Assessment of marine benthic diatom communities: insights from a combined morphological–metabarcoding approach in Mediterranean shallow coastal waters.

Javier Pérez-Burillo[1,2], Greta Valoti[3], Andrzej Witkowski[4], Patricia Prado[1], David G. Mann[1,5] & Rosa Trobajo[1*]

[1]IRTA-Institute for Food and Agricultural Research and Technology, Marine and Continental Waters Programme. Ctra de Poble Nou Km 5.5, E43540, Sant Carles de la Ràpita, Tarragona, Spain

[2]Departament de Geografia, Universitat Rovira i Virgili, C/ Joanot Martorell 15, E43500, Vila-seca, Tarragona, Spain

[3]Università Politecnica delle Marche, Piazza Roma, 22 , IT60131, Ancona, Italy

[4]Institute of Marine and Environmental Sciences, University of Szczecin, Mickiewicza 16a, 70-383 Szczecin, Poland

[5]Royal Botanic Garden Edinburgh, Edinburgh, EH3 5LR, Scotland, UK

*corresponding author: rosa.trobajo@irta.cat

Abstract

We investigated the advantages and disadvantages of light microscope (LM)-based identifications and DNA metabarcoding, based on a 312-bp *rbcL* marker, for examining benthic diatom communities from Mediterranean shallow coastal environments. For this, we used biofilm samples collected from different substrata in the Ebro delta bays. We show that 1) Ebro delta bays harbour high-diversity diatom communities [LM identified 249 taxa] and 2) DNA metabarcoding effectively reflects this diversity at genus-but not species level, because of the incompleteness of the DNA reference library. Nevertheless, DNA metabarcoding offers new opportunities for detecting small, delicate and rare diatom species missed by LM and diatoms that lack silica frustules. The primers used, though designed for diatoms, successfully amplified rarely reported members of other stramenopile groups. Combining LM and DNA approaches offers stronger support for ecological studies of benthic microalgal communities in shallow coastal environments than using either approach on its own.

Key words:

Diversity, environmental DNA, epibiotic, microalgae, Stramenopiles, *rbcL*.

1. Introduction

Coastal ecosystems are ecologically important because they are highly productive areas that harbour a great diversity, which is reflected in many types of ecological communities found in these systems, such as seagrass beds, sandflat communities, coral and bivalve reefs (Cloern et al., 2013). These ecosystems are also important from a socio-economic point of view as they provide numerous ecosystem services and contribute importantly to the global total (Costanza et al., 2014). Benthic diatom communities constitute an important component of these systems because of their large contribution to total production (MacIntyre et al., 1996). A recent study of seagrass beds in shallow systems (Cox et al. 2020) has shown that the contribution of diatoms can be over 80% of benthic production and that without them the seagrass beds can be net heterotrophic. In addition, they have a major role in the stabilization of sediments thanks to the production of extracellular polymeric substances (EPS) and consequently, they regulate on of nutrient fluxes and other biogeochemical processes (Cahoon et al., 1999, Sundbäck & Granéli, 1988; Sundbäck et al., 1991; Triska & Oremland, 1981; Trobajo & Sullivan, 2010). They can be found in or attached to different substrata, such as the surface of sediments (epipelon), sand grains (epipsammon), seagrasses, macroalgae, and microalgae (epiphyton), or the surface of animals including the shells of molluscs (epizoon). Each of these community types can be very species-rich (Round, 1971) and it has been shown that within communities, different species can play different roles; for instance, in tidal habitats epipelic species show differences in photophysiology and migration activity (Underwood et al., 2005). Hence, it is crucial to combine system-wide estimates of benthic diatom contributions to primary production with an understanding of the roles and functioning of the species comprising these communities. However, the morphological identifications of diatoms at the species level are difficult and require expertise in taxonomy. This is especially true for shallow coastal environments (Mann et al., 2016; Trobajo et al., 2004), despite their ecological and economic importance.

DNA metabarcoding has proved to be a reliable method for studying species diversity from environmental samples (Deiner et al., 2017) and has emerged as an alternative to light microscope-based identifications (LM) due to its speed, reproducibility, and cost (Kermarrec et al., 2014; Zimmermann et al., 2015). It has been broadly tested for freshwater ecological assessment based on benthic diatoms (e.g. Bailet et al., 2019; Kelly et al., 2020; Mortágua et al., 2019; Pérez-Burillo et al. 2020, Vasselon et al., 2017) and for biodiversity studies (e.g. Stoof-Leichsenring et al., 2020; Rimet et al., 2018). DNA metabarcoding has also been applied in marine environments, especially in phytoplankton studies (e.g. De Luca et al.2021; Malviya et al. 2016; Piredda et al., 2018), but rarely to the phytobenthos of coastal areas, which are very productive and species rich. Exceptions include studies of US saltmarshes (Plante et al. 2021a, b), intertidal sediments in Korea (An et al. 2020), a eutrophic estuary in South Africa (Nunes et al., 2021), and sea turtle biofilms (Rivera et al. 2018).

In the context of ongoing research into the biodiversity and functioning of Mediterranean shallow coastal habitats (e.g. Benito et al., 2015; Carballeira et al., 2017; Prado, 2018, 2020; Rovira et al. 2009), we set out to study the benthic diatom communities in these poorly known systems through the combined use of DNA metabarcoding, based on a 312-bp *rbcL* marker, and LM-based identifications. Sampling was aimed for selection of the different benthic communities dwelling on sediments, seaweeds, seagrasses and molluscs (i.e. epipelic, epiphytic and epizoic/epilithic communities) in coastal areas of the Ebro delta. In particular, Ebro Delta bays sustain a very important shellfish aquaculture of Japanese oyster and Mediterranean mussel (Ramón et al., 2005), providing important substrata for biofilm development. Besides, the area holds one of the last remaining populations of fan mussel (*Pinna nobilis*) after major mass mortality events throughout the Mediterranean (Prado et al. 2014, 2021). Moreover, beds of the seagrasses *Cymodocea nodosa* and *Zostera noltii* are present in the area. In this paper we evaluate the advantages and disadvantages of each survey approach – morphological and molecular – and we assess whether the

66 *rbcL* marker, which was originally developed for diatom biomonitoring of freshwaters, is equally
67 useful in marine environments, where the diversity of related groups of ochrophyte microalgae and
68 macroalgae is much greater.

69

70 ## 2.Material and Methods

71 ### 2.1. Study area and sampling collection

72 Nine biofilm samples were taken in Alfacs and Fangar bays offshore from the Ebro Delta on the
73 Mediterranean coast of the Iberian Peninsula (Fig. 1). The bays constitute semi-enclosed estuarine
74 water bodies that receive freshwater inputs, rich in nutrients and organic matter, from rice fields that
75 border both bays, which have led to eutrophication (Llebot et al., 2011; Prado, 2018). Alfacs Bay
76 encompasses an area of 50km^2 with an average depth of ~3 m and a maximum of 6 m. Sampling in
77 this bay was conducted by wading in a semi-sheltered at ca. 60 cm depth near the southern shore,
78 where the seagrass *Cymodocea nodosa* and/or the macroalga *Caulerpa prolifera* constitute the
79 dominant benthic habitat, and where there is also an important population of *Pinna nobilis* (Prado et
80 al., 2014, 2020, 2021). Fangar Bay is smaller, occupying 12 km^2, and has an average depth of 2 m
81 and a maximum of 4 m. Sampling in this bay was conducted within farms of the introduced Pacific
82 oyster *Crassostrea gigas*, located in the southern area of the Bay. Physicochemical information at
83 each sampling site is shown in Table 1.

84 Samples were collected in March 2020 and comprised seven biofilm samples and two
85 sediment samples. Five of the biofilms were taken from the shell surfaces of *P. nobilis* (three
86 different individuals from Alfacs Bay separated by distances in the order of 10s of metres) and from
87 *Crassostrea gigas* (two individuals from Fangar Bay). The other two biofilm samples were taken
88 from the surfaces of *Caulerpa prolifera* and *Cymodocea nodosa*, respectively, both from the same
89 area of Alfacs Bay as those of *P. nobilis*. Finally, the two sediment samples were collected from the
90 surface sediments (ca. 1-2 cm) immediately adjacent to specimens of *P. nobilis* and transported to

the laboratory within small containers. For collecting the biofilm samples, the surface of the organisms was scraped using a different toothbrush for each specimen. Each sample was divided into two aliquots and preserved in ethanol (to a final concentration ≥ 70%). One was used for morphological examinations and the other for DNA metabarcoding analysis.

2.2 Microscopical analysis

Samples for morphological analysis were cleaned using concentrated (37%) hydrogen peroxide (H_2O_2). However, prior to hydrogen peroxide samples were treated with few millilitres of 10% HCl to remove carbonates. After the reaction with carbonates ceased, samples were washed several times with deionized water. Thereafter samples were boiled with hydrogen peroxide for a few hours to oxidize the organic matter and then washed several times with deionized water at 24 h intervals. Cleaned diatomaceous suspension was dropped onto cover slips and left at room temperature overnight to dry. Permanent slides were mounted with Naphrax (Brunel Microscopes: http://www.brunelmicroscopes.co.uk/), which has a high refractive index. Diatom analysis was performed using a Leica DMLB microscope equipped with 100× PlanApo objective (n.a. 1.4). Approximately 300 to 400 valves were counted in each sample. Problems in identification were resolved with scanning electron microscopy (SEM). For SEM examination, a drop of the cleaned sample was filtered onto Whatman Nuclepore polycarbonate membranes (Fisher Scientific, Schwerte, Germany). Filters were air-dried overnight, mounted onto aluminium stubs and coated with 5 nm of gold. Samples were analysed with an ultra-high field emission Hitachi SU 8020 instrument at West Pomeranian University of Technology in Szczecin.

2.3 DNA extraction, PCR amplification and high-throughput sequencing (HTS) library preparation

A volume of 2 mL of each sample was centrifuged at 4 °C and 11,000 g for 20 min. Ethanol present in the supernatant was removed and the DNA contained in the remaining pellet was extracted using

the commercial DNA kit Macheray–Nagel NucleoSpin® Soil extraction kit (MN-Soil). A short *rbcL* region of 312 bp constituted the DNA marker and this was amplified by PCR using an equimolar mix of the modified versions of the Diat_rbcL_708F (forward) and R3 (reverse) primers given by Vasselon et al. (2017). In order to prepare the HTS library using a 2-step PCR strategy, a part of the P5 (TCGTCGGCAGCGTCAG ATGTGTATAAGAGACAG) and P7 (GTCTCGTGGGCTCGGAGATGTGTATA AGAGACA) Illumina adapters were included at the 5′ end of the forward and reverse primers respectively. PCR1 reactions for each DNA sample were performed in triplicate using 1 μL of the extracted DNA in a final volume of 25 μL. The conditions and the reaction mix of the PCR1 were as described in Vasselon et al. (2017). All three PCR1 replicates were pooled and sent to "Plateforme Génome Transcriptome" (PGTB, Bordeaux, France), where the PCR1 products were purified and used as a template for a second round of PCR (PCR2), with Illumina-tailed primers targeting the half of P5 and P7 adapters. The resulting dual-indexed amplicons were pooled for sequencing on an Illumina MiSeq platform using a V2 paired-end sequencing kit (250 bp × 2).

