Protist, Vol. 170, 121–140, April 2019
http://www.elsevier.de/protis
Published online date 13 October 2018

Protist

ORIGINAL PAPER

Extensive Cryptic Diversity in the Terrestrial Diatom *Pinnularia borealis* (Bacillariophyceae)

Eveline Pinseel[a,b,c,1,2], Jana Kulichová[d,2], Vojtěch Scharfen[d], Pavla Urbánková[d], Bart Van de Vijver[b,c,3], and Wim Vyverman[a,3]

[a]Protistology & Aquatic Ecology (PAE), Department of Biology, Faculty of Science, Ghent University, Krijgslaan 281-S8, B–9000 Ghent, Belgium
[b]Research Department, Botanic Garden Meise, Nieuwelaan 38, B–1860 Meise, Belgium
[c]Ecosystem Management Research Group (ECOBE), Department of Biology, Faculty of Science, University of Antwerp, Universiteitsplein 1, B–2610 Wilrijk, Antwerp, Belgium
[d]Department of Botany, Faculty of Science, Charles University in Prague, Benátská 2, CZ–12801 Prague 2, Czech Republic

Submitted April 30, 2018; Accepted October 2, 2018
Monitoring Editor: Wiebe H. C. F. Kooistra

With the increasing application of molecular techniques for diatom species discovery and identification, it is important both from a taxonomic as well as an ecological and applied perspective, to understand in which groups morphological species delimitation is congruent with molecular approaches, or needs reconsideration. Moreover, such studies can improve our understanding of morphological trait evolution in this important group of microalgae. In this study, we used morphometric analysis on light microscopy (LM) micrographs in SHERPA, detailed scanning electron microscopy (SEM), and cytological observations in LM to examine 70 clones belonging to eight distinct molecular lineages of the cosmopolitan terrestrial diatom *Pinnularia borealis*. Due to high within-lineage variation, no conclusive morphological separation in LM nor SEM could be detected. Morphological stasis due to the "low-morphology" problem or stabilizing selection, as well as parallel/convergent evolution, phenotypic plasticity and structural inheritance are discussed as potential drivers for the observations. Altogether, *P. borealis* is truly cryptic, in contrast to the majority of other diatom species complexes which turned out to be pseudo-cryptic following detailed morphological analysis.

Key words: Diatoms; LSU rDNA; molecular phylogenies; morphometrics; *rbc*L; shape.

Introduction

Diatoms (Bacillariophyta) are highly diverse and, with an estimated number of species ranging between 30,000 and 200,000 (Mann 1999; Mann and Vanormelingen 2013), are generally considered to be the most species-rich group of algae on

[1]Corresponding author.
[2]These authors contributed equally and should be considered shared first authors
[3]These authors contributed equally and should be considered shared last authors
e-mail eveline.pinseel@gmail.com (E. Pinseel).

https://doi.org/10.1016/j.protis.2018.10.001
1434-4610/

Earth. Diatoms are widely used in various applied sciences such as water quality biomonitoring (Kelly 1998) and paleo-ecological reconstructions (Smol and Stoermer 2010), mainly because of the possibility of long-term preservation of their siliceous cell wall in marine, lake and peat sediments and the specific ecological preferences of individual taxa. These applications are based on the assumption that individual species can be identified based on the morphological features of the cell wall discernible in light microscopy (LM). Nevertheless, an increasing number of molecular studies has shown that many, often common, diatom morphospecies harbour distinct molecular variation, which likely corresponds to species-level differentiation (Beszteri et al. 2007; Kermarrec et al. 2013; Pinseel et al 2017b; Quijano-Scheggia et al. 2009; Souffreau et al. 2013b; Vanelslander et al. 2009; Vanormelingen et al. 2013). The presence of (pseudo)cryptic diversity not only hampers the use of diatom morphologies in applied sciences, but also influences our understanding of species distributions, niche differentiation and diatom evolution. However, in recent years, the development of various tools for quantitative morphometric analysis of diatom valves, such as landmark analysis (Bookstein 1997) or SHERPA (Kloster et al. 2014, 2016), has revealed that several members of these morphospecies complexes are morphologically differentiated in a subtle way, often involving small differences in valve shape (Beszteri et al. 2005a; Kloster et al. 2018; Mann et al. 2004; Poulíčková et al. 2010; Veselá et al. 2009). Some studies found a link between morphological and ecological differentiation (Kulichová and Fialová 2016; Potapova and Hamilton 2007; Urbánková et al. 2016). In other cases, morphometric analyses were used to distinguish or describe species with highly similar morphologies (Fránková et al. 2009; Mann et al. 2004; Theriot 1992; Van de Vijver et al. 2013). Morphometric methods can thus be a powerful tool to study the morphological variation in diatoms, and to reveal potential links among morphological, molecular and eco(physio)logical differentiation between and within closely related diatom taxa. This could have important implications for (paleo)environmental inferences where morphologically similar and often difficult to identify diatom taxa are usually lumped together, thereby ignoring the potential presence of different ecotypes.

One diatom species complex that has gained interest over the last years is *Pinnularia borealis* Ehrenberg. It is a common, cosmopolitan component of (semi)terrestrial habitats such as soils and mosses (Krammer 2000; Krammer and Lange-Bertalot 1988; Souffreau et al. 2013b), which is well-adapted to living in extreme environments (Hejduková et al. 2017; Souffreau et al. 2013a; Stock et al. 2018). Occasionally, *P. borealis* can also be found living in the littoral zones of freshwater lakes and ponds (Pinseel et al. 2017a; Rumrich et al. 2000). Based on morphological features of the cell wall discernible in LM, taxonomists separated about 66 different varieties and formae within *P. borealis* (Kociolek et al. 2018), although most of these considerably overlap in their morphological features resulting in an uncertain taxonomic status. Recent molecular phylogenetic analyses based on the nuclear encoded D1–D3 region of the LSU rDNA (28S) and the plastid gene *rbc*L revealed that *P. borealis* consists of several distinct lineages that likely separated several millions of years ago (Souffreau et al. 2013b). Several of these lineages showed niche-differentiation regarding optimal growth temperature and upper temperature limits for growth, altogether suggesting they represent distinct species (Souffreau et al. 2013b). However, the molecular lineages identified by Souffreau et al. (2013b) did not show obvious morphological discontinuities in LM preventing their separation using traditional measurements of their valve morphologies, such as length, width and stria densities. Nevertheless, although they could shed light on the true morphological diversity of this species complex, detailed morphometric techniques were not used, nor was ultrastructure examined. The possibility to reliably identify lineages within *P. borealis* using morphological features would be valuable in light of recent interests to use terrestrial diatom species for monitoring of soil ecological quality (Barragán et al. 2018) or to provide the much desired link between molecular studies and original type material for which no genetic material is available. Furthermore, considering the amount of already available data on *P. borealis*, this taxon forms an ideal model system to study the (morphological) evolution and differentiation of species-level lineages within diatoms.

In this study, monoclonal cultures of *P. borealis* were established, originating from various (sub-)polar, temperate and Mediterranean sites in the Northern and Southern hemisphere. First, using their sequences of the nuclear encoded D1–D3 LSU rDNA (28S) and the plastid gene *rbc*L, the phylogenetic position of these newly obtained strains was assessed in relation to the already existing *P. borealis* phylogeny (Pinseel et al. 2017a; Souffreau et al. 2013b) and additional new lineages within *P. borealis* were delineated. The choice for D1–D3

28S was motivated by the fact that this marker proved to be highly variable in the genus *Pinnularia* Ehrenberg, highlighting its usefulness to distinguish between closely related *Pinnularia* species (Kollár et al. 2018). Although *rbc*L is distinctly less variable in the genus *Pinnularia* (Kollár et al. 2018), it is a useful marker to provide a good resolution at deeper nodes in the phylogeny. Second, morphometric analyses and ultrastructural analysis in scanning electron microscopy (SEM) were performed on all newly obtained strains as well as a selected subset of previously studied strains (Souffreau et al. 2013b), to check the hypothesis that different molecular lineages within *P. borealis* can be distinguished using detailed morphological information. Finally, cytological observations were carried out in LM to test for divergence in plastid arrangement and general morphology, and nucleus morphology. The results of the morphometric analysis were assessed within the context of the phylogenetic lineage to which the strains belong.

Results

Taxon Sampling

In total 48 new strains of *Pinnularia borealis* were established (Fig. 1, Supplementary Material Table S1). The dataset was complemented with culture material from 22 *P. borealis* strains already published in Souffreau et al. (2013b) (Fig. 1, Supplementary Material Table S1), resulting in a total of 70 strains of *P. borealis* to be examined for their morphology in this study. For the phylogenetic analysis (see below), all published strains were analysed to put the results in a wider context (Supplementary Material Table S2). All selected strains showed highly similar valve morphologies (Fig. 2). None of the cultures showed growth abnormalities during culturing, with the rare exception of an occasionally malformed valve or raphe. Such malformed valves were not selected for the morphometric analysis. Although strains differed in their average cell length, approximately the same size range was sampled for

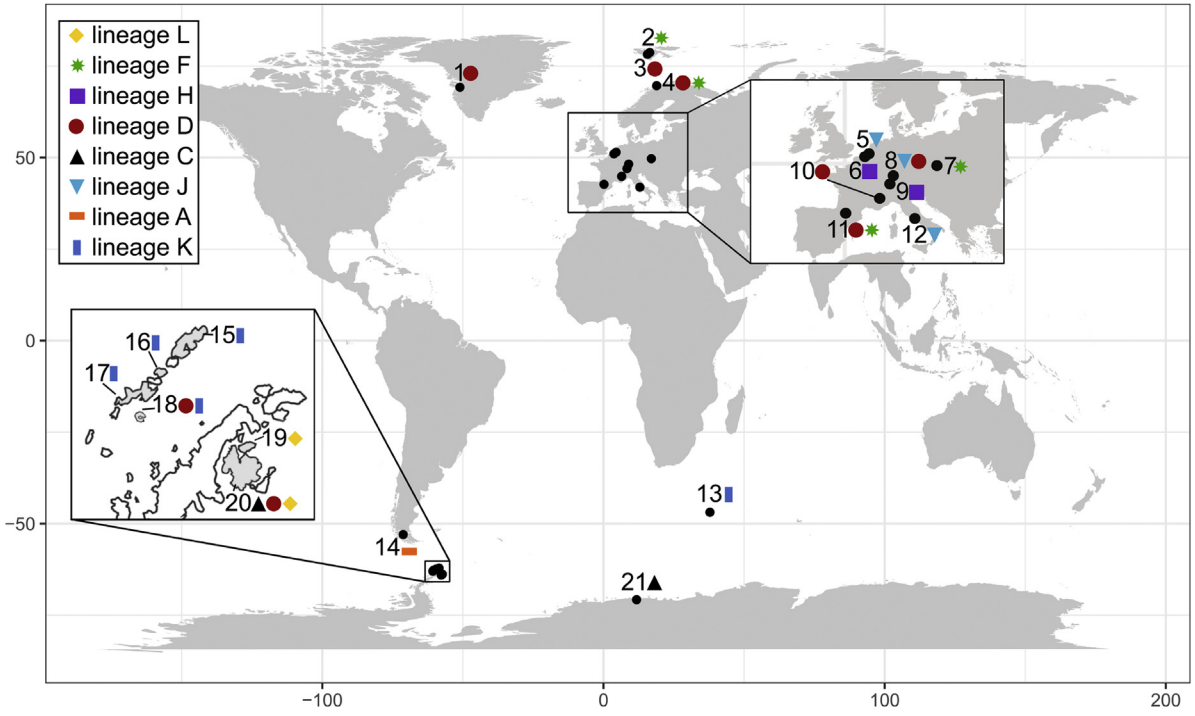

Figure 1. Map with indication of the sampling localities (black dots) of the here examined strains. The symbols indicate the molecular identity of the cultures found at each sampling location. Legend of the localities: 1. Greenland – Illulisat, 2. Spitsbergen – Petuniabukta, 3. Spitsbergen – Longyearbyen, 4. Norway – Tromsø, 5. Belgium – Essen, 6. Belgium – Gent, 7. Czech Republic – Podkova, 8. Germany – Swabian Jura, 9. Switzerland – Lucerne, 10. France – Ailfroide, 11. Spain – the Pyrenees, 12. Italy – Monte Cerella, 13. Sub-Antarctica – Marion Island, 14. Chile – Torres del Paine, 15. Maritime Antarctica – King George Island, 16. Maritime Antarctica – Robert Island, 17. Maritime Antarctica – Livingston Island, 18. Maritime Antarctica – Deception Island, 19. Maritime Antarctica – Vega Island, 20. Maritime Antarctica – James Ross Island, and 21. Continental Antarctica – Schirmacher Oasis.

