonitoring in the United Kingdom and France (Vasselon et al. 2017, Kelly et al. 2018). Therefore, based on the data at hand, *P. lacustrigibba* seems to be unambiguously distinguishable from its sister species with the *rbc*L regions currently being used in diatom metabarcoding. Given that, if *P. parvulissima* were to be used as the name of either clade, it would be more appropriate for the wider-valved clade J, we propose the name *Pinnularia lacustrigibba* Poulíčková, D.G. Mann et Kollár sp. nov. for the clade K clones and their source populations.

It might be argued that, since the clones and natural populations of *P. lacustrigibba* are very similar to *P. subgibba* var. *undulata*, overlapping with it in dimensions and valve morphology, they should be assigned to the same taxon and var. *undulata* consequently elevated in status to a separate species. However, there are no molecular data to confirm that Krammer's *P. subgibba* var. *undulata* and *P. lacustrigibba* are truly conspecific (the possibility of cryptic or semicryptic species in this group is already evident from the molecular differentiation between *P. lacustrigibba* and the *parvulissima*-like clade J revealed by Kollár et al.), and in addition the name '*Pinnularia undulata*' has already been given to a different diatom by Gregory (1854). Furthermore, synonymizing *P. subgibba* var. *undulata* and *P. lacustrigibba* would imply that the geographical distribution (hence the known ecological range) of *P. lacustrigibba* includes the localities from

which var. *undulata* is recorded in Franconia, Germany (the type locality at 'Erlangen, Franken, Bavaria' and other 'diverse records from Franken': Krammer 2000), which is potentially testable by new isolations from Franconia or metabarcoding but cannot currently be confirmed. We therefore see no advantage at present in making an explicit nomenclatural link between *P. subgibba* var. *undulata* and *P. lacustrigibba*, i.e., we reject the idea of placing *P. subgibba* var. *undulata* in synonymy with *lacustrigibba* and hence also the option of indicating the new taxon as '*stat. nov., nom. nov.*'

Species concepts and geographical distribution in relation to global diversity are among the most discussed problems in relation to the taxonomy, distribution and dispersal of microorganisms. In contrast to non-sexual microorganisms such as cyanobacteria (Dvořák et al. 2015), the biological species concept (Mayr 1942, 1946, Mann 1999) can potentially be applied to most diatoms, since most diatoms reproduce sexually and the main principles of their life cycle have been known for almost 150 years (Pfitzer 1869). However, field observations of sexual reproduction are rare because this phase occupies a short period (weeks) within a very long (several years) life cycle (Mann 1988, 2011). The same type of heterothallic reproduction was observed in clones isolated from three lakes and test crosses showed compatibility between them (Poulíčková et al. 2007), consistent with the molecular evidence that all belong to a single clade. In this case, the phylogenetic and biological species concepts lead to identical conclusions concerning species status.

Finally, our observations of *P. lacustrigibba* illustrate two points of more general interest in relation to *Pinnularia* morphology. The first is that many published illustrations of *Pinnularia* species show valves in which the stria pores are clearly visible from the outside of the valve (e.g., in Krammer 1992, 2000, Pillsbury & Slavik 2006, Pereira et al. 2014, Pinseel et al. 2016, Zahng et al. 2016). In these, the finest structures and pore occlusions have been destroyed. In valves still in almost pristine condition, such as those of *P. lacustrigibba* shown here in Figs 18, 19, each of the small round poroids in the multiseriate striae is occluded externally by a hymen (sieve plate) perforated by tiny (< 10 nm diameter) circular holes (Fig. 19). Because the hymenes are placed very close to the external apertures of the poroids, they combine with the almost planar external surfaces of the striae and transapical ribs to form what appears to be a smooth external 'skin' to the valves, which is evident in Fig. 18 and also in some published illustrations of other taxa (e.g., Van de Vijver & Zidarova 2011, Zidarova et al. 2016). It is only when the hymenes are destroyed by over-zealous cleaning or through dissolution during storage that the poroids become clearly visible in external views (Fig. 16).

The second point illustrated by *P. lacustrigibba* is that two morphological features often used to help characterize taxa of *Pinnularia*, the intermissio and the fascia, can vary within a group that, according to the phylogenetic or biological species concepts, comprises a single species. The intermissio is a break in the raphe visible internally at the centre. Some *Pinnularia* species have no intermissio and the internal raphe fissure therefore appears continuous from pole to pole, the raphe endings being hidden by silica overgrowth. Krammer (2000) illustrates this for *P. subgibba* in his pl. 72, figs 3 and 5) and there are many other examples (e.g., the three species studied by Luís et al. 2012). Other *Pinnularia* species have a more or less distinct intermissio internally, separating the raphe into two branches, the ends of which are hooked towards the primary side of the valve at the centre. Examples are *P. wuyiensis* (Zahng et al. 2016) and *P. catenaborealis* (Pinseel et al. 2016). In *P. lacustrigibba* this character varies, even within the same clone (particularly in clone PIN598A), though the intermissio is never wide. It will be important, therefore, not to place much emphasis on this character in *Pinnularia* taxonomy, except where the degree of phenotypic plasticity can be clearly demonstrated. Similarly, the existence of a few valves without a fascia in clones of *P. lacustrigibba* indicates the need for caution in the use of this character too, and indicates that some *Pinnularia* species could indeed be heterogeneous with respect to the presence of a fascia, as in Krammer's (1992, 2000) revisions. Variation in fascia development and internal raphe structure within a species may be stochastic, but could also be a response to particular environmental conditions (perhaps silicate availability, since production of a fascia would utilize more Si than formation of striae), reflected in the arrest of valve development at slightly different stages (cf. the ontogeny of a *Pinnularia* species documented by Cox 1999). Unfortunately, as Cox (2014) has pointed out, 'experimental studies have been conducted on only a tiny proportion of diatoms, so we have little idea of the extent to which morphology could be modified in the majority of taxa, or of the potential environmental signals reflected in morphology.'

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Internal Grant agency of Palacký University under [grant number PrF–2018–001]; Grant agency of the Czech Republic under [grant number 206/07/0115] (AP) and Endowment Fund of Palacký University in Olomouc (JK).

Supplemental data

Supplemental data for this article can be accessed at https://doi.org/10.1080/0269249X.2018.1513869.

References

Barber H.G. & Haworth E.Y. 1981. *A guide to the morphology of the diatom frustule with a key to the British freshwater genera*. Freshwater Biological Association, Ambleside. Scientific Publication 41. 112 pp.

Bey M.-Y. & Ector L. 2013. *Atlas des diatomées des cours d'eau de la région Rhône-Alpes.Tome 4. Naviculacées: naviculoidées*. Direction régionale de l'Environnement, de l'Aménageement et du Logement Rhône-Alpes, Lyon. 531–750 pp. Available from: http://www.auvergne-rhone-alpes.developpement-durable.gouv.fr/atlas-des-diatomees-a3480.html.

Cox E.J. 1999. Variation in patterns of valve morphogenesis between representatives of six biraphid diatom genera (Bacillariophyceae). *Journal of Phycology* 35: 1297–1312.

Cox E.J. 2014. Diatom identification in the face of changing species concepts and evidence of phenotypic plasticity. *Journal of Micropaleontology* 33: 111–120.

Dvořák P., Casamatta D.A., Hašler P. & Poulíčková A. 2015. Putative extremely long evolutionary stasis in bacteria might be explained by serial convergence. *Proceedings of the National Academy of Sciences of the United States of America* 112: E2559. doi:10.11646/phytotaxa.170.1.3.

Geitler L. 1973. Auxosporenbildung und Systematik bei pennaten Diatomeen und die Cytologie von *Cocconeis*-Sippen. *Österreichische Botanische Zeitschrift* 122: 299–321.

Germain H. 1981. *Flore des diatomées*. Boubée, Paris. 444 pp.

Gregory W. 1854. Notice of the new forms and varieties of known forms occurring in the diatomaceous earth of Mull; with remarks on the classification of the Diatomaceae. *Quarterly Journal of Microscopical Science* 2: 90–100.

Guiry M.D. & Guiry G.M. 2018. AlgaeBase [online]. Available from: http://www.algaebase.org [Accessed 16 January 2018].

Hartley B. 1986. A check-list of the freshwater, brackish and marine diatoms of the British Isles and adjoining coastal waters. *Journal of the Marine Biological Assocation of the United Kingdom* 66: 531–610.

Hartley B., Barber H.G., Carter J.R. & Sims P.A. 1996. *An atlas of British diatoms*. Biopress, Bristol. 601 pp.

Hustedt F. 1930. Bacillariophyta. In: *Die Süsswasser-Flora Mitteleuropas* (Ed. by A. Pascher), Vol. 10, G. Fischer, Jena, 466 pp.

Kaczmarska I., Poulíčková A., Sato S., Edlund M.B., Idei M., Watanabe T & Mann D.G. 2013. Proposals for a terminology for diatom sexual reproduction, auxospores and resting stages. *Diatom Research* 28: 263–294.

Kelly M., Boonham N., Juggins S., Kille P., Mann D., Pass D., Sapp M., Sato S. & Glover R. 2018. *A DNA based diatom metabarcoding approach for Water Framework Directive classification of rivers*. SC140024/R, Environment Agency, Bristol. ISBN: 978-1-84911-406-6.

Kociolek J.P., Balasubramanian K., Blanco S., Coste M., Ector L., Liu Y., Kulikovskiy M., Lundholm N., Ludwig T., Potapova M., Rimet F., Sabbe K., Sala S., Sar E., Taylor J., Van de Vijver B., Wetzel C.E., Williams D.M., Witkowski A. & Witkowski J. 2018. DiatomBase [online]. Available from: http://www.diatombase.org [Accessed 25 May 2018].

Kollár J., Fránková M., Hašler P., Letáková M. & Poulíčková A. 2015. Epiphytic diatoms in lotic and lentic waters – diversity and representation of species complexes. *Fottea* 15: 259–271.

Kollár J., Pinseel E., Vanormelingen P., Poulíčková A., Souffreau C., Dvořák P. & Vyverman W. (in press). A polyphasic approach to the delimitation of diatom species: a case study for the genus *Pinnularia* (Bacillariophyta). *Journal of Phycology*.

Krammer K. 1992. *Pinnularia*: eine Monographie der europäischen Taxa. *Bibliotheca diatomologica* 26: 1–353.

Krammer K. 2000. The genus *Pinnularia*. In: *Diatoms of Europe* (Ed. by H. Lange-Bertalot), Vol. 1, pp. 1–704. A.R.G. Gantner, Ruggell.

Krammer K. & Lange-Bertalot H. 1986. Bacillariophyceae 1. Teil: Naviculaceae. In: *Süsswasserflora von Mitteleuropa* (Ed. by H. Ettl, J. Gerloff, H. Heynig & D. Mollenhauer), Vol. 2/1. G.Fischer, Stuttgart & New York. 876 pp.

Luís A.T., Novais M.H., Van de Vijver B., Aleida S.F.P., Ferreira da Silva E., Hofmann L. & Ector L. 2012. *Pinnularia aljustrelica* sp. nov. (Bacillariophyceae), a new diatom species found in acidic waters in the Aljustrel mining area (Portugal) and further observations on the taxonomy and ecology of *P. acidophila* Hofmann et Krammer and *P. acoricola* Hustedt. *Fottea* 12: 27–40.

Mann D.G. 1981. Sieves and flaps: siliceous minutiae in the pores of raphid diatoms. In: *Proceedings of the 6th symposium on recent and fossil diatoms* (Ed. by R. Ross), pp. 279–300. O. Koeltz, Koenigstein.

Mann D.G. 1988. Why didn't Lund see sex in *Asterionella*? A discussion of the diatom life cycle in nature. In: *Algae and the aquatic environment* (Ed. by F.E. Round), pp. 383–412. Biopress, Bristol.

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

Mann D.G. 2011. Size and sex. In: *The diatom world* (Ed. by J.P. Kociolek & J. Seckbach), pp. 147–166. Springer, Dordrecht.

Mayr E. 1942. *Systematics and the origin of species from the viewpoint of a zoologist*. Columbia University Press, New York, 334 pp.

Mayr E. 1946. What is a species, and what is not? *Philosophy of Science* 63: 262–277

Palmer M.A. & Roy D.B. 2001. *An estimate of the extent of dystrophic, oligotrophic, mesotrophic and eutrophic standing fresh water in Great Britain*. JNCC Report 317. Joint Nature Conservation Committee, Peterborough, UK.

Pereira A.C., Torgan L.C. & Melo S. 2014. Four new *Pinnularia* Ehrenberg (Bacillariophyta, Pinnulariaceae) species from Amazonian black water (Tupé Lake, Amazonas State, Brazil). *Phytotaxa* 158: 154–168.

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

Pillsbury R.W. & Slavik K.A. 2006. *Pinnularia aldenii* sp. nov., a diatom from acidic habitats in northern Michigan. *Diatom Research* 21: 365–370.

Pinseel E., Hejduková E., Vanormelingen P., Kopalová K., Vyverman W. & Van de Vijver B. 2016. *Pinnularia catenaborealis* sp. nov. (Bacillariophyceae), a unique chain-forming diatom species from James Ross Island and Vega Island (Maritime Antarctica). *Phycologia* 56: 94–107.

Poulíčková A., Dvořák P., Mazalová P. & Hašler P. 2014. Epipelic microphototrophs: an overlooked assemblage in lake ecosystems. *Freshwater Science* 33: 513–523.

Poulíčková A., Hájková P., Kintrová K., Baťková R., Czudková M. & Hájek M. 2013. Tracing decadal environmental change in ombrotrophic bogs using diatoms from herbarium collections and transfer functions. *Environmental Pollution* 179: 201–209.

Poulíčková A., Letáková M., Hašler P., Cox E. & Duchoslav M. 2017. Species complexes within epiphytic diatoms and their relevance for the bioindication of trophic status. *Science of the Total Environment* 599–600: 820–833.

Poulíčková A., Mayama S., Chepurnov V.A. & Mann D.G. 2007. Heterothallic auxosporulation, incunabula and perizonium in *Pinnularia* (Bacillariophyceae). *European Journal of Phycology* 42: 1–24.

Ross R. 1947. Freshwater Diatomeae (Bacillariophyta). *Bulletin of the National Museum of Canada* 97: 178–233.

Ross R., Cox E.J., Karayeva N.I., Mann D.G., Paddock T.B.B., Simonsen R. & Sims P.A. 1979. An emended terminology for the siliceous components of the diatom cell. *Nova Hedwigia, Beiheft* 64: 513–530.

Round F.E. 1953. An investigation of two benthic algal communities in Malham Tarn, Yorkshire. *Journal of Ecology* 41: 174–179.

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

Souffreau C., Verbruggen H., Wolfe A.P., Vanormelingen P., Siver P.A., Cox E.J., Mann D.G., Van der 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.

Tynni R. 1976. Über Finnlands rezente und subfossile Diatomeen, IX. *Bulletin of the Geological Survey of Finland* 284: 1–37.

Van de Vijver B. & Zidarova R. 2011. Five new taxa in the genus *Pinnularia* sectio *Distantes* (Bacillariophyta) from Livingston Island (South Shetland Islands). *Phytotaxa* 24: 39–50.

Vasselon V., Rimet F., Tapolczai K. & Bouchez A. 2017. Assessing ecological status with diatoms DNA metabarcoding: scaling up on a WFD monitoring network (Mayotte island, France). *Ecological Indicators* 82: 1–12.