2.4 Bioinformatic analysis

The sequencing facilities performed the demultiplexing of all the samples providing two fastq files per sample, one corresponding to the forward reads (R1) and one to the reverse reads (R2). Primers from all the demultiplexed MiSeq reads were removed by cutadapt (Martin, 2011) and the resulting R1 and R2 reads were processed together using the R package DADA2 (Callahan *et al.*, 2016). R1 reads were truncated to 225 bases and R2 to 180 bases based on their quality profile (median quality score < 30). Reads with ambiguities or an expected error (maxEE) > 2 were discarded. The DADA2 denoising algorithm was then applied to determine an error rates model to infer Amplicon sequence variants (ASVs). Chimeric ASVs were detected and discarded using the "removeBimeraDenovo" function. The taxonomic affiliations of the ASVs was determined using the database "A ready-to-

use database for DADA2: Diat.barcode_rbcL_312bp_DADA2" (Chonova et al., 2020), which is derived from the curated diatom reference library Diat.barcode v9 (Rimet et al., 2019, available at https://www6.inra.fr/carrtel-collection_eng/Barcoding-database and at https://data.inrae.fr/file.xhtml?persistentId=doi:10.15454/TOMBYZ/IEGUXB&version=10.0); the naïve Bayesian classifier method (Wang et al., 2007) was used, with 85% set as the minimum confidence threshold. The taxonomy of unclassified ASVs was checked using the Basic Local Alignment Search Tool (BLAST) against the Nucleotide database of NCBI GenBank, with standard settings (Camacho *et al.*, 2009). Taxonomy was assigned keeping taxa with a percentage of identity higher than 97%. To allow inter-sample comparisons, all samples were resampled to the minimum number of reads recorded in any single sample (5427 reads) using the R package *phyloseq* (McMurdie & Holmes, 2013).

2.5 Data analyses

For assessing the effectiveness of the two methods in identifying taxa, the percentages of reads or cells identified to species and genus were determined. Furthermore, the percentages of species and genera recorded molecularly that were also identified by the morphological approach and vice versa were calculated. For other statistical analyses, the rarefied molecular inventory was used. To compare diatom diversity between methods and sampling sites, the Shannon–Wiener index was calculated (based on natural logarithms), using the relative abundances of taxa from the corresponding morphological and molecular inventories. The Sørensen index, based on presence/absence data, was also calculated to evaluate the similarities in diatom communities between samples. To visualize patterns in taxon composition (in LM and DNA metabarcoding inventories) among samples non-metric multidimensional scaling (NMDS) was used, based on Bray–Curtis dissimilarity matrices on ASV, species and genus relative abundance. The correlation between the distance matrices generated by both methods, using diatom species relative

abundances, was evaluated by computing a Mantel test (with 999 permutations). Statistically significant differences in diatom community composition at the ASV- , species- and genus level regarding the type of substratum (i.e. biofilm samples taken from *P. nobilis*, , *Crassostrea gigas, Caulerpa prolifera and Cymodocea nodosa* and samples collected from the sediment adjacent to *P. nobilis*) were evaluated through a permutation multivariate analysis of variance (PERMANOVA). To identify the taxa that accounted for most of the dissimilarities between the LM and DNA metabarcoding inventories, an analysis of similarity percentages (SIMPER) was performed on both species and genus relative abundance. The R package *vegan* (Oksanen et al., 2020) was used for performing all these analyses.

2.6. Phylogenetic analyses of non-diatom ASVs

Although the primers used here were designed specifically for freshwater diatom biomonitoring (Vasselon et al. 2017), they do nevertheless sometimes amplify *rbcL* from other groups of algae. For example, the ASV with most reads in the 2017 Catalan rivers dataset used by Pérez-Burillo et al. (2020, 2021) was an unknown green alga related to *Nautococcus* and *Oophila* (Chlorococcaceae), which was present in 116 of 164 samples analysed; Ochrophyta classes (sensu Adl et al. 2019) were also represented, including Xanthophyceae (e.g., *Vaucheria*) and Eustigmatophyceae (e.g., *Neomonodus*). In marine habitats the diversity of ochrophytes and red algae is much greater than in freshwaters and different green algal groups are present. Indeed, preliminary blastn analysis of our reads that were not assigned to any diatom taxon by the Bayesian classifier indicated that some ASVs belonged to different classes or phyla of algae. The majority (both in terms of ASVs and reads) were ochrophytes and we therefore performed phylogenetic analyses of the non-diatom ASVs together with GenBank sequences of selected ochrophytes to elucidate their affiliations and phylogeny. To do this, we assembled the sets of *rbcL* sequences used by Graf et al. (2020) and Wetherbee et al. (2021) and added representatives of other ochrophyte

191 classes (particularly Chrysophyceae and Synurophyceae) to provide a wide coverage of the group.
192 We also added further Phaeophyceae that blastn analysis indicated were close to some ASVs. The
193 sequences were aligned by eye in Mega X (Kumar et al. 2018) after initial use of Muscle (Edgar,
194 2004), truncated to remove ragged ends and regions poorly represented among the taxa analysed,
195 and exported for phylogenetic analysis to RAxML (Stamatakis 2014), as implemented in raxmlGUI
196 v. 2.0 (Edler et al. 2021).A Maximum Likelihood (ML) tree was constructed with the alignment
197 partitioned by codon position, using a GRT-Gamma model; 1000 replicates were made for the
198 bootstrap analysis. The tree was visualized, midpoint-rooted, and prepared for publication using
199 iTOL (https://itol.embl.de) (Letunic et al., 2021).

200 The affiliations of the few non-ochrophyte ASVs (Chlorophyta and Rhodophyta) were
201 assessed by blast of NCBI GenBank.

202

203 2.7. Trait classification

204 Alongside analyses of diatom communities based on species composition we also classified for the
205 different diatom taxa identified (either microscopically or molecularly) according to their ecological
206 guilds and growth-forms. For this we largely followed Passy et al. (2007) and Rimet & Bouchez
207 (2012) but we split the original euplanktonic group defined by Passy et al. (2007) into planktonic
208 and tychoplanktonic groups. Thus, the resulting growth-forms were: high-profile, low-profile,
209 motile, planktonic and tychoplanktonic. For some taxa, Passy and Rimet & Bouchez provided no
210 information (their focus was on freshwater communities) and for these the growth-form was
211 inferred on the basis of information in Round et al. (1990) and expert knowledge.

212

3 Results

3.1 Morphological inventory

A total of 249 diatom taxa (including varieties, forms, and species) were identified, the number per sample ranging from 40 to 75, with an average of 58.9. The most abundant diatom taxa were *Navicula* sp. 4, *Amphora helenensis, Amphora* cf. *helenensis, Cocconeis scutellum* var. *posidoniae, Navicula normaloides, Nanofrustulum shiloi, Cyclotella choctawhatcheeana, Navicula normalis, Cocconeis scutellum* and *Berkeleya fennica* (Supplementary Table 1). The 249 taxa recorded represented 73 different genera and 128 different species. The number of species identified per sample ranged from 25 to 47, with an average of 36.4. A total of 122 taxa (49%) could not be identified at species level but only at genus level.

Low profile and motile growth forms, mainly represented by species from *Amphora, Cocconeis, Navicula* and *Nitzschia*, were the predominant groups in all the samples, followed by the high-profile group (Fig. 2a), in which *Berkeleya* was the most abundant genus (Supplementary Table 1). The planktonic and tychoplanktonic groups were less represented and not identified in all the samples. They were more abundant in *P. nobilis* samples (Fig 2a) and were represented mainly by *Cyclotella*.

3.2 Molecular inventory

MiSeq Illumina sequencing produced a total of 176,248 raw DNA reads from the nine samples. After processing the reads through the DADA2 pipeline, 139,815 reads remained, belonging to 682 ASVs, with an average of 145.1 ASV per sample (Supplementary Table 2). The maximum number of ASVs per sample was 214 (in sample E12 – *Pinna nobilis* biofilm) and the minimum 72 (in sample E15 – *Caulerpa prolifera*). 127 ASVs were classified at diatom species or genus level using the Bayesian classifier, on the basis of Diat.barcode v9 (i.e., bootstrap values at species level ≥ 85%). The taxonomic positions of 46 further ASVs that were not allocated to species by the

Bayesian classifier (i.e., their bootstrap values at species level were < 85%) were resolved by blastn on GenBank and allocated to species using a percentage of identity >97% as threshold. Finally, an additional 181 ASVs that did not fulfil either of the two previous criteria were classified at genus level by using a combination of expert knowledge and examination of the most similar sequences in GenBank.

Altogether, the three approaches described above allowed a total of 354 of the 682 ASVs to be classified to species or genera of diatoms, with 69 species and 73 genera identified. After rarefaction was applied, these numbers were very slightly reduced (the total number of 354 diatom ASVs was reduced to 338 ASVs, accounting for 51.2% of the total of rarefied reads, comprising fully identified 69 species and 71 genera) (Supplementary Table 2). The number of species per sample ranged from 21 to 54, with an average of 37.3, and the ten most abundant diatom taxa in the inventory were: *Thalassiosira profunda, Achnanthes longipes, Berkeleya fennica, Nanofrustulum shiloi, Navicula* sp., *Cyclotella* sp., *Haslea howeana, Seminavis* cf. *robusta, Craspedostauros constricta* and *Licmophora paradoxa.*

Among the ASVs were several genera and species that were missed or poorly represented in the morphological dataset. One important factor was the lower detection limit of metabarcoding: even in the least productive sample (*Caulerpa* epiphytes) >5000 reads were obtained, offering the possibility to detect rare species undetectable among the c. 400 specimens per sample identified morphologically. It was noticeable too that some ASVs represented species that have very delicate or small cells. Several of these are rarely evident in any cleaned material, such as *Cylindrotheca* and some species of Cymatosirales (comprising *Arcocellulus, Extubocellulus, Papiliocellulus* and *Minutocellus* in our material). *Cylindrotheca* species are very lightly silicified and often destroyed by oxidative cleaning (Round et al. 1990). Only one sample was recorded to contain *Cylindrotheca* by LM analysis but eleven ASVs were assigned to *Cylindrotheca* by the classifier, one or more occurring in each of the nine samples.

Processing with DADA2 does not remove all artifactual sequences and examination of the sequences of rare diatom ASVs revealed some that could not represent functional genes since they contained one or more stop codons. The most abundant of these was ASV0569, with a total of six reads and occurring in just one of the nine samples. However, rare ASVs were not necessarily artifactual. ASV0645, with just three reads, was an exact match to GenBank accession DQ813818 of *Pseudo-nitzschia delicatissima* (see also section 3.4 below).