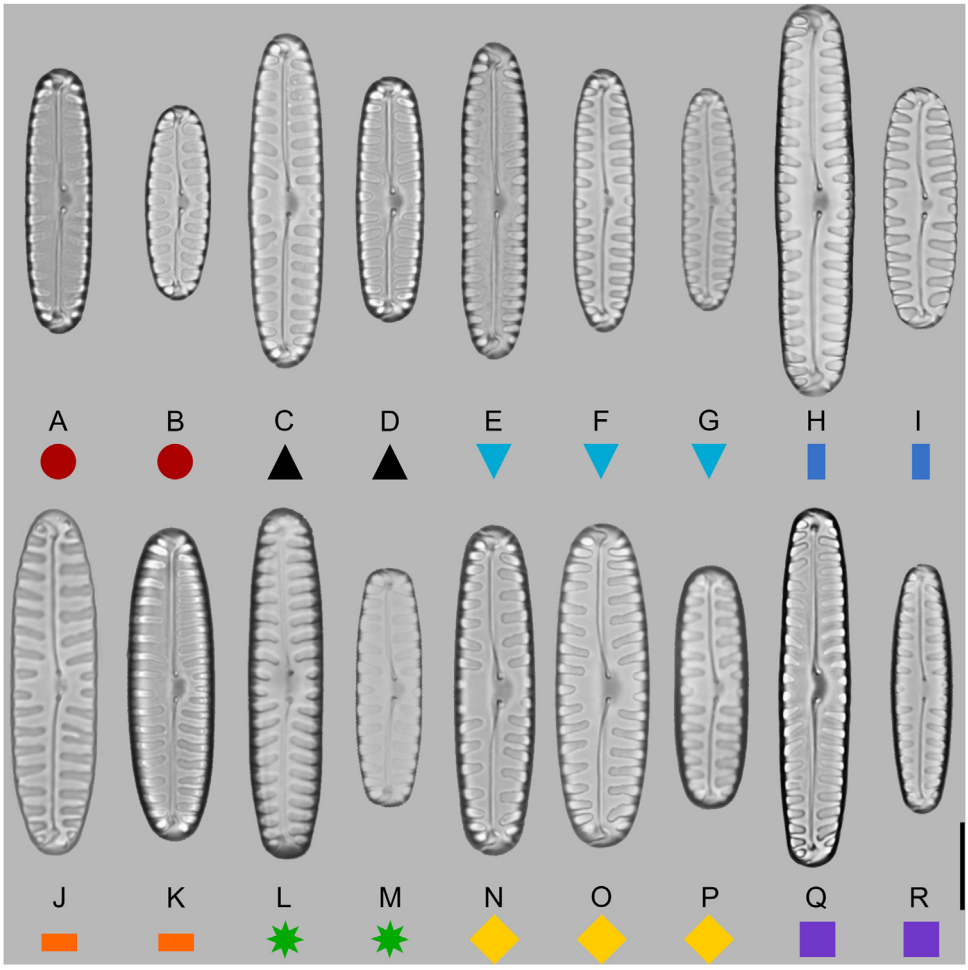

Figure 2. LM micrographs of oxidised material of a subset of the strains analysed in this study. The symbols indicate the molecular identity of the cultures (see Fig. 1 for a legend). Strains: **A**) NW14/2-C5; **B**) JRI15/18b-02; **C**) SCHIR-P10; **D**) JRI15/18b-09; **E**) BE14/E03-02; **F**) RO15-03-09; **G**) DE15/20-05; **H**) Mit7-15; **I**) SSI15/M18-11; **J**) (Tor12)d; **K**) (Tor12)h; **L**) SP14/T003-07; **M**) NW14/2-C3; **N**) JRI15/10-01; **O**) JRI15/32-13; **P**) VEGA-L35C-19; **Q**) (Sterre1)d; **R**) LU15-09-07. Scale bar = 10 μm.

the different lineages (Supplementary Material Fig. S1). We chose not to identify the different strains upon morphovariety level using the morphological literature as unambiguous identification using the standard literature proved difficult. This is due to the distinct overlap in morphological characteristics between multiple morphovarieties (Krammer 2000).

Molecular Phylogenetic Analysis

Maximum Likelihood (ML) and Bayesian Inference (BI) phylogenetic analysis on the concatenated dataset of 28S and *rbc*L produced identical results and therefore only the ML phylogeny is shown (Fig. 3). All eight lineages defined by Souffreau et al. (2013b), as well as the lineage of *Pinnularia catenaborealis* (Pinseel et al. 2017a) were recovered with (close to) maximal bootstrap and PP support (Fig. 3). Four of the lineages identified by Souffreau et al. (2013b) were also recovered from other regions (Fig. 3, lineages C, D, F and H). Apart from the already known lineages, three new monophyletic clusters of sequences preceded by a relatively long branch with maximal statistical support were recovered (Fig. 3, lineages J, K and L), suggesting these new clusters correspond to distinct species. Some lineages were found to occur sympatrically within the same island, i.e. strains from Deception Island (strain code 'DEC') and James Ross Island (strain code 'JRI'), or even sample, i.e. strains JRI15/18b-02 and JRI15/18b-

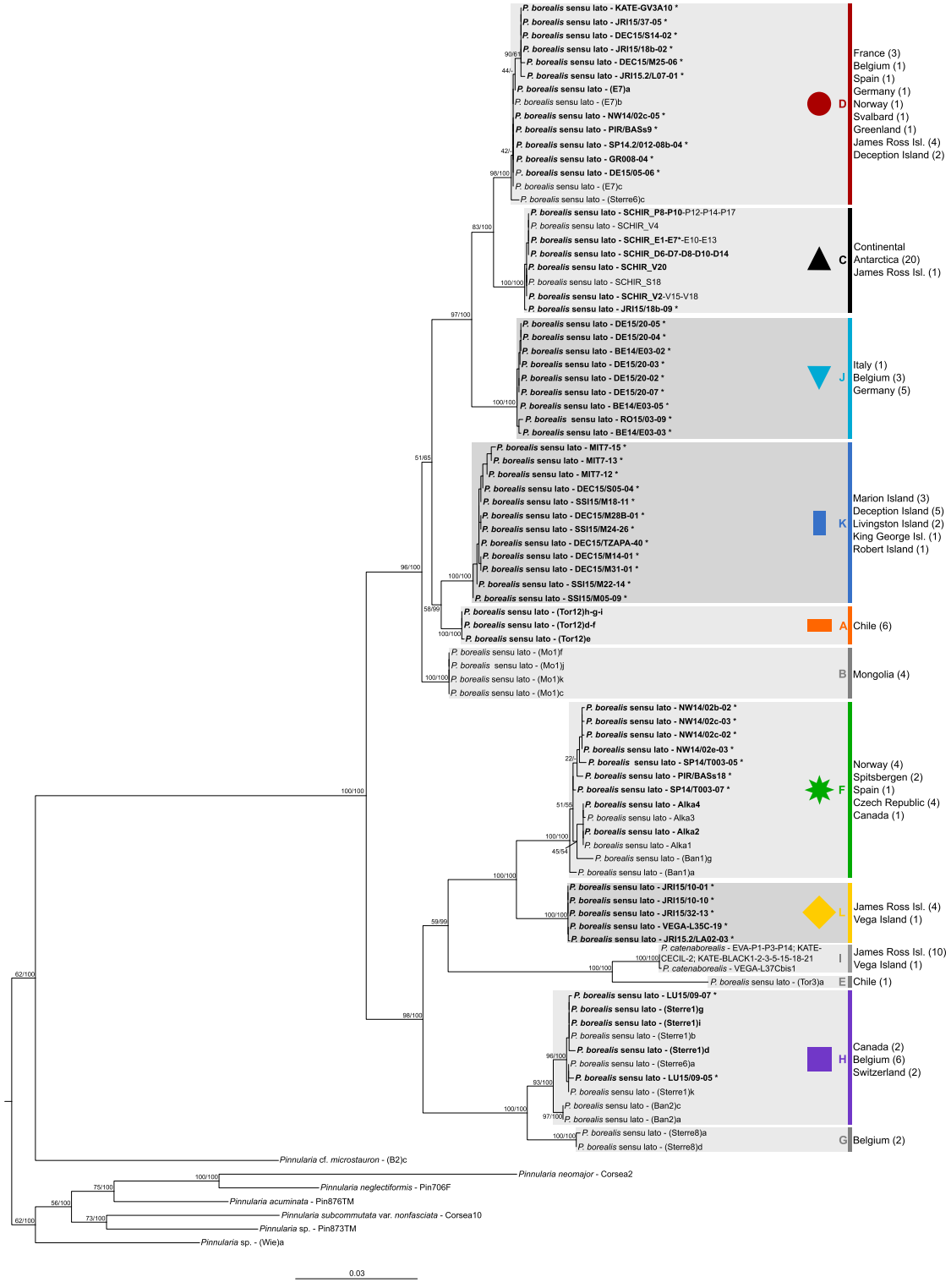

Figure 3. Maximum likelihood phylogeny of *P. borealis* based on the concatenated dataset of the nuclear-encoded D1-D3 LSU rDNA and the plastid gene *rbc*L with indication of statistical support (ML bootstrap proportions/BI posterior probabilities). The different lineages are indicated with grey boxes, with the dark grey boxes indicating lineages newly discovered in this study. Lineage numbers are based on Souffreau et al. (2013b) and Pinseel et al. (2017a). Lineages examined in the morphometric analysis are indicated by symbols. Strains in bold were analysed in the morphometric analysis. Strains indicated with an asterisk were newly obtained for this study.

09 of respectively lineages D and C, and strains PIR/BASs9 and PIR/BASs18 of respectively lineages D and F.

The nuclear encoded 28S gene was the most variable marker in this study (Supplementary Material Tables S3–S4). Strains differed from each other by a maximum of 114 bp (18.8%) for 28S, whereas this variation was much lower for *rbc*L, with a maximum sequence divergence of 43 bp (3.5%). When only taking into account the strains used for the morphometric analysis, maximum sequence divergence equalled 92 bp (10.3%) for 28S (Supplementary Material Table S3) and 38 bp (2.7%) for *rbc*L (Supplementary Material Table S4). The two most closely related lineages (lineages C and D) differed from each other by 13–17 bp (1.4–1.9%) for 28S (Supplementary Material Table S3) and 9–14 bp (0.7–1.0%) for *rbc*L (Supplementary Material Table S4). All lineages examined by morphometry showed intralineage sequence variation in 28S and/or *rbc*L. For *rbc*L, only lineage C did not show any divergence, whereas for 28S, all strains belonging to lineages A and L had identical sequences. Intralineage sequence divergence was highest in lineage F, with strains of this lineage differing up to 8 bp (1.3%) for 28S (Supplementary Material Table S3) and 3 bp (0.2%) for *rbc*L (Supplementary Material Table S4). Lineage D had the widest geographic distribution of all lineages in this study, which was also (partially) reflected in the sequence divergence in lineage D (Supplementary Material Tables S3–S4). In contrast, lineage K, which had a relatively restricted distribution, being retrieved from a series of islands in Antarctica, showed distinct intra-lineage sequence divergence (Supplementary Material Tables S3–S4). Phylogeographic structuring was evident in lineages D, F and K (Fig. 3), which contained separate subclades with strains from respectively Antarctica, Czech Republic and Sub-Antarctica. Specifically for lineage D, and taking into account all strains used for the phylogenetic analysis, the Antarctic strains differed by 2–6 bp for 28S and 1–3 bp for *rbc*L from their relatives in the Northern Hemisphere, although within hemisphere strain differentiation also existed for 28S. Despite relatively high sequence divergence, phylogeographic structuring was not evident in lineage H, where multiple intermediate haplotypes were detected.