Wingfield R.A., Murphy K.J., Hollingsworth P. & Gaywood M.J. 2004. *The ecology of* Najas flexilis. Scottish Natural Heritage Commissioned Report No. 017. Scottish Natural Heritage, Edinburgh.

Zahng W., Pereira A.C., Kociolek J.P., Liu C., Xu X. & Wang L. 2016. *Pinnularia wuyiensis* sp. nov., a new diatom (Bacillariophyceae, Naviculales) from the north region of Wuyi Mountains, Jiangxi Province, China. *Phytotaxa* 267: 121–134.

Zidarova R., Kopalová K. & Van de Vijver B. 2016. Ten new Bacillariophyta species from James Ross Island and the South Shetland Islands (Maritime Antarctic Region). *Phytotaxa* 272: 37–62.

Appendix 4

Paper IV

On the Relativity of Species, or the Probabilistic Solution to the Species Problem

Jan Kollár*, Aloisie Poulíčková and Petr Dvořák

Department of Botany, Faculty of Science, Palacký University Olomouc, Šlechtitelů 27, CZ-78371 Olomouc, CZ

*Corresponding author: jan.kollar@upol.cz

Running title: On the Relativity of Species

Abstract

For centuries, both scientists and philosophers have discussed the nature of 'species'. Since half of the last century, it is increasingly clear that this so-called species problem is multi-layered, meaning that multiple incomparable kinds of phenomena have been covered under the term 'species'. The intermingling of these phenomena by different authors, among other things, was a source of much controversy and, to date, at least 27 species concepts were proposed. However, none of them incorporated recent advances in evolutionary genomics. By doing so, we concluded that 'species' existence as a natural discontinuous entity is relative to its dimensionality and that the most accurate way to reflect the complex species reality is thus probabilistic. We realized that due to a continual nature of evolution (regardless of its rate and constancy), species are inevitably undefinable by natural means (with exception of those originating in saltational speciation) whenever time dimension is taken into consideration. A direct consequence of the species relativity is the duality of incipient species meaning that, in a given time, they may be perceived as both being and not being a species. Therefore, we developed a probabilistic approach to the delimitation of species and explained its application using real population-genomic datasets of both prokaryotic and eukaryotic model organisms. Our results demonstrate that in the light of evolutionary genomics species reality is more complex than previously thought. In the same time, however, they demonstrate that there is a universal delimitation criterion applicable to all domains of life. We anticipate that this radically different view of species facilitates one of the most serious and universal issues of current taxonomy (the problem of incipient species) and, subsequently, may be a starting point for unification of the species-level taxonomy across the tree of life potentially influencing research in many other fields of life sciences.

Key words:

incipient species, speciation, species concept, species delimitation, species problem

A Brief History of the Species Problem

"*I have just been comparing definitions of species… It is really laughable to see what different ideas are prominent in various naturalists' minds, when they speak of 'species'… It all comes, I believe, from trying to define the undefinable.*" (C. Darwin's letter to J. D. Hooker from the 24th December 1856; Darwin 1887, p. 88)

The 'species' has always been one of the most widely used concepts in biology. At the same time, however, it is one of the most controversial concepts which potentially introduces ambiguity and confusion into many branches of life sciences. In general, one of the sources of the controversy is that different authors discussed different kinds of often incomparable phenomena using the term 'species'. One of the results is that, to date, at least 27 species concepts were proposed (Wilkins 2011) and many of them exhaustively discussed (e.g., Mayr 1982, Wheeler and Meier 2000, de Queiroz 2007, Wilkins 2009). However, as Mayr (2004) correctly pointed out, the vast majority is focused on the problem of species delimitation (i.e., on the problem of species taxa) rather than the problem of species definition (i.e., the problem of species category). Therefore, we will concise ourselves exclusively to the two broad categories of species concepts concerned with the latter, i.e., the species concepts based on reproductive isolation and the lineage-based species concepts.

The species concepts based on reproductive isolation were developed and advocated by many naturalists, the most prominent being Mayr (1942, 1963, 1982). Even though he had adjusted his famous Biological Species Concept (BSC) over the years, its core remained unchanged – *biological species are groups of interbreeding populations which are reproductively isolated from other such groups*. For the sake of species delimitation, reproductive compatibility is operational criterion. In practice, however, it is usually inferred indirectly based on other evidence at hand (e.g., morphology, ecology, behaviour; Mayr 2004) because the direct testing of the reproductive compatibility in most groups of sexual organisms is an intricate task. On the other hand, laborious crossing experiments allowed for an application of the BSC even in several groups of eukaryotic sexual microorganisms including ciliate *Paramecium* (Sonneborn 1957), filamentous fungi *Neurospora* (Dettman et al. 2008, Menkis et al. 2009), and diatoms such as *Sellaphora* (Behnke et al. 2004, Mann et al. 2004) and *Pinnularia* (Poulíčková et al. 2007).

Recent advances in evolutionary genomics provided more insight into isolation mechanisms and speciation at the genetic level. Speciation without gene flow (either allopatric or sympatric;

Richards et al. 2018) is no longer considered dominant mode of speciation. In most of the studied systems, some degree of gene flow at some phase was observed. In eukaryotes, some species emerged while facing gene flow for entire time of the speciation process (e.g., cichlids, *Heliconius* butterflies; Fan et al. 2012, Martin et al. 2013), while some made secondary contact after the speciation event (e.g., *Neurospora discreta*; Gladieux et al. 2015). Furthermore, Wu (2001) proposed genic model of speciation suggesting that the genetic divergence between species does not increase evenly over the genome. Instead, loci under positive selection and loci of the genes responsible for reproductive isolation diverge faster. The divergence of the rest of the genome is shaped rather by genetic drift, introgressive hybridization and other factors (Malinsky et al. 2015, Han et al. 2017).

One of the most serious obstacles to the universal application of the BSC is its definitional exclusion of asexual lineages. At first, Mayr (1982, p. 273) tried to resolve the issue by adjusting the concept into "*A species is a reproductive community of populations (reproductively isolated from others) that occupies a specific niche in nature.*" thus effectively merging it with those from the category of the ecological species concepts (e.g., Van Valen 1976, Andersson 1990). Nevertheless, he always considered reproductive isolation primary for the definition of the species category (Mayr 1982) and, in the end, concluded that species of sexual organisms and species of strictly asexual organisms (if such exist) are incomparable kinds of phenomena (Mayr 2004). However, it was recently shown that other than sex-mediated gene flow may play a significant role in the evolution of lineages. For example, it was shown that the frequency of the horizontal gene transfer between the closely related pair of strictly asexual prokaryotic lineages continually decreases as they genetically diverge in time. On the other hand, when the lineages are still genetically closely related, the frequency of the horizontal gene transfer can be so high that, effectively, the lineages evolve as if the sexual reproduction was involved (e.g., Dvořák et al. 2015, Bobay and Ochman 2017). Therefore, in a case of horizontal species (see below), a shift of the focus from reproduction to gene flow finally allows for extended applicability of the Biological Species Concept even to strictly asexual lineages (e.g., bacterial).

One of the most important and often neglected property of the Biological Species Concept is its 'nondimensionality'. By that Mayr (e.g., 1982, 2004) meant that his concept is truly applicable exclusively if time dimension is excluded from considerations, i.e., on species distributed over space, not through time (in our opinion, the term 'horizontal' is more appropriate adjective than Mayr's 'nondimensional' because the species as a natural units must always exist in some (physical or other; see below) dimensions). The time dimension is

incorporated into the lineage-based (or evolutionary) species concepts (Simpson 1951, 1961, Wiley 1978, Wiley and Mayden 2000, de Queiroz 2007). Let the definition of Wiley and Mayden (2000) serve as an example: "*An evolutionary species is an entity composed of organisms that maintains its identity from other such entities through time and over space and that has its own independent evolutionary fate and historical tendencies*". In other words, species are segments of separately evolving metapopulation lineages (de Queiroz 2007). However, none of them specify (1) what exactly makes the lineages evolutionarily independent, (2) where in the time dimension is the borderline between the two subsequent species and (3) how species lineages differ from other lineages which reflect metapopulation structure rather than separate species.

A Thorny Path towards the Probabilistic Solution

In our opinion, Darwin was partly right in his letter from 1856 (see above). While the end of the terminal species (i.e., the end of the species lineage) is marked by an extinction event, its origin is hidden in the so-called grey zone of a speciation event where the evolutionary change of one species into the other is continual (with exception of saltational speciation). Therefore, due to a continual nature of evolution (regardless of its rate and constancy), species is inevitably undefinable as natural discontinuous unit by natural means (except for species originating in saltational speciation event) whenever time dimension is taken into consideration. In other words, species cannot exist as natural units in time (except for species originating in saltational speciation event) because an incorporation of time dimension inevitably transforms more or less discontinuous horizontal natural entities into continual evolutionary lineages. The existence or nonexistence of species as a natural unit is thus relative to its dimensionality (Figure 1).

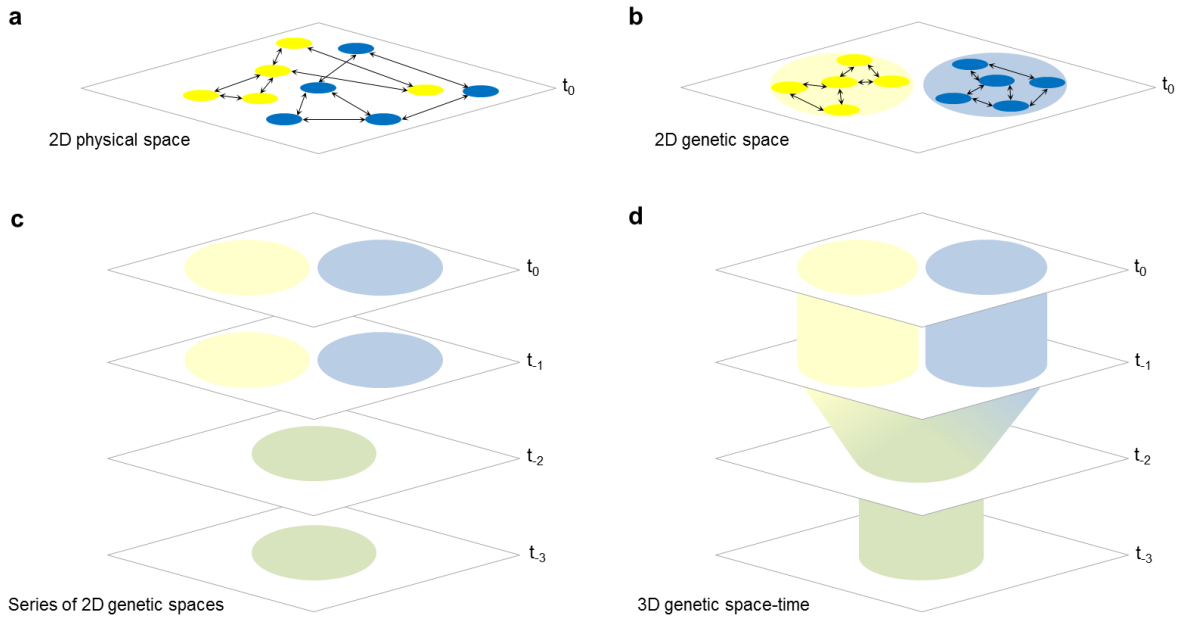


Figure 1. A simplified model of species in physical (a) and genetic space (b-d). (a) At present time (t_0), local populations (spots) of two separate metapopulations (yellow and blue) are dispersed in physical space and connected by gene flow (arrows). (b) If genetic space is considered instead of physical space, local populations of each metapopulation are grouped together. (c-d) Adding time dimension will reveal two separately evolving metapopulation lineages derived from ancestral lineage due to a restriction of gene flow between sets of local populations in t_{-2}. While it is relatively easy to delimit species in certain time plane or time period after their features sufficiently diverged (period between t_{-1} and t_0), it is difficult to delimit species lineages that split relatively recently (period between t_{-2} and t_{-1}) due to incomplete lineage sorting, possible hybridization, and/or absence of operational criteria.

The relativity of species' existence may be difficult to comprehend at first. However, Mayr (1982, p. 273) himself was aware that his 'biological species' are nonsense whenever the time dimension is incorporated into considerations. We simply extend this awareness by stating that the view of species as natural discontinuous entities in time is logically invalid and thus nonsense concept in general (except species originating in the saltational speciation event). The meaningful concept in time is that of evolutionary lineages. Luckily, in most biological fields working with species (e.g., biodiversity, ecology, ethology, conservation biology, population genetics, medicine, some branches of evolutionary biology), the units of interest are horizontal species (e.g., Stamos 2004). Nevertheless, the study of species through time is of great importance for some branches of evolutionary biology (e.g., speciation, palaeontology, natural history).

Despite species-level taxonomists are presumably dealing mostly with horizontal species, they are often bitterly exposed to the species relativity whenever they deal with incipient species. The incipient species are those which are at a given time somewhere between the levels of subpopulation and independent species. Concerning possible evolutionary tendencies, they can either further diverge from the rest of the parental metapopulation or undergo backward merging and thus never achieving relative independency of a fully developed species. However, at a given time, it is usually impossible to predict their future with reasonable certainty. Therefore, for sake of further considerations, they can be viewed as being virtually in a dual state of being and not being species at the same time.

In our opinion, the most accurate way to reflect both the species relativity and the duality of incipient species is probabilistic. In other words, incipient species under certain conditions have certain probability that they will continue to diverge from the rest of the parental metapopulation and, eventually, achieve an independency of fully developed species. In theory, this probability is computable and thus have a potential to guide even the delimitation of incipient species. Therefore, we developed a probabilistic model of speciation which we call the Universal Probabilistic Concept of Evolutionary Lineages (UPCEL; Figure 2). The UPCEL can form the theoretical basis for the further development of probabilistic species delimitation methods based on the operational criterion, which is the fundamental cause of species' inner coherence, i.e. on gene flow. In theory, UPCEL should be universally applicable to all domains of life (with exception of lineages without any gene flow, if such exist) and also to incipient species. However, to avoid possible future terminological disputes so characteristic for the species problem, we consider important to explicitly define key components of the model:

evolutionary unit = any unit that is subject to evolutionary processes, e.g., genetic marker (in the sense of concrete physical segment of nucleic acid or protein), genome, individual, population, or species.

gene flow = any transfer of genetic material (carrying genetic information) between evolutionary units.

genetic space = n-dimensional extent where evolutionary units have relative position and velocity (consisting of direction and speed). Each axis represents one dimension of the genetic space and is perpendicular to all other axes. Genetic dimensions are measured using the same units as genetic distance. The relative position of given evolutionary units can be defined by

genetic distances among these units. The relative velocity of the evolutionary units is defined by frequency of the gene flow. In other words, the frequency of the gene flow is the vector driving evolutionary units closer (when increasing) or further (when decreasing) in genetic space-time (see below).

genetic space-time = genetic space after inclusion of the time dimension.

evolutionary lineage = ancestral-descendant series of evolutionary units connected by gene flow.