Motile and planktonic growth forms predominated in most of the samples in the molecular analyses and were primarily represented by *Nitzschia* and *Navicula* (motile) and *Thalassiosira* (planktonic) (Fig. 2b and Supplementary Table 2). The exceptions were the samples taken from *Crassostrea gigas* shells, where high profile forms were dominant, and *Cymodocea nodosa* (Fig. 2b), which had approximately equal proportions of high profile and motile forms. The high-profile group was mainly represented by *Achnanthes* and *Berkeleya* species. Conversely to LM, planktonic and tychoplanktonic forms were recorded in all the samples (Fig. 2a and b), while the low-profile group was much less represented; *Nanofrustulum* and *Amphora* genera were the most important representatives for the low-profile group.

The most striking feature of the molecular data was the abundance in most samples (except *Crassostrea*) of *Thalassiosira profunda*, a species for which only three specimens were identified by LM (Supplementary Tables 1 and 2). Because of the systematic bias introduced by this species, we recalculated the relative abundances of the growth forms excluding *T. profunda*. The resulting graphs (Fig. 2c) showed closer agreement with the morphological data.

3.3 Comparative analyses of samples from different substrata

Taken together, the two approaches identified a total of 102 different genera, of which 43 were identified in both inventories (43.4%), and each of both methods recorded exclusively 28 different genera. At species level, both approaches identified a total of 176 different species, of which 19

(10.9%) were identified in both inventories; 106 and 50 were exclusively detected in the morphological and molecular inventories respectively (Supplementary Table 3).

The Shannon diversity index calculated on taxa relative abundances differed between inventories. For almost all the samples, the index values were higher in the LM inventory (Table 2) and the averages obtained for the LM and DNA metabarcoding inventory were 3.29 and 2.31 respectively. Both approaches agreed that the highest diversity was in a sample from a shell of *P. nobilis* (LM = 3.74, DNA metabarcoding = 3.04; Table 2) but disagreed for the lowest diversity; in the LM inventory this was in the sample from *Cymodocea nodosa* (2.61) but in the DNA one it was in the sample from *Caulerpa prolifera* (1.59) (Table 2). A Mantel test indicated that DNA metabarcoding and LM distance matrices calculated on diatom species relative abundances were not significantly correlated (Mantel r = 0.31; *p* value = 0.077).

The NMDS and Sørensen similarity index based on DNA metabarcoding data showed a tendency for community composition to be more similar among samples taken from the same host (Fig. 3a; Supplementary Table 4); this tendency was still evident after the *Thalassiosira profunda* ASVs were removed and the NMDS recalculated (Supplementary Fig. 1). However, these tendencies were not as obvious when NMDS and the Sørensen index were calculated using LM data (Fig. 3b; Supplementary Table 4). In particular, the two samples of *C. gigas* were widely separated from each other in the LM-based analyses but very close and separated from the rest in the DNA metabarcoding-based ones.

PERMANOVA confirmed the previous tendencies observed, with statistically significant differences in the community composition among different substrata for the DNA metabarcoding inventory (PERMANOVA using ASVs: $F_{4,4} = 2.7965$, $p = 0.012$; using species: $F_{4,4} = 3.3896$, $p = 0.01$; and using genera: $F_{4,4} = 3.5155$, $p = 0.007$) and for the LM inventory at species level (PERMANOVA: $F_{4,4} = 1.362$, $p = 0.032$) but not at genus level though differences were close to being statistically significant (PERMANOVA: $F_{4,4} = 1.6881$, $p = 0.056$).

According to the SIMPER analyses, the five genera that contributed most to the discrepancy between the morphological and molecular approaches were *Thalassiosira* (18.54%), *Navicula* (10.79%), *Amphora* (9.78%), *Cocconeis* (5.80%) and *Achnanthes* (5.61%). Below the genus level, the taxon that most influenced the dissimilarity was *T. profunda*, which was identified only by DNA metabarcoding. This species appeared in all samples analysed and was responsible for 14.37% of the discrepancy between the two inventories (Table 3). The second most important taxon was *Navicula* sp.4, contributing 4.88% of the dissimilarity. It was identified only by LM, and it appeared in most of the samples (Table 3). Next was *Amphora helenensis*, which was identified by both methods, but it was much more abundant in the LM inventory. The opposite case was exemplified by *Achnanthes longipes*, a large-celled species with many chloroplasts that was much more abundant in the DNA metabarcoding inventory; it was the fourth species most influencing the dissimilarities between the two inventories (4.15%) (Table 3). A total of 83 species identified by LM lacked representative sequences in the reference library and they accounted for 16.32% of the total dissimilarities between inventories. Noteworthy among these were *Cocconeis scutellum* var. *posidoniae*, *C. scutellum*, *Navicula normaloides* and *N. normalis*, which together accounted for 6.10% of the total discrepancies (Table 3).

3.4 Diversity and phylogenetic analyses of non-diatom ASVs

Blastn and phylogenetic analyses allowed us to classify many of the non-diatom DNA reads to a class of algae and in some cases to a genus or species. In total, 41 non-diatom ASVs were analysed and allocated, a few of them with considerable hesitation, to an alga class or division. Ten of them were assigned to the Chlorophyta (and were easily recognized in the ASV alignment because all had an extra amino acid relative to the ochrophyte and red algal sequences), mostly with low similarity to any named taxon, except for *Umbraulva*, *Ulvella* and marine *Ulothrix* (the kleptoplasts of *Strombidium* sequenced for GenBank AY257112 are presumably from this genus:

Supplementary Table 5); one sequence was apparently related closely to an uncultured *Picochlorum* (Supplementary Table 5). The three red algal ASVs were placed more definitively, as species (or relatives) of the genera *Grania* and *Acrochaetium*, which both grow as branching filaments, and the crustose *Pneophyllum*.

Most of the remaining non-diatom ASVs could be assigned with varying degrees of confidence to one of 10 classes of Ochrophyta (sensu Adl et al. 2019) (Fig. 4, Supplementary Table 5): Chrysophyceae (1 ASV, with low confidence), Synchromophyceae (2 ASVs), Pinguiophyceae (1 ASV), Eustigmatophyceae (1 ASV), Dictyochophyceae (2 ASVs), Pelagophyceae (7 ASVs), Raphidophyceae (2 ASVs), Xanthophyceae (1 ASV), Chrysomeridophyceae (2 ASVs) and Phaeophyceae (10 ASVs). The Phaeophyceae ASVs were mostly allied to species with simple or branched filaments, either in the Ectocarpales (*Hincksia*, *Myrionema*, *Streblonema*, *Elachista*, *Nemacystus*) or the Sphacelariales (*Sphacelaria*). Eight ochrophyte classes were unrepresented in the dataset: the predominantly freshwater Synurophyceae, the picoplanktonic Bolidophyceae, and the Olisthodiscophyceae, Aureanophyceae, Phaeothamniophyceae, Phaeosacciophyceae, Chrysoparadoxophyceae and Schizocladiophyceae.

The only non-diatom ASV that could almost certainly be discounted as an artifact was the very rare ASV0657, which had a low similarity to *Tetraselmis* (c. 80%: Supplementary Table 5) and was represented by just three reads; this sequence contained two stop codons. However, three even rarer non-diatom ASVs, each represented by two reads (ASV0677–0679, belonging to the Rhodophyta and Synchromophyceae), were clearly not artifactual, judging by blastn assignment or phylogenetic analysis (Fig. 4, Supplementary Table 5).

None of the non-diatom ASVs were abundant, the only one exceeding 1% of reads in any sample being an ectocarpalean brown alga (ASV0078, related to *Nemacystus decipiens*) in one of the *Pinna* biofilms (E12 – *Pinna nobilis* biofilm). The most widespread was ASV0183, an unclassified eustigmatophyte that was found in all five *Pinna* biofilm and sediment samples but

363 nowhere else (Supplementary Table 5). Another possibly significant association was between the
364 brown alga *Streblonema maculans* and *Crassostrea*. *Pinna* biofilms were a rich source of non-
365 diatom ASVs, especially in sample E12 (*Pinna nobilis* biofilm.

366

367 4. Discussion

368 4.1. Diatom diversity in shallow coastal environments is very high and DNA metabarcoding is a
369 promising tool for studying it.

370 Our results demonstrate that the shallow coastal ecosystems studied here harbour a very rich diatom
371 community. A total of 126 species were identified by morphology (LM); this is remarkably high
372 when compared with previous studies on coastal environments based on a much larger sampling
373 effort (e.g., 68-328 diatom taxa from 21-165 samples; Lobban et al., 2012; Facca & Sfriso, 2007;
374 Kanjer et al., 2019; Virta et al., 2019). To these 126 species identified by LM, an extra 50 were
375 added by metabarcoding. Furthermore, the large number of taxa identified only at generic level or
376 above in both LM and DNA metabarcoding, may indicate that the total number of diatom species in
377 the study area is very much higher. Comparisons with freshwater benthic communities are also
378 instructive. The average diversities of our nine samples exceeded those of periphyton samples from
379 Catalan rivers (for which we had many more samples, Supplementary Table 6), whether the
380 approach taken was metabarcoding or microscopical analysis, emphasizing how rich the diatom
381 communities of the marine benthos can be.

382 Hence, this first survey of some of the substrata in the Ebro Bays suggests the area is a hot-
383 spot of diatom biodiversity and provides a first step towards understanding how this biodiversity
384 originates and is maintained and the ecological roles that it performs. For instance, diatom biofilms
385 in shallow coastal ecosystems are known to play a major role in sediment stabilization and in
386 providing habitat and food for other organisms (references in Trobajo & Sullivan, 2010; see also
387 Camps-Castellà et al.2020 for a relevant example from Ebro Bays); moreover, a recent study of

benthic diatoms in the Baltic Sea has shown that high diatom diversity supports high ecosystem productivity (Virta et al., 2019). Our study demonstrates that DNA metabarcoding based on the short 312-bp *rbcL* marker also constitutes an efficient method for surveying diatom biodiversity in coastal ecosystems. The effectiveness of the method was reflected in that 1) it recorded the same number of genera as the LM method did, 2) a high proportion of the genera (43.4%) identified by LM were also recorded by DNA metabarcoding and 3) a high number of genera (43) were exclusively identified by DNA metabarcoding. Nevertheless, the LM method showed a greater efficiency for identifying taxa at species level, which was mainly caused by the lack of representative sequences in the DNA reference library for many common species, which consequently could not be retrieved when the molecular approach was applied (further discussion in section 4.3). Despite this limitation, similarity analyses calculated on the DNA metabarcoding inventory (at ASV, species and genus level) revealed a highly-structured molecular signal, suggesting therefore that our *rbcL*-based metabarcoding was able to discriminate different shallow coastal habitats, as in some other recent studies using other DNA markers (e.g. Bombin et al., 2021; Jeunen et al. 2018; Plante et al., 2021a). In addition, the habitat preferences hinted at by DNA metabarcoding could also indicate a degree of host-specificity among diatom taxa. However, our study was not designed to investigate these aspects in detail but to explore the feasibility of using *rbcL* metabarcoding to study the benthic diatom communities of shallow coastal habitats. Therefore, another study with a greater sampling effort and strategy (e.g., to have matching samples from the two Ebro bays for those hosts present in both; more replication etc) will be needed before any further conclusions can be drawn.