Morphometric Analyses

The morphometric analysis was based on 20 micrographs per strain, photographed using bright field light microscopy (BF). Euclidean-distance based ANISOM analysis on the morphometric parameters measured in this study revealed that the morphology of valves within lineages was more similar than the morphology of valves belonging to different lineages ($P < 0.01$), except for lineages H-D and A–K (Supplementary Material Fig. S2). Nevertheless, it proved impossible to reliably discriminate the lineages, since all are highly variable in morphology and partially or completely overlap in their morphological features (Fig. 4, Table 1). Nevertheless, some centroid of strains clearly cluster separately in the PCA graphs. The strains with the smallest perimeters (DEC15/M25-06, DEC15/S14-02), the largest strains in width (JRI15/32-13, JRI15.2/LA02-03) and the strains with a combination of the largest perimeter and area (Mit7-12, Mit7-13, Mit7-15) are separated from all other strains along the first two PC axes (Fig. 4A). These three clusters represented both different lineages (D, L, K) and sampling sites: i.e. Deception Island (strain code 'DEC'), James Ross Island (strain code 'JRI') and Marion Island (strain code 'Mit'). However, strains from other geographic areas, but representing the same lineages, or strains from the same geographic region, but from different lineages, are dispersed in the PCA plot (Fig. 4A). Distinct variability in valve shape and size was also found among strains from the same sampling area and lineage, i.e. in lineage C from Continental Antarctica (strain code 'SCHIR'). Only three lineages, represented by strains from different sampling sites, form relatively separate clusters along the PC1-PC2 axis (Fig. 4A): lineage J and lineage L, and PC1-PC4 axis (Fig. 4B): lineage H. Width, rectangularity, and PCAF (Percent Concave Area Fraction) are mostly different among these lineages (linear mixed effect models: $P < 0.001$). However, it is impossible to discriminate individual valves and/or strains from these three lineages as some strains are morphologically more similar to strains from a different lineage than to strains from their own lineage. Linear mixed effect models did not support the separation of lineages L and H from other lineages using width, rectangularity and/or PCAF, but a combination of these variables enabled separation of lineages F, J, and K (Supplementary Material Table S5). Nevertheless, as mentioned above, the ranges of these parameters are not discrete among lineages (Table 1). To estimate the success of discrimination among all eight lineages the gradient boosting algorithm XGBoost (Chen and Guestrin 2016) was used. The resulting confusion matrix showed a classification power between 56% (for lineage A) and 78.8% (for lineage H) (Supplementary Material Table S6).

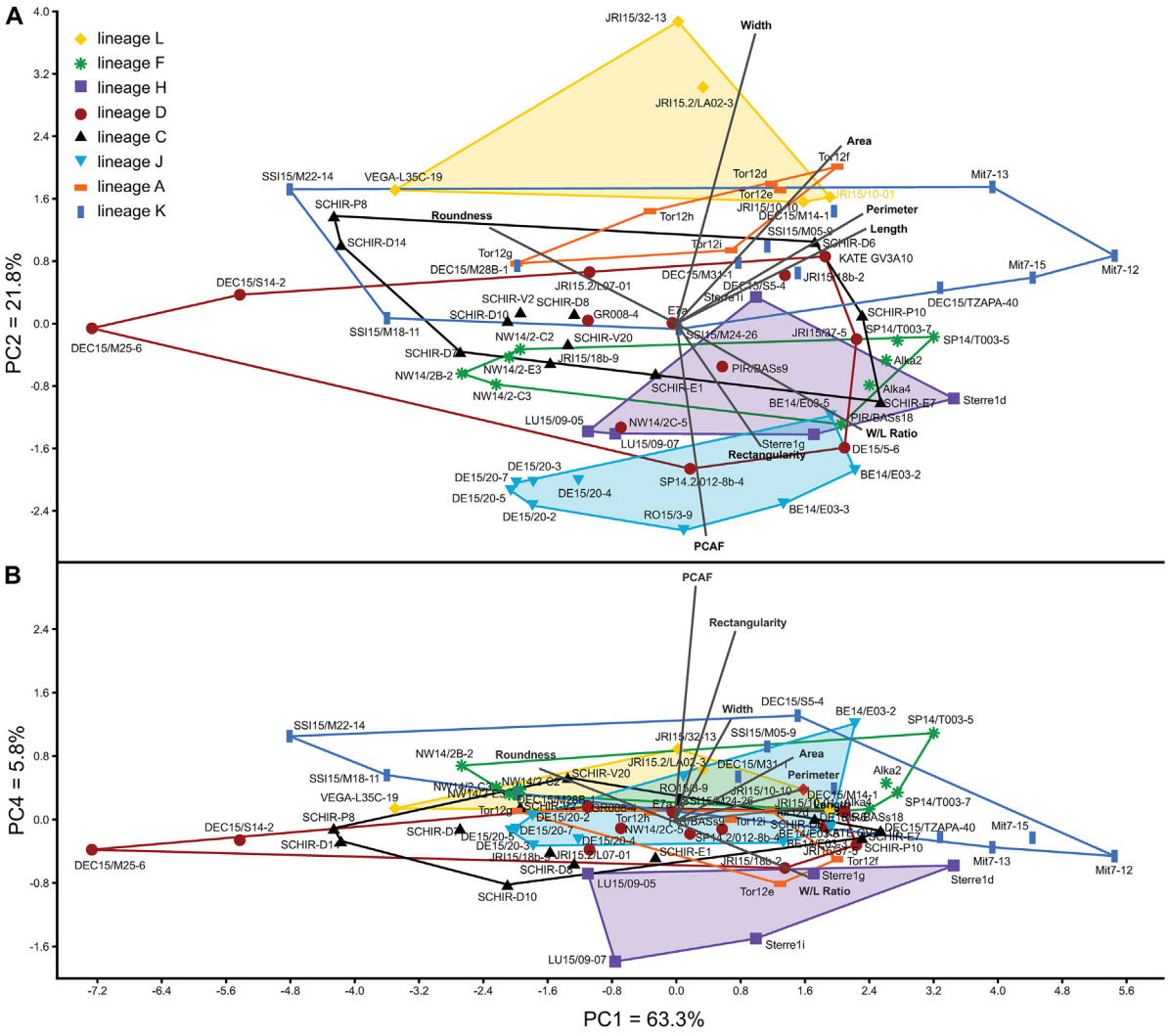

Figure 4. Principal component analysis (PCA) ordination biplots. The different symbols represent the different molecular lineages. Individual symbols represent centroids of strains. Coloured polygons form relatively separated clusters along PC1-PC2 axis (lineages L and J) or PC1-PC4 axis (lineage H). **A**) PC axis 1 and 2; **B**) PC axis 1 and 4.

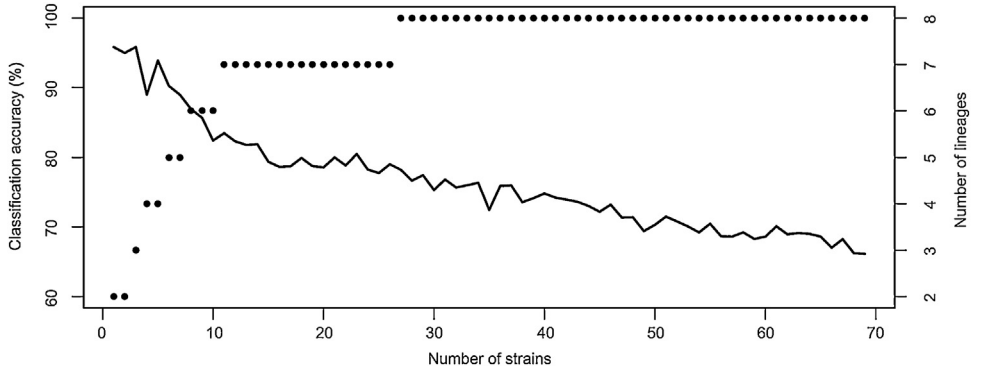

Figure 5. Plot visualising the effect of the dataset size on classification accuracy. The black line represents the mean classification accuracy in the dataset of randomly sampled strains. The black dots represent the mean number of lineages present in the dataset.

Table 1. Ranges and mean values of morphological variables measured in the eight lineages. Abbreviations: Var. = variable; Lin. = lineage; N = number of observations; Min-Max = minimum and maximum values; SD = standard deviation; PCAF = Percent Concave Area Fraction; AA-apex = width of the axial area near the apices; AA-CA = width of the axial area near the central area; AA-CA/W = ratio between the width of the axial area near the central area and the width of the valve.

Var.	Lin.	N	Min-Max	Mean ± SD	Var.	Lin.	N	Min-Max	Mean ± SD
Length (μm)	L	100	23.9–41.4	35.6 ± 4.7	Rectangularity	L	100	0.81–0.87	0.84 ± 0.01
	F	180	24.5–42.9	33.2 ± 6.0		F	180	0.83–0.90	0.86 ± 0.02
	H	100	25.8–42.1	33.5 ± 4.8		H	100	0.80–0.86	0.83 ± 0.01
	D	240	14.8–40.7	31.2 ± 6.4		D	240	0.80–0.89	0.84 ± 0.02
	C	240	22.8–41.8	30.4 ± 5.1		C	240	0.80–0.87	0.83 ± 0.01
	J	180	23.2–38.6	29.3 ± 4.6		J	180	0.83–0.89	0.86 ± 0.01
	A	120	27.0–45.9	35.6 ± 3.9		A	120	0.80–0.86	0.83 ± 0.01
	K	240	21.4–49.9	36.6 ± 7.6		K	240	0.81–0.88	0.85 ± 0.01
Width (μm)	L	100	8.0–11.7	10.1 ± 0.8	Compactness	L	100	0.50–0.64	0.55 ± 0.03
	F	180	7.0–9.8	8.3 ± 0.5		F	180	0.48–0.61	0.53 ± 0.04
	H	100	6.8–9.3	7.9 ± 0.6		H	100	0.46–0.55	0.50 ± 0.02
	D	240	6.2–10.7	8.2 ± 0.8		D	240	0.48–0.69	0.54 ± 0.05
	C	240	7.0–10.4	8.4 ± 0.6		C	240	0.47–0.64	0.55 ± 0.04
	J	180	6.3–9.0	7.2 ± 0.6		J	180	0.47–0.58	0.52 ± 0.02
	A	120	8.2–10.8	9.5 ± 0.6		A	120	0.49–0.60	0.53 ± 0.02
	K	240	7.6–10.7	9.2 ± 0.6		K	240	0.45–0.66	0.53 ± 0.05
Area (μm^2)	L	100	169–386	305 ± 56	Roundness	L	100	0.25–0.41	0.31 ± 0.04
	F	180	160–361	240 ± 57		F	180	0.23–0.37	0.28 ± 0.04
	H	100	150–304	222 ± 46		H	100	0.21–0.30	0.25 ± 0.02
	D	240	76–356	219 ± 59		D	240	0.23–0.48	0.29 ± 0.05
	C	240	156–340	215 ± 47		C	240	0.22–0.41	0.30 ± 0.04
	J	180	128–295	184 ± 44		J	180	0.22–0.33	0.27 ± 0.03
	A	120	186–398	284 ± 43		A	120	0.24–0.36	0.29 ± 0.02
	K	240	145–415	289 ± 72		K	240	0.20–0.44	0.28 ± 0.05
Perimeter (μm)	L	100	58–98	85 ±11	PCAF	L	100	0.54–1.28	0.78 ± 0.13
	F	180	59–103	78 ± 13		F	180	0.62–2.05	1.03 ± 0.21
	H	100	60–97	78 ± 11		H	100	0.65–1.50	1.05 ± 0.18
	D	240	37–94	73 ± 14		D	240	0.61–1.72	0.99 ± 0.20
	C	240	57–96	72 ± 11		C	240	0.64–1.97	0.99 ± 0.21
	J	180	55–90	68 ±10		J	180	0.69–2.19	1.17 ± 0.25
	A	120	65–107	84 ±9		A	120	0.60–1.80	0.90 ± 0.19
	K	240	53–113	86 ± 16		K	240	0.55–1.64	0.93 ± 0.21
W/L ratio	L	100	2.56–4.28	3.53 ± 0.43	Formfactor	L	100	0.45–0.66	0.53 ± 0.05
	F	180	2.82–4.83	3.99 ± 0.57		F	180	0.40–0.62	0.49 ± 0.05
	H	100	3.53–5.09	4.23 ± 0.40		H	100	0.39–0.54	0.46 ± 0.04
	D	240	2.14–4.85	3.77 ± 0.61		D	240	0.42–0.73	0.52 ± 0.07
	C	240	2.46–4.93	3.61 ± 0.52		C	240	0.42–0.65	0.53 ± 0.05
	J	180	3.25–4.91	4.03 ± 0.39		J	180	0.42–0.59	0.49 ± 0.04
	A	120	2.95–4.46	3.73 ± 0.29		A	120	0.44–0.60	0.51 ± 0.03
	K	240	2.41–5.49	3.96 ± 0.70		K	240	0.38–0.69	0.50 ± 0.07
AA-apex (μm)	L	200	0.83–2.73	1.65 ± 0.31	AA-CA/W ratio	L	200	0.14–0.30	0.21 ± 0.03
	F	360	0.70–2.13	1.17 ± 0.22		F	360	0.09–0.29	0.17 ± 0.03
	H	120	0.81–1.69	1.16 ± 0.16		H	120	0.13–0.28	0.20 ± 0.03
	D	120	0.88–2.04	1.39 ± 0.21		D	120	0.12–0.35	0.21 ± 0.04
	C	120	0.82–2.76	1.38 ± 0.33		C	120	0.10–0.33	0.18 ± 0.04
	J	120	0.91–2.20	1.47 ± 0.26		J	120	0.17–0.51	0.29 ± 0.06
	A	40	0.84–1.69	1.31 ± 0.18		A	40	0.11–0.20	0.16 ± 0.02
	K	120	1.11–2.49	1.86 ± 0.23		K	120	0.14–0.35	0.24 ± 0.04

