

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 VYVERMAN² & Aloisie POULÍČKOVÁ¹

¹Department of Botany, Faculty of Science, Palacký University Olomouc, Šlechtitelů 27, CZ-78371 Olomouc, CZ; *Corresponding author e-mail: jan.kollar@upol.cz

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

³Research 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

Abstract: Many diatom groups are known for widespread (pseudo)cryptic species diversity and *Pinnularia gibba* group is one of them. Recently, KOLLÁR et al. (2019) delimited species within the group 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. 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. 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.

Key words: 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. EHRENBURG

(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 *rbcL* 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 et al. (2020) 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: 2.5 µl of 10× PCR buffer (Tris–HCl, (NH₄)₂SO₄, KCl, 15 mM MgCl₂, pH 8.7 at 20 °C; “Buffer I”, Applied Biosystems, Foster City, USA), 200 µM of deoxynucleoside triphosphates at a concentration of 2 mM each, 0.5 µM of each primer (Table S1 – primers), 0.4 µg.mL^{−1} of bovine serum albumin (BSA), 1.25U of Taq polymerase (AmpliTaq, Perkin–Elmer, Wellesley, USA) and 1 µl of template DNA. The total reaction volume was adjusted to 25 µl 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

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

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 et al. 2020), (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 & 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 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 et al. (2020; see the source data file of their Suppl. Fig. 6c for full details). 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 (Fig. 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 $\pm 2.5 \mu\text{m}$ and $\pm 0.7 \mu\text{m}$ 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 1 μ M 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: [dx.doi.org/10.5883/DS-PIN3](https://doi.org/10.5883/DS-PIN3)) 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 1st/2nd codon sites, and the most variable 3rd 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 1st/2nd codon sites of *rbcL*, 0.072 for the 3rd codon sites of *rbcL*, 0.061 for the 1st/2nd

codon sites of *cox1*, and 0.349 for the 3rd 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 1st/2nd codon sites of *rbcL*, 0.704 and 0.506 for the 3rd codon sites of *rbcL*, 0.665 and 0.474 for the 1st/2nd codon sites of *cox1*, and 0.629 and 0.480 for the 3rd 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 (Fig. 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

Fig. 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 (Figs 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 \pm 9 Ma ago (Figure 2). The common ancestor of the terminal clade (i.e., F–G–H–I–S2–S3–S4) diverged 17 \pm 4 Ma ago. The Scottish sister species J (*P. cf. parvulissima*) and K (*P. lacustrigibba*) diverged 7 \pm 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 $70.2 \mu\text{m}$ in the basal lineages and $30.4 \mu\text{m}$ in the terminal clade. The average valve widths were $10.6 \mu\text{m}$ in the basal lineages and $5.9 \mu\text{m}$ 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 $791.9 \mu\text{m}^2$ 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