Regarding diatom composition, it was noticeable that a higher proportion of the taxa identified by LM corresponded to the low profile and motile groups, the genera *Amphora*, *Navicula*, *Cocconeis*, *Nitzschia* and *Halamphora* being particularly abundant. These have been recorded as important members of epiphyte communities in the Mediterranean (Car et al., 2019, Mabrouk et al.,

2014) or as epizoic on the shells of *P. nobilis*, *C. gigas* and other molluscs (e.g., Andriana et al., 2021; Barillé et al., 2017; D'Alelio et al., 2011; Totti et al., 2011). Conversely to LM, DNA metabarcoding better represented the planktonic group. Some planktonic taxa that were only recorded in our benthic samples by DNA metabarcoding have been previously reported as 'epizoic' or 'epiphytic' in other studies. Examples are *Actinoptychus octonarius*, which has been reported elsewhere as occurring on *Pinna nobilis* (Politis 1949, cited by Round, 1971), and *Asterionellopsis*, found on the seagrass *Posidonia oceanica* (Mabrouk et al., 2014). However, although it is possible that these two are genuinely benthic or tychoplanktonic, it is also possible that the cells represent sedimented phytoplankton: the much lower limit of detection of the metabarcoding approach makes it much easier to detect occasional cells or colonies displaced from their normal habitat. Another planktonic species, *Thalassiosira profunda*, is considered in detail below.

4.2 Opportunities of DNA Metabarcoding

4.2.1 Detection and discrimination of tiny or delicate species

One advantage of the metabarcoding approach is that it is more likely to pick up small and/or delicate species. We recorded *Cylindrotheca* and Cymatosirales species in the metabarcoding dataset, but almost none in the morphological inventory. The *Cylindrotheca closterium* complex is commonly reported from coastal marine phytoplankton samples and it is probably mostly tychoplanktonic. However, these samples are often counted without prior cleaning, whereas benthic samples are not generally examined before oxidative cleaning because of the difficulty of observing sufficient frustule detail in intact biofilms or sediment samples. Hence many records of *Cyindrotheca* have probably been lost. One of the Cymatosirales that we recorded with metabarcoding, *Papiliocellulus simplex*, was first described from intertidal sand at two localities in Great Britain (Gardner & Crawford, 1992) and has subsequently been recorded only planktonically from the Liguro-Provencal basin of the Mediterranean (where it was 'extremely rare' at two

stations: Percopo et al., 2011) and from several localities around Australia, mostly from metagenomic data [GBIF query 19 July 2021]. A final example of a small species easily missed or misidentified in light microscopy is *Gedaniella panicellus*, which was detected by DNA at frequencies of 0.1–1% in all but one sample but was not recorded at all in LM. This species was recently described by Li et al. (2018), who noted the difficulties of unambiguous identification by LM. Our ASV differed from the reference sequence (MF092953) by 1 bp and our record extends the known range to Europe from S. Africa and China, and from muddy rockpools to epizoic and epiphytic diatom communities.

4.2.2 Rare species

Another benefit of metabarcoding is the possibility of detecting species that are too rare to be found in routine LM cell counts. It is usually unrealistic to count more than a few 100s of valves per sample in LM, but it is common to obtain 1000s or 10000s of reads with metabarcoding. However, rare ASVs need to be treated with some caution, because amplification and sequencing can generate errors. Indeed, a few of our ASVs seem to be artifacts, despite the error modelling and correction incorporated in DADA2, since they contained stop codons. It is much more difficult to detect errors (e.g., in the third codon position) that do not affect the amino acid coded for. Errors can be minimized by imposing an arbitrary criterion – like a minimum number of reads – to try to avoid including artifactual sequences. However, our data illustrate (e.g., the rare ASVs identified as *Pseudo-nitzschia delicatissima*, *Acrochaetium*, *Pneophyllum* and *Synchroma*: see sections 3.2, 3.4) what would be lost by imposing such a limit. The rarest ASVs can be genuine.

460 4.2.3 Primers designed for diatoms successfully amplify some non-diatom species.

461 The primers we used were originally designed for use in freshwaters, for biomonitoring and
462 biodiversity studies of diatoms (Vasselon et al. 2017). Our study is one of the first to apply the same
463 primers in marine environments and reveals that the 'diatom-specific' primers do in fact amplify a
464 wide variety of other marine microalgae, and even some green and red algae. The phylogenetic
465 analyses revealed ASVs belonging to nine different non-diatom classes in the Ochrophyta. Some of
466 these, not surprisingly, were brown algae (various Ectocarpales and Sphacelariales), which probably
467 formed part of the macroscopic structure of the attached communities (though perhaps surprisingly,
468 some also occurred in the sediment samples), but others were microalgae, including some that are
469 seldom recorded. Rather than a weakness of the HTS protocols, as suggested by Grant et al. (2021),
470 we would argue that this 'contamination' of the diatom data is not only tolerable (since the
471 proportion of non-diatom reads was low – c. 1% of the total; for comparison, 25% of the Ion
472 Torrent 18S rDNA reads obtained by Plante et al., 2021b, were non-diatoms) but a valuable bonus,
473 especially because many of the microalgae recorded are probably small and morphologically simple
474 (judging by the nearest relatives that can be identified in GenBank) and will therefore be easily
475 overlooked using microscopy or culturing. Especially interesting was the discovery of ASVs related
476 to the amoeboid alga *Synchroma pusillum* (recently described by Schmidt et al., 2012) and the
477 coccoid *Pinguiococcus pyrenoidosus* (which is difficult to identify in LM due to its small cell
478 diameter of 3-8 μm: Andersen et al., 2002), and lineages of Pelagophyceae and
479 Chrysomeridophyceae. The phylogenetic analyses also revealed one ASV closely related to the
480 raphidophyte species *Chattonella subsalsa*. This species has been reported, among other species of
481 *Chattonella*, to produce red tides and fish kills (Lewitus et al., 2008); the reads came from one of
482 the sediment samples but could perhaps represent stray cells from the bay phytoplankton. Thus,
483 these analyses illustrate the potential of DNA metabarcoding, even when based on primers designed

for diatoms, for identifying at least some of the other microeukaryote taxa also present in the community, including ecologically or economically relevant taxa.

4.3 Discrepancies between the LM and metabarcoding results

4.3.1 The case of *Thalassiosira profunda*.

The greatest dissimilarities between the results obtained by the two methods, morphological and metabarcoding, were attributable to one particularly small, delicate species, the centric *Thalassiosira profunda*. This was by far the most abundant species recorded by DNA metabarcoding but only three specimens were identified by LM (2 and 1 respectively in the *C. prolifera* and *C. nodosa* samples). These were found after an additional exhaustive examination of the samples was performed, beyond the normal 300–400 count and prompted by the metabarcoding data, to be sure that this species had not been overlooked in LM. *Thalassiosira profunda* is an extremely small species (valve diameter 1.25–5.5 µm), generally regarded as planktonic, which is very widely distributed (Percopo et al., 2011, Li et al., 2013, Park et al., 2016, Guiry 2021). The almost complete absence of this species from the morphological counts, which can alternatively be described as gross overrepresentation in the metabarcoding dataset, requires special explanation, because such overrepresentation is generally associated with large-celled species, such as *Ulnaria ulna*, *Pinnularia viridiformis* or *Navicula lanceolata* (Vasselon et al., 2018, Kelly et al., 2020), because they have a larger number of copies of *rbcL* per cell.

Several hypotheses might explain why *T. profunda* was abundant in the DNA reads but extremely rare from the LM inventory. None of them can be discounted entirely; all of them have consequences for planning and interpreting morphological and metabarcoding studies of marine benthic diatoms.

1. *T. profunda* could be detected almost exclusively by metabarcoding because diatoms with tiny valves are easily overlooked and difficult to identify in LM (e.g., see Belcher & Swale, 1986). In the present case, such an explanation can be discounted, given the abundance implied by the molecular data (especially taking into account the likely low *rbcL* copy number per cell) and given that all slides were examined in detail using a ×100 objective. Furthermore, our re-examination of the slide preparations after analysing the metabarcoding data confirmed the almost complete absence of *T. profunda*, while no other *Thalassiosira* species were recorded as abundant in LM. However, the greater number of very small-celled and delicate diatoms (e.g. of Thalassiosirales or Cymatosirales) in many marine habitats, relative to freshwaters, means that there is greater potential for discrepancies between molecular and morphological datasets in marine studies.
2. Valves of delicate species like *T. profunda* can be destroyed during preparation for LM, as has been reported for the weakly-silicified cells of the freshwater *Fistulifera saprophila* (Kelly et al.,2020; Pérez-Burillo et al., 2020; Zgrundo et al., 2013). Small size also predisposes them to be lost, since centrifugation and sedimentation during washing steps will be less effective (e.g., Andrews, 1972). The solution is clearly to examine material before it is cleaned or retain aliquots for examination later. Unfortunately, we did not do this, but the detection of a few intact valves of *T. profunda* in the *Caulerpa prolifera* and *Cymodocea nodosa* samples, following an exhaustive search for the species, undermines destruction as a reason for 'overrepresentation' in the molecular dataset.
3. The molecular signal captured for *T. profunda* may not be contemporary with the morphologically characterized benthic communities but come from an earlier bloom. Some planktonic species form resting stages following a bloom (McQuoid & Hobson 1996; Inoue & Taniguchi, 1999; Sugie & Kuma, 2008), leading to the deposition of a large numbers of resting spores in the sediment. These might be detectable using DNA but more difficult by LM due to morphological differences from the vegetative cells (cf.

Kuwata & Takahashi, 1999). However, this strategy is not known to occur in *T. profunda*. In any case, diatom resting spores and resting cells are usually more robust than vegetative cells (Krawczyk et al., 2012) and should have been found in our material if present. Alternatively, an earlier bloom of *T. profunda* could perhaps have left a molecular trace even though the frustules had redissolved *in situ*. A moderate abundance of DNA reads of *Thalassiosira* and other planktonic species, including *T. profunda*, was recently reported in saltmarsh sediments in S Carolina, USA, by Plante et al. (2021a), who suggested this could reflect deposition of faecal pellets or recent phytoplankton blooms. However, their study did not include accompanying cell counts from LM.