Table 1 (*Continued*)

Var.	Lin.	N	Min-Max	Mean ± SD	Var.	Lin.	N	Min-Max	Mean ± SD
AA-CA (μm)	L	200	1.26–3.34	2.17 ± 0.36	Areola density	L	113	0.15–0.29	0.22 ± 0.03
	F	360	0.71–2.21	1.37 ± 0.29		F	158	0.15–0.31	0.22 ± 0.04
	H	120	1.06–2.13	1.52 ± 0.23		H	71	0.18–0.33	0.25 ± 0.04
	D	120	0.89–2.45	1.65 ± 0.28		D	37	0.18–0.28	0.22 ± 0.03
	C	120	0.82–2.73	1.47 ± 0.35		C	64	0.15–0.29	0.22 ± 0.03
	J	120	1.23–3.67	2.07 ± 0.46		J	55	0.24–0.31	0.28 ± 0.02
	A	40	1.05–1.79	1.46 ± 0.21		A	26	0.24–0.28	0.26 ± 0.01
	K	120	1.29–3.45	2.27 ± 0.35		K	68	0.19–0.36	0.28 ± 0.04

Since a visual inspection of the LM micrographs showed variations in size of the axial area, the width of the axial area near the central area and the apices was measured for the subset of thirty strains also examined in SEM (Supplementary Material Table S1). An overall significant lineage effect was found ($P < 0.001$) for the width of the axial area near both the central area and the apices. Although there were clearly significant differences between the lineages (Supplementary Material Table S5), there was overlap in the width of the axial area between all lineages (Table 1), indicative of incomplete differentiation. The classification accuracy of the XGBoost algorithm for the subset of thirty strains, analysed with and without axial area measurements, was 80.6% and 79.6%, respectively. The difference was not significant between the datasets. The classification algorithm found width and rectangularity to be important for both datasets (Supplementary Material Fig. S3). The classification accuracy of the algorithm decreased with the increasing size of the dataset (Fig. 5). When the entire dataset was used, the accuracy dropped to 67.1%.

Lineage F from Europe and the High Arctic possessed most intraspecific molecular variation (Supplementary Material Tables S3–S4), and was also one of the lineages which showed most variation in the morphometric dataset (Supplementary Material Fig. S2). Lineage D, which has the widest geographic distribution in this study, being retrieved from both hemispheres, was morphologically seen also one of the most diverse lineages (Supplementary Material Fig. S2). Nevertheless, no (clear) relation between geographic origin, molecular differentation and morphology could be found in this lineage, with strains of the same geographic area (e.g. Maritime Antarctica) being very variable in their morphologies (Fig. 4). Lineage K had a relatively restricted distribution in Antarctica, but nevertheless showed a relatively large variation in both the molecular (Fig. 3) and morphological dataset (Supplementary Material Fig. S2). On the other hand, lineage H, which also showed distinct molecular variation (Supplementary Material Table S3), was one of the two least variable lineages in the morphometric dataset (Supplementary Material Fig. S2). Nevertheless, even lineages with low molecular differentiation, e.g., lineages C, J and L, showed intermediate morphological differentiation (Supplementary Material Fig. S2).

A strong sample effect (geographic origin) and a small effect of lineage identity on the morphology of the strains was found. Strain identity explained 11.5% of the variability in the morphometric data. The joint effect of strain, sample and/or lineage identity explained 74.5% of the variability: 52.0% of the variability was explained by covariation in strain and sample, 0.8% by covariation in strain and lineage identity, and 21.7% by covariation in strain, sample origin and lineage identity. Unexplained variability in the morphometric data (14.0%) was mostly associated with intrastrain variability that was relatively high in about one-third of the strains (Supplementary Material Fig. S4). Increased Euclidean distances between individuals of these strains were not correlated with the life cycle (valve length) as almost an equal number of strains had smaller or larger valves than the average length of all valves (Supplementary Material Fig. S1).

Landmark-based geometric morphometrics on a selection of 15 strains from two lineages could not reveal any inter-lineage differentiation (data not shown), confirming the SHERPA results. A PROTEST test to assess the differences between ordinations based on SHERPA and landmark-based geometric morphometrics revealed a highly significant concordance between both datasets for the first four PC aces ($P < 0.001$, $m_{12} = 0.43$, correlation in a symmetric Procrustes rotation = 0.75).

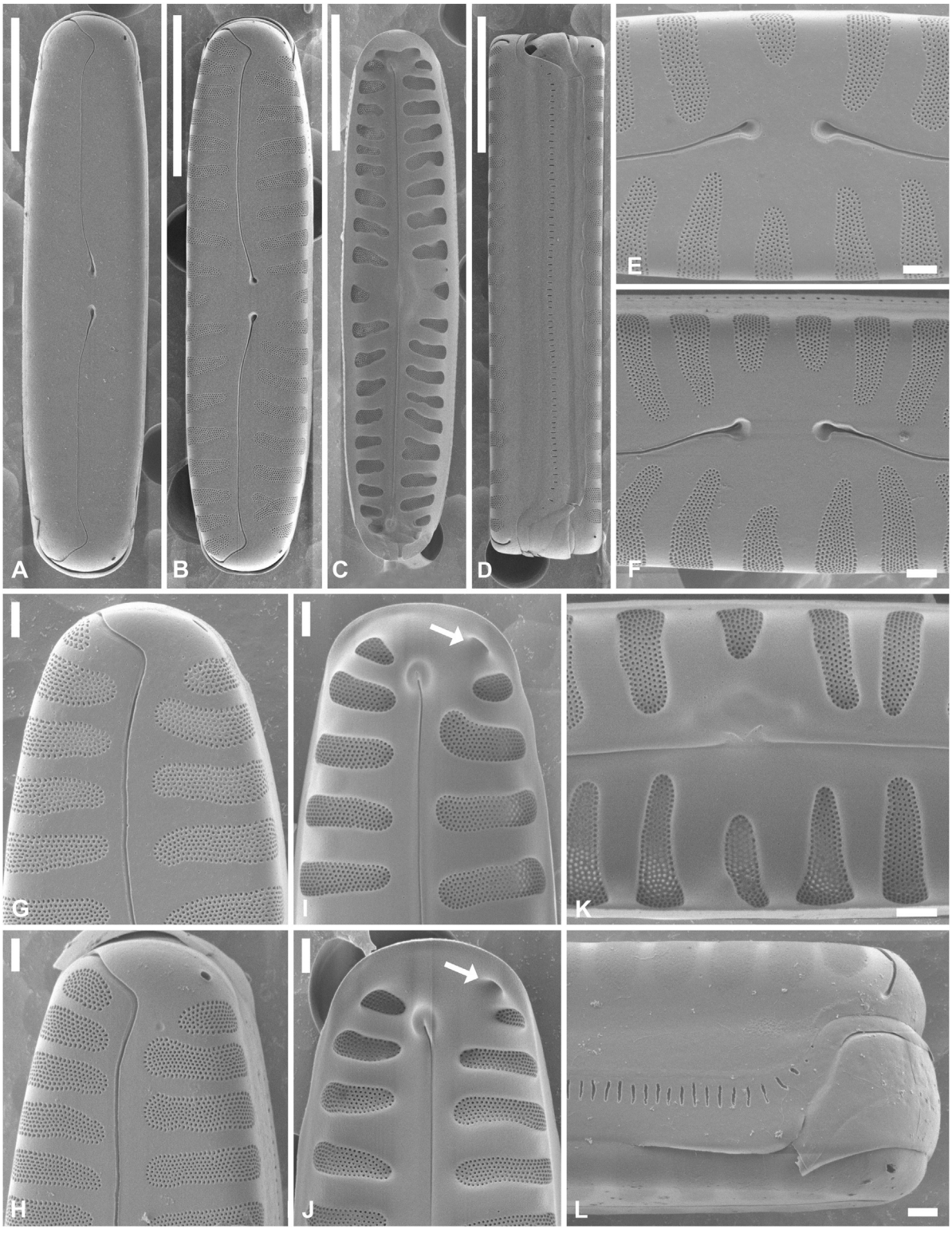

Figure 6. SEM micrographs of oxidised material of strains belonging to lineage F. Strains: **A, L**) SP15/T003-05; **B, D, J**) PIR/BASs18; **C**) Alka4; **E**) NW15/2-B2; **F, H**) SP14/T003-07; **G**) NW14/2-C3; **I**) NW14/2-E3; **K**) NW14/2-C2; **A-B**) external valve view; **C**) internal valve view, including valvocopula; **D**) side view of an entire valve; **E-F**) external details of the central area; **G-H**) external details of the apices; **I-J**) internal details of the

Ultrastructural Analysis

All strains examined in SEM showed the same ultrastructure (Fig. 6, Supplementary Material Figs. S5–S11). The striae were composed of one large alveolus. Normally, the areolae are covered by silicified hymenes (Fig. 6A, L). Due to the here used oxidation method, the hymenes are usually destroyed, allowing the observation of the underlying structure of the alveoli (e.g. Fig. 6B). Each alveolus usually contained 7–13 rows of small regularly oriented areolae (e.g. Fig. 6E–F). The shape of the areolae varied between round to irregular (see for example Fig. 6G–H), with both types occurring in all lineages. The alveoli continued onto the mantle edge, ending in a round or pointy outline (e.g. Fig. 6D, L) with both types present within the same valves and all lineages. Internally, the alveoli were open (e.g. Fig. 6C, I, J, K). The external terminal raphe fissures were clearly curved towards the secondary side of the valve, sickle shaped and continuing onto the mantle as a straight line downward, terminating shortly before the mantle edge (e.g. Fig. 6A, B, G, H). The internal terminal raphe endings terminated onto small, eccentrically situated, weakly raised helictoglossae (e.g. Fig. 6C, I, J). The external central raphe fissures were spatulate, expanded and unilaterally deflected (e.g. Fig. 6E, F). The internal central raphe fissures were straight with short, bent, central raphe endings which terminated on a well-developed unilaterally inflated central nodule, internally interrupting the raphe branch (intermissio) (e.g. Fig. 6C, K). Externally, small apical pores with a round to irregular formed shape were visible near the primary side of the apices, opposite of the terminal raphe fissures. Both round and irregular forms were observed in all lineages (e.g. Fig. 6A, B, G, H, L). In lineage H, areolae were sometimes observed adjacent to the apical pore (Supplementary Material Fig. S8G), although the majority of the apical pores were simple holes (Supplementary Material Fig. S8C, F, K). Internally, the apical pores are visible as simple round openings (e.g. Fig. 6I, J, arrows). The girdle was composed of three open girlde bands (e.g. Fig. 6D, L). All girlde bands bore a row of transapically elongated, slitlike areolae (e.g. Fig. 6D, L). On the advalvar part of the valvocopula, a series of inwardly pointing projections was present, corresponding to the internal virgae (e.g. Fig. 6C, Supplementary Material Figs. S8B, S11H).