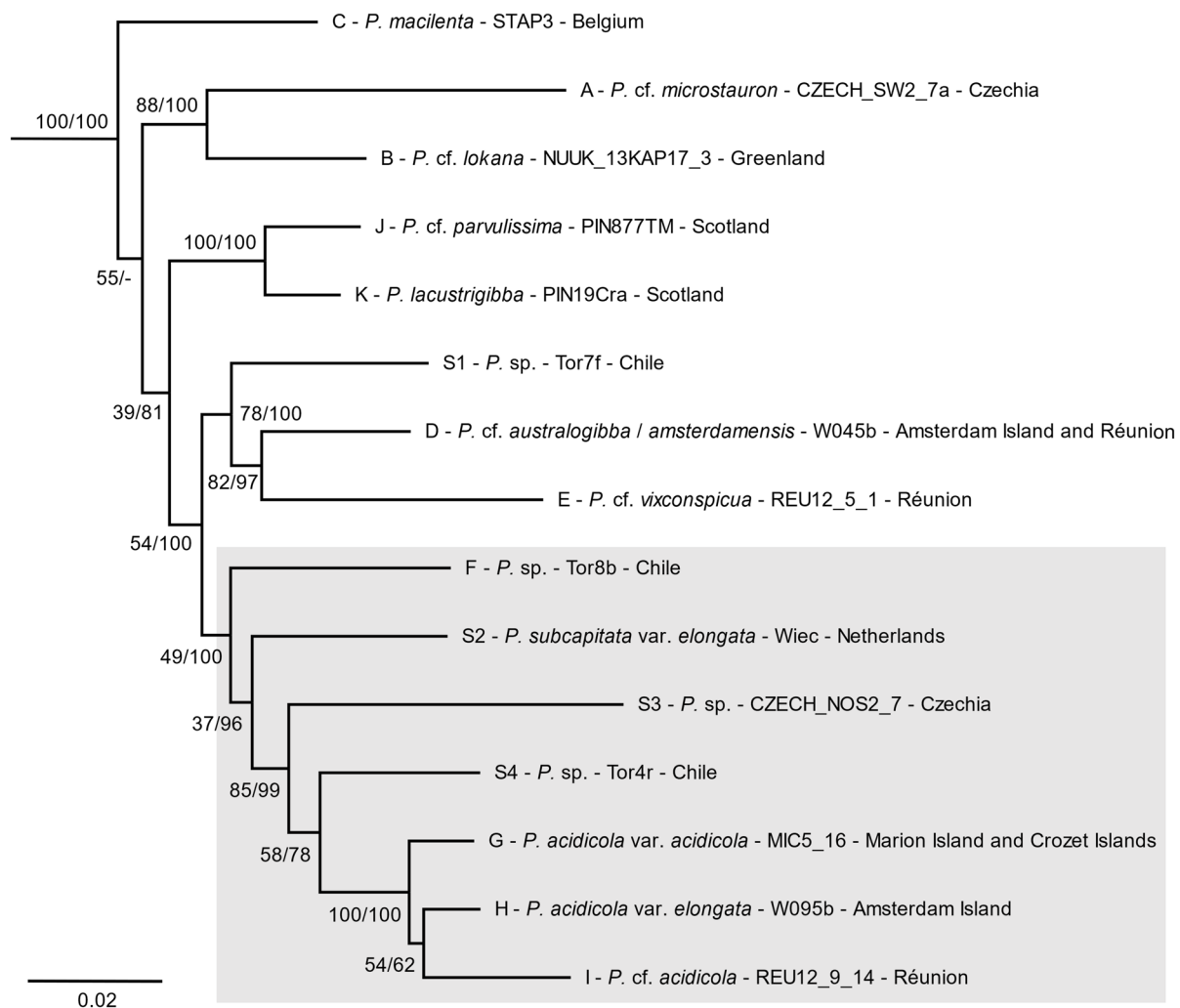


Fig. 1. Rooted phylogeny of the *Pinnularia gibba* group inferred from a three-locus DNA alignment using ML under a partitioned model. Values at nodes indicate statistical supports given as percentages, ML ultrafast bootstrap approximation / BI posterior probabilities. Whenever the ML and BI topologies 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.