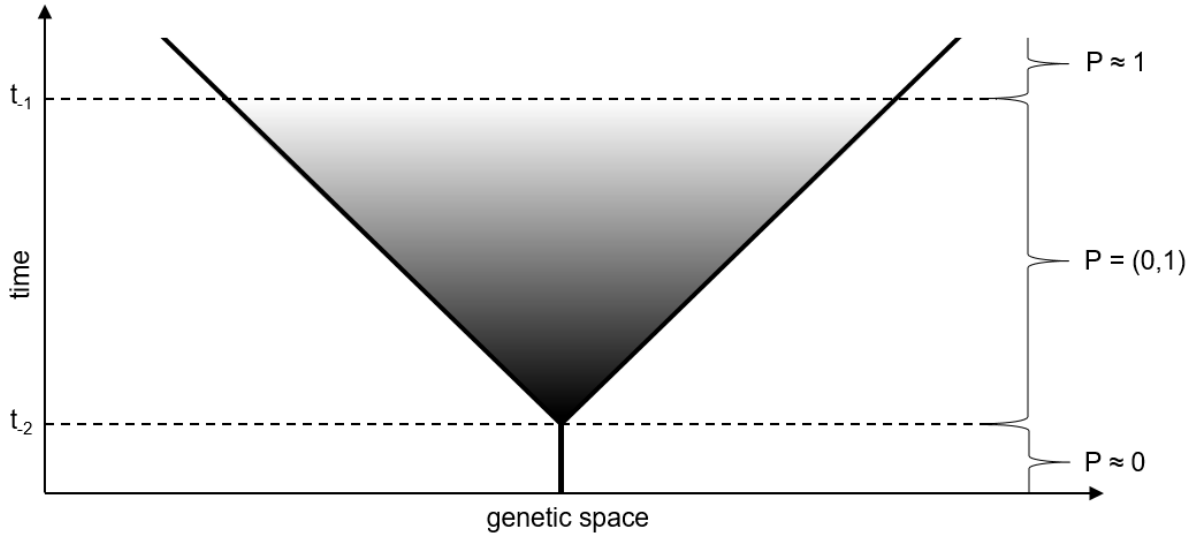


Figure 2. A simplified model of the divergence of evolutionary lineage in the genetic space-time. Continually decreasing gene flow cause parental population to split (at t_{-2}) and filial subpopulations to further diverge in the genetic space-time. In given time planes between t_{-2} and t_{-1}, the degrees of grey indicate intensity of gene flow between the two incipient species. The incipient species under certain conditions have certain probability (P) that they will continue to diverge from each other and, eventually, achieve an independency of fully developed species (at t_{-1}). 'Full' independency is achieved when the frequency of interspecific gene flow decreases to such an extent that it is no longer significant for subsequent evolution of the filial lineages. Notice that gene flow intensity does not have to be absolutely zero to be effectively insignificant to the evolution of the lineages (i.e., P approaches 1 asymptotically).

Therefore, in terms of UPCEL:

species is the cluster of populations connected with gene-flow of the frequency sufficiently high to maintain the cluster coherent and separated from other such clusters in the genetic space-

time. In the terms of UPCEL, the process of speciation is irreversible, i.e. the evolutionary unit (population or metapopulation) qualifies as species only if the probability of its backward merging with its sister unit is reasonably close to zero. The synonym for a species as defined here may be 'fully developed species'. For the sake of species delimitation, the species can be also viewed as the separate cluster of populations (or individuals) where genetic distances (in genetic space) among these populations are significantly lower than the genetic distances between any of these populations and a population from other such cluster.

and

protospecies is a population (or metapopulation) that still have not-negligible probability of backward merging with its sister population (or metapopulation). In terms of UPCEL, the pair of populations (or metapopulations) can be called protospecies when the P is within the interval (0, 1). The synonym for a protospecies as defined here is 'incipient species'. Notice that 'protospecies' is an arbitrary term allowing arbitrary discretization of continual reality introduced for the sake of clarity in potential discussions.

We further suggest that using the frequency of gene flow as a proxy for the evolutionary relationship of the two lineages may allow to infer in which phase of the potential speciation process the lineages are. Consequently, this knowledge may guide the potential taxonomic decisions. However, notice that when lineages are found to be in the protospecific phase, the taxonomic decision on whether they should be formally described as species (or, alternatively, ranked into some of the subspecific categories) must be inevitably guided by arbitrarily chosen threshold (e.g., <5% probability of backward merging of the pair of protospecific lineages in question) due to the non-existence of species (and thus also of their boundaries) in time. Nevertheless, an advantage is that such arbitrarily chosen threshold would be universally applicable throughout the tree of life because it would be computed based on one of the fundamental properties of life – the gene flow. Therefore, the unification of the species-level taxonomy could be achieved.

Beyond a theory – an application of UPCEL using population genomic data

A potential application of UPCEL may be demonstrated on the results of published evolutionary genomics studies. The gene flow between lineages can be estimated using population genomic

data. Better precision can be achieved by mapping nucleotide diversity from population samples over complete or almost complete reference genome with a physical map. This approach is laborious and expensive, but it provides excellent data for estimation of gene flow frequency and location. Here, we consider the gene flow as an introgression which leaves signatures in the genome due to recombination. This can be observed in most of sexually reproducing organisms with an exception of clonal and selfing species. Recombination is indeed widespread also among the prokaryotes (González-Torres et al. 2019). Although they lack sexual reproduction, recombination is so frequent that it appears to leave signatures similar to sexual organisms (sometimes called quasi-sexuality; e.g., Bobay and Ochman 2017). In the following lines, we will demonstrate two examples (one eukaryotic and one prokaryotic) of diverging lineages with ongoing speciation with gene flow. We will show that we can recognise different stages of speciation which have quantifiable gene flow frequency and, by doing so, exemplify assumptions of UPCEL.

Martin et al. (2013) investigated speciation continuum in the lineages of *Heliconius* butterflies. These butterflies represent an excellent study system of speciation, because their biology have been intensively studied and they possess easily distinguishable morphological features, which seem to be relevant for natural selection. Moreover, the complete mapped genome was obtained (*Heliconius* genome consortium 2012). Martin and colleagues used two metrics to quantify gene flow. Firstly, they estimated relative nucleotide diversity over the genome, pairwise, between lineages using F_{st} (e.g., Holsinger and Weir 2015). The F_{st} values range between 0 (low divergence of populations - high gene flow frequency) and 1 (high divergence of populations - low gene flow frequency). The F_{st} values density peaks showed continuum of gene flow levels (Martin et al. 2013, Figure 5). All races and species (both allopatric and sympatric) are still in contact and the barriers of gene flow are leaky. Some lineage pairs were more diverged than others which means that some have more gene flow than other. This method provides only relative evidence of difference in gene flow. An estimation of absolute gene flow frequency can be calculated by a second used metric – ABBA-BABA test (also called D-statistics; Martin et al. 2015). This statistics tests for a deviation from strictly bifurcating relationships between lineages. Martin et al. (2013) were able to quantify also the change of the gene flow fraction of the genome through different time periods between species pairs *H. timareta – H. amaryllis* and *H. cydno – H. rosina*. Between the 2 Mya and present, the fractions decreased from 15.6% to 2.1% and from 29.7% to 4.4% in the first and the second pair, respectively. Inverted fraction of the genome with gene flow could represent a distance in the genetic space in terms of UPCEL.

Papers focused on reconstruction of the gene flow frequency among prokaryotes are relatively rare. Most of the papers used only small datasets which do not have population level sampling. Thus, we cannot present example as specific as above. However, Bobay and Ochman (2017) analysed 91 bacterial species which were represented by at least 15 different genotypes (identical or nearly identical genome sequences were removed). They proposed that a ratio of homoplasic and nonhomoplasic polymorphisms is indicative of presence or absence of gene flow from other species. The key idea is that strains which have reduced homoplasic and nonhomoplasic polymorphisms ratio should have reduced gene flow. Bobay and Ochman (2017) found that there is a scale of the gene flow frequency among the bacterial species. There are species with frequent gene flow within and outside the species and therefore they are probably incipient species, i.e. they are at the beginning of speciation process. On the other hand, many species do not show signatures of recombination outside the species and they recombine only within the species. Tests such as ABBA-BABA were not extensively applied to prokaryotes and, therefore, we cannot, at the moment, present any specific estimates of the genome fraction exhibiting the gene flow.

Many layers, one solution?

The complexity of the species problem is caused in part by the fact that multiple often incomparable kinds of phenomena have been discussed under the term species. In this regard, the species problem may be considered multi-layered. For example, the distinction of species as taxon and species as category may be considered one of the layers, dividing the problem of species into the problem of species delimitation and the problem of species definition (e.g., Mayr 1982). In this regard, an analysis of one of the most neglected layers of the species problem, the dimensionality of species (horizontal vs vertical species), along with the incorporation of the recent findings in the field of evolutionary genomics, allowed us to identify yet unknown complexity of the species reality – the relativity of species existence and the duality of incipient species. Taking these into account, we suggest that the most accurate model of speciation is probabilistic and that such a model may provide the theoretical basis for development of a new kind of species delimitation methods. We further suggest that this radically different view of species as evolutionary lineages with changing probabilities of continuing divergence may facilitate one of the most serious and universal issues of current taxonomy (the problem of incipient species). Consequently, it may even be a starting point

for the unification of the species-level taxonomy across the tree of life thus potentially influencing research in many other fields of life sciences.

References

Andersson, L. 1990. The Driving Force: Species Concepts and Ecology. *Taxon*. 39:375–82.

Behnke, A., Friedl, T., Chepurnov, V.A. & Mann, D.G. 2004. Reproductive compatibility and rDNA sequence analyses in the Sellaphora pupula species complex (Bacillariophyta). *J. Phycol.* 40:193–208.

Bobay, L.-M. & Ochman, H. 2017. Biological Species Are Universal across Life's Domains. *Genome Biol. Evol.* 9:491–501.

de Queiroz, K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56:879–86.

Dettman, J.R., Anderson, J.B. & Kohn, L.M. 2008. Divergent adaptation promotes reproductive isolation among experimental populations of the filamentous fungus Neurospora. *BMC Evol. Biol.* 8:35.

Dvořák, P., Poulíčková, A., Hašler, P., Belli, M., Casamatta, D.A. & Papini, A. 2015. Species concepts and speciation factors in cyanobacteria, with connection to the problems of diversity and classificatio. *Biodivers. Conserv.* 24:739–57.

Fan, S., Elmer, K.R. & Meyer, A. 2012. Genomics of adaptation and speciation in cichlid fishes: Recent advances and analyses in African and neotropical lineages. *Philos. Trans. R. Soc. B Biol. Sci.* 367:385–94.

Gladieux, P., Wilson, B.A., Perraudeau, F., Montoya, L.A., Kowbel, D., Hann-Soden, C., Fischer, M. et al. 2015. Genomic sequencing reveals historical, demographic and selective factors associated with the diversification of the fire-associated fungus Neurospora discreta. *Mol. Ecol.* 24:5657–75.

González-Torres, P., Rodríguez-Mateos, F., Antón, J. & Gabaldón, T. 2019. Impact of homologous recombination on the evolution of prokaryotic core genomes. *MBio*. 10:1–17.

Han, F., Lamichhaney, S., Grant, B.R., Grant, P.R., Andersson, L. & Webster, M.T. 2017.

Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Res.* 1–12.

Holsinger, K.E. & Weir, B.S. 2015. Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nat. Rev. Genet.* 10:639–50.

Malinsky, M., Challis, R.J., Tyers, A.M., Schiffels, S., Terai, Y., Ngatunga, B.P., Miska, E.A. et al. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science (80-.).* 350:1493–8.

Mann, D.G., McDonald, S.M., Bayer, M.M., Droop, S.J.M., Chepurnov, V.A., Loke, R.E., Ciobanu, A. et al. 2004. The Sellaphora pupula species complex (Bacillariophyceae): Morphometric analysis, ultrastructure and mating data provide evidence for five new species. *Phycologia*. 43:459–82.

Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., Blaxter, M. et al. 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Res.* 23:1817–28.

Martin, S.H., Davey, J.W. & Jiggins, C.D. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* 32:244–57.

Mayr, E. 1942. Systematics and the Origin of Species. Columbia University Press, New York. 334 pp.

Mayr, E. 1963. Animal Species and Evolution. Harvard University Press, Cambridge, MA. 797 pp.

Mayr, E. 1982. The Growth of Biological Thought revisited. 11th ed. Harvard University Press, Cambridge, MA. 974 pp.

Mayr, E. 2004. What Makes Biology Unique? Cambridge University Press, New York. 248 pp.

Menkis, A., Bastiaans, E., Jacobson, D.J. & Johannesson, H. 2009. Phylogenetic and biological species diversity within the Neurospora tetrasperma complex. *J. Evol. Biol.* 22:1923–36.

Poulíčková, A., Mayama, S., Chepurnov, V.A. & Mann, D.G. 2007. Heterothallic

auxosporulation, incunabula and perizonium in Pinnularia (Bacillariophyceae). *Eur. J. Phycol.* 42:367–90.

Richards, E.J., Poelstra, J.W. & Martin, C.H. 2018. Don't throw out the sympatric speciation with the crater lake water: fine-scale investigation of introgression provides equivocal support for causal role of secondary gene flow in one of the clearest examples of sympatric speciation. *Evol. Lett.* 2:524–40.

Simpson, G.G. 1951. The species concept. *Evolution (N. Y).* 5:285–298.

Simpson, G.G. 1961. Principles of animal taxonomy. Columbia University Press, New York. 247 pp.

Sonneborn, T.M. 1957. Breeding systems, reproductive methods, and species problems in protozoa. *In* Mayr, E. [Ed.] *The Species Problem*. merican Association for the Advancement of Science, Washington DC, pp. 155–324.

Van Valen, L. 1976. Ecological Species, Multispecies, and Oaks. *Taxon*. 25:233–9.

Wheeler, Q.D. & Meier, R. 2000. Species concepts and phylogenetic theory: a debate. Columbia University Press, New York. 230 pp.

Wiley, E.O. 1978. The evolutionary species concept reconsidered. *Syst. Zool.* 27:17–26.

Wiley, E.O. & Mayden, R.L. 2000. The evolutionary species concept. *In* Wheeler, Q. D. & Meier, R. [Eds.] *Species Concepts and Phylogenetic Theory*. Columbia University Press, New York, pp. 70–89.

Wilkins, J.S. 2009. Species: a history of the idea. 1st ed. University of California Press, Berkeley. 320 pp.

Wilkins, J.S. 2011. Philosophically speaking, how many species concepts are there? *Zootaxa*. 2765:58-60.

Wu, C.I. 2001. The genic view of the process of speciation. *J. Evol. Biol.* 14:851–65.