4. Finally, it is possible that intact *T. profunda* cells were present in the samples but lacked frustules, so that they were undetectable in material prepared for microscopy. As far as we know there is no confirmed report of *free-living* diatoms lacking a silica cell wall, apart from some morphotypes of *Phaeodactylum* (Round et al. 1990), but this does not mean that none exist. But wall-less diatoms certainly do occur as endosymbionts, for example in some foraminifera (Lee 2011) and dinoflagellates (Yamada et al., 2020), while other foraminifera and dinoflagellates ingest diatoms and jettison the frustules, retaining their chloroplasts (as 'kleptoplasts') for days or months afterwards (e.g., Pillet et al., 2011, Yamada et al., 2019), and hence also, perhaps, their *rbcL*. In freshwaters, some Thalassiosirales are known to be endosymbionts of dinoflagellates (e.g., Takano et al., 2007; You et al., 2015), while in marine environments chloroplasts of Thalassiosirales are retained by some foraminifera, e.g., *Elphidium* (Pillet et al., 2011) and *Nonionellina* species (Jauffrais et al., 2019), at least one of which (*Elphidium*) occurs in the Ebro Bays (Benito et al., 2016).

The possibility that *T. profunda* could have been living endosymbiotically or as kleptoplasts in the communities we sampled receives further specific support from the study of Schmidt et al. (2018), who looked at the endosymbionts of the benthic

foraminifera *Pararotalia calcarioformata* growing in the East Mediterranean. Among the endosymbiotic diatoms they extracted and cultured from *P. calcarioformata* was one species identified and illustrated as *Minidiscus* sp. (Schmidt et al., 2018), which we would identify instead, from the specimen illustrated (op. cit., fig. 4.12), as *T. profunda*. In any case, it is clear that several Thalassiosirales do occur without frustules in marine environments, providing a possible reconciliation of our molecular and morphological data. The potential of kleptoplasts to confuse metabarcoding results extends beyond the photic zone, since functioning diatom kleptoplasts (again from Thalassiosirales, though apparently of *Skeletonema* not *Thalassiosira*) have been recorded, with intact *rbcL*, from depths of more than 500 m in the foraminifer *Nonionella stella* (Gomaa et al., 2021).

4.3.2 The effects of taxonomic resolution, reference library and gene copy number

Another important reason for the discrepancies observed between methods was the impossibility of identifying some taxa at species level. Marine littoral diatoms have been much less studied than their freshwater counterparts in rivers and lakes, so that several species in our samples – some of them abundant – remained unidentified at species level in LM, despite the great taxonomic effort and resources applied (including thorough LM identifications supported by SEM and TEM). The number of taxa identified as sp. or confer (cf) or affinis (aff) illustrates the incompleteness of the taxonomy underlying the morphological approach. This was particularly true for *Amphora* and also for *Navicula*, since a total of eight *Navicula* species could not be identified to known species. One of these, *Navicula* sp. 4 was very abundant and contributed greatly to the discrepancies between the LM and metabarcoding outputs. The prevalence of *Navicula* spp. without a species assignation in epibiotic communities has been shown also in other recent studies (e.g., Andriana et al., 2021; Car et al., 2019; Kanjer et al., 2019; Medlin & Juggins, 2018).

583 Concerning the metabarcoding inventories, the impossibility of reaching species-level
584 identifications of the ASVs was often due to the incompleteness of the reference library since many
585 species identified by LM lack representative DNA sequences, again mainly due to understudied
586 environments. Diat.barcode (Rimet et al., 2019) aims to list and check all available *rbcL* sequences
587 for diatoms, whether marine, coastal, or freshwater, but it depends to a considerable extent on what
588 sequences have been deposited in NCBI GenBank, which reflects historical trends in systematic and
589 other research. Overall, the data available show a strong bias towards freshwater diatoms, which
590 account for around 60% of the *rbcL* entries in Diat.barcode v. 9 (>4500 sequences), despite the
591 greater diversity of marine diatoms. Perhaps not surprisingly, therefore, many of our ASVs were not
592 assigned to a species, or even a genus, by the Bayesian classifier. The contrast with freshwater
593 biomonitoring analyses is illustrated in Supplementary Table 7 which shows (for the marine
594 samples and two campaigns of metabarcoding in Catalan rivers) the proportions of the ASVs that
595 find exact matches with reference sequences or matches at 95–99% similarity levels: the marine
596 analysis lags well behind.

597 The incompleteness of the reference database explained c. 16% of dissimilarities between
598 methods and some of the species missed are known for being important components in the epizoic
599 and epiphytic diatom communities, so they could be considered priorities for future barcoding.
600 Among them are several *Navicula* species, including *N. normaloides, N. normalis* and *N. subagnita*;
601 these taxa were identified in all or most of the samples, indicating their importance in the study
602 area. *N. normaloides* and *N. subagnita* have also been recorded as epiphytic on leaves from
603 *Posidonia oceanica* and *Caulerpa taxifolia* in the Adriatic Sea (Kanjer et al., 2019; Car et al.,
604 2019). It is important to emphasize, however, that in many cases the lack of representative
605 sequences only partially prevents interpretation of metabarcoding data, though it reduces the
606 resolution achieved. For example, despite the lack of representative sequences for the particular
607 *Navicula* species known (from LM) to occur in our samples, the coverage of the genus in the

reference dataset was sufficient for the classifier to assign many ASVs at genus level. These assignments could then be checked by individual blastn searches and can be examined further in future by phylogenetic approaches, as with the non-diatom ASVs.

Other important species underestimated by DNA metabarcoding due to lack of reference sequences were *Cocconeis* species, notably *C. scutellum* and *C. scutellum* var. *posidoniae*. Both are cosmopolitan taxa and important components of the attached diatom communities (e.g., De Stefano et al., 2008, Polifrone et al., 2020; Ryabushko & Ryabushko, 2000; Witak et al., 2020). Overall, the genus *Cocconeis* was very poorly represented by DNA metabarcoding despite the very high diversity of *Cocconeis* species revealed by LM. Furthermore, and contrary to what we found with *Navicula* species, only a small proportion of reads and ASVs unclassified at the species level by the Bayesian classifier could be convincingly related to *Cocconeis* even at genus level. This can be explained by the fact that the reference database contains few *Cocconeis* (and almost all of them are freshwater species whose relationship to the marine species remains unknown), making it impossible for the classifier to assign ASVs to *Cocconeis* or related genera at any level. A few ASVs were tentatively identified as possible *Cocconeis* or Cocconeidaceae species on the basis of the spread of hits from blastn interrogation of GenBank, but overall it seems that the reference library is currently the main limitation to study *Cocconeis* diversity by DNA metabarcoding. Due to the importance of these diatoms in marine attached communities, further efforts should be made to increasing their representation in the DNA reference library. A further and more worrying possibility is that some *Cocconeis* taxa may carry mutations in critical parts of one or both primer regions, but this too cannot be known without long reference sequences of the marine species. In contrast to *Cocconeis*, genera like *Pseudo-nitzschia*, *Haslea* and *Achnanthes* are well represented in the DNA reference database.

Finally, some discrepancies between methods can probably be attributed to variation among species in the *rbcL* copy number per cell, as noted previously by Vasselon et al. (2018), Kelly et al.

633 (2020) and Pérez-Burillo et al. (2020). This variation depends on the number of gene copies per
634 chloroplast and the number of chloroplasts per cell. A correlation between the *rbcL* copy number
635 and cell biovolume has been reported, leading to much higher relative abundances for high
636 biovolume species, e.g., *Ulnaria ulna*, large *Pinnularia* or *Pleurosira laevis*, in metabarcoding
637 outputs (Vasselon et al., 2018). This very likely explains the higher abundances obtained by the
638 DNA method for *Achnanthes longipes* and *Pleurosigma*. These taxa are characterized by high
639 biovolume and either high numbers of chloroplasts per cell (*A. longipes*) or highly complex, large
640 chloroplasts (*Pleurosigma*).

641 **Concluding remarks**

642 As mentioned earlier, diatoms can contribute well over 50% of primary benthic production in
643 marine habitats where the only obvious photosynthetic organisms are seagrasses (Cox et al., 2020),
644 while on apparently bare sediments lacking macrophytes, diatoms generally dominate, except in
645 summer when cyanobacteria are often important (e.g., Scholz & Liebezeit, 2012b). Furthermore,
646 individual diatom species, including species that coexist, can exhibit different responses to
647 macronutrients (e.g., Underwood & Provot, 2000), different vertical migration patterns (Underwood
648 et al., 2005), and different seasonality (e.g., Scholz & Liebezeit, 2012a). Hence, to understand
649 marine benthic communities, it is important to identify and quantify species, and hence to have
650 resources that facilitate consistent accurate identification.

651 The main aim of this paper was to use a small dataset to examine the advantages and disadvantages
652 of metabarcoding and morphological approaches to study the benthic diatom communities of
653 shallow coastal environments. Our results show that both approaches are more difficult to
654 implement than in freshwater environments and in both cases the cause is essentially the same:
655 marine microphytobenthic communities have been greatly understudied, despite their ecological
656 and economic importance. As a result, the traditional morphology-based taxonomy has yet to
657 advance to the level achieved for freshwaters, while the lack of reference sequences limits the

658 resolution achievable with metabarcoding, though this did not prevent the molecular method from
659 separating the samples according to the type of substratum. There are also special features of the
660 marine benthos – such as the presence of a wider range of related microalgal groups – that offer
661 extra opportunities for studying non-diatom diversity but also mean that the reference database
662 needs to be more inclusive than in freshwaters for efficient identification of ASVs. Clearly, then,
663 both approaches, morphological and metabarcoding, are in some senses incomplete for marine
664 benthic diatom communities, but together they offer a strong foundation for ecological and
665 biogeographical studies. We suggest that the way forward, for the moment, is to develop
666 metabarcoding and morphological approaches in parallel and exploit their particular strengths and
667 complementarity: for example, the far greater resolution and sensitivity of metabarcoding (and the
668 albeit limited capacity to detect non-diatom components), combined with the insights into life-form,
669 cell surface area: volume relationships and functional group membership that are inherent in the
670 morphological approach and can never be fully realized with metabarcoding, even when the
671 reference database is complete and allows all ASVs to be allocated to known species.