Inter-lineage difference in areola density was tested using a linear mixed effect model, indicating that the different lineages do not differ significantly for this parameter ($P = 0.054$). Nevertheless, the data show an indication for the existence of two morphogroups: lineages C, D, F and L showed overall lower areola densities, compared to lineages A, H, J and K (Table 1). Differences in areola density did not follow a phylogenetic signal: both groups are not monophyletic. Nevertheless, the minimum–maximum range of all lineages (partially) overlapped, indicating incomplete differentiation.

Cytological Analysis

BF and epifluorescence (FM) microscopy on a subset of strains belonging to six molecular lineages here examined revealed that these lineages had the same general chloroplast arrangement and morphology (Supplementary Material Fig. S12). All examined cells showed two linear strip-like parallel plastids located on each side of the apical axis. Although usually not visible, each plastid most likely possessed one centrally located cushion-like pyrenoid (Supplementary Material Fig. S12). Examination of DAPI-stained cultures in FM showed that all examined lineages have the same general nucleus morphology, showing a round to oval shape (Supplementary Material Fig. S13).

Discussion

Our comprehensive morphometric and ultrastructural data of eight phylogenetically distinct lineages in the *Pinnularia borealis* species complex could not reveal conclusive morphological differentiation, although subtle morphological differences between most lineages existed. It is clear that members of the *P. borealis* species complex share a broad series of morphological characteristics, i.e. a general 'Bauplan'. Valves are heavily silicified with large striae that are clearly visible at low magnifications. The striae are broad and composed of one alveolus. The areolae are covered by hymenes and have a round to irregular shape. All valves show apical pores at the primary side of the valve and share the same raphe structure, girdle band morphology as well as plastid morphology and arrangement, and nucleus morphology. In the here examined dataset, morphological variations

apices with indication of the internal opening of the apical pore (arrows); **K**) internal detail of the central area; **L**) side view with detail of the apex. Scale bar = 10 μm in figures A-D, and 1 μm in figures E–L.

between and within lineages are confined to valve outline and size, the width of the axial area and the areola density. Although for the majority of lineage pairs it is possible to find morphological features that enable to separate the lineages statistically, morphological traits strongly overlapped for most lineages due to large variation within all lineages. Moreover, the more strains are added to the classification model, the less clear the morphological signal between lineages became. None of the lineages thus possessed a unique morphological character or character combination. As a result, it proved impossible to reliably assign any given valve to a molecular lineage. Altogether, this indicates that, despite the existence of some differentiation, the *P. borealis* lineages here studied are truly indistinguishable, i.e. cryptic. Although several studies initially reported cryptic diversity in diatoms (Degerlund et al. 2012; Kooistra et al. 2010), detailed morphological, morphometric and/or ultrastructural examinations often revealed subtle but clearly distinct morphological differentiation among species, i.e. pseudocryptic diversity (Amato et al. 2007; Chamnansinp et al. 2013; Lundholm et al. 2012; Vanormelingen et al. 2008), although not always (Beszteri et al. 2005b; Kooistra et al. 2008).

The observation of large morphological variation within a single lineage or strain could be the result of structural inheritance, which typically occurs when slight morphological deviations are passed on to the next generations, increasing shape variation in a (monoclonal) population when average cell size decreases (Woodard et al. 2016). Nevertheless, some of the strains with a cell length above average, and thus less further down the cell size-reduction trajectory, showed some of the largest variations in valve morphology. This effect was independent of lineage identity. Perhaps, some strains have been influenced more by the culture conditions than others, increasing the accumulation of small morphological deviations. Since natural populations were not studied, it is difficult to assess the potential impact of the culture conditions on valve morphology. Studying natural valves of *P. borealis* is largely hampered by the fact that multiple, potentially morphologically indistinguishable lineages can occur within a single locality. Nevertheless, malformed valves were only very rarely observed in the cultures here examined. Furthermore, even after 13 months of culturing, which was distinctly longer than the cultures used in this study, the closely related species *P. catenaborealis* still showed identical valve morphologies in cultures and natural samples (Pinseel et al. 2017a). This suggests that overall valve morphology, independent from life-cycle associated changes, is relatively stable in cultured *P. borealis*. It is not unlikely that morphological variation within a monoclonal line also happens in the natural environment. Interestingly, many taxa in the *P. borealis* species complex have (almost) exclusively been described based on subtle, often overlapping, differences in valve shape and valve dimensions. This includes the commonly cited varieties *scalaris, sublinearis, islandica* and *subislandica* (Krammer 2000). Here, the taxonomic value of these characteristics was assessed, making it clear that both valve outline and size do not discriminate confidently between phylogenetic lineages, although they represent the main characters which are variable among lineages. This indicates that at least some of the varieties described in literature do not correspond to biological meaningful units (= taxa), but rather represent interspecies morphotypes. This confirms the phylogenetic analysis of Souffreau et al. (2013b) who showed that different morphological forms cluster together in the same lineages, whereas the same morphotypes can be found in multiple lineages.

Despite being the oldest known species complex to date (Souffreau et al. 2013b), the morphology of the *P. borealis* lineages here examined is remarkably stable and one can ask why this is the case. Morphological discrimination can be hampered by several drivers. First, the "low-morphology" problem implies that there might simply not exist sufficient morphological characteristics to distinguish between all members in a species complex, meaning that the amount of genetic variation exceeds the maximum amount of potential morphological variation (Verbruggen 2014). This issue is especially eminent in morphologically simple species complexes which have only a few discriminating features (Verbruggen et al. 2009). Indeed, *P. borealis* is structurally not a very complex taxon compared to many other diatoms. Second, our observations could result from stabilizing selection. Mutations constantly introduce new variation in natural populations (Drake et al. 1998; Lynch 2010). When considering macroevolutionary timescales, natural selection and genetic drift acting on these mutations (Coyne and Orr 2004) as well as mutations acting as direct drivers for phenotypic change (Nei 2007), have a great potential to change morphological characters (e.g. Hoekstra et al. 2006; Wray 2007). However, in presence of stabilizing selection, the evolution of novel morphologies which are located outside of an adaptive optimum is prevented. Stabilizing selection has been suggested to be the dominant mode of evolution (Gould and Eldredge 1993; Hansen 1997). If stabilizing

selection plays a role in *P. borealis,* it raises the question why the adaptive optimum itself is conserved (Hansen and Houle 2004). Perhaps, the generally (semi)terrestrial life style of *P. borealis* and the potentially associated morphological adaptions, although not understood, could be key in this sense. The subtle inter- and intraspecific morphological variations can be interpreted as variations within the boundaries of an adaptive optimum. Third, it can be argued that instead of morphological stasis, the examined *P. borealis* lineages have independently evolved highly similar valve structures by parallel/convergent evolution, potentially to cope with the same environment. To test this hypothesis, ancestral state reconstructions should be carried out on phylogenies covering the majority of the global molecular and morphological diversity in this complex. Based on the phylogeny presented here, a case of parallel/convergent evolution of the overall *P. borealis* Bauplan seems unlikely since all lineages, including *P. catenaborealis*, share the same Bauplan. In this sense it is interesting to note that the valve morphology of several strains is more similar to strains of other (distantly related) lineages than to conspecific strains. Environmentally induced phenotypic plasticity and/or structural inheritance could explain this effect. Alternatively, it could also suggest that within a general framework of morphological stasis, parallel/convergent evolution, either adaptive or neutral, might be acting to generate similar morphological variations in distantly related lineages. Finally, past hybridisation and incomplete lineage sorting should be mentioned as potential causes for the introduction of similar morphotypes in distant lineages for the sake of completeness, although there is currently no indication for this phenomenon.

Our study showed that when examining new strains belonging to the *P. borealis* species complex, additional molecular lineages are likely to be found, indicating that much larger sampling efforts will be needed to reveal the true lineage diversity in this group. Interestingly, four lineages originally identified by Souffreau et al. (2013b) were now also recovered from other geographic regions. These lineages were mostly found in other regions within the same geographic area (Europe, Antarctica), resulting in sequence clusters unique for the northern or southern hemisphere and pointing towards restricted geographic distributions of species in this complex. However, one previously considered European lineage was now also obtained from regions as far apart as the High Arctic and Maritime Antarctica. Although these data indicate that at least one *P. borealis* lineage might have a, close to, cosmopolitan distribution, the Maritime Antarctic strains formed a, although statistically poorly supported, monophyletic subgroup relative to the northern hemisphere strains, indicative of intra-lineage phylogeographic structuring. This could point to ongoing speciation events or even the presence of distinct species. In general, intra-lineage genetic variation was common, having been detected in one or both genes in all examined lineages. In all but one lineage, maximum intra-lineage *rbc*L variation equalled 1 bp. This is identical to the genetic divergence detected between strains of *P. catenaborealis* (Pinseel et al. 2017a). Although maximum sequence divergence in *rbc*L was slightly higher for lineage F (3 bp), even higher intraspecific differences in *rbc*L (6 bp) have been reported in *Sellaphora* (Vanormelingen et al. 2013). For D1–D3 28S, genetic divergence reached a maximum of 8 bp. Similar levels of D1–D3 28S divergence have been suggested to correspond to species boundaries in *Achnanthidium* (Pinseel et al. 2017b), whereas two *Pseudo-nitzschia* species did not at all differ in their D1–D3 28S sequences (Amato et al. 2007). In contrast, Beszteri et al. (2005b) found intraspecific differences in D1–D2 28S up to 3 bp in the *Cyclotella meneghiniana* complex and 5 bp in *C. scaldensis*. Overall D1–D3 28S divergence in *Pinnularia* is relatively high, exceeding the genetic divergence reported in *rbc*L and the mitochondrial *cox*1 (Kollár et al. 2018). This suggests that intraspecific divergence in D1–D3 28S is likely to be relatively high as well. However, we cannot entirely exclude the possibility of the existence of multiple species (or ongoing speciation) within some of the lineages here examined. The internal transcribed spacer (ITS2) would be an interesting candidate gene to shed light on the genetic divergence within *P. borealis* 28S-lineages, as it has been successfully applied to reveal closely related species in for example *Pseudo-nitzschia* (Amato et al. 2007), *Navicula cryptocephala* (Poulíčková et al. 2010) and *Eunotia bilunaris* (Vanormelingen et al. 2008). However, the presence of extensive intragenomic variation prevents routine PCR amplification of ITS in *P. borealis*. Nevertheless, using the overall interspecific differentiation of different genes among various diatom genera as a general proxy for the minimal sequence divergence required to detect species level differentiation in diatoms in general is optimistic at the very least. Given the tremendous diversity of the diatoms, combined with their relatively long evolutionary history, it is to be expected that shifts in diversification rate between the genomes and individual genes of different gen-

era and species are common (see also Nakov et al. 2018). Genes that are thus highly useful for species level detection in one genus, may be too conserved in another. For example, in *Seminavis robusta*, different mating groups with differentiation in *rbc*L were not detected by D1–D3 28S nor ITS1-5.8S-ITS2 (De Decker et al. 2018), whereas ITS2 was selected as the ideal marker for *Pseudo-nitzschia* (Amato et al. 2007; Casteleyn et al. 2009) and D1–D4 28S was successfully used in *Chaetoceros* (Degerlund et al. 2012). Altogether, the real question here is not whether we correctly identified all potential lineages in our dataset, but rather how the presence of additional undetected species-level diversity within these lineages could have affected our conclusions on the morphological analysis. From Fig. 4 it is evident that even if some of the lineages with high sequence variability would fall apart in distinct species when more evidence would become available, their morphospace would still largely overlap with other lineages within the *P. borealis* species complex.