Appendix 5

Author’s curriculum vitae

Curriculum vitae

Jan Kollár

E-mail: jan.kollar@upol.cz

Address: Radiměř 439, 569 07, CZ

Phone: +420 731 905 056

Research experience:

Ghent University — Jul 2014 – Sep 2014

Erasmus intern

Topic: Development of a DNA barcode database for identification and discovery of diatom species

Supervisor: Dr Pieter Vanormelingen

Academy of Sciences of the Czech Republic — Mar 2015

Intern

Topic: Morphological study of diatoms

Supervisors: Dr Markéta Fránková and Dr Petr Marvan

Ghent University — Nov 2015

Research fellow

Topic: Phylogenetic species delimitation in diatoms

Supervisor: Dr Wim Vyverman

Natural History Museum London — Apr 2018 – Sep 2018

Research fellow

Topic: Phylogenetics and systematics of diatoms

Supervisor: Dr David M. Williams

Selected experience with science popularization:

UP Crowd, non-profit organization devoted to the popularization of science — Nov 2015 – Present

Coordinator and lecturer

Fort Science, Palacký University interactive science center — Apr 2015 – Present

Lecturer

Education:

Systematic biology and ecology (BSc) Sep 2011 – May 2014

Palacký University in Olomouc

Hydrobiology (MSc, Graduated with Honours) Jun 2014 – Jun 2016

Palacký University in Olomouc

Botany (PhD) Sep 2016 – Present

Palacký University in Olomouc

Supervisor: Dr Aloisie Poulíčková

Awards and Achievements:

Jul 2014: Scholarship by Erasmus programme

Jun 2015: Scholarship by the Endownment fund of Palacký University for support of excellent junior research projects

Mar 2018: Award "Cena děkana PřF UP" (winner of the PhD Biology section)

Apr 2018: Scholarship by Erasmus+ programme

Apr 2020: One of the national finalists of FameLab 2020

Scientific societies:

The Czech Phycological Society

The International Society for Diatom Research

The Phycological Society of America

Academic conferences:

2015: Meeting of the Czech Phycological Society in České Budějovice, CZ

2016: Meeting of the Czech Phycological Society in Prague, CZ

2016: Workshop Phytobentos in Vyškovec, CZ (lecturer)

2017: Central European Diatom Meeting in Prague, CZ (poster)

2017: Meeting of the Czech Phycological Society in Ostrava, CZ (presentation)

2018: International Diatom Symposium in Berlin, DE (presentation)

2019: International Barcode of Life Conference in Trondheim, NO

Publications in international peer-reviewed journals with impact factor:

- Kollár J., Fránková M., Hašler P., Letáková M. & Poulíčková A. 2015: Epiphytic diatoms in lotic and lentic waters - diversity and representation of species complexes. *Fottea* 15(2): 259-271. (IF_{2014} = 2.190; Q2)

- Poulíčková A., Kollár J., Hašler P., Dvořák P. & Mann D. G. 2018: A new species *Pinnularia lacustrigibba* sp. nov. within the *Pinnularia subgibba* group (Bacillariophyceae). *Diatom Res.* 33, 273–282. (IF_{2018} = 1.169; Q3)

- Kollár J., Pinseel E., Vanormelingen P., Poulíčková A., Souffreau C., Dvořák P. & Vyverman W. 2019: A Polyphasic approach to the delimitation of diatom species: a case study for the genus *Pinnularia* (Bacillariophyta). *J. Phycol.* 55, 365–379. (IF_{2018} = 2,831; Q1)

- Kollár J., Pinseel E., Vyverman W. & Poulíčková A.: A time-calibrated phylogeny provides an insight into the evolution, taxonomy and DNA barcoding of the *Pinnularia subgibba* group (Bacillariophyta). *Fottea*, (*accepted*). (IF_{2018} = 1,727; Q2)

- Kollár J., Poulíčková A. & Dvořák P.: On the relativity of species, or the probabilistic solution to the species problem. *Syst. Biol.* (*submitted*). (IF_{2018} = 10,266; Q1)

Publications in pop-science magazines:

- Kollár J. (2014): Je na Lamarckismu něco pravdy? Vesmír 93, 115, 2014/2.

- Kollár J. (2014): Malý krok pro rybu, velký pro čtvernožce. Vesmír 93, 244, 2014/4.

- Kollár J. (2014): Stará nová velryba. Vesmír 93, 245, 2014/4.

- Kollár J. (2015): Přímo před našima očima. Vesmír 94, 71, 2015/2.

- Kollár J. (2015): Na velikosti záleží! Vesmír 94, 191, 2015/4.

- Kollár J. (2015): Vidět se zkrátka vyplatí. Vesmír 94, 471, 2015/9.

- Kollár J. (2016): Evoluce v přímém přenosu? Proč ne! Vesmír, 95, 387, 2016/7-8.

- Kollár J. (2016): Čtvrtý druh. Vesmír 95, 614, 2016/11.

- Kollár J. (2018): Víc škody než užitku? Vesmír, 97, 394, 2018/7.

- Kollár J. (2019): Utvářela náš genom i neznámá lidská linie? Vesmír, 98, 201, 2019/4.

- Kollár J. (2020): Geny vznikající de novo, další přelom evoluční biologie. Vesmír 99, 76, 2020/2.

Palacký University Olomouc

Faculty of Science

Department of Botany

Polyphasic approach to the study of diatom diversity

Summary of the Doctoral Thesis

by

Mgr. Jan Kollár

Supervisor: prof. RNDr. Aloisie Poulíčková, CSc.

Olomouc 2020

This doctoral thesis was carried out within the framework of internal doctoral studies at the Department of Botany, Faculty of Science, Palacký University Olomouc during the years 2016-2020.

Aspirant: **Mgr. Jan Kollár**

Department of Botany, Faculty of Science, Palacký University Olomouc

Supervisor: **prof. RNDr. Aloisie Poulíčková, CSc**.

Department of Botany, Faculty of Science, Palacký University Olomouc

Opponents: **prof. RNDr. Jiří Neustupa, Ph.D.**

Department of Botany, Faculty of Science, Charles University

doc. RNDr. Linda Nedbalová, Ph.D.

Department of Ecology, Faculty of Science, Charles University

Mgr. Markéta Fránková, Ph.D.

Department of Vegetation Ecology, Czech Academy of Sciences

The summary of the doctoral thesis has been written by Department of Botany, Faculty of Science, Palacký University Olomouc.

The summary of the doctoral thesis was sent out on …….

The oral defence of the thesis takes place in front of the Commission for doctoral thesis for the Study Subject Botany in the conference room at the Department of Botany, Faculty of Science, Palacký University, Olomouc – Holice, Šlechtitelů 27, on ……

The doctoral thesis is available in the Library of Biological Deartment of Faculty of Science at Palcký University, Olomouc – Holice, Šlechtitelů 27.

prof. Ing. Aleš Lebeda, DrSc.
Chairman of the Commission for the Ph.D.
Theses for Study Subject Botany
Faculty of Science, Palacký University Olomouc

CONTENTS

1. INTRODUCTION

Diatoms (Bacillariophyta) are one of the most diverse groups of protists with an estimated number of extant species ranging between 30,000 and 200,000 (Andersen 1992, Mann and Vanormelingen 2013). Over 12,000 diatom species have been formally described (Guiry 2012). As a group, diatoms are cosmopolitan and adapted to almost all aquatic and some terrestrial habitats. Consequently, they are of great significance not only to fundamental research in the fields of biodiversity, evolutionary biology, ecology, and paleolimnology (e.g., Lewis et al. 2008), but also for applied disciplines such as biomonitoring (e.g., Smol and Stoermer 2010), biotechnology (e.g., Hu et al. 2008), and nanotechnology (e.g., Drum and Gordon 2003). However, an enormous diversity of diatoms leads not only to their significance but also to many difficulties with their systematics.

Traditionally, the descriptions of diatom species have been based on discontinuities in frustule morphology. Therefore, the majority of traditionally described and currently accepted diatom species may be considered morphospecies, i.e., groups of organisms distinguishable by morphology (Cronquist 1978), and their underlying biological identity remains questionable. After developement of the biological species concept (Mayr 1942, Mayr and Ashlock 1991), the differences in morphology were considered to reflect reproductive isolation (Mann 1999). However, even though this assumption may be correct for some species, it cannot be generalised as was proven by the widespread occurrence of cryptic and pseudocryptic species diversity in diatoms (e.g., Mann et al. 2004, Sarno et al. 2005, Evans et al. 2007, Vanormelingen et al. 2008a, Trobajo et al. 2009, Vanelslander et al. 2009, Poulíčková et al. 2010, Souffreau et al. 2013, Pinseel et al. 2017a, Pinseel et al. 2018, Amato et al. 2019, Pinseel et al. 2019).

The cryptic and pseudocryptic species diversity means that species are often morphologically indistinguishable or distinguishable only after employment of methods with higher resolution than light-microscopical observations alone (e.g., geometric morphometry, statistical analysis of morphological data, thorough examination in the electron microscope), respectively (Mann and Evans 2008). A disentanglement of diatom (pseudo)cryptic species complexes is a vital step towards the fine-grained taxonomy reflecting the evolutionary history of diatom lineages. Moreover, correct taxonomical identifications are vital for many fundamental and applied purposes. For example, the data from autecological studies of the (pseudo)cryptic species

increase the resolution of the diatom-based bioindicative systems of water quality assessments such as the one implemented through the European Union's Water Framework Directive (e.g., Poulíčková et al. 2017).

In my opinion, when intending to study the species diversity of any group of organisms we cannot avoid asking the fundamental question of 'What is a species?' Without a doubt, the identity of species is one of the most controversial topics of post-Darwinian biology (e.g., Dobzhansky 1937, Mayr 1942, 1963, 1982, Mayden 1997, Mann 1999, Wheeler and Meier 2000, de Queiroz 2007, Wilkins 2011). To date, this so-called species problem lead to a development of at least 27 more or less different species concepts (Wilkins 2011). An exhaustive discussion on different species concepts may be found elsewhere (e.g., Mayr 1982, Wheeler and Meier 2000, Wilkins 2009) and more details are given also in the doctoral thesis. Here, I will confine myself to state that this doctoral research project was carried out within the framework of the Unified Species Concept (USC; de Queiroz 1998, 2007).

The major contribution of the USC was clear separation of the theoretical concept of what species is (i.e., the definition of species category) and the operational criteria used to delimit species taxa (e.g., morphological discontinuity, genetic discontinuity, reproductive discontinuity). The theoretical component of the unified species concept belongs to the category of evolutionary species concepts (e.g., Simpson 1951, 1961, Wiley 1978, Wiley and Mayden 2000) – *species are segments of separately evolving metapopulation lineages*. On the other hand, operational criteria emerge in time as a result of the evolutionary separation. The problem is that in different pairs of lineages different operational criteria emerge in different times, order or may not emerge at all (in case of morphological discontinuity leading to a cryptic species). Therefore, it should be borne in mind that these operational criteria may be considered just imperfect reflections of the underlying species reality which can merely help us formulate hypotheses on the numbers and limits of lineages, i.e., on their gene flow interconnection state. Since we cannot efficiently (i.e., in sufficient detail under reasonable time, effort, and financial expenses) study the gene flow directly in diatoms, we should, in the spirit of science, at least consider as many operational criteria as possible before formulating a hypothesis on the species limits. This approach to systematics is sometimes called polyphasic (Colwell 1970, Vandamme et al. 1996, Komárek 2016). It spread from bacteriologists (e.g., Colwell 1970) and cyanobacteriologists (e.g., Komárek 2016)

to protistologists (e.g., Škaloud 2008). In groups of macroorganisms this practice is called an integrative taxonomy (e.g., Will et al. 2005, Puillandre et al. 2012).

Genus *Pinnularia* Ehrenberg was selected as a focus of this research. Two of the main reasons follow. Firstly, it is among the most taxon-rich genera of raphid diatoms, with 723 currently accepted names in AlgaeBase (Guiry and Guiry 2018) and 1527 species names listed in DiatomBase (Kociolek et al. 2018), although, the taxonomic/nomenclatural status of the vast majority of the latter taxa is unassessed. In addition, the genus *Caloneis* Cleve was shown to be polyphyletic within *Pinnularia* (Bruder et al. 2008, Souffreau et al. 2011) raising the total number of accepted taxa in the *Pinnularia–Caloneis* complex (i.e., *Pinnularia* sensu lato) to 931 according to AlgaeBase and potentially making it, at the current state of knowledge, the second most taxon-rich diatom genus (outnumbered only by *Navicula* Bory with 1270 taxa, and followed by *Nitzschia* Hassall with 788 taxa; Guiry and Guiry 2018). And secondly, it contains several presumably (pseudo)cryptic species complexes. In recent years many new *Pinnularia* species have been described using the traditional morphological approach (e.g., Van de Vijver et al. 2011, Luís et al. 2012, Pereira et al. 2014) or using both molecular and morphological data (e.g., Pinseel et al. 2016). However, molecular studies suggest that many traditional *Pinnularia* morphospecies are species complexes with cryptic or pseudocryptic diversity (e.g., Souffreau et al. 2011, Pinseel et al. 2016).

The genus is cosmopolitan, occurring in a broad range of both terrestrial (e.g., Pinseel et al. 2020) and freshwater environments within assemblages such as the epipelon (e.g., Poulíčková et al. 2014), epibryon (e.g., Poulíčková et al. 2013) and epiphyton (e.g., Kollár et al. 2015, Poulíčková et al. 2017). Although their ecological demands vary, most species prefer low electrolyte content (Round et al. 1990, Krammer 2000). The genus is estimated to have originated in Paleocene some 60±15 Ma ago (Souffreau et al. 2011). To date, two DNA-based intrageneric phylogenetic hypotheses were proposed using concatenated alignments of three (SSU rDNA, LSU rDNA, and *rbc*L; Bruder et al. 2008) and five genetic markers (*cox*1, SSU rDNA, LSU rDNA, *rbc*L and *psb*A; Souffreau et al. 2011). Both revealed three major clades to be present within the genus. Clade B sensu Souffreau et al. (2011) consists of three distinct subclades called “grunowii”, “nodosa”, and “subgibba”. Because the whole genus would be too challenging for a single doctoral research project, subclade “subgibba” was selected as a model. Henceforth, it will be referred to as the “*Pinnularia subgibba* group” or “*Pinnularia gibba* group”.

The *P. gibba* group is estimated to originate in the upper Eocene around 35±10 Ma ago (Souffreau et al. 2011). Both the entire clade B and the *P. gibba* group are very well statistically supported (ML bootstrap proportions-BI posterior probabilities = 100-100) and can be distinguished by a certain combination of morphological features (Souffreau et al. 2011). Specimens of clade B are characterised by small linear cells with a drop-like central raphe endings and large alveolar openings. In addition, specimens of the *Pinnularia gibba* group are usually elongated in shape with two girdle-appressed parallel plastids and a broad non-porous central area called the fascia (Souffreau et al. 2011). Several species of the *Pinnularia gibba* group (e.g., *Pinnularia gibba* Ehrenberg, *Pinnularia parvulissima* Krammer, *Pinnularia subgibba* Krammer) are very uniform in morphology and are generally considered to be cosmopolitan (Krammer 2000). However, studies focusing on molecular species delimitations in the *P. gibba* group are entirely lacking. In the light shed by recent discoveries in the field of diatom biogeography, namely that many diatom species were shown to be geographicaly restricted or even endemic (e.g., Vyverman et al. 2007, Vanormelingen et al. 2008b, Verleyen et al. 2009, Pinseel et al. 2017b), this may suggest an underestimation of the species diversity within the group.

2. AIMS

The ultimate aim of this doctoral research project was investigating species diversity and evolution of the selected (pseudo)cryptic species complex of the genus *Pinnularia* (i.e., the *P. gibba* group). Firstly, species diversity was explored by means of a polyphasic approach. This means that both primary (by using multiple genetic markers and automated species delimitation methods) and secondary (by considering all available morphological, environmental, geographic and reproductive data associated with the studied strains) species delimitations were performed. To date, it was the first (explicit, in the case of polyphasic approach) application of both the automated DNA-based species delimitation methods and the polyphasic approach in diatoms. Secondly, the evolution of the group was investigated by inferring its time-calibrated multi-gene phylogeny. Finally, newly generated knowledge on species limits and phylogentic relationships was used to investigate whether DNA barcode markers proposed for routine identification of diatoms can unambiguously distinguish even closely related (pseudo)cryptic species of the group.