672

673 **Acknowledgments**

674 We acknowledge the Erasmus+ programme of the European Union which supported the traineeship
675 of Greta Valoti (Università Politecnica delle Marche, Italy) in IRTA. J. Pérez-Burillo acknowledges
676 IRTA and the Universitat Rovira i Virgili for his Martí Franqués PhD grant (2018PMF-PIPF-22).
677 The Royal Botanic Garden Edinburgh is supported by the Scottish Government's Rural and
678 Environment Science and Analytical Services Division. This article was facilitated by COST Action
679 DNAqua-Net (CA15219), supported by the COST (European Cooperation in Science and
680 Technology) program. We also acknowledge support from the CERCA Programme/ Generalitat de
681 Catalunya. P Prado was contracted under the INIA-CCAA cooperative research program for
682 postdoctoral incorporation from the Spanish National Institute for Agricultural and Food Research

683 and Technology (INIA). The authors would like to thank the Biodiversity Foundation from the
684 Spanish Ministry for Ecological Transition for providing additional support for fieldwork sampling
685 in the context of the Recupera Pinna project. Prof. Rafał J. Wróbel (West Pomeranian University of
686 Technology) is acknowledged for his help with SEM examination of the diatomaceous samples. G.
687 Valoti acknowledges Prof. Cecilia Totti (Università Politecnica delle Marche) for her support with
688 respect to GV's IRTA visit. The authors would like to thank Vanessa Castan, IRTA technician, who
689 collected the individuals of *Crassotrea gigas*. The authors state no conflicts of interest.

690

691 References

692 Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C.,
693 Brown, M.W., Burki, F., Cárdenas, P., Čepička, I., Chistyakova, L., del Campo, J., Dunthorn,
694 M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., Heiss, A.A., Hoppenrath, M., James,
695 T.Y., Karnkowska, A., Karpov, S., Kim, E., Kolisko, M., Kudryavtsev, A., Lahr, D.J.G., Lara,
696 E., Le Gall, L., Lynn, D.H., Mann, D.G., Massana, R., Mitchell, E.A.D., Morrow, C., Park,
697 J.S., Pawlowski, J.W., Powell, M.J., Richter, D.J., Rueckert, S., Shadwick, L., Shimano, S.,
698 Spiegel, F.W., Torruella, G., Youssef, N., Zlatogursky, V. & Zhang, Q. (2019). Revisions to
699 the classification, nomenclature, and diversity of eukaryotes. *Journal of Eukaryotic*
700 *Microbiology* 66: 4–119. https://doi.org/10.1111/jeu.12691

701 An, S.M., Choi, D.H. & Noh, J.H. 2020. High-throughput sequencing analysis reveals dynamic
702 seasonal succession of diatom assemblages in a temperate tidal flat. *Estuar. Coast. Shelf Sci*.
703 237. https://doi.org/10.1016/j.ecss.2020.106686

704 Andersen, R.A., Potter, D. & Bailey, J.C. 2002. *Pinguiococcus pyrenoidosus* gen. et sp. nov.
705 (Pinguiophyceae), a new marine coccoid alga. *Phycol. Res*. 50: 57-65.
706 https://dx.doi.org/10.1111/j.1440-1835.2002.tb00136.x

Andrews, G.W. 1972. Some fallacies of quantitative diatom paleontology. *Nova Hedwigia, Beih.* 39: 285–295.

Andriana, R., Engel, F.G., Gusmao, J.B., & Eriksson, B.K. 2021. Intertidal mussel reefs change the composition and size distribution of diatoms in the biofilm. *Mar. Biol*. 168: 24. https://doi.org/10.1007/s00227-020-03819-2

Bailet, B., Bouchez, A., Franc, A., Frigerio, J.-M., Keck, F., Karjalainen, S.-M., Rimet, F., Schneider, S. & Kahlert, M. 2019. Molecular versus morphological data for benthic diatoms biomonitoring in Northern Europe freshwater and consequences for ecological status. *Metabarcoding and Metagenomics* 3: 21–35. https://doi.org/10.3897/mbmg.3.34002

Barille, L., Le Bris, A., Méléder, V., Launeau, P., Robin, M., Louvrou, I. & Ribeiro, L. 2017. Photosynthetic epibionts and endobionts of Pacific oyster shells from oyster reefs in rocky versus mudflat shores. *PLoS ONE* 12: e0185187. https://doi.org/10.1371/journal.pone.0185187

Belcher J.H. & Swale E.M.F. 1986. Notes on some small *Thalassiosira* species (Bacillariophyceae) from the plankton of the lower Thames and other British estuaries (identified by transmission electron microscopy). *Br. Phycol. J*. 21: 139–145. https://doi.org/10.1080/00071618600650161

Benito, X., Trobajo, R. & Ibáñez, C. 2015. Benthic diatoms in a Mediterranean delta: ecological indicators and a conductivity transfer function for paleoenvironmental studies. *J. Paleolimnol*. 54: 171–188. https://doi.org/10.1007/s10933-015-9845-3

Benito, X., Trobajo, R., Cearreta, A. & Ibáñez, C. 2016. Benthic foraminifera as indicators of habitat in a Mediterranean delta: implications for ecological and palaeoenvironmental studies. *Estuar. Coast. Shelf Sci*. 180: 97–113. https://doi.org/10.1016/j.ecss.2016.06.001

730 Bombin, S., Wysor, B. & Lopez-Bautista, J.M. 2021. Assessment of littoral algal diversity from the
731 northern Gulf of Mexico using environmental DNA metabarcoding. *J. Phycol*. 57: 269–278.
732 https://doi.org/10.1111/jpy.13087

733 Cahoon, L.B. 1999. The role of benthic microalgae in neritic ecosystems. *Oceanogr. Mar. Biol.*
734 *Ann. Rev*. 37: 47–86. https://doi.org/10.1201/9781482298550-4

735 Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. & Holmes, S.P. 2016.
736 DADA2: High resolution sample inference from Illumina amplicon data. *Nat. Methods*. 13:
737 581–583. https://doi.org/10.1038/nmeth.3869

738 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J. & Bealer, K. 2009. BLAST
739 plus: architecture and applications. *BMC Bioinform*. 10: 421. https://doi.org/10.1186/1471-
740 2105-10-421

741 Camps-Castellà, J., Romero, J. & Prado P. 2020. Trophic plasticity in the sea urchin Paracentrotus
742 lividus, as a function of resource availability and habitat features. *Mar Ecol Prog Ser*. 637:71-
743 85. https://doi.org/10.3354/meps13235

744 Car, A., Witkowski, A., Dobosz, S., Jasprica, N., Ljubimir, S. & Zgłobicka, I. 2019. Epiphytic
745 diatom assemblages on invasive *Caulerpa taxifolia* and autochthonous *Halimeda tuna* and
746 *Padina* sp. seaweeds in the Adriatic Sea – summer/autumn aspect. *Oceanol Hidrobiol Stud.*
747 48: 209–226. https://doi.org/10.2478/ohs-2019-0019

748 Carballeira, R., Trobajo, R., Leira, M., Benito, X., Sato, S. & Mann, D.G. 2017. A combined
749 morphological and molecular approach to *Nitzschia varelae* sp. nov., with discussion of
750 symmetry in Bacillariaceae. *Eur. J. Phycol*. 52: 342–359, DOI:
751 https://doi.org/10.1080/09670262.2017.1309575

752 Chonova, T., Keck, F., Bouchez, A. & Rimet, F. 2020. A ready-to-use database for DADA2:
753 Diat.barcode_rbcL_263bp_DADA2 based on Diat.barcode v9. Portail Data INRAE, V2.
754 https://doi.org/10.15454/QBLSXP

755 Cloern, J.E., Foster, S.Q. & Kleckner, A.E. 2013. Review: phytoplankton primary production in the
756 world's estuarine-coastal ecosystems. *Biogeosci. Discuss*. 10:17725–17783.
757 https://doi.org/10.5194/bg-11-2477-2014

758 Cox, T.E., Cebrian, J., Tabor, M., West, L. & Krause, J.W. 2020. Do diatoms dominate benthic
759 production in shallow systems? A case study from a mixed seagrass bed. *Limnol Oceanogr*. 5:
760 425-434. https://doi.org/10.1002/lol2.10167

761 Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S., Kubiszewski, I., Farber, S. &
762 Turner, R. 2014. Changes in the global value of ecosystem services. *Global Environ. Change*
763 26:152–158. https://doi.org/10.1016/j.gloenvcha.2014.04.002

764 D'Alelio, D., Cante, M.T., Russo, G.F., Totti, C. & De Stefano, M. 2011. Epizoic diatoms on
765 gastropod shells: when substrate complexity selects for microcommunity complexity. *In*
766 Dubinsky, Z. & Seckbach, J. [eds.] *All Flesh Is Grass. Plant-animal interrelationships*.
767 Springer, Netherlands, pp. 349–364.

768 Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S.,
769 Bista, I., Lodge, D.M., Vere, N., Pfrender, M.E. & Bernatchez, L. 2017. Environmental DNA
770 metabarcoding: Transforming how we survey animal and plant communities. *Mol. Ecol*. 26:
771 5872–5895. https://doi.org/10.1111/mec.14350

772 De Luca, D., Piredda, R., Sarno, D. & Kooistra, W.H.C.F. 2021. Resolving cryptic species
773 complexes in marine protists: phylogenetic haplotype networks meet global DNA
774 metabarcoding datasets. *ISME J*. https://doi.org/10.1038/s41396-021-00895-0

De Stefano, M., Romero, O.E. & Totti, C. 2008. A comparative study of *Cocconeis scutellum* Ehrenberg and its varieties (Bacillariophyta). *Bot. Mar*. 51: 506–536. https://doi.org/10.1515/BOT.2008.058

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res*. 32: 1792–1797. https://doi.org/10.1093/nar/gkh340

Edler, D., Klein, J., Antonelli, A. & Silvestro, D. 2021. raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods Ecol. Evol.* 12: 373– 377. https//doi.org/10.1111/2041-210X.13512

Facca, C.A. & Sfriso, A. 2007. Epipelic diatom spatial and temporal distribution and relationship with the main environmental parameters in coastal waters. *Estuar. Coast. Shelf Sci*. 75: 35–49. https://doi.org/10.1016/j.ecss.2007.03.033.

Gomaa, F., Utter, D.R., Powers, C., Beaudoin, D.J., Edgcomb, V.P., Filipsson, H.L., Hansel, C.M., Wankel, 7:S.D., Zhang, Y. & Bernhard, J.M. 2021. Multiple integrated metabolic strategies allow foraminiferan protists to thrive in anoxic marine sediments. *Sci. Adv*. 7: eabf1586. https://doi.org/10.1126/sciadv.abf1586

Gardner, C & Crawford, R.M. 1992. A description of the diatom Papiliocellulus simplex sp. nov. (Cymatosiraceae, Bacillariophyta) using light and electron microscopy. *Phycologia*. 31: 246-252. https://doi.org/10.2216/i0031-8884-31-3-4-246.1

Graf, L., Yang, E.C., Han, K.Y., Küpper, F.C., Benes, K.M., Oyadomari, J.K., Herbert, R.J.H., Verbruggen, H., Wetherbee, R., Andersen, R.A. & Yoon, H.S. 2020. Multigene phylogeny, morphological observation and re-examination of the literature lead to the description of the Phaeosacciophyceae classis nova and four new species of the Heterokontophyta SI clade. *Protist* 171: 125781. https://doi.org/10.1016/j.protis.2020.125781

798 Grant, D.M., Brodnicke, O.B., Evankow, A.M., Ferreira, A.O., Fontes, J.T., Hansen, A.K.,
799 Jensen,,M.R., Kalaycı, T.E., Leeper, A., Patil,,S.K., Prati, S., Reunamo, A., Roberts, A.J.,
800 Shigdel, R., Tyukosova, V., Bendiksby, M., Blaalid, R., Costa, F.O., Hollingsworth, P.M.,
801 Stur, E. & Ekrem, T. 2021. The future of DNA barcoding: reflections from early career
802 researchers. *Diversity* 2021, 13: 313. https://doi.org/10.3390/d13070313

803 Guiry, G.M. 2021a. *Thalassiosira profunda*. In *AlgaeBase* (Ed. by M.D Guiry & G.M. Guiry).
804 World-wide electronic publication, National University of Ireland, Galway.
805 http://www.algaebase.org; searched on 16 July 2021.