At last, one can argue that if *P. catenaborealis* merits species level, the other lineages delineated in this study should deserve the same fate. Although species descriptions would be a logic consequence of our work, we refrain here from doing so. A taxonomic revision of a (pseudo)cryptic species complex of this scale should ideally be done with as much information as possible. That is by capturing most of the (molecular) diversity on a global scale, together with comprehensive information regarding species distributions, ecology, etc. Ideally this also includes examinations of type material covering the various morphotypes and species already described in the literature as to provide a link between past and present work, whenever possible. Given the fact that (i) increasing the sampling with a small number of strains already resulted in the findings of three additional lineages and, (ii) our sampling can be considered far from complete since many regions and continents were not covered, it is highly likely that further sampling will reveal substantial additional phylogenetic diversity at the inter- and intraspecific level, increasing phylogenetic complexity. A similar situation occurred in the *Skeletonema costatum* complex where increased sampling following species descriptions rendered ambiguous interpretations of morphology and species boundaries (Ellegaard et al. 2008; Kooistra et al. 2008). The current molecular information of *P. borealis* is limited to two genes, and due to the still relatively low intra-lineage taxon sampling, the potential prevalence of paraphyletic species in *P. borealis* is difficult to assess (Alverson 2008). Altogether, describing these *P. borealis* lineages as species at current state of knowledge risks to increase the confusion surrounding this species complex, rather than solve any taxonomic problems. In *P. catenaborealis*, the situation was different as this species could be clearly defined based on multiple lines of evidence, i.e. genes, morphology and ecology.

In conclusion, we have shown that detailed morphological examinations in LM and SEM could not reveal conclusive differentiation among several closely and distantly related lineages in the *P. borealis* species complex. It is thus recommended that morphological features are carefully assessed for the presence of a phylogenetic signal before they are used to describe new species in this complex. The interpretation of morphometric data in the absence of secondary information, such as molecular phylogenies, ecophysiology or mating experiments, should be done with care as the absence of a morphological signal might not be sufficient evidence for the absence of multiple taxa. Furthermore, when analysing small datasets of geographically distant populations, clear morphological clustering might not necessarily indicate the presence of multiple species, but could also point towards the presence of intraspecific morphological variation that is not interpreted as such due to the absence of intermediate valve morphologies. Further work should focus on increasing the lineage sampling in this complex to (i) get a better idea on the molecular evolution and total diversity, and the potential influence of hybridisation and incomplete lineage sorting, (ii) test hypotheses on ancestral morphology, parallel/convergent evolution and morphological stasis, and (iii) build a comprehensive dataset allowing for a complex-wide taxonomic revision. Although conclusive morphological differentiation could not be detected, screenings of (eco)physiology and the life cycle have the potential to gain a deeper understanding of the nature of the species boundaries in *P. borealis*.

Methods

Taxon sampling: Environmental samples from soils (top layer, upper 2 cm), mosses and epilithon from the littoral zones of lakes and ponds were collected between 2011 and 2015 from various locations worldwide and transferred to falcon tubes or sampling bags (Fig. 1, Supplementary Material Table S1). All samples were kept dark, and if possible, cool (< 10 °C) during transport. Upon arrival in the laboratory, small quantities of the natural material were incubated for several weeks in WC medium (Guillard and Lorenzen 1972) without pH adjustment or vitamin addition and with double silica concentrations, at 4 °C, 5–10 μmol photons m^{-2} s^{-1} and a 12:12 h (light:dark)

cycle. *Pinnularia borealis* cells were isolated from the natural samples under an Olympus SZX9 stereomicroscope (Olympus, Tokyo, Japan) using a needle and a micropipette and grown in WC medium at standard culture conditions of 18 °C, 5–10 μmol photons m^{-2} s^{-1} and a 12:12 h (light:dark) cycle. The cultures were reinoculated when reaching late exponential phase. *Pinnularia borealis* is a slow grower and divides about once every 3–10 days when cultured at the above mentioned culture conditions. In order to obtain sufficient cell biomass, cells were grown for 4–6 months prior to subsampling for DNA and morphology. Although this is a relatively long time period, the slow division time of *P. borealis* ensured that at the moment of cell harvesting, the number of divisions as well as the associated cell size reduction was relatively small.

Molecular phylogenetic analysis: Subcultures (2 mL) for molecular characterization were harvested in late exponential phase, centrifuged and frozen at −80 °C prior to DNA extraction. DNA extraction was performed following a bead-beating method with phenol extraction and ethanol precipitation (Zwart et al. 1998). All extracted DNA is stored at −80 °C and available upon request. The D1–D3 region of the nuclear-encoded large subunit LSU rDNA (28S) and the plastid gene *rbc*L were amplified by polymerase chain reaction (PCR) using the universal primers DIR-f and T24U (Hamsher et al. 2011; Scholin et al. 1994) and DPrbcL1-F and DPrbcL7-R (Jones et al. 2005), respectively. PCR reaction mixtures contained per sample 1 μL of the template DNA, 0.2 μM of each primer, 200 μM of each deoxynucleoside triphosphate, 0.4 μg μL-1 of bovine serum albumin, 2.5 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, California USA) and 1.25U of Taq polymerase (AmpliTaq, Perkin–Elmer, Wellesley, Massachusetts USA). The mixtures were adjusted to a final volume of 25 μL with high performance liquid chromatography water (Sigma, St. Louis, Missouri USA). For 28S, the PCR conditions were as follows: 35 cycles (1 min at 94 °C, 1 min at 55 °C and 1 min at 74 °C) with an initial denaturing step of 5 min at 95 °C and a final step of 10 min at 72 °C. For *rbc*L, the following PCR conditions were used: 40 cycles (1 min at 94 °C, 1 min at 55 °C and 1.5 min at 72 °C) with an initial denaturing step of 3 min at 94 °C and a final step of 5 min at 72 °C. The PCR products were sequenced with their respective PCR primers and the additional primers D2C-r and T16N (Scholin et al. 1994; Hamsher et al. 2011) for 28S and RbcL13-F and RbcL-17R (Daugbjerg and Andersen 1997; Jones et al. 2005) for *rbc*L. PCR products were sent for sequencing to MACROGEN, Inc. The obtained chromatograms were individually edited using BioNumerics v3.5 (Applied Maths, Kortrijk, Belgium). All sequences are available on GenBank and the Barcode of Life Database (Supplementary Material Table S1) (Ratnasingham and Hebert 2007). The latter also includes environmental information (sampling location, GPS coordinates, material sampled, and sampling information).

Outgroup sequences for the phylogeny were based on Souffreau et al. (2013b) and included several *Pinnularia* taxa. All previously published sequences of *P. borealis* (Pinseel et al. 2017a; Souffreau et al. 2013b) were included in the analysis (see Supplementary Material Table S2 for BOLD and GenBank accession numbers). All sequences were automatically aligned using the CLUSTAL-W algorithm (Thompson et al. 1994) as implemented in BioEdit v7.2.5 (Hall 1999) and the ends were trimmed to ensure at least 70% site-coverage. All *rbc*L sequences aligned unambiguously without any gaps. Due to the presence of highly variable regions in the 28S alignment, the latter was manually corrected allowing all variable regions to be included in the final alignment (doi:10.17632/jsw2x5k45m.1). All phylogenetic analyses were performed on the concatenated dataset of both genes. For the phylogeny reconstructions, 1020 and 1395 positions were used for 28S and *rbc*L respectively. The best-fit models for nucleotide substitution were assessed simultaneously with the most optimal partition scheme in PartitionFinder v1.1.0 (Lanfaer et al. 2012) using the BIC selection criterion and a greedy search scheme. The best possible partition scheme and substitution model was then chosen for each downstream phylogenetic analysis, depending on the available models in the given programs for phylogenetic inference. Maximum likelihood analysis (ML) was performed in RAxML (Stamatakis et al. 2008) as implemented in the CIPRES Science Gateway v3.3 (Miller et al. 2010) with 10 independent runs and 1,000 pseudoreplicates. The analysis was divided into four partitions to allow for different parameter estimates: 28S was analysed as a single partition, whereas *rbc*L was partitioned by codon. All partitions were analysed using a GTR + I + G model. Bayesian phylogenetic inference (BI) was performed in MrBayes v.3.2.6 (Ronquist et al. 2012). 28S was analysed as a single partition under the GTR + I + G model whereas *rbc*L was partitioned by codon (F81 + I model for codon 1, JC + I model for codon 2 and GTR + G model for codon 3). Two runs of four (three heated and one cold) Metropolis-coupled Monte–Carlo Markov Chains were completed on all datasets for 8 million generations and sampled every 1000th generation. Convergence and stationarity of the log-likelihood and parameter values were assessed using Tracer v.1.6 (Rambaut et al. 2014). The first 25% of the generations were discarded as burn-in after which the sumt command in MrBayes was used to summarize the post-burn-in trees and calculate the posterior-probabilities (PPs). Pairwise distance calculations, i.e. number of base pair (bp) differences and p-distance, between the sequences were calculated in MEGA 7.0 (Kumar et al. 2016) taking into account both transitions and transversions and with pairwise deletion for treatment of gaps and missing data.

Morphological analysis (BF): Subcultures (3 mL) for morphological analysis of the newly obtained strains were harvested in late exponential phase and cleaned by adding 65% nitric acid and heating to 60 °C for 7 days. All samples were subsequently washed eight times with distilled water and mounted in Naphrax®. Oxidised culture material and microscope slides are available upon request.

Twenty valves from each strain were examined in bright field (BF) at ×1,000 magnification under oil immersion using a Zeiss Axioplan 2 light microscope (Carl Zeiss, Jena, Germany) equipped with an AxioCam Mrm camera and an Olympus BX-51 microscope (Olympus, Tokyo, Japan) equipped with an Olympus DP72 camera. All pictures are available upon request. Microphotographs were processed in Adobe Photoshop CS4 (Adobe Systems, San Jose, CA, USA) as follows: the colour mode of the microphotographs was changed into greyscale, the scale bar of the pictures of both microscopes was unified, and all objects were removed using the lasso tool with exception of a single complete *P. borealis* valve per image (Fig. 2). The images were loaded into the image analysis software SHERPA v1.0 (Kloster et al. 2016) after which the success of retrieving the contour data was checked. SHERPA outcomes included shape measures (e.g., length, width, area, width/length ratio), heuristic descriptors (e.g., rectangularity, compactness, formfactor), Elliptic Fourier Descriptors, and convexity measures (e.g., CFD – Convexity Defection Factor, PCAF – Percent Concave Area Fraction). Ten univariate characteristics, which were variable in their min-max range and standard deviation within the dataset, were selected: width, length, area, perimeter, width/length ratio, rectangularity, compactness, roundness, formfactor, and PCAF. Correlations between selected parameters are shown in Supplementary Material Table S7. For details about the variables,

see the documentation of SHERPA (Kloster et al. 2014, 2016). Stria density, a common parameter used to distinguish diatom species, is not measured by SHERPA. We chose not to include this parameter in the analysis since the measurements by Souffreau et al. (2013b) did not show any notable differences in stria density among several lineages of *P. borealis*, nor did our preliminary measurements of the newly added strains show any such indication.