To summarize, main aims of this doctoral research project were

(1) exploring species limits within the *Pinnularia gibba* group,
(2) inferring its fossil-guided time-calibrated multi-gene phylogeny and
(3) testing the taxonomic resolution of the proposed diatom DNA barcode markers on the species of the group.

In addition, newly generated knowledge along with the available morphological, environmental, geographic and reproductive data allowed for

(4) formulating and testing of some evolutionary hypotheses and
(5) initiating the taxonomic review of the group.

3. RESULTS

3.1 On the diversity of the *P. gibba* group

(this chapter is based on Kollár et al. 2019 - reprinted as Appendix 2 of the doctoral thesis)

The species limits within the *Pinnularia gibba* group were investigated using a data set of 105 strains with world-wide origin (Figure 1) by means of a polyphasic approach (Kollár et al. 2019). Firstly, a set of nuclear, chloroplast and mitochondrial markers was examined for their sequence variability. Secondly, two most variable markers (LSU rDNA and *cox*1) and three automated molecular species delimitation methods were used to formulate the primary (i.e., DNA-based) hypothesis on species limits within the group (Figure 2). The most variable markers were used because they have the highest chance of covering intraspecific variability required by applied coalescence-based species delimitation methods. Finally, DNA-based species limits were confronted with other available lines of evidence such as morphology studied under both LM and SEM (e.g., Figure 3), geographic, environmental and/or reproductive data (all available evidence is summarized in Table 1).

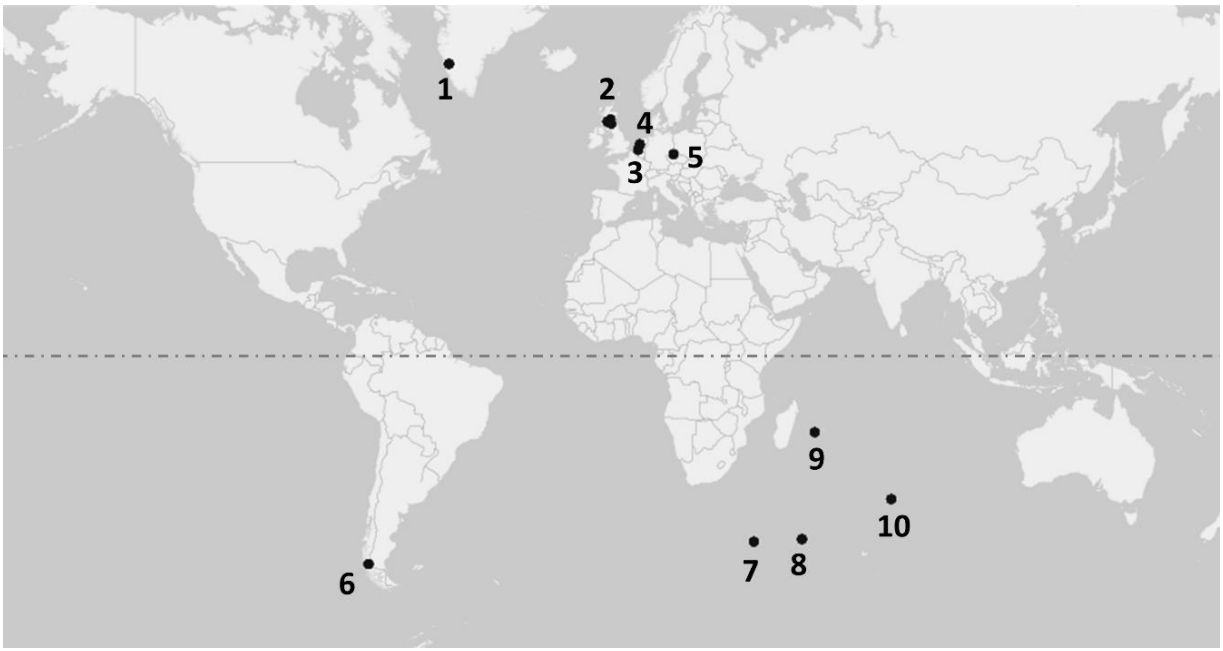


Figure 1. Geographic origin of the 105 strains of the *P. gibba* group used in this thesis. (1) Greenland – Kapisilit, (2) Scotland – Lake Menteith, Loch Achray, Loch of Craiglush, and Threipmuir reservoir, (3) Belgium – Stappersven, (4) Netherlands – De Wieden, (5) Czech Republic – Lake Mácha, (6) Chile – Torres del Paine, (7) Prince Edward Islands – Marion Island, (8) Crozet Islands – Île de la Possession, (9) Réunion, (10) Amsterdam Island (this figure was included in Kollár et al. 2019).

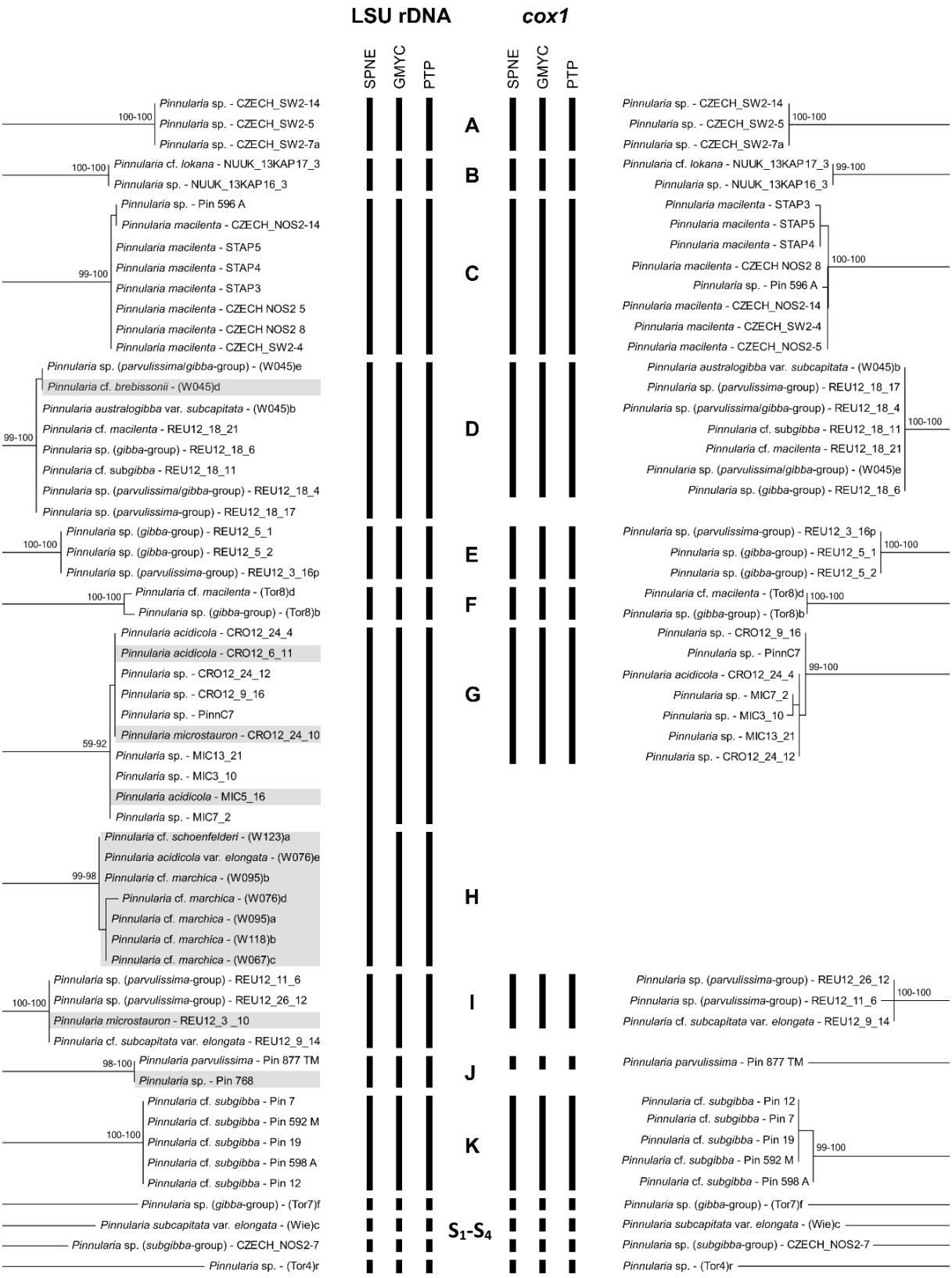


Figure 2. Comparison of species limits within the *P. gibba* group based on *cox*1 and LSU rDNA. Species boundaries were delimited using three methods (SPNE, sGMYC, and PTP) as indicated with black bars. Strains for which *cox*1 could not be amplified, are indicated in grey in the LSU rDNA data set. The visualisation is based on a ML phylogeny. The numbers at the nodes represent ML bootstrap supports - BI posterior probabilities. The

four species represented by single strains (depicted in the bottom of the diagram) are coded S_1-S_4 (this figure was included in Kollár et al. 2019 - edited).

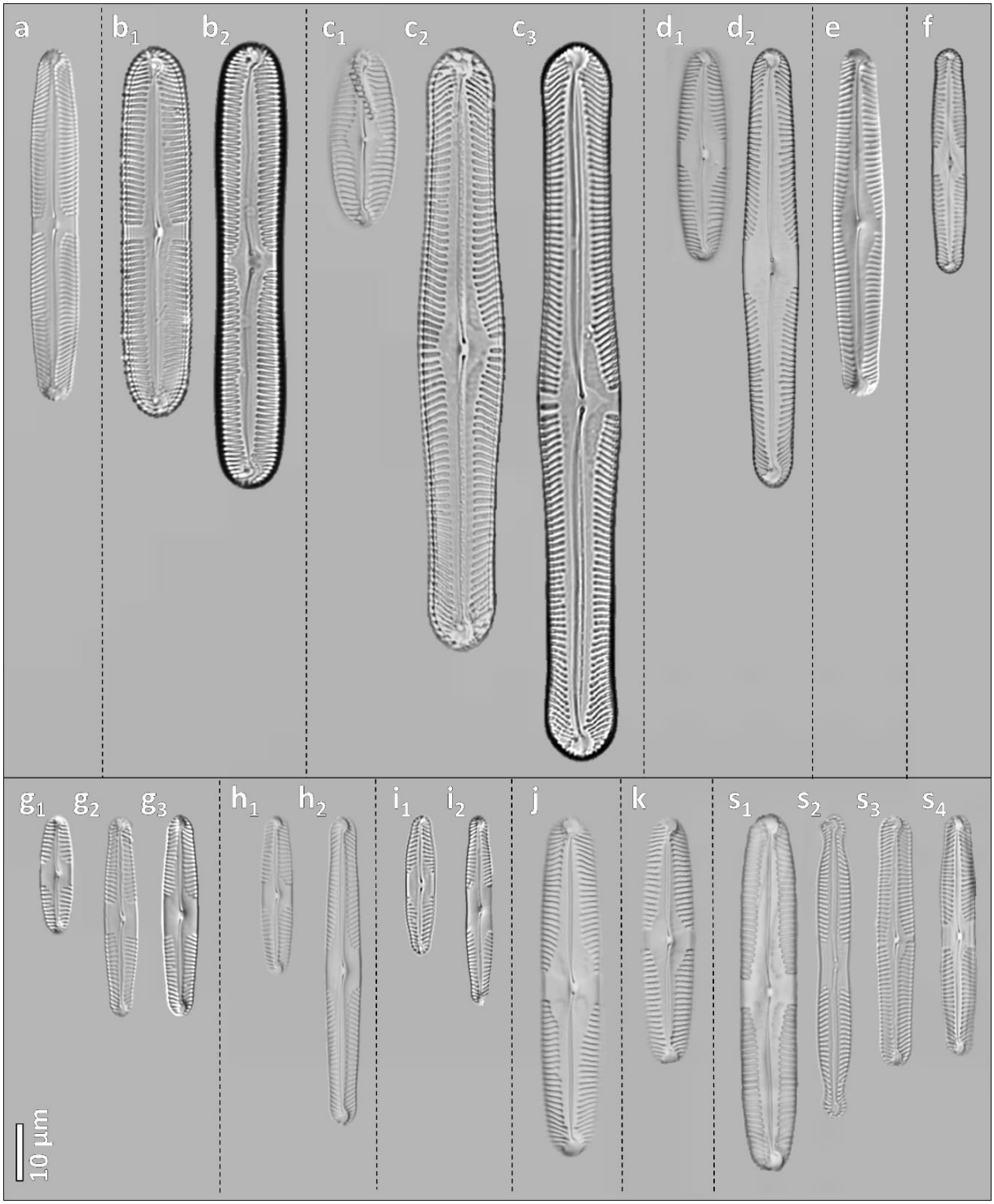


Figure 3. Morphological variability of species within the *P. gibba* group as delimited by both *cox*1 and LSU rDNA. Representatives are shown for each clade (a-k) and singleton strain (s). Clade A is represented by strain CZECH_SW2_7a (a); clade B by NUUK_13KAP17_3 (b_1), and NUUK_13KAP16_3 (b_2); clade C by PIN596A (c_1),

CZECH_NOS2_5 (c_2), and STAP4 (c_3); clade D by W045d (d_1), and W045e (d_2); clade E by REU12_5_1 (e); clade F by Tor8b (f); clade G by MIC7_2 (g_1), CRO12_9_16 (g_2), and MIC5_16 (g_3); clade H by W076d (h_1), and W076e (h_2); clade I by REU12_3_10 (i_1), and REU12_9_14 (i_2); clade J by PIN877TM (j); clade K by PIN19Cra (k); singleton strains are Tor7f (s_1), Wiec (s_2), CZECH_NOS2_7 (s_3), and Tor4r (s_4). Species D-S_4 (with exception of S_2) are members of the (pseudo)cryptic species complex *gibba-subgibba-parvulissima*. More LM and SEM microphotographs may be found in the BoLD (this figure was included in Kollár et al. 2019).

Table 1. Summary of available evidence for each operational taxonomic unit (OTU), i.e. clade, singleton strain, or morphogroup (1 = available for all strains; 0 = unavailable for all strains; 7/8 = available for 7 of the total 8 strains). Evidence include primary species delimitation results based on LSU rDNA and *cox*1, monophyly based on both molecular markers, morphological analysis, coherent biogeography (geographical origin of strains) and ecology (including living mode and in some cases also environmental factors such as pH, conductivity, altitude, water temperature, and salinity) and reproductive compatibility. Colours indicate falsifying (dark grey) and unevaluable or ambiguous evidence (light grey) to the hypothesis that the OTU is a single distinct species or, in the case of the three morphogroups (i.e., DEF, GHI, and JK), that the group is natural (this table was included in the Supplementery material of Kollár et al. 2019).