806 Inoue, T. & Taniguchi, A. 1999. Seasonal distribution of vegetative cells and resting spores of the
807 arcto-boreal diatom *Thalassiosira nordenskioeldii* Cleve in Onagawa Bay, northeastern Japan.
808 In Mayama, I. & Koizumi, I. [eds.] *Proceedings of the 14th International Diatom Symposium*.
809 Tokyo. pp. 263–276.

810 Jauffrais, T., LeKieffre, C., Schweizer, M., Geslin, E., Metzger, E., Bernhard, J.M., Jesus, B.,
811 Filipsson, H.L., Mare, O. & Meiborn, A. 2019. Kleptoplastidic benthic formanifera from
812 aphotic habitats: insights into assimilation of inorganic C, N and S studied with sub-cellular
813 resolution. *Env. Microbiol*. 21: 125–141. https://doi.org/10.1111/1462-2920.14433

814 Jeunen, G-J., Knapp, M., Spencer, H.G., Lamare, M.D., Taylor, H.R. Stat, M., Bunce, M. &
815 Gemmell, N.J. 2018. Environmental DNA (eDNA) metabarcoding reveals strong
816 discrimination among diverse marine habitats connected by water movement. *Mol. Ecol.*
817 *Resour*. 19: 426– 438. https://doi.org/10.1111/1755-0998.12982

818 Kanjer, L., Mucko, M., Car, A. & Bosak, S. 2019. Epiphytic diatoms on *Posidonia oceanica* (L.)
819 Delile leaves from eastern Adriatic Sea. *Nat. Croat.* 28: 1–20.
820 https://doi.org/10.20302/NC.2019.28.1

821 Kelly, M.G., Juggins, S., Mann, D.G., Sato, S., Glover, R., Boonham, N., Sapp, M., Lewis, E.,
822 Hany, U., Kille, P., Jones, T. & Walsh, K. 2020. Development of a novel metric for
823 evaluating diatom assemblages in rivers using DNA metabarcoding. *Ecol. Indic*. 118: 106725.
824 https://doi.org/10.1016/j.ecolind.2020.106725

825 Kermarrec, L., Franc, A., Rimet, F., Chaumeil, P., Frigerio, J.M., Humbert, J.F. & Bouchez, A.
826 2014. A next-generation sequencing approach to river biomonitoring using benthic diatoms.
827 *Freshw. Sci*. 33: 349–363. https://doi.org/10.1086/675079.

828 Krawczyk, D.W., Witkowski, A., Wroniecki, M., Waniek, J., Kurzydłowski, K.J. & Płociński, T.
829 2012. Reinterpretation of two diatom species from the West Greenland margin —
830 *Thalassiosira kushirensis* and *Thalassiosira antarctica* var. *borealis* – hydrological
831 consequences. *Mar. Micropaleontol*. 88–89: 1–14.
832 https://doi.org/10.1016/j.marmicro.2012.02.004.

833 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. 2018. MEGA X: Molecular Evolutionary
834 Genetics Analysis across computing platforms. *Mol. Biol. Evol*. 35: 1547–1549.
835 https://doi.org/10.1093/molbev/msy096

836 Kuwata, A. & Takahashi, M. 1999. Survival and recovery of resting spores and resting cells of the
837 marine planktonic diatom *Chaetoceros pseudocurvisetus* under fluctuating nitrate condition.
838 *Mar. Biol*. 134: 471–478. https://doi.org/10.1007/s002270050563

839 Lee, J.J. 2011. Diatoms as endosymbionts. In Seckbach, J & Kociolek, P. [eds.] *The diatom world.*
840 *Cellular Origin, Life in Extreme Habitats and Astrobiology. Springer,* Dordrecht. pp. 439 – 464.
841 https://doi.org/10.1007/978-94-007-1327-7_20

842 Letunic, I. & Bork, P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
843 display and annotation. *Nucl. Acids Res.* 49: W293–W296,
844 https://doi.org/10.1093/nar/gkab301

845 Lewitus, A.J., Brock, L.M., Burke, M.K., DeMattio, K.A. & Wilde, S.B. 2008. Lagoonal
846 stormwater detention ponds as promoters of harmful algal blooms and eutrophication along
847 the South Carolina coast. *Harmful Algae*. 8: 60–65. https://doi.org/10.1016/j.hal.2008.08.012

848 Li, Ch-L., Witkowski, A., Ashworth, M.P., Dąbek, P., Sato, S., Zgłobicka, I., Witak, M., Khim, J.S.
849 & Kwon, C.J. 2018. The morphology and molecular phylogenetics of some marine diatom
850 taxa within the Fragilariaceae, including twenty undescribed species and their relationship to
851 *Nanofrustulum*, *Opephora* and *Pseudostaurosira*. *Phytotaxa.* 355: 1–104.
852 https://doi.org/10.11646/phytotaxa.355.1.1

853 Li, Y., Zhao, Q. & Lü, S. 2013. The genus *Thalassiosira* off the Guangdong coast, South China
854 Sea. *Bot. Mar*. 56: 83-110. https://doi.org/10.1515/bot-2011-0045

855 Llebot, C., Solé, J., Delgado, M., Fernández-Tejedor, M., Campa, J. & Estrada, M. 2011.
856 Hydrographical forcing and phytoplankton variability in two semi-enclosed estuarine bays. *J.*
857 *Mar. Syst*. 86: 69–86. https://doi.org/10.1016/j.jmarsys.2011.01.004

858 Lobban, C.S., Schefter, M., Jordan, R.W., Arai, Y., Sasaki, A., Theriot, E.C., Ashworth, M., Ruck,
859 E. & Chiara, P. 2012. Coral-reef diatoms (Bacillariophyta) from Guam: New records and
860 preliminary checklist, with emphasis on epiphytic species from farmer-fish territories.
861 *Micronesica*. 43: 237–479.

862 Mabrouk, L., Ben Brahim, M., Hamza, A., Mahfoudhi, M. & Bradai, M.N. 2014. A comparison of
863 abundance and diversity of epiphytic microalgal assemblages on the leaves of the seagrasses
864 *Posidonia oceanica* (L.) and *Cymodocea nodosa* (Ucria) Asch in Eastern Tunisia. *J. Mar. Biol.*
865 2014: 1–10. https://doi.org/10.1155/2014/275305

866 MacIntyre, H.L., Geider, R.J. & Miller, D.C. 1996. Microphytobenthos: the ecological role of the secret
867 garden of unvegetated, shallow-water marine habitats. I. Distribution, abundance and primary
868 production. *Estuaries* 12:186–201. https://doi.org/10.2307/1352224

869 Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, J., Ludicone,
870 D., De Vargas, C., Bittner, L., Zingone, A. & Bowler, C. 2016. Insights into global diatom
871 distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci*. 113:1516–1525.
872 https://doi.org/10.1073/pnas.1509523113

873 Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.
874 *EMBnet. J*. 17: 10–12. https://doi.org/10.14806/ej.17.1.200

875 Mann, D.G., Crawford, R.M. & Round, F.E. 2016. Bacillariophyta. In: Archibald, J.M., Simpson,
876 A.G.B., Slamovits, C.H., Margulis, L., Melkonian, M., Chapman, D.J. & Corliss, J.O. [Eds.]
877 *Handbook of the Protists*. Springer, Cham, New York, pp. 1–62. https://doi.org/10.1007/978-
878 3-319-32669-6_29-1.

879 McMurdie, P.J. & Holmes, S. 2013. phyloseq: An R package for reproducible interactive analysis
880 and graphics of microbiome census data. *PLOS ONE* 8: e61217.
881 https://doi.org/10.1371/journal.pone.0061217

882 McQuoid, M.R. & Hobson, L.A. 1996. Diatom resting stages. *J. Phycol*. 32: 889–902.
883 https://doi.org/10.1111/j.0022-3646.1996.00889.x

884 Medlin, L.K. & Juggins, S. 2018. Multivariate analyses document host specificity, differences in
885 the diatom metaphyton vs. epiphyton, and seasonality that structure the epiphytic diatom
886 community. *Estuar. Coast. Shelf Sci*. 213: 314–330.
887 https://doi.org/10.1016/j.ecss.2018.06.011

888 Mortágua, A., Vasselon, V., Oliveira, R., Elias, C., Chardon, C., Bouchez, A., Rimet, F., João Feio,
889 M., & Almeida, S.F. 2019. Applicability of DNA metabarcoding approach in the
890 bioassessment of Portuguese rivers using diatoms. *Ecol. Indic*. 106: 105470.
891 https://doi.org/10.1016/j.ecolind.2019.105470.