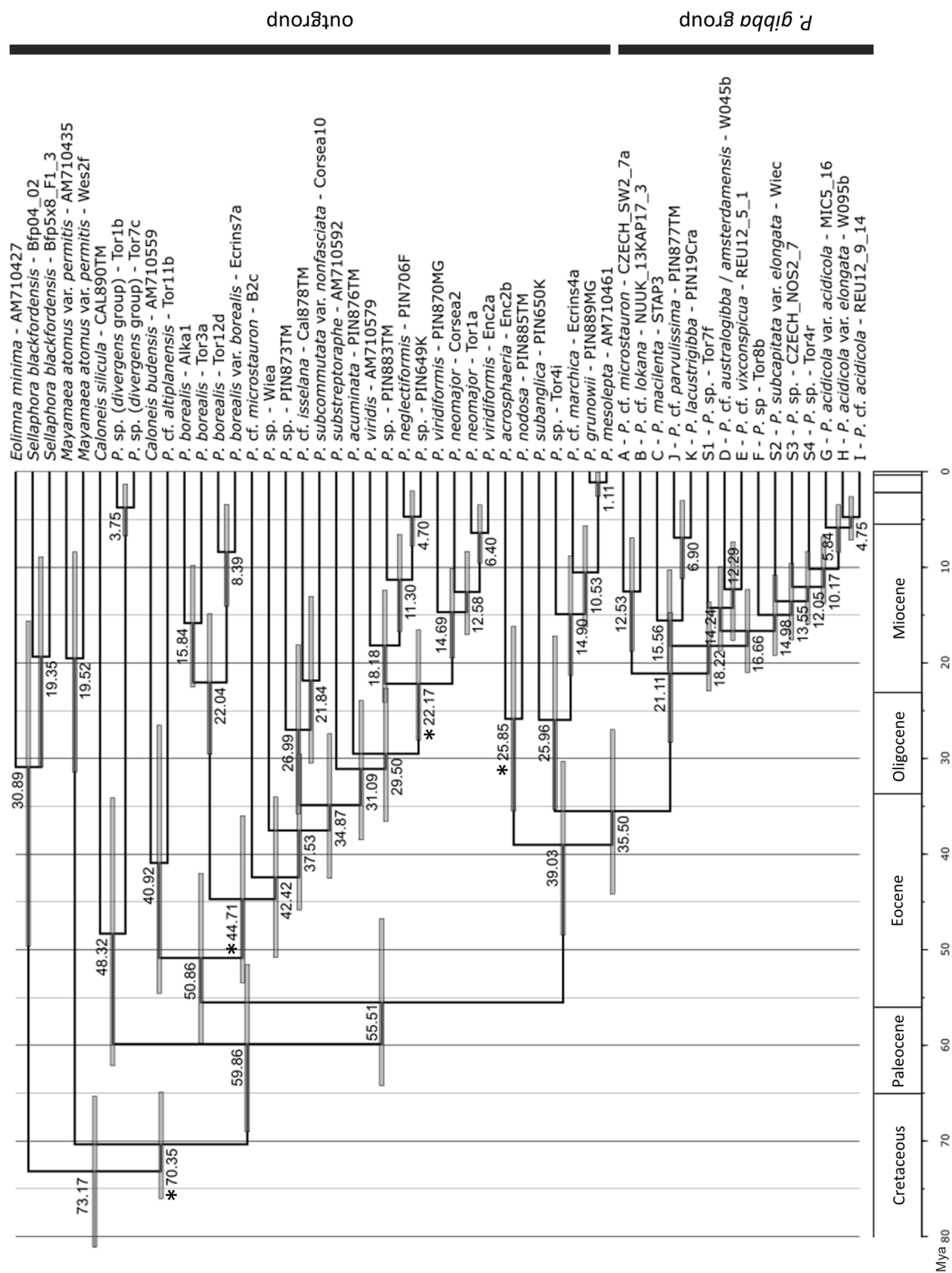


Fig. 2. A time-calibrated phylogeny of the *Pinnularia gibba* group inferred from a three-locus DNA alignment using Bayesian relaxed molecular clock analysis with four fossil-based time constraints (indicated by asterisks). Values at nodes are mean node ages. Grey bars are 95% HPD (highest posterior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

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 \pm standard deviation.