OTU	LSU	*cox*1	monophyly	morphology	biogeography	ecology	rep.
Clade A	1	1	1	1	1	1	0
Clade B	1	1	1	1	1	1	0
Clade C	1	1	1	1	1	4/8	0
Clade D	1	7/8	1	1	1	0	0
Clade E	1	1	1	1	1	0	0
Clade F	1	1	1	1	1	0	0
Clade G	1	8/10	1	1	1	2/10	0
Clade H	1	0	1	1	1	0	0
Clade I	1	3/4	1	1	1	0	0
Clade J	1	1/2	1	1	1	1	0
Clade K	1	1	1	1	1	1	1
Tor4r	1	1	0	1	1	0	0
CZECH_NOS2_7	1	1	0	1	1	1	0
Tor7f	1	1	0	1	1	0	0
Wiec	1	1	0	1	1	0	0
group DEF	1	12/13	1	1	1	0	0
group GHI	1	11/21	1	1	1	19/21	0
group JK	1	6/7	1	1	1	0	1/2

Automated molecular species delimitations based on two most variable markers suggested the presence of 15 species in our data set and showed complete congruence between *cox*1 and LSU rDNA (Figure 2). We then confronted our primary hypothesis with other lines of evidence. Whereas it is relatively easy to distinguish some of the potential species (clades A, B, C) based on morphological characteristics, it is not straightforward to distinguish others (clades D, E, F, G, H, I, J, and K, and singleton strains), especially the small specimens. However, the relative size of the fascia seems promising for the recognition of both higher natural species clusters (e.g., JK and GHI) and possibly even some species within these

clusters (e.g., H) because it seems to be more stable for both small and large specimens compared to valve outline or length-to-width ratio. Nevertheless, we conclude that using morphology as the sole line of evidence for species delimitations and consequent taxonomic changes may lead to underestimation of diatom species diversity.

The delimited species have largely unknown biogeographies and seem mostly restricted to a single location. Several are geographically very distant from the sampling localities of the presumed corresponding type specimens (e.g., species A - *Pinnularia* cf. *microstauron*; Czech Republic vs. Brasil), suggesting that these taxa either have very wide geographic distributions, or that they do not correspond to single species, but rather could represent members of (pseudo)cryptic species complexes with restricted geographic distributions. It is recently becoming clear that many former cosmopolitan diatom taxa actually show hidden diversity with their members showing (relatively) restricted geographic distributions (Vanormelingen et al. 2008b), and/or niche-differentiation (Vanelslander et al. 2009, Souffreau et al. 2013). However, significant increase of the sampling coverage is necessary to study the presence and nature of potential biogeographical patterns in the *P. gibba* group.

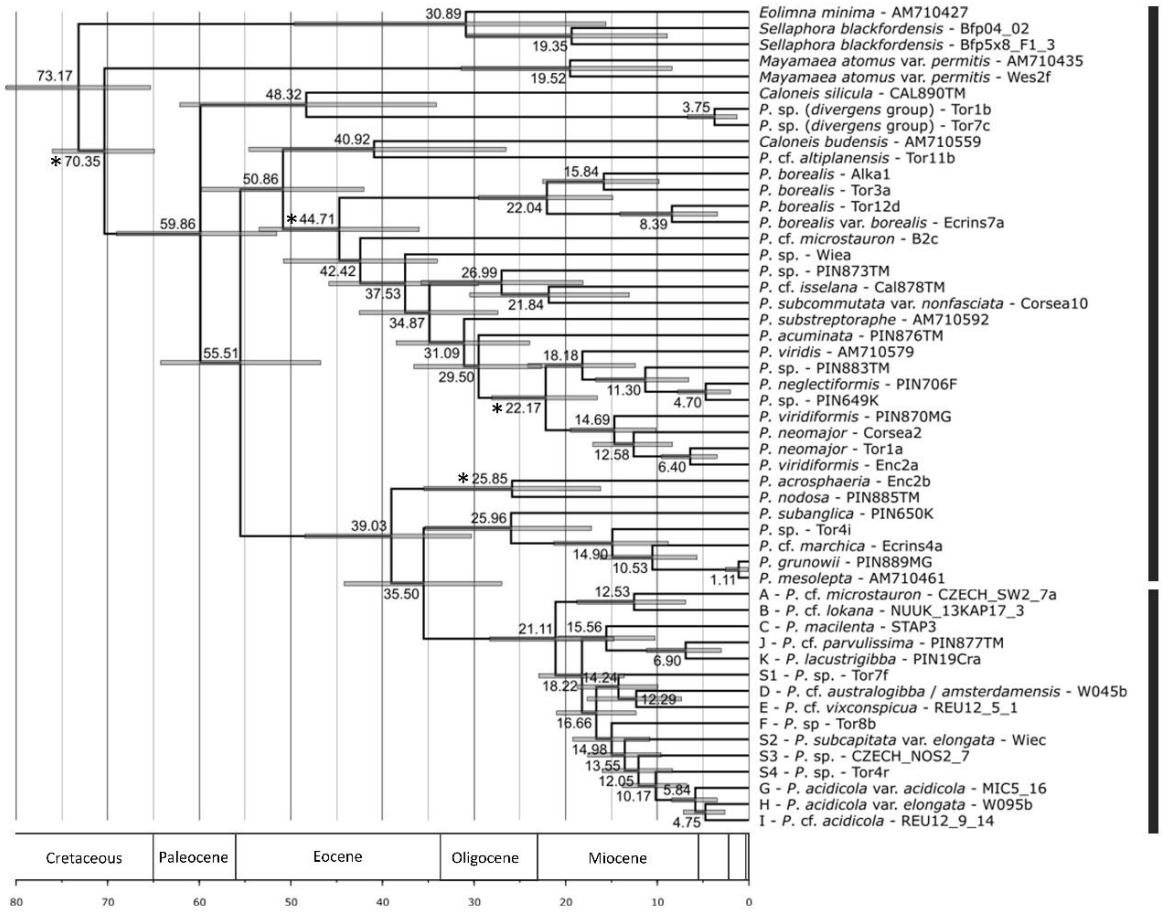
3.2 On the evolution of the *P. gibba* group

(this chapter is based on Kollár et al. *accepted* - reprinted as Appendix 3)

While the study design in Kollár et al. (2019) was optimized for species delimitation, namely that the data set was built in a way to maximize the probability of covering all the available intraspecific variability, the next step focused on phylogeny inference (Kollár et al. *accepted*). To this end, we added the more conserved plastid marker *rbc*L and information from the fossil record to infer a time-calibrated multi-gene phylogeny of the group (see Figure 4). In addition, the phylogenetic results allowed for formulation and (in some cases) statistical testing of evolutionary hypotheses. For example, visual comparison of the phylogenetic results with the valve morphology of the different species (e.g., Figure 3), suggested that the mean cell-sizes of the strains of the basal lineages (i.e., A-B, C, J-K, and S1-D-E) are larger than those of the strains of the lineages in the terminal clade (i.e., F-S2-S3-S4-G-H-I). We used the full data set of Kollár et al. (2019) and the randomization test to statistically test the null hypothesis (H_0: *there is no difference in the mean cell-size of the strains of the basal lineages and the mean cell-size of the strains of the lineages in the terminal clade*) and it was

rejected. In other words, we found that the two groups in the *P. gibba* group (the terminal clade, and the basal grade) showed significant difference in cell-size.

In microbes, cell-size plays a crucial role in their physiology and ecology (Marshall et al. 2012). However, an investigation of available environmental and geographic data associated with the *P. gibba* group strains (Kollár et al. 2019) revealed no connection between species within neither the terminal clade nor the basal grade. In addition, the size of diatom cells usually reduce during the cell-cycle due to the mechanism of frustule inheritance (MacDonald 1869, Pfitzer 1869). However, a hypothesis that strains of any of the two tested groups (i.e., terminal clade or basal grade) are more similar (relative to the strains of the second group) in cell-size because, by chance, they were measured in the similar phase of the cell-cycle, is



refused by the statistical analysis itself. The observed pattern was significantly non-random. Therefore, we concluded that, at the current state of knowledge, observed cell-size difference is best explained by evolutionary origin, i.e., relatively diminutive appearance was probably inherited from the common ancestor of the species of the terminal clade.

Figure 4. A fossil-guided time-calibrated phylogeny of the *Pinnularia gibba* group inferred from a three-locus DNA alignment using Bayesian relaxed molecular clock analysis with four fossil-based time constraints (indicated by asterisks). Values at nodes are mean node ages. Grey bars are 95% HPD (highest posterior density) intervals. Ingroup tip labels correspond to the delimited species sensu Kollár et al. (2019): species code - taxon – strain (this figure was included in Kollár et al. accepted).

3.3 On the taxonomy of the *P. gibba* group

(this chapter is based on Kollár et al. *accepted* and Poulíčková et al. 2018 - reprinted as Appendices 3 and 4)

As mentioned above (see Introduction), unambiguous identification of diatom species is important not only for fundamental research but also for some applied disciplines. For example, more fine-grained species-level taxonomy increase the resolution of the diatom-based bioindicative systems of water quality assessments such as the one implemented through the European Union's Water Framework Directive (e.g., Poulíčková et al. 2017). To this end, however, diatom species must be (1) properly delimited and (2) unambiguously and routinely identifiable. In the *P. gibba* group, species limits were investigated through our polyphasic species delimitation (Kollár et al. 2019). However, most delimited species are members of large (pseudo)cryptic species complex *gibba-subgibba-parvulissima* (members are indicated in the legend of the Figure 3) and their morphology-based identification is either challenging or imposible. Therefore, the diagnostic characters allowing for unambiguous and routine identification of species must be sought outside of morphology. For this reason, we tested whether the proposed diatom DNA barcode markers in *rbc*L and SSU rDNA have sufficient resolution to distinguish even closely related species in the *P. gibba* group.

The test revealed that both complete and partial *rbc*L (which is proposed for diatom-based biomonitoring in the United Kingdom and France; Vasselon et al. 2017, Kelly et al. 2018) can unambiguously distinguish between different species, with a minimum divergence of 5 and 3 bp, respectively (Kollár et al. accepted). Following the results of *rbc*L, complete and partial SSU rDNA (also proposed for DNA barcoding of diatoms; Zimmermann et al. 2011) was also able to distinguish between all *P. gibba* species (with a minimum divergence of 5 and 2 bp respectively), suggesting that both *rbc*L and SSU rDNA could be good barcode markers to distinguish even closely-related (pseudo)cryptic *P. gibba* group species.

One of the contributions of this thesis to the taxonomy is an initiation of the taxonomical review of the *P. gibba* group by formal description of the species K sensu Kollár et al. (2019) as *Pinnularia lacustrigibba* sp. nov. (Poulíčková et al. 2018). The strains were earlier studied (including their reproductive compatibility) and identified as *Pinnularia* cf. *gibba* by Poulíčková et al. (2007). Comparison with morphologically similar taxa lead to the conlusion that, at current state of knowledge, it seems impossible to distinguish with reasonable certainty *P. lacustrigibba*, species J sensu Kollar et al. (2019; identified as *P.* cf. *parvulissima*), *P. parvulissima*, *P. subgibba* var. *subgibba*, *P. subgibba* var. *sublinearis* and *P. subgibba* var. *undulata* solely on the basis of LM (see Supplementary Table 1 of Poulíčková et al. 2018). Nevertheless, analysis of species covered by data set of Kollár et al. (2019) revealed that they can be distinguished using genetic markers. For example, the genetic difference between sequences of sister species K (*P. lacustrigibba* sp. nov.) and J was found to be 2.3% (20 bp), 5.5% (36 bp) and 0.4% (6 bp) for LSU rDNA, *cox*1 and *rbc*L, respectively. Several other taxonomic considerations were formulated and/or discussed in the doctoral thesis and/or the papers upon which it is based (especially in Kollár et al. *accepted*).

3.4 On the relativity of species

(this chapter is based on Kollár et al. *submitted* - reprinted as Appendix 5)

The work on this thesis, coupled with thorough study of the species problem, lead to a developement of a novel probabilistic model of speciation and species itself. The model is based on several realizations (see the dissertation and Kollár et al. *submitted* for more details). Some of the most important follow. Firstly, the existence or nonexistence of species as a natural discontinuous unit is relative to its dimensionality, namely on whether the time dimesnion is or is not incorporated into the considerations. This relativity of species' existence may be difficult to comprehend at first. However, Mayr (1982, p. 273) himself was aware that his 'biological species' are nonsense whenever time dimension is incorporated into considerations. In Kollár et al. (*submitted*), we simply extend this awareness by stating that the view of species as the natural discontinuous entities in time is nonsense concept in general (with exception of species originated in saltational speciation).

Secondly, the so-called incipient species (i.e., species which are evolutionarily somewhere between the levels of subpopulation and independent species) may be theoretically viewed as

being virtually in a dual state of being and not being species at the same time. Finally, the most accurate way to reflect both the species relativity and the duality of incipient species is probabilistic. Therefore, we developed the Universal Probabilistic Concept of Evolutionary Lineages (UPCEL; Figure 5), a probabilistic view of speciation which might potentially form a theoretical basis for the development of probabilistic species delimitation methods based on the operational criterion which is the fundamental cause of species' inner coherence, i.e., on the gene flow between the lineages in question (Kollár et al. *submitted*). In theory, its advantages are that it should be universally applicable to all domains of life (with exception of lineages without any form gene flow, if such exist) and also to incipient species.

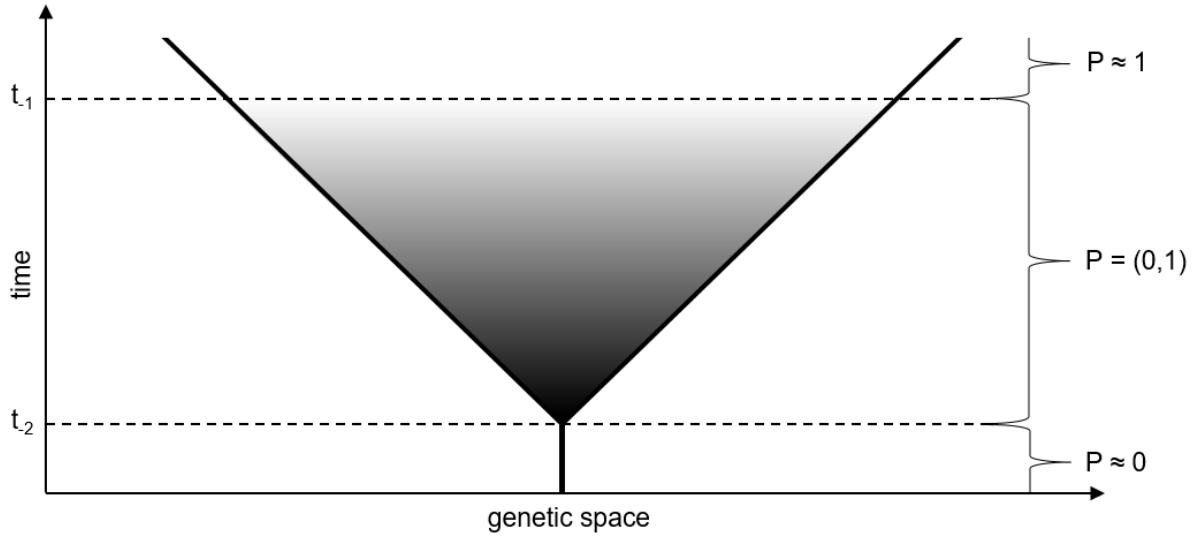


Figure 5. A probabilistic model of the divergence of evolutionary lineage in the genetic space-time. Continually decreasing gene flow cause parental population to split (at t_{-2}) and filial subpopulations to further diverge in the genetic space-time. In a given time planes between t_{-2} and t_{-1}, the degrees of grey indicate intensity of gene flow between the two incipient species. The incipient species under certain conditions have certain probability (P) that they will continue to diverge from each other and, eventually, achieve an independency of fully developed species (at t_{-1}). 'Full' independency is achieved when the frequency of interspecific gene flow decreases to such an extent that it is no longer significant for subsequent evolution of the filial lineages. Notice that gene flow intensity does not have to be absolutely zero to be effectively insignificant to the evolution of the lineages (P approaches 1 asymptotically).