Principal component analysis (PCA) was performed on the standardized ten parameters using PAST v3.4 (Hammer et al. 2001). Strain positions (represented by centroids of clusters) were visualized along the first five PC axes. Variance in morphometric data was partitioned into six components: variance uniquely described by one factor (strain, lineage, or sample), variance jointly described by two or three factors (covariation), and unexplained variability. Strain name and lineage identity were coded as qualitative variables while the factor 'sample' included both sample identity and geographic distances among samples (calculation according to Kulichová and Fialová 2016). The Euclidean distances of standardised morphometric values were used to create a distance matrix of the dependent variable. Variation partitioning was calculated by permutational multivariate analysis of variance (Per-MANOVA) using the function *adonis* (vegan package, Oksanen et al. 2011) in R v3.4.3. The significance of Per-MANOVA was evaluated by randomization tests based on 999 permutations. Analysis of similarity (ANOSIM) using Euclidean distances in PAST v.3.4 was used to visualise the variability within lineages or strains and to test if valves within groups are more similar in morphology than valves from different groups.

The width of the axial area was measured for the subset of thirty strains also examined in SEM (Supplementary Material Table S1). The axial area was measured on two different localities near the central area and the apices, resulting in four observations per valve. To correct for the effect of valve width, the ratio between axial area width and total valve width was measured as well. For the selected strains, all twenty LM micrographs were examined. In total, 600 valves were measured. The resulting data (original and corrected data) were analysed using a linear mixed effect model (function *nlme* of the nlme package; Pinheiro et al. 2018) in R. To correct for variation specific to the strains and the individual valves and thus not the factor of interest, i.e. 'lineage identity', 'strain identity' and 'valve individual' were included as random effects in the model, with 'valve individual' nested within 'strain identity' which was on its turn nested into 'lineage identity'. The axial area widths of the central area and the apices were analysed separately. Model significance was tested using ANOVA. For each possible effect combination in the final model, post hoc testing was performed using the *glht*() function (multcomp package, Hothorn et al. 2008) which allowed obtaining P-values corrected for multiple testing. The same statistical model with post hoc testing was used for testing the inter-lineage differences in width, rectangularity, and PCAF (using the original dataset of 70 strains) as the variability of these characters was associated with inter-lineage variability (PCA and XGBoost analysis). Because only one measurement per individual was obtained for these parameters, 'valve individual' was not included in the model. For PCAF, the data showed large deviations from normality as well as an indication for heteroscedasticity. Therefore, different normalization methods using the function *bestNormalize* (bestNormalize package, Peterson 2017) were tested. Subsequently, a Yeo–Johnson transformation of the data was performed using the function *yeo.johnson* (VGAM package, Yee 2015).

To estimate the success of discrimination among all eight lineages we used a gradient boosting algorithm XGBoost (Chen and Guestrin 2016, R package xgboost). The dataset was split 70/30 into a training/test set. The probabilistic classification (objective = *multi:softprob*) with multiclass *logloss* evaluation metric (eval_metric = mlogloss) was used for a 5-fold cross-validation with maximum number of iterations (nruns) set to 50. Other parameters were left as default. Training and evaluation on the test set was repeated 10 times for each of the 3 datasets: all 70 strains (1), the subset of 30 strains with (2) and without (3) axial area measurements. For each, mean classification accuracy and mean feature gain were calculated. The feature gain describes the contribution of features to the improvement of the classification accuracy. The classification accuracy for the dataset with and without axial area parameters was compared using t-test in R.

In the morphometric analyses, datasets of different size were used (all 70 strains, 15 strains in landmark-based GM, 30 strains in axial area measurements) and this changed the success of differentiation between lineages. Therefore, the effect of the dataset size on the classification accuracy was assessed. Strains were randomly sampled from the entire dataset for training of the classification algorithm (setting are described above). Subsequently, the changes in its classification accuracy were tracked together with the increase of the species number. The random sampling of the strains was repeated 10 times.

SHERPA is a relatively novel development in the field of morphometrics. In order to assess the impact of our choice for SHERPA, landmark-based geometric morphometrics (Bookstein 1997), which has been widely used for morphometrics in diatoms (Beszteri et al. 2005a; Poulíčková et al. 2010; Van de Vijver et al. 2013; Veselá et al. 2009), was performed on a subset of 15 strains (10 cells/strain) representing lineages D (strains DE15/05-06, E7a, GR008-04, NW14/2-C5, PIR/BASs9 and SP14.2/012-08b-04) and F (strains Alka2, Alka4, PIR/BASs18, NW14/2-C2, NW14/2-C3, NW14/2-E3, NW14/2-B2, SP14/T003-05 and SP14/T003-07), excluding the data of the axial area. Outlines of valves for the landmark analysis were obtained from the SHERPA outputs. All images were rotated so that the apical axis of each valve was in vertical position. Landmark-based morphometrics were performed using tpsUtil v.1.6. (Rohlf 2015). Fifty-six equally distanced landmarks were positioned along the valve outline. One fixed landmark was placed on the top apex of the valve, whereas all other landmarks were allowed to slide along the valve outline. Shape coordinates were computed by Generalised Procrustes Analysis of landmarks (TpsRelw v.1.46). The scores from the relative warp analysis (RWA) were used in PCA analysis. The multivariate datasets (scores at first four PCA axis) obtained by both SHERPA and landmark analysis were compared using the on Procrustes statistics based PROTEST test (Peres-Neto and Jackson 2001) in R using the function *protest* (vegan package). Prior to this analysis, the SHERPA dataset was reduced so that only the cells used for the landmark analysis were taken into account.

Morphological analysis (SEM): For the ultrastructural analysis in SEM, five drops of the oxidised material were filtered through a 5-μm Isopore^{TM} polycarbonate membrane filter (Merck Millipore, Darmstadt, Germany). The stubs were sputter-coated with a gold-palladium layer of approximately 10 nm and studied in a JEOL-JSM-7100F (Tokyo, Japan). A subset of thirty representative strains was chosen for the ultrastructural analysis (Supplementary Table S1). All strains of the genetically most diverse lineage (F) as well as one of the least diverse lineages (L) were chosen for an initial screening of ultrastructural features. This allowed assessing which features, if any, showed minimal and maximal differentiation between and within lineages. For all other lineages, three strains/lineage

were examined, except for lineage A for which oxidised material was only available for one strain (Supplementary Material Table S1). Strain selection was based on molecular divergence as well as geographic origin: strains with maximal molecular divergence in 28S and/or *rbc*L and originating from different samples were preferred. Over 1000 SEM micrographs were taken and examined. None of the specimens and micrographs have been tilted.

The density of the areolae per alveolus was measured using the Analyze Particle function in FIJI (Schindelin et al. 2012). The areola density was measured on detailed SEM pictures of the central area (see for example Fig. 6E-F). All completely visible alveoli in a picture were measured. For these measurements, all strains for which SEM micrographs were obtained were analysed, except for strain SP14/T003-05 in which the hymen covered the alveoli in all valves (Fig. 6A, L), preventing measurements of areola density, and strain (E7)a for which the sample contained no clean valves. At least two pictures per strain were used. In total, 592 alveoli were measured. A linear mixed effect model (function *nlme*) in R was used to assess the between lineage significance, applying the same model structure as for the data of the axial area (see above).

Cytological observations: Cytological observations were obtained for a subset of fourteen strains (Supplementary Material Table S1). For lineages A and C, none of the here examined strains were alive when the study was initiated, and therefore no cytological observations of these lineages were made. Non-synchronized live diatom cultures in exponential or early stationary phase were examined for general chloroplast arrangement and morphology. Live cultures were photographed using BF and epifluorescence microscopy (FM) at $\times$1,000 magnification under oil immersion. BF was carried out on a Leitz Diaplan microscope (Leitz, Wetzlar, Germany) equipped with a Nikon dsFi2 digital camera (Nikon, Tokyo, Japan). FM was carried out on a Zeiss Axiophot 2 Universal microscope (Zeiss, Jena, Germany) equipped with an AxioCam MRm camera (Zeiss, Jena, Germany).

For observations on nuclear cytology, diatom cultures were dark-synchronized after which 1 mL of dens cultures were fixed in 75% ethanol. Two weeks after fixation, the fixed cells were washed three times with distilled water to remove the ethanol, and stained with DAPI (4,6-diamino-2-phenylindole.HCl; Sigma, St. Louis, MO, USA) at a final concentration of 1 μg/mL. DAPI-stained cells were photographed on a Zeiss Axiophot 2 Universal microscope (Zeiss, Jena, Germany) equipped with an AxioCam MRm camera (Zeiss, Jena, Germany) and using a Zeiss 49 DAPI BP filter. Only interphase cells, recognized by the two parallel plastids at each side of the nucleus in valve view, were photographed.

Declarations of interest

None.

Acknowledgements

This research was financially supported by the Fund For Scientific Research – Flanders (FWO-Flanders, Belgium, funding of EP as a PhD student), the Czech Science Foundation 16-17346Y, and Charles University Research Centre program No. 204069. Part of the culture work was performed under the facilities of the Belspo funded BCCM/DCG diatom culture collection (PAE, Ghent University). We are grateful to Myriam de Haan for her help with the SEM observations. Sofie D'hondt is thanked for her help with the molecular analysis. Willem Stock is thanked for the discussions concerning the statistics. Dr. Pieter Vanormelingen and Dr. Caroline Souffreau are thanked for the valuable discussions, particularly in the early stages of this project. We thank Dr. Barbora Chattová, Dr. Elisabeth M. Biersma, Dr. Peter Convey, Dr. Pieter Vanormelingen, Dr. Kateřina Kopalová, Dr. Elie Verleyen, Dr. Wim Van Nieuwenhuyze, Dr. Tyler Kohler, Dr. Lorenz Meire, Dr. Maxime Sweetlove, Joris Pinseel, Claas G. Steigüber, Berthold Mariën and Joachim Mariën for providing samples. Sampling on Marion Island was funded by the Belspo CCAMBIO Project (SD/BA/03) and FWO-Flanders, upon invitation of Dr. Steven Chown. We thank the editor and two anonymous reviewers for providing constructive comments that helped improving the manuscript.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.protis.2018.10.001.

References

Alverson AJ (2008) Molecular systematics and the diatom species. Protist **159**:339–353

Amato A, Kooistra WHCF, Levialdi GJH, Mann DG, Pröschold T, Montresor M (2007) Reproductive isolation among sympatric cryptic species in marine diatoms. Protist **158**:193–207

Barragán C, Wetzel CE, Ector L (2018) A standard method for the routine sampling of terrestrial diatom communities for soil quality assessment. J Appl Phycol **30**:1095–1113

Beszteri B, É Ács, Medlin LK (2005a) Conventional and geometric morphometric studies of valve ultrastructural variation in two closely related *Cyclotella* species (Bacillariophyta). Eur J Phycol **40**:89–103

Beszteri B, É Ács, Medlin LK (2005b) Ribosomal DNA sequence variation among sympatric strains of the *Cyclotella meneghiniana* complex (Bacillariophyceae) reveals cryptic diversity. Protist **156**:317–333

Beszteri B, John U, Medlin LK (2007) An assessment of cryptic genetic diversity within the *Cyclotella meneghiniana* species complex (Bacillariophyta) based on nuclear and plastid genes, and amplified fragment length polymorphisms. Eur J Phycol **42**:47–60

Bookstein FL (1997) Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Med Image Anal **1**:225–243

Casteleyn G, Adams NG, Vanormelingen P, Debeer A-E, Sabbe K, Vyverman W (2009) Natural hybrids in the marine diatom *Pseudo-nitzschia pungens* (Bacillariophyceae): genetic and morphological evidence. Protist **160**:343–354

Chamnansinp A, Li Y, Lundholm N, Moestrup Ø (2013) Global diversity of two widespread, colony-forming diatoms of the marine plankton, *Chaetoceros socialis* (synØC. radians) and *Chaetoceros gelidus* sp. nov. J Phycol **49**:1041–1128

Chen T, Guestrin C (2016) XgBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 785–794

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Inc., Sunderland, 480 p

Daugbjerg N, Andersen RA (1997) A molecular phylogeny of the heterokont algae based on analyses of chloroplast-encoded rbcL sequence data. J Phycol **33**:1031–1041