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

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 cytoplasmic 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 & 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 et al. 2020). 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, figs 10–12) by KRAMMER (2000). Given that the character is probably homologous in all three subclades (together they form clade B sensu SOUFFREAU 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 prove 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. This research was financially supported by the Internal grant agency of the Faculty of Science at Palacký University in Olomouc IGA PrF-2020-003 and PrF-2021-001 (JK and AP). 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.; BRUGEROLLE, G.; FARMER, M.A.; KARPOV, S.; KOLISKO, M.; LANE, C.E.; LODGE, D.J.; MANN, D.G.; MEISTERFELD, R.; MENDOZA, L.; MOESTRUP, Ø.; MOZLEY-STANDRIDGE, S.E.; SMIRNOV, A.V. & SPIEGEL, F. (2007): Diversity, nomenclature, and taxonomy of protists. – *Syst. Biol.* 56:684–689.
- 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.* 1:267–292.
- ATKINSON, K.M. (1972): Birds as transporters of algae. – *Br. Phycol. J.* 7:319–321.
- 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* 17:209–221.
- BOFKIN, L. & GOLDMAN, N. (2007): Variation in evolutionary processes at different codon positions. – *Mol. Biol. Evol.* 24:513–521.
- BOUCKAERT, R.; VAUGHAN, T.G.; BARIDO-SOTTANI, J.; DUCHÊNE, S.; FOURMENT, M.; GAVRYUSHKINA, A.; HELED, J.; JONES, G.; KÜHNERT, D.; DE MAIO, N.; MATSCHINER, M.; MENDES, F.K.; MÜLLER, N.F.; OGILVIE, H.A.; DU PLESSIS, L.; POPINGA, A.; RAMBAUT, A.; RASMUSSEN, D.; SIVERONI, I.; SUCHARD, M.A.; WU, C.H.; XIE, D.; ZHANG, C.; STADLER, T. & DRUMMOND, A.J. (2019): BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. – *PLoS Comput. Biol.* 15:1–28.
- 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.* 13:581–583.
- CASTRESANA, J. (2000): Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. – *Mol. Biol. Evol.* 17:240–252.
- CHRISTENHUSZ, M.J.M. & BYNG, J.W. (2016): The number of known plants species in the world and its annual increase. – *Phytotaxa* 261:201–217.
- 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.* 44:124–131.
- COX, E.J. (2009): What's in a name? – Diatom classification should reflect systematic relationships. – *Acta Bot. Croat.* 68:443–454.
- CROLL, D.A. & HOLMES, R.W. (1982): A Note on the Occurrence of Diatoms on the Feathers of Diving Seabirds. – *Auk* 99:765–766.
- DRUM, R.W. & GORDON, R. (2003): Star Trek replicators and diatom nanotechnology. – *Trends Biotechnol.* 21:325–328.
- EDGAR, R.C. (2004): MUSCLE: multiple sequence alignment with improved accuracy and speed. – *Proceedings. 2004 IEEE Comput. Syst. Bioinforma. Conf.* 2004. CSB 2004. 689–690.
- EDGAR, R.C. (2016): UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. – *bioRxiv*. 81257. DOI: <https://doi.org/10.1101/081257>.
- 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: 43–49.
- 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: 8927–8932.
- 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: 368–381.
- 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: 525–529.
- HALL, K.; MEIKLEJOHN, I. & BUMBY, A. (2011): Marion Island volcanism and glaciation. – *Antarct. Sci.* 23: 155–163.
- HAMSHER, S.E.; EVANS, K.M.; MANN, D.G.; POULÍČKOVÁ, A. & SAUNDERS, G.W. (2011): Barcoding diatoms: Exploring alternatives to COI–5P. – *Protist* 162: 405–422.
- HARTMANN, D.L. (2015): Global physical climatology (2nd ed.). – 498 pp., Newnes, Boston.
- HASEGAWA, M.; KISHINO, H. & YANO, T. (1985): Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. – *J. Mol. Evol.* 22: 160–174.
- 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: 313–321.
- 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. – *The Plant J.* 54: 621–639.
- KELLY, M.; BOONHAM, N.; JUGGINS, S.; KILLE, P.; MANN, D.G.; PASS, D.; SAPP, M.; SHINYA, S. & GLOVER, R. (2018): A