Nevertheless, it seems clear that tremendously long and difficult journey stands between this first report of an interesting idea and its potential application. In order to describe this view, we also developed concepts of the genetic space and the genetic space-time (Figure 6). The difference between the UPCEL and the Unified Species Concept (USC) is graphically summarized in Figure 7.

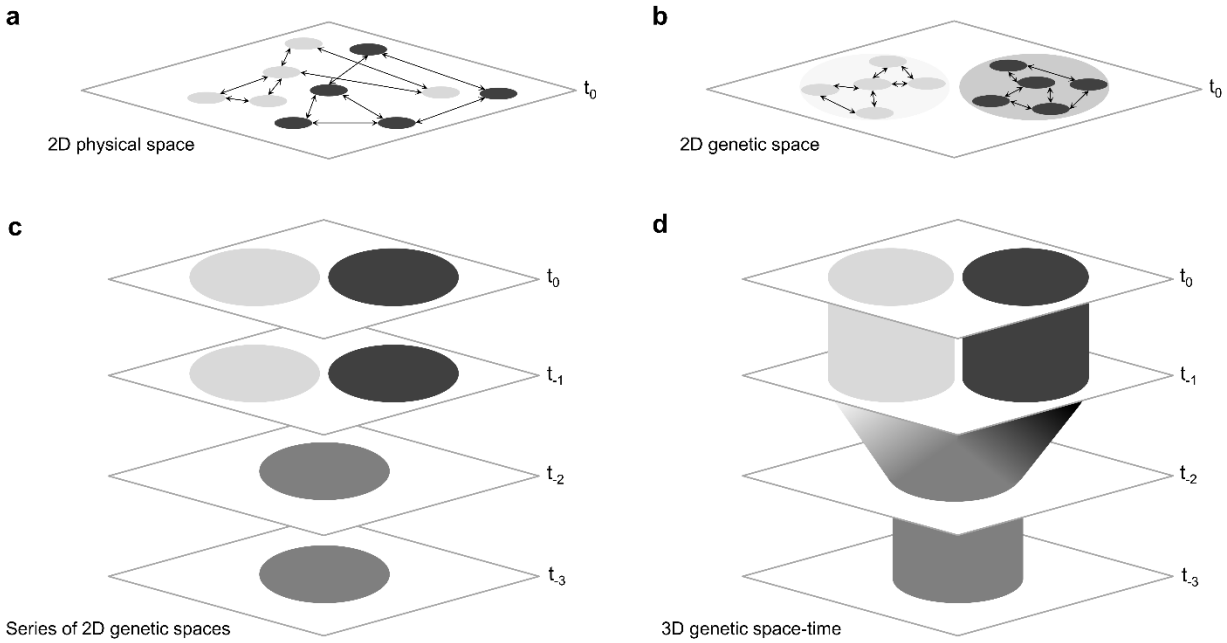


Figure 6. A simplified model of species in physical (a) and genetic space (b-d). (a) At present time (t_0), local populations (spots) of two separate metapopulations (light and dark) are dispersed in physical space and connected by gene flow (arrows). (b) If genetic space is considered instead of physical space, local populations of each metapopulation are grouped together. (c-d) Adding time dimension will reveal two separately evolving metapopulation lineages derived from ancestral lineage due to a restriction of gene flow between sets of local populations in t_{-2}. While it is relatively easy to delimit species in certain time plane or time period after their features sufficiently diverged (period between t_{-1} and t_0), it is difficult to delimit species lineages that split relatively recently (period between t_{-2} and t_{-1}) due to incomplete lineage sorting, possible hybridization, and/or absence of operational criteria (the definitions of the terms such as genetic space may be found in Kollár et al. *submitted* included as Appendix 5 of the thesis).

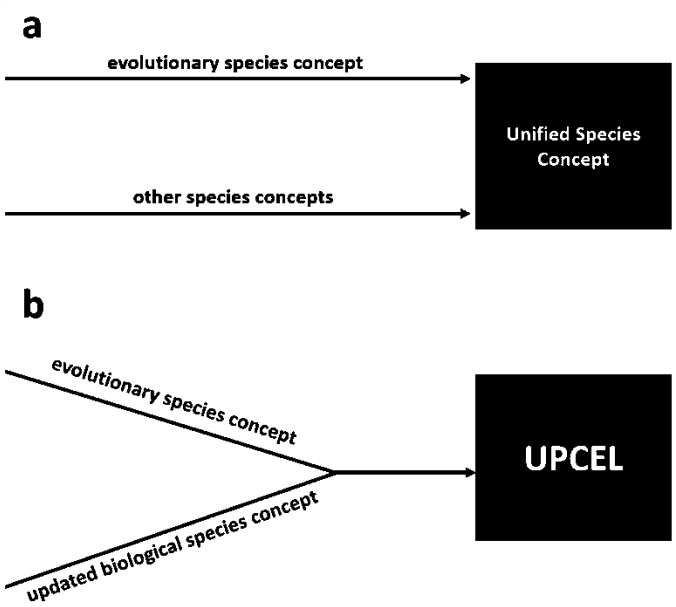


Figure 7. A comparison of the Unified Species Concept (USC) and the Universal Probabilistic Concept of Evolutionary Lineages (UPCEL). Within the freamework of USC, the definition of species category falls into the category of evolutionary species concepts (i.e., time-incorporating lineage-based concepts) while operational criteria for species delimitation merely reflect the species reality but do not cause it and, therefore, are kept separately. Within the framework of UPCEL, vertical species are defined using time-incorporating lineage-based evolutionary species concept and horizontal species are defined by an updated version of the biological species concept (updated in a way that reproductive isolation is exchanged for the evolutionarily significant decrease in gene flow frequency which allowed the UPCEL to cover those asexual lineages where other means of gene flow, e.g., mediated through the horizontal gene transfer, were observed). The most accurate way of combining both definitions, e.g., for the sake of species delimitation, is probabilistic.

4. CONCLUSIONS

The contributions to science made in this doctoral thesis may be summarized as follows:

(1) The species diversity covered by the data set of 105 clonal strains with worldwide origin (representing the selected *P. gibba* group) was studied by means of a polyphasic approach. Primary (i.e., DNA-based) species delimitation based on LSU rDNA and *cox*1 uncovered 15 species within the group. This primary hypothesis was further confronted with other available lines of evidence (i.e., morphological, geographic, environmental and/or reproductive data) coming to the secondary hypothesis on species limits within the group, namely that our data set indeed covered 15 species of the group and that it is difficult or imposible to distinguish majority of them solely on the basis of morphology.

(2) The fossil-guided time-calibrated multi-gene phylogeny of the *P. gibba* group was inferred using genetic markers from nuclear (LSU rDNA), plastid (*rbc*L) and mitochondrial (*cox*1) genomes.

(3) The knowledge on the limits of species combined with the knowledge on their phylogenetic relationships was used to design the test of taxonomic resolution of DNA barcode markers proposed for unambiguous and routine identification of diatoms. We have concluded that proposed DNA barcode markers of both *rbc*L and SSU rDNA are variable enough to unambiguously distinguish even sister cryptic species of the group.

(4) The phylogenetic results further allowed for a formulation and (in some cases) statistical testing of some evolutionary hypothesis. For example, we have found significant differences in cell-size in different parts of the tree which, at current state of knowledge, are best explained by an evolutionary origin. In other words, relatively diminutive cell-size of species in the terminal clade was inherited from its common ancestor.

(5) The taxonomic review of the group was initiated by describing one of the delimited species as *Pinnularia lacustrigibba* sp. nov. Several other uncovered species are new to science and should be formally decribed in the future.

(6) The work on this thesis, coupled with thorough study of the species problem, lead to a development of a novel probabilistic view of the species, the Universal Probabilistic Concept of Evolutionary Lineages (UPCEL). In the future, this view may potentially proof useful in developing probabilistic species delimitation methods (perhaps based on

gene flow interconnection state between evolutionary lineages) and in taxonomic decisions.

5. REFERENCES

Andersen, R.A. 1992. Diversity of eukaryotic algae. *Biodivers. Conserv.* 1:267–92.

Bruder, K., Sato, S. & Medlin, L.K. 2008. Morphological and molecular investigations of naviculoid diatoms IV. Pinnularia vs. Caloneis. *Diatom*. 24:8–24.

Colwell, R.R. 1970. Polyphasic taxonomy of the genus Vibrio: numerical taxonomy of Vibrio cholerae, Vibrio parahaemolyticus, and related Vibrio species. *J. Bacteriol.* 104:410–33.

Cronquist, A. 1978. Once again, what is a species? *In* Romberger, J. A. [Ed.] *Biosystematics in Agriculture*. Allanheld and Osmun, Montclair, pp. 3–20.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. *In* Howard, D. J. & Berlocher, S. H. [Eds.] *Endless Forms: Species and Speciation*. Oxford University Press, New York, pp. 57–75.

de Queiroz, K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56:879–86.

Dobzhansky, T. 1937. Genetics and the Origin of Species. Columbia University Press, New York. 364 pp.

Drum, R.W. & Gordon, R. 2003. Star Trek replicators and diatom nanotechnology. *Trends Biotechnol.* 21:325–8.

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*. 158:349–64.

Guiry, M.D. 2012. How many species of algae are there? *J. Phycol.* 48:1057–63.

Guiry, M.D. & Guiry, G.M. 2018. AlgaeBase. Available At: www.algaebase.org (last accessed November 13, 2018).

Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M. & Darzins, A. 2008. Microalgal triacylglycerols as feedstocks for biofuel production: Perspectives and advances. *Plant J.* 54:621–39.

Kelly, M., Boonham, N., Juggins, S., Kille, P., Mann, D.G., Pass, D., Sapp, M. et al. 2018. A DNA based diatom metabarcoding approach for Water Framework Directive classification of rivers. *Environ. Agency*. Bristol.

Kociolek, J.P., Balasubramanian, K., Blanco, S., Coste, M., Ector, L., Liu, Y. & Kulikovskiy, M. 2018. Pinnularia C. G. Ehrneberg 1943. Available At: www.diatombase.org/aphia.php?p=taxdetails&id=149208 (last accessed May 25, 2018).

Kollár, J., Fránková, M., Hašler, P., Letáková, M. & Poulíčková, A. 2015. Epiphytic diatoms in lotic and lentic waters - Diversity and representation of species complexes. *Fottea*. 15:259–71.

Kollár, J., Pinseel, E., Vanormelingen, P., Poulíčková, A., Souffreau, C., Dvořák, P. & Vyverman, W. 2019. A polyphasic approach to the delimitation of diatom species: a case study for the genus Pinnularia

(Bacillariophyta). *J. Phycol.* 55:365–79.

Kollár, J., Pinseel, E., Vyverman, W. & Poulíčková, A. n.d. A time-calibrated multi-gene phylogeny provides insights into the evolution, taxonomy and DNA barcoding of the Pinnularia gibba group (Bacillariophyta). *Fottea*.

Kollár, J., Poulíčková, A. & Dvořák, P. n.d. On the Relativity of Species, or the Probabilistic Solution to the Species Problem. *Syst. Biol.*

Komárek, J. 2016. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. *Eur. J. Phycol.* 51:346–53.

Krammer, K. 2000. The genus Pinnularia. A.R.G. Ganter Verlag K.G., Ruggell. 703 pp.

Lewis, A.R., Marchant, D.R., Ashworth, A.C., Hedenäs, L., Hemming, S.R., Johnson, J. V, Leng, M.J. et al. 2008. Mid-Miocene cooling and the extinction of tundra in continental Antarctica. *Proc. Natl. Acad. Sci.* 105:10676–80.

Luís, A.T., Novais, M.H., van de Vijver, B., Almeida, S.F.P., da Silva, E.A.F., Hoffmann, L. & Ector, L. 2012. Pinnularia aljustrelica sp. nov. (Bacillariophyceae), a new diatom species found in acidic waters in the Aljustrel mining area (Portugal) and further observations on the taxonomy and ecology of P. acidophila Hofmann et Krammer and P. acoricola Hustedt. *Fottea*. 12:27–40.

MacDonald, J.D. 1869. On the structure of the diatomaceous frustule, and its genetic cycle. *Ann. Mag. Nat. Hist. Ser.* 4:1–8.

Mann, D.G. 1999. The species concept in diatoms. *Phycologia*. 38:437–95.

Mann, D.G. & Evans, K.M. 2008. The species concept and cryptic diversity. *In Proceedings of the 12th International Conference on Harmful Algae*. International Society for the Study of Harmful Algae and Intergovernmental Oceanographic Commission of UNESCO, Copenhagen, pp. 262–8.

Mann, D.G. & Vanormelingen, P. 2013. An inordinate fondness? the number, distributions, and origins of diatom species. *J. Eukaryot. Microbiol.* 60:414–20.

Mayden, R.L. 1997. A hierarchy of species concepts: the denoument in the saga of the species problem. *In* Claridge, M. F., Dawah, H. A. & Wilson, M. R. [Eds.] *Species: The Units of Diversity*. Chapman and Hall, London, pp. 381–423.

Mayr, E. 1942. Systematics and the Origin of Species. Columbia University Press, New York. 334 pp.

Mayr, E. 1963. Animal Species and Evolution. Harvard University Press, Cambridge, MA. 797 pp.

Mayr, E. 1982. The Growth of Biological Thought revisited. 11th ed. Harvard University Press, Cambridge, MA. 974 pp.

Mayr, E. & Ashlock, P.D. 1991. Principles of Systematic Zoology. McGraw-Hill Inc., New York. 475 pp.

Pereira, A.C., Torgan, L.C. & Melo, S. 2014. Four new Pinnularia Ehrenberg (Bacillariophyta,

Pinnulariaceae) species from Amazonian black water (Tupé Lake, Amazonas State, Brazil). *Phytotaxa*. 158:154–68.

Pfitzer, E. 1869. Über den Bau und die Zellteilung der Diatomeen. *Bot. Zeitung*. 27:774–6.

Pinseel, E., Hejduková, E., Vanormelingen, P., Kopalová, K., Vyverman, W. & de Vijver, B. Van 2016. Pinnularia catenaborealis sp. nov . (Bacillariophyceae), a unique chain-forming diatom species from James Ross Island and Vega Island (Maritime Antarctica). *Phycologia*. 56:94–107.

Pinseel, E., Janssens, S.B., Verleyen, E., Vanormelingen, P., Kohler, T.J., Biersma, E.M., Sabbe, K. et al. 2020. Global radiation in a rare biosphere soil diatom. *Nat. Commun.* 11:1–12.

Pinseel, E., Van de Vijver, B., Kavan, J., Verleyen, E., Kopalovï¿½, K. & Kopalová, K. 2017a. Diversity, ecology and community structure of the freshwater littoral diatom flora from Petuniabukta (Spitsbergen). *Polar Biol.* 40:533–51.