De Decker S, Vanormelingen P, Pinseel E, Sefbom J, Audoor S, Sabbe K, Vyverman W (2018) Incomplete reproductive isolation between genetically distinct sympatric clades of the pennate model diatom *Seminavis robusta*. Protist **169**:569–583

Degerlund M, Huseby S, Zingone A, Sarno D, Landfald B (2012) Functional diversity in cryptic species of *Chaetoceros socialis* Lauder (Bacillariophyceae). J Plankton Res **34**:416–431

Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics **148**:1667–1686

Ellegaard M, Godhe A, Härnström K, McQuoid M (2008) The species concept in a marine diatom: LSU rDNA-based phylogenetic differentiation in *Skeletonema marinoi/dohrnii* (Bacillariophyceae) is not reflected in morphology. Phycologia **47**:156–167

Fránková M, Poulíčková A, Neustupa J, Pichrtová M, Marvan P (2009) Geometric morphometrics – a sensitive method to distinguish diatom morphospecies: a case study on the sympatric populations of *Reimeria sinuata* and *Gomphonema tergestinum* (Bacillariophyceae) from the River Bečva, Czech Republic. Nov Hedwigia **88**:81–95

Gould SJ, Eldredge N (1993) Punctuated equilibrium comes of age. Nature **366**:223–227

Guillard RRL, Lorenzen CL (1972) Yellow–green algae with chlorophyllide c. J Phycol **8**:10–14

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser **41**:95–98

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: paleontological statistics software package for education and data analysis. Palaeontol Electron **4**:1–9

Hamsher SE, Evans KM, Mann DG, Poulíčková A, Saunders GW (2011) Barcoding diatoms: Exploring alternatives to COI-5P. Protist **162**:405–422

Hansen TF (1997) Stabilizing selection and the comparative analysis of adaptation. Evolution **51**:1341–1351

Hansen TF, Houle D (2004) Evolvability, Stabilizing Selection, and the Problem of Stasis. In Pigliucci M, Preston K (eds) Phenotypic Integration. Studying the Ecology and Evolution of Complex Phenotypes. Oxford University Press, New York, pp 130–154

Hejduková E, Pinseel E, Vanormelingen P, Nedbalová L, Elster J (2017) Tolerance of pennate diatoms to experimental freezing: comparison of polar and temperate strains. 11th Central European Diatom Meeting: Programme & Abstract Book, p 129

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse colour pattern. Science **313**:101–104

Hothorn T, Bretz F, WestFall P (2008) Simultaneous inference in general parametric models. Biom J **50**:346–363

Jones HM, Simpson GE, Stickle AJ, Mann DG (2005) Life history and systematics of *Petroneis* (Bacillariophyta), with special reference to British waters. Eur J Phycol **40**:61–87

Kelly MG (1998) Use of the trophic diatom index to monitor eutrophication in rivers. Water Res **32**:236–242

Kermarrec L, Bouchez A, Rimet F, Humbert JF (2013) First evidence of the existence of semi-cryptic species and of a phylogeographic structure in the *Gomphonema parvulum* (Kützing) Kützing complex (Bacillariophyta). Protist **164**:686–705

Kloster M, Kauer G, Beszteri B (2014) SHERPA: an image segmentation and outline feature extraction tool for diatoms and other objects. BMC Bioinformatics **15**:218

Kloster M, Kauer G, Beszteri B (2016) Morphometry with SHERPA – a novel software for automated image processing and shape analysis of diatom valves and other objects In SHERPA Workshop, Macquarie University Sydney, Australia, 13 May 2016

Kloster M, Kauer G, Esper O, Fuchs N, Beszteri B (2018) Morphometry of the diatom *Fragilariopsis kerguelensis* from Southern Ocean sediment: high-throughput measurements show second morphotype occurring during glacials. Mar Mic, http://dx.doi.org/10.1016/j.marmicro.2018.07.002

Kociolek JP, 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 CE, Williams DM, Witkowski A, Witkowski J (2018) DiatomBase. Accessed at http://www.diatombase.org/ on 2018-02-28

Kollár J, Pinseel E, Vanormelingen P, Poulíčková A, Souffreau C, Dvořák P, Vyverman W (2018) A polyphasic approach to the delimitation of diatom species: a case study for the genus *Pinnularia* (Bacillariophyta). Abstracts of the 25th International Diatom Symposium, p 80

Kooistra WHCF, Sarno D, Balzano S, Gu H, Andersen RA, Zingone A (2008) Global diversity and biogeography of *Skeletonema* species (Bacillariophyta). Protist **159**:177–193

Kooistra WHCF, Sarno D, Hernández-Becerril DU, Assmy P, Di Prisco C, Montresor M (2010) Comparative molecular and morphological phylogenetic analyses of taxa in the *Chaetocerotaceae* (Bacillariophyta). Phycologia **49**:471–500

Krammer K (2000) The Genus *Pinnularia*. In Lange-Bertalot H (ed) Diatoms of Europe, Diatoms of the European Inland Waters and Comparable Habitats **1**. ARG Gantner Verlag KG, Ruggell, 703 p

Krammer K, Lange-Bertalot H (1988) Bacillariophyceae 2. Teil: Bacillariaceae, Epithemiaceae, Surirellaceae. In Ettl H, Gerloff J, Heynig H, Mollenhauer D (eds) Süsswasserflora von Mitteleuropa **2**. VEB Gustav Fischer Verlag, Jena, 596 p

Kulichová J, Fialová M (2016) Correspondence between morphology and ecology: Morphological variation of the *Frustulia crassinervia-saxonica* species complex (Bacillariophyta) reflects the ombro-minerotrophic gradient. Cryptogam Algol **37**:15–28

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7. 0 for bigger datasets. Mol Biol Evol **33**:1870–1874

Lanfaer R, Calcott B, Ho SYW, Guindo S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol **29**:1695–1701

Lundholm N, Bates SS, Baugh KA, Bill BD, Connell LB, Léger C, Trainer VL (2012) Cryptic and pseudo-cryptic diversity in diatoms–with descriptions of *Pseudo-nitzschia hasleana* sp. nov. and *P. fryxelliana* sp. nov. J Phycol **48**:436–454

Lynch M (2010) Evolution of the mutation rate. Trends Genet **26**:345–352

Mann DG (1999) The species concept in diatoms. Phycologia **38**:437–495

Mann DG, Vanormelingen P (2013) An inordinate fondness? The number, distributions, and origins of diatom species. J Eukaryot Microbiol **60**:414–420

Mann DG, McDonald SM, Bayer MM, Droop SJM, Chepurno VA, Loke RE, Ciobanu A, Du Buf JMH (2004) The *Sellaphora pupula* species complex (Bacillariophyceae): morphometric analysis, ultrastructure and mating data provide evidence for five new species. Phycologia **43**:459–482

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the Gateway Computing Environments Workshop (GCE): pp 1–8

Nakov T, Beaulieu JM, Alverson AJ (2018) Accelerated diversification is related to life history and locomotion in a hyperdiverse lineage of microbial eukaryotes (Diatoms, Bacillariophyta). New Phytol **219**:462–473

Nei M (2007) The new mutation theory of phenotypic evolution. Proc Natl Acad Sci USA **104**:12235–12242

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymus P, Stevens MHH, Wagner H (2011) Vegan: community ecology package. R Foundation for Statistical Computing, Vienna

Peterson RA (2017) Estimating normalization transformations with bestNormalize. Available at: https://github.com/petersonR/bestNormalize

Peres-Neto PR, Jackson DA (2001) How well do multivariate data sets match? The advantages of a procrustean superimposition approach over the Mantel test. Oecologia **129**:169–178

Pinheiro J, Bates D, Debroy S, Sarkar D, R Core Team (2018) nlme: linear and nonlinear mixed effect models. R package version 3.1-137. Available at: https://CRAN.R-project.org/package=nlme

Pinseel E, Hejduková E, Vanormelingen P, Kopalová K, Vyverman W, Van de Vijver B (2017a) *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, Vanormelingen P, Hamilton PB, 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**:264–280

Potapova M, Hamilton PB (2007) Morphological and ecological variation within the *Achnanthidium minutissimum* (Bacillariophyceae) species complex. J Phycol **43**:561–575

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

Quijano-Scheggia SI, Garcés E, Lundholm N, Ø Moestrup, 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

Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v16, available at: http://beastbioedacuk/Tracer

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (www. barcodinglife.org). Mol Ecol Notes **7**:355–364

Rohlf FJ (2015) The tps series of software. Hystrix **26**:9–12

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3. 2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol **61**:539–542

Rumrich U, Lange-Bertalot H, Rumrich M (2000) Diatomeen der Anden (von Venezuela bis Patagonien/Tierra del Fuego). In Lange-Bertalot H (ed) Iconographia Diatomologica **9**. Koeltz Scientific Books, Königstein, 649 p

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. Nat Meth **9**:676–682

Scholin CA, Herzog M, Sogin M, Anderson DM (1994) Identification of group–specific and strain-specific genetic-markers for globally distributed *Alexandrium* (Dinophyceae). II. Sequence analysis of a fragment of the LSU ribosomal RNA gene. J Phycol **30**:999–1011

Smol JP, Stoermer EF (2010) The Diatoms: Applications for the Environmental and Earth Sciences. Cambridge University Press, New York, 686 p

Souffreau C, Vanormelingen P, Sabbe K, Vyverman W (2013a) Tolerance of resting cells of freshwater and terres-

trial 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

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. Syst Biol **57**:758–771

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

Theriot E (1992) Clusters, species concepts, and morphological evolution of diatoms. Syst Biol **41**:141–157

Thompson JD, Higgins DG, Gibson TJ (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res **22**:4673–4680

Urbánková P, Scharfen V, Kulichová J (2016) Molecular and automated identification of the diatom genus *Frustulia* in northern Europe. Diatom Res **31**:217–229

Van de Vijver B, Moravcová A, Kusber W-H, Neustupa J (2013) Analysis of the type material of *Pinnularia divergentissima* (Grunow in Van Heurck) Cleve (Bacillariophyceae). Fottea **13**:1–14

Vanelslander B, Créach V, Vanormelingen P, Ernst A, Chepurnov VA, Sahan E, Muyzer G, Stal LJ, Vyverman W, Sabbe K (2009) Ecological differentiation between sympatric pseudocryptic species in the estuarine benthic diatom *Navicula phyllepta* (Bacillariophyceae). J Phycol **45**:1278–1289

Vanormelingen P, Chepurnov VA, Mann DG, 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

Vanormelingen P, Evans KM, Chepurnov VA, Vyverman W, Mann DG (2013) Molecular species discovery in the diatom *Sellaphora* and its congruence with mating trials. Fottea **13**:133–148

Verbruggen H (2014) Morphological complexity, plasticity, and species diagnosability in the application of old species names in DNA-based taxonomies. J Phycol **50**:26–31

Verbruggen H, Vlaeminck C, Sauvage T, Sherwood AR, Leliaert F, De Clerck O (2009) Phylogenetic analysis of *Pseudochlorodesmis* strains reveals cryptic diversity above the family level in the siphonous green algae (Bryopsidales Chlorophyta). J Phycol **45**:726–731

Veselá J, Neustupa J, Pichrtová M, Poulíčková A (2009) Morphometric study of *Navicula* morphospecies (Bacillariophyta) with respect to diatom life cycle. Fottea **9**:307–316

Woodard K, Kulichová J, Poláčková T, Neustupa J (2016) Morphometric allometry of representatives of three naviculoid genera throughout their life cycle. Diatom Res **31**:231–242

Wray GA (2007) The evolutionary significance of *cis*-regulatory mutations. Nat Rev Genet **8**:206–216

Yee TW (2015) Vector Generalized Linear and Additive Models: With an Implementation in R. Springer, New York, 532 p

Zwart G, Huismans R, van Agterveld MP, Van de Peer Y, De Rijk P, Eenhoor H, Muyzer G, van Hannen EJ, Gons HJ, Laanbroek HJ (1998) Divergent members of the bacterial division Verrucomicrobiales in a temperate freshwater lake. FEMS Microbiol Ecol **25**:159–169