- DNA based diatom metabarcoding approach for Water Framework Directive classification of rivers. – Science Report SC140024/R, Environment Agency, Bristol.
- KIMURA, M. (1981): Estimation of evolutionary distances between homologous nucleotide sequences b. – *Proc. Natl. Acad. Sci. U. S. A.* 78: 454–458.
- 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). – 31 pp., Palacký University Olomouc, Olomouc.
- 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.
- KRAMMER, K. (1992): *Pinnularia* eine Monographie de europäischen Taxa. – *Bibl. Diatomol.* 26: 1–353.
- KRAMMER, K. (2000): The genus *Pinnularia*. – 703 pp., A.R.G. Ganter Verlag K.G., Ruggell.
- KUMAR, S.; STECHER, G.; LI, M.; KNYAZ, C. & TAMURA, K. (2018): MEGA X: Molecular evolutionary genetics analysis across computing platforms. – *Mol. Biol. Evol.* 35: 1547–1549.
- 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: 179–196.
- LEMEX, P.; SALEMI, M. & VANDAMME, A.M. (eds; 2009): *The Phylogenetic Handbook* (2nd ed.). – 723 pp., 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.; MACHLUS, M.L.; NEWTON, A.E.; RAINE, J.I.; WILENBRING, J.K.; WILLIAMS, M. & WOLFE, A.P. (2008): Mid–Miocene cooling and the extinction of tundra in continental Antarctica. – *Proc. Natl. Acad. Sci.* 105: 10676–10680.
- MANN, D.G.; McDONALD, S.M.; BAYER, M.M.; DROOP, S.J.M.; CHEPURNOV, V.A.; LOKE, R.E.; CIOBANU, A. & DU BUF, J.M.H. (2004): The *Sellaphora pupula* species complex (Bacillariophyceae): Morphometric analysis, ultrastructure and mating data provide evidence for five new species. – *Phycologia* 43: 459–482.
- MANN, D.G. & EVANS, K.M. (2008): The species concept and cryptic diversity. – In: MOESTRUP, Ø.; DOUCETTE, G.; ENEVOLDSEN, H.; GODHE, A.; HALLEGRAEFF, G.; LUCKAS, B.; LUNDHOLM, N.; LEWIS, J.; RENGEFORS, K.; SELLNER, K.; STEIDINGER, K.; TESTER, P. & ZINGONE, A. (eds): *Proceedings of the 12th International Conference on Harmful Algae*. – pp. 262–268, International Society for the Study of Harmful Algae and Intergovernmental Oceanographic Commission of UNESCO, Copenhagen.
- 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. Algal.* 31: 557–577.
- MANN, D.G. & VANORMELINGEN, P. (2013): An inordinate fondness? The number, distributions, and origins of diatom species. – *J. Eukaryot. Microbiol.* 60: 414–420.
- 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: 1188–1195.
- 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.; 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.
- 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: 462–473.
- NASA (1976): U.S. Standard Atmosphere. – 241 pp., National Aeronautics and Space Administration, Washington D.C.
- 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: 330–347.
- PINSEEL, E.; VANORMELINGEN, P.; HAMILTON, P.B.; VYVERMAN, W.; VAN DE VIJVER, B. & KOPALOVA, K. (2017): Molecular and morphological characterization of the *Achnanthisidium minutissimum* complex (Bacillariophyta) in Petuniabukta (Spitsbergen, High Arctic) including the description of *A. digitatum* sp. nov. – *Eur. J. Phycol.* 52: 264–280.
- 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: 121–140.
- PINSEEL, E.; JANSSENS, S.B.; VERLEYEN, E.; VANORMELINGEN, P.; KOHLER, T.J.; BIERMA, E.M.; SABBE, K.; VAN DE VIJVER, B. & VYVERMAN, W. (2020): Global Radiation in a Rare Biosphere Soil Diatom. – *Nat. Commun.* 11: 1–12.
- POSADA, D. (2003): Using Modeltest and PAUP* to Select a Model of Nucleotide Substitution. In: *Current Protocols in Bioinformatics*. – pp. 6.5.1–6.5.14, John Wiley & Sons, Inc., New York.