Pinseel, E., Vanormelingen, P., Hamilton, P.B., Vyverman, W., Van de Vijver, B. & Kopalova, K. 2017b. Molecular and morphological characterization of the Achnanthidium minutissimum complex (Bacillariophyta) in Petuniabukta (Spitsbergen, High Arctic) including the description of A. digitatum sp. nov. *Eur. J. Phycol.* 52:264–80.

Poulíčková, A., Dvořák, P., Mazalová, P. & Hašler, P. 2014. Epipelic microphototrophs: an overlooked assemblage in lake ecosystems. *Freshw. Sci.* 33:513–23.

Poulíčková, A., Hájková, P., Kintrová, K., Baťková, R., Czudková, M. & Hájek, M. 2013. Tracing decadal environmental change in ombrotrophic bogs using diatoms from herbarium collections and transfer functions. *Environ. Pollut.* 179:201–9.

Poulíčková, A., Kollár, J., Hašler, P., Dvořák, P. & Mann, D.G. 2018. A new species Pinnularia lacustrigibba sp. nov. within the Pinnularia subgibba group (Bacillariophyceae). *Diatom Res.* 33:273–82.

Poulíčková, A., Letáková, M., Hašler, P., Cox, E. & Duchoslav, M. 2017. Species complexes within epiphytic diatoms and their relevance for the bioindication of trophic status. *Sci. Total Environ.* 599–600:820–33.

Poulíčková, A., Mayama, S., Chepurnov, V.A. & Mann, D.G. 2007. Heterothallic auxosporulation, incunabula and perizonium in Pinnularia (Bacillariophyceae). *Eur. J. Phycol.* 42:367–90.

Poulíčková, A., Veselá, J., Neustupa, J. & Škaloud, P. 2010. Pseudocryptic Diversity versus cosmopolitanism in diatoms: A case study on Navicula cryptocephala Kütz. (Bacillariophyceae) and morphologically similar taxa. *Protist*. 161:353–69.

Puillandre, N., Modica, M. V., Zhang, Y., Sirovich, L., Boisselier, M.C., Cruaud, C., Holford, M. et al. 2012. Large-scale species delimitation method for hyperdiverse groups. *Mol. Ecol.* 21:2671–91.

Round, F.E., Crawford, R.M. & Mann, D.G. 1990. The Diatoms - Biology and Morphology of the Genera. Cambridge University Press, Cambridge, UK. 747 pp.

Sarno, D., Kooistra, W.H.C.F., Medlin, L.K., Percopo, I. & Zingone, A. 2005. Diversity in the genus Skeletonema (Bacillariophyceae). II. An assessment of the taxonomy of S. costatum-like species with the description of four new species. *J. Phycol.* 41:151–76.

Simpson, G.G. 1951. The species concept. *Evolution (N. Y).* 5:285–298.

Simpson, G.G. 1961. Principles of animal taxonomy. Columbia University Press, New York. 247 pp.

Škaloud, P. 2008. Polyphasic approaches in the taxonomy of green aerophytic algae. Charles University in Prague, 111 pp.

Smol, J.P. & Stoermer, E.F. 2010. The Diatoms: Applications for the Environmental and Earth Sciences. 2nd ed. Cambridge University Press, Cambridge. 667 pp.

Souffreau, C., Vanormelingen, P., Van de Vijver, B., Isheva, T., Verleyen, E., Sabbe, K. & Vyverman, W. 2013. Molecular Evidence for Distinct Antarctic Lineages in the Cosmopolitan Terrestrial Diatoms Pinnularia borealis and Hantzschia amphioxys. *Protist*. 164:101–15.

Souffreau, C., Verbruggen, H., Wolfe, A.P., Vanormelingen, P., Siver, P.A., Cox, E.J., Mann, D.G. et al. 2011. A time-calibrated multi-gene phylogeny of the diatom genus Pinnularia. *Mol. Phylogenet. Evol.* 61:866–79.

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 (Bacillariophyceae). *Phycologia*. 48:443–59.

Van de Vijver, B., Zidarova, R. & de Haan, M. 2011. Four new Luticola taxa (Bacillariophyta) from the South Shetland Islands and James Ross Island (Maritime Antarctic Region). *Nov. Hedwigia*. 92:137–58.

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. & Swings, J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60:407–38.

Vanelslander, B., Créach, V., Vanormelingen, P., Ernst, A., Chepurnov, V.A., Sahan, E., Muyzer, G. et al. 2009. Ecological differentiation between sympatric pseudocryptic species in the estuarine benthic diatom navicula phyllepta (bacillariophyceae). *J. Phycol.* 45:1278–89.

Vanormelingen, P., Chepurnov, V.A., Mann, D.G., Sabbe, K. & Vyverman, W. 2008a. Genetic Divergence and Reproductive Barriers among Morphologically Heterogeneous Sympatric Clones of Eunotia bilunaris Sensu Lato (Bacillariophyta). *Protist*. 159:73–90.

Vanormelingen, P., Verleyen, E. & Vyverman, W. 2008b. The diversity and distribution of diatoms: From cosmopolitanism to narrow endemism. *Biodivers. Conserv.* 17:393–405.

Vasselon, V., Rimet, F., Tapolczai, K. & Bouchez, A. 2017. Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island, France). *Ecol. Indic.* 82:1–12.

Verleyen, E., Vyverman, W., Sterken, M., Hodgson, D.A., De Wever, A., Juggins, S., Van De Vijver, B. et al. 2009. The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities. *Oikos*. 118:1239–49.

Vyverman, W., Verleyen, E., Sabbe, K., Vanhoutte, Koenradd, Sterken, M., Hodgson, D.A., Mann, D.G. et al. 2007. Historical Processes Constrain Patterns in Global Diatom Diversity. *Ecology*. 88:1924–31.

Wheeler, Q.D. & Meier, R. 2000. Species concepts and phylogenetic theory: a debate. Columbia University Press, New York. 230 pp.

Wiley, E.O. 1978. The evolutionary species concept reconsidered. *Syst. Zool.* 27:17–26.

Wiley, E.O. & Mayden, R.L. 2000. The evolutionary species concept. *In* Wheeler, Q. D. & Meier, R. [Eds.] *Species Concepts and Phylogenetic Theory*. Columbia University Press, New York, pp. 70–89.

Wilkins, J.S. 2009. Species: a history of the idea. 1st ed. University of California Press, Berkeley. 320 pp.

Wilkins, J.S. 2011. Philosophically speaking, how many species concepts are there? *Zootaxa*. 2765:58-60.

Will, K.W., Mishler, B.D. & Wheeler, Q.D. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* 54:844–51.

Zimmermann, J., Jahn, R. & Gemeinholzer, B. 2011. Barcoding diatoms: Evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Org. Divers. Evol.* 11:173–92.

6. ASPIRANT'S CURRICULUM VITAE

Research experience:

Ghent University Jul 2014 – Sep 2014

Erasmus intern

Topic: Development of a DNA barcode database for identification and discovery of diatom species

Supervisor: Dr Pieter Vanormelingen

Academy of Sciences of the Czech Republic Mar 2015

Intern

Topic: Morphological study of diatoms

Supervisors: Dr Markéta Fránková and Dr Petr Marvan

Ghent University Nov 2015

Research fellow

Topic: Phylogenetic species delimitation in diatoms

Supervisor: Dr Wim Vyverman

Natural History Museum London Apr 2018 – Sep 2018

Research fellow

Topic: Phylogenetics and systematics of diatoms

Supervisor: Dr David M. Williams

Selected experience with science popularization:

UP Crowd Nov 2015 – Present

non-profit organization devoted to the popularization of science

Coordinator and lecturer

Fort Science Apr 2015 – Present

Palacký University interactive science center

Lecturer

Education:

Systematic biology and ecology (BSc) Sep 2011 – May 2014

Palacký University in Olomouc

Hydrobiology (MSc, Graduated with Honours) Jun 2014 – Jun 2016

Palacký University in Olomouc

Botany (PhD) Sep 2016 – Present

Palacký University in Olomouc

Supervisor: Dr Aloisie Poulíčková

Awards and Achievements:

Jul 2014: Scholarship by Erasmus programme

Jun 2015: Scholarship by the Endownment fund of Palacký University for support of excellent junior research projects

Mar 2018: Award "Cena děkana PřF UP" (winner of the PhD Biology section)

Apr 2018: Scholarship by Erasmus+ programme

Apr 2020: One of the national finalists of FameLab 2020

Scientific societies:

The Czech Phycological Society

The International Society for Diatom Research

The Phycological Society of America

Academic conferences:

2015: Meeting of the Czech Phycological Society in České Budějovice, CZ

2016: Meeting of the Czech Phycological Society in Prague, CZ

2016: Workshop Phytobentos in Vyškovec, CZ (lecturer)

2017: Central European Diatom Meeting in Prague, CZ (poster)

2017: Meeting of the Czech Phycological Society in Ostrava, CZ (presentation)

2018: International Diatom Symposium in Berlin, DE (presentation)

2019: International Barcode of Life Conference in Trondheim, NO

Publications in international peer-reviewed journals with impact factor:

- Kollár J., Fránková M., Hašler P., Letáková M. & Poulíčková A. 2015: Epiphytic diatoms in lotic and lentic waters – diversity and representation of species complexes. *Fottea* 15(2): 259-271. (IF_{2014} = 2.190; Q2)

- Poulíčková A., Kollár J., Hašler P., Dvořák P. & Mann D. G. 2018: A new species *Pinnularia lacustrigibba* sp. nov. within the *Pinnularia subgibba* group (Bacillariophyceae). *Diatom Res.* 33, 273–282. (IF_{2018} = 1.169; Q3)

- Kollár J., Pinseel E., Vanormelingen P., Poulíčková A., Souffreau C., Dvořák P. & Vyverman W. 2019: A Polyphasic approach to the delimitation of diatom species: a case study for the genus *Pinnularia* (Bacillariophyta). *J. Phycol.* 55, 365–379. (IF_{2018} = 2,831; Q1)

- Kollár J., Pinseel E., Vyverman W. & Poulíčková A.: A time-calibrated phylogeny provides an insight into the evolution, taxonomy and DNA barcoding of the *Pinnularia gibba* group (Bacillariophyta). *Fottea*, (*accepted*). (IF_{2018} = 1,727; Q2)

- Kollár J., Poulíčková A. & Dvořák P.: On the relativity of species, or the probabilistic solution to the species problem. *Syst. Biol.* (*submitted*). (IF_{2018} = 10,266; Q1)

Publications in pop-science magazines:

- Kollár J. (2014): Je na Lamarckismu něco pravdy? Vesmír 93, 115, 2014/2.

- Kollár J. (2014): Malý krok pro rybu, velký pro čtvernožce. Vesmír 93, 244, 2014/4.

- Kollár J. (2014): Stará nová velryba. Vesmír 93, 245, 2014/4.

- Kollár J. (2015): Přímo před našima očima. Vesmír 94, 71, 2015/2.

- Kollár J. (2015): Na velikosti záleží! Vesmír 94, 191, 2015/4.

- Kollár J. (2015): Vidět se zkrátka vyplatí. Vesmír 94, 471, 2015/9.

- Kollár J. (2016): Evoluce v přímém přenosu? Proč ne! Vesmír, 95, 387, 2016/7-8.

- Kollár J. (2016): Čtvrtý druh. Vesmír 95, 614, 2016/11.

- Kollár J. (2018): Víc škody než užitku? Vesmír, 97, 394, 2018/7.

- Kollár J. (2019): Utvářela náš genom i neznámá lidská linie? Vesmír, 98, 201, 2019/4.

- Kollár J. (2020): Geny vznikající de novo, další přelom evoluční biologie. Vesmír 99, 76, 2020/2.

7. SOUHRN (SUMMARY IN CZECH)

Rozsivky (Bacillariophyta) jsou jednou z druhově nejbohatších skupin řas. Formálně popsáno je v současnosti více než 12 000 druhů, ale výpočty odhadované druhové bohatosti predikují, že jejich skutečný počet bude mezi 30 000 a 200 000. Rozsivky jsou kosmopolitní a adaptovány na život téměř ve všech vodních a mnohých terestrických habitatech. Jsou využívány v mnoha odvětvích základního (např. biodiverzita, evoluční biologie, ekologie, paleolimnologie) i aplikovaného výzkumu (nebo přímo jeho aplikacích; např. biomonitoring, biotechnologie, nanotechnologie). Pro mnohé z nich je klíčové pracovat se správným taxonomickým určením, které odráží evoluci daných evolučních linií.

Jedním z nejvážnějších problémů, jímž se zabývají nejen diatomologové, je problém kryptické a pseudokryptické druhové diverzity. (Pseudo)kryptické druhy jsou takové, které nejsou spolehlivě odlišitelné čistě na základě morfologie/anatomie (respektive jsou odlišitelné až po sofistikovanější morfologické analýze; v případě rozsivek například po detailním zkoumání pod elektronovým mikroskopem, či po statistické analýze morfologických dat). V posledních desetiletích se zejména díky analýze sekvenčních a reprodukčních dat ukázalo, že (pseudo)kryptická diverzita je velmi rozšířená napříč celým kládem rozsivek a že mnohé tradičně popsané rozsivkové druhy jsou ve skutečnosti komplexy několika (pseudo)kryptických druhů.

Rod *Pinnularia* je dle našich současných znalostí druhým druhově nejbohatším rodem rozsivek (po rodu *Navicula*) a obsahuje několik předpokládaných (pseudo)kryptických druhových komplexů. Jedním z nich je tzv. skupina *P. gibba*. V této disertační práci jsem shromáždil dataset 105 kmenů skupiny *P. gibba* a polyfázicky v něm delimitoval druhy. Primární (tj. založená na DNA) delimitace druhů založená na dvou genetických markerech a třech automatizovaných molekulárně-delimitačních metodách odhalila, že dataset pokryl 15 druhů skupiny *P. gibba*. Sekundární delimitace spočívala v konfrontaci této primární hypotézy s veškerými morfologickými, geografickými, environmentálními a reprodukčními daty dostupnými k daným kmenům. Finální hypotézou je, že dataset skutečně zahrnuje 15 druhů. Většina z nich je (pseudo)kryptická a vykazuje omezené geografické rozšíření.

Po přidání dalšího genetického markeru a dat z fosilního záznamu jsem zkonstruoval časově kalibrovaný multigenový evoluční strom, který odhalil nejen fylogenetické vztahy mezi delimitovanými druhy, ale i geologické časy divergence evolučních linií. Tato datovaná

fylogeneze dále umožnila formulaci a (v některých případech i) statistické testování evolučních/biogografickcých hypotéz. Mimo jiné byl v různých částech stromu objeven signifikantní rozdíl ve velikosti frustul daných druhů, který je nejlépe vysvětlitelný evolučním původem. Bylo též zjištěno, že genetické markery navržené pro DNA barkóding rozsivek (*rbc*L a SSU rDNA) mají dostatečné taxonomické rozlišení, aby jednoznačně rozlišili dokonce i sesterské (pseudo)kryptické druhy skupiny *P. gibba*. Taxnomická revize skupiny byla zahájena popisen jednoho z delimitovaných druhů jako *P. lacustrigibba* sp nov. Tato disertační práce rovněž vedla k vývoji nového probabilistického modelu speciace, který by v budoucnu mohl posloužit jako teoretický základ k vývoji nových probabilistických delimitačních metod. Ty by teoreticky mohly být aplikovatelné nejen na mladé druhy (z nichž mnohé mohou být kryptické), ale i na většinu asexuálních linií.