- POULÍČKOVÁ, A.; MAYAMA, S.; CHEPURNOV, V.A. & MANN, D.G. (2007): Heterothallic auxosporulation, incunabula and perizonium in *Pinnularia* (Bacillariophyceae). – *Eur. J. Phycol.* 42: 367–390.
- 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: 353–369.
- 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.
- R CORE TEAM (2016): R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria (<https://www.R-project.org/>).
- 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: 901–904.
- RONQUIST, F.; TESLENKO, M.; VAN DER MARK, P.; AYRES, D.L.; DARLING, A.; HÖHNA, S.; LARGET, B.; LIU, L.; SUCHARD, M.A. & HUESENBECK, J.P. (2012): Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. – *Syst. Biol.* 61: 539–542.
- 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.). – 667 pp., Cambridge University Press, Cambridge.
- SOUFFREAU, C. (2011): *Biogeography, species diversity and stress tolerance of aquatic and terrestrial diatoms* (Doctoral Thesis). – 285 pp., Ghent University, Ghent.
- 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: 309–324.
- SOUFFREAU, C.; VERBRUGGEN, H.; WOLFE, A.P.; VANORMELINGEN, P.; SIVER, P.A.; COX, E.J.; MANN, D.G.; VAN DE VIJVER, B.; SABBE, K. & VYVERMAN, W. (2011): A time-calibrated multi-gene phylogeny of the diatom genus *Pinnularia*. – *Mol. Phylogenet. Evol.* 61: 866–879.
- 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: 246–255.
- 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: 101–115.
- STOCK, W.; PINSEEL, E.; DE DECKER, S.; SEFBOM, J.; BLOMMAERT, L.; CHEPURNOVA, O.; SABBE, K. & VYVERMAN, W. (2018): Expanding the toolbox for cryopreservation of marine and freshwater diatoms. – *Sci. Rep.* 8: 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 HAESLER, A. & MINH, B.Q. (2016): W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. – *Nucleic Acids Res.* 44: 232–235.
- 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: 443–459.
- VAN DE VIJVER, B.; CHATTOVÁ, B.; METZELTIN, D. & LÉBOUVIER, M. (2012): The genus *Pinnularia* (Bacillariophyta) on Ile Amsterdam (TAAF, Southern Indian Ocean). – *Nova Hedwigia* 141: 201–236.
- 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: 1791–1798.
- 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: 2478–2484.
- 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* Ssensu 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: 229–252.
- VYVERMAN, W.; VERLEYEN, E.; WILMOTTE, A.; HODGSON, D.A.; WILLEMS, A.; PEETERS, K.; VAN DE VIJVER, B.; DE WEVER, A.; LELIAERT, F. & SABBE, K. (2010): Evidence for widespread endemism among Antarctic micro-organisms. – *Polar Sci.* 4: 103–113.
- XIA, X. (2013): DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. – *Mol. Biol. Evol.* 30: 1720–1728.
- 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: 173–192.

Supplementary material

The following supplementary material is available for this article:

- Table S1. The primer sequences used in this study.
- Table S2. List of 36 outgroup strains used in the molecular clocks analysis with Genbank identifiers.
- Fig. S1. Saturation curves for protein-coding (A-D) and ribosomal (E) markers.
- Fig. S2. Rooted phylogeny of the *Pinnularia gibba* group inferred from a LSU rDNA alignment using ML under TIM3+F+I+G4 substitution model.
- Fig. S3. Rooted phylogeny of the *Pinnularia gibba* group inferred from a rbcL alignment using ML under K3Pu+F+I+G4 substitution model.
- Fig. S4. Rooted phylogeny of the *Pinnularia gibba* group inferred from a cox1 alignment using ML under GTR+F+I+G4 substitution model.

This material is available as part of the online article (<http://fottea.czechphycology.cz/contents>)