892 Nunes, M., Lemley, D.A., Matcher, G.F. & Adams, J.B. 2021. The influence of estuary
893 eutrophication on the benthic diatom community: a molecular approach. *Afr. J. Mar. Sci*. 43:
894 171–186. https://doi.org/10.2989/1814232X.2021.1897039

895 Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin,
896 P.R., O'Hara, B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. & Wagner, H.
897 2020. Vegan: Community Ecology Package. R package, version 2.5-7. https://CRAN.R-
898 project.org/package=vegan

899 Park, J.S., Jung, S.W., Lee, J.H., Yun, S.M. & Lee, J.H. 2016. Species diversity of the genus
900 *Thalassiosira* (Thalassiosirales, Bacillariophyta) in South Korea and its biogeographical
901 distribution in the world. *Phycologia* 55: 403–423. https://doi.org/10.2216/15-66.1

902 Passy, S.I. 2007. Diatom ecological guilds display distinct and predictable behavior along nutrient
903 and disturbance gradients in running waters. *Aquat. Bot*. 86: 171–178.
904 https://doi.org/10.1016/j.aquabot.2006.09.018

905 Percopo, I., Siano, R., Cerino, F., Sarno, D. & Zingone, A. 2011. Phytoplankton diversity during the
906 spring bloom in the northwestern Mediterranean Sea. *Bot. Mar.* 54: 243–267.
907 https://doi.org/10.1515/bot.2011.033

908 Pérez-Burillo, J., Trobajo, R., Vasselon, V., Rimet, F., Bouchez, A. & Mann, D.G. 2020. Evaluation
909 and sensitivity analysis of diatom DNA metabarcoding for WFD bioassessment of
910 Mediterranean rivers. *Sci. Total Environ*. 727: 138445.
911 https://doi.org/10.1016/j.scitotenv.2020.138445

912 Pérez-Burillo, J., Trobajo, R., Leira, M., Keck, F., Rimet, F., Sigró, J. & Mann, D.G. 2021. DNA
913 metabarcoding reveals differences in distribution patterns and ecological preferences among
914 genetic variants within some key freshwater diatom species. *Sci. Total Environ*. 728: 149029.
915 https://doi.org/10.1016/j.scitotenv.2021.149029

916 Pillet, L., de Vargas, C. & Pawlowski, J. 2011. Molecular identification of sequestered diatom
917 chloroplasts and kleptoplastidy in foraminifera. *Protist* 162: 394–404.
918 https://doi.org/10.1016/j.protis.2010.10.001

919 Piredda, R., Claverie, J.-M., Decelle, J., De Vargas, C., Dunthorn, M., Edvardsen, B., Eikrem, W.,
920 Forster, D., Kooistra, W.H.C.F., Logares, R., Massana, R., Montresor, M., Not, F., Ogata, H.,
921 Pawlowski, J., Romac, S., Sarno, D., Stoeck, T. & Zingone, A. 2018. Diatom diversity
922 through HTS-metabarcoding in coastal European seas. *Sci. Rep.* 8: 18059.
923 https://doi.org/10.1038/s41598-018-36345-9

924 Plante, C.J., Hill-Spanik, K., Cook, M. & Graham, C. 2021a. Environmental and spatial influences
925 on biogeography and community structure of saltmarsh benthic diatoms. *Estuaries Coasts*.
926 44: 147–161. https://doi.org/10.1007/s12237-020-00779-0

927 Plante, C.J., Hill-Spanik, K. & Lowry, J. 2021b. Controls on diatom biogeography on South
928 Carolina (USA) barrier island beaches. *Mar. Ecol. Progr. Ser.* 661: 17–33. DOI:
929 https://doi.org/10.3354/meps13598

930 Polifrone, M., Viera-Rodríguez, M.A., Pennesi, C., Conte, M.T., Del Pino, A.S., Stroobant, M. &
931 De Stefano, M. 2020. Epiphytic diatoms on Gelidiales (Rhodophyta) from Gran Canaria
932 (Spain). *Eur. J. Phycol*. 55: 404–411. https://doi.org/10.1080/09670262.2020.1737967

933 Prado, P., Caiola, N. & Ibañez, C. 2014. Habitat use by a large population of *Pinna nobilis* in
934 shallow waters. *Sci. Mar*. 78: 555–565. https://doi.org/10.3989/scimar.04087.03A

935 Prado, P. 2018. Seagrass epiphytic assemblages are strong indicators of agricultural discharge but
936 weak indicators of host features. *Estuar. Coast. Shelf Sci*. 204: 140–148.
937 https://doi.org/10.1016/j.ecss.2018.02.026

938 Prado, P., Andree, K.B., Trigos, S., Carrasco, N., Caiola, N., García-March, J.R., Tena, J.,
939 Fernández-Tejedor, M. & Carella, F. 2020. Breeding, planktonic and settlement factors shape

940 recruitment patterns of one of the last remaining major population of *Pinna nobilis* within
941 Spanish waters. *Hydrobiologia*. 847: 771–786. https://doi.org/10.1007/s10750-019-04137-5

942 Prado, P., Grau, A., Catanese, G., Cabanes, P., Carella, F., Fernández-Tejedor, M., Andree, K.B.,
943 Añón, T., Hernandis, S., Tena, J., García-March, J.R. 2021. *Pinna nobilis* in suboptimal
944 environments are more tolerant to disease but more vulnerable to severe weather phenomena.
945 Mar. Environ. Res. 163: 105220. https://doi.org/10.1016/j.marenvres.2020.105220

946 Ramón, M., Cano, J., Peña, J.B., & Campos, M.J. 2005. Current status and perspectives of mollusc
947 (bivalves and gastropods) culture in the Spanish Mediterranean. Bol. Inst. Esp. Oceanogr. 21:
948 361–373.

949 Rimet, F. & A. Bouchez. 2012. Life-forms, cell-sizes and ecological guilds of diatoms in European
950 rivers. *Knowl. Manag. Aquat. Ecosyst*. 406: 1–14. doi:10.1051/kmae/2012018

951 Rimet, F., Vasselon, V., A.-Keszte, B. & Bouchez, A. 2018. Do we similarly assess diversity with
952 microscopy and high-throughput sequencing? Case of microalgae in lakes. Org. Divers. Evol.
953 18: 51–62. https://doi.org/10.1007/s13127-018-0359-5.

954 Rimet, F., Gusev, E., Kahlert, M., Kelly, M.G., Kulikovskiy, M., Maltsev, Y., Mann, D.G.,
955 Pfannkuchen, M., Trobajo, R., Vasselon, V., Zimmermann, J. & Bouchez, A. 2019.
956 Diat.barcode, an open-access curated barcode library for diatoms. *Sci.Rep.* 9: 1–12.
957 https://doi.org/10.1038/s41598-019-51500-6.

958 Rivera, S.F., Vasselon, V., Ballorain, K., Carpentier, A., Wetzel, C.E., Ector, L., Bouchez, A. &
959 Rimet, F. 2018. DNA metabarcoding and microscopic analyses of sea turtles biofilms:
960 complementary to understand turtle behavior. *PLoS ONE*, 13(4): e0195770.
961 https://doi.org/10.1371/journal.pone.0195770

962 Round, F.E. 1971. Benthic marine diatoms. *Oceanogr. Mar. Biol. Ann. Rev.* 9:83–139.

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

965 Rovira, L., Trobajo, R. & Ibañez, C. 2009. Periphytic diatom community in a Mediterranean salt
966 wedge estuary: the Ebro Estuary (NE Iberian Peninsula). *Acta Bot. Croat.* 68: 285–300.

967 Ryabushko, L.I. & Ryabushko, V.I. 2000. Communities of diatoms on the shells of mollusks of the
968 genus *Mytilus* L. *Int. J. Algae*. 2: 15–22. https://doi.org/10.1615/InterJAlgae.v2.i2.20

969 Schmidt, C., Morard, R., Romero, O. & Kucera, M. 2018. Diverse Internal Symbiont Community in
970 the Endosymbiotic Foraminifera *Pararotalia calcariformata*: Implications for Symbiont
971 Shuffling Under Thermal Stress. *Front Microbiol*. 9: 2018.
972 https://doi.org/10.3389/fmicb.2018.02018

973 Schmidt, M., Horn, S., Flieger, K., Ehlers, K., Wilhelm, C. & Schnetter, R. 2012. *Synchroma*
974 *pusillum* sp. nov. and other new algal isolates with chloroplast complexes confirm the
975 Synchromophyceae (Ochrophyta) as a widely distributed group of amoeboid algae. *Protist*.
976 163: 544–559. https://doi.org/10.1016/j.protis.2011.11.009

977 Scholz, B. & Liebezeit, G. 2012a. Microphytobenthic dynamics in a Wadden Sea intertidal flat –
978 Part I: seasonal and spatial variation of diatom commnities in relation to macronutrient
979 supply. *Eur. J. Phycol.* 47: 105–119. https://doi.org/10.1080/09670262.2012.663793

980 Scholz, B. & Liebezeit, G. 2012b. Microphytobenthic dynamics in a Wadden Sea intertidal flat –
981 Part II: seasonal and spatial variability of non-diatom community components in relation to
982 abiotic parameters. *Eur. J. Phycol.* 47: 120–137.
983 https://doi.org/10.1080/09670262.2012.665251

984 Stamatakis, A. 2014. Raxml version 8: a tool for phylogenetic analysis andpost-analysis of large
985 phylogenies. *Bioinformatics* 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stoof-Leichsenring, K.R., Pestryakova, L.A., Epp, L.S. & Herzschuh, U. 2020. Phylogenetic diversity and environment form assembly rules for Arctic diatom genera—A study on recent and ancient sedimentary DNA. *J. Biogeogr.* 47: 1166–1179. https://doi.org/10.1111/jbi.13786

Sugie, K. & Kuma, K. 2008. Resting spore formation in the marine diatom *Thalassiosira nordenskioeldii* under iron- and nitrogen-limited conditions, *J. Plankton. Res*. 30: 1245–1255, https://doi.org/10.1093/plankt/fbn080

Sundbäck, K. & Granéli, W. 1988. Influence of microphytobenthos on the nutrient flux between sediment and water: A laboratory study. *Mar. Ecol. Prog. Ser*. 43: 63–69. https://doi.org/10.3354/meps043063

Sundbäck, K., Enoksson, V., Granéli, W. & Pettersson, K. 1991. Influence of sublittoral microphytobenthos on the oxygen and nutrient flux between sediment and water: A laboratory continuous-flow study. *Mar. Ecol. Prog. Ser*. 74: 263–279. https://doi.org/10.3354/meps074263

Takano, Y., Hansen, G., Fujita, D. & Horiguchi, T. 2007. Serial replacement of diatom endosymbionts in two freshwater dinoflagellates, *Peridiniopsis* spp. (Peridiniales, Dinophyceae). *Phycologia* 47: 41–53. https://doi.org/10.2216/07-36.1

Totti, C., Romagnoli, T., De Stefano, M., Di Camillo, C.G. & Bavestrello, G. 2011. The diversity of epizoic diatoms: relationships between diatoms and marine invertebrates. *In* Dubinsky, Z. & Seckbach, J. [eds.] *All Flesh Is Grass. Plant-Animal Interrelationships*. Springer, Netherlands, pp. 327–343.

Triska, F.J. & Oremland, R.S. 1981. Denitrification associated with periphyton communities. *Appl. Environ. Microbiol*. 42: 745–748. https://doi.org/10.1128/aem.42.4.745-748.1981

Trobajo, R. & Sullivan, M.J. 2010. Applied diatom studies in estuaries and shallow coastal environments. In Smol, J.P. & Stoermer, E.F.[eds.] *The Diatoms: Applications for the*

1010 *Environmental and Earth Sciences*. Cambridge University Press, Cambridge, UK. pp 309–
1011 323.

1012 Trobajo, R., Quintana, X.D. & Sabater, S. 2004. Factors affecting the periphytic diatom community
1013 in Mediterranean coastal wetlands (Empordà wetlands, NE Spain). *Arch. Hydrobiol*. 160:
1014 375–399. https://doi.org/10.1127/0003-9136/2004/0160-0375

1015 Underwood, G.J.C. & Provot, L. 2000. Determining the environmental preferences of four epipelic
1016 diatom taxa: growth across a range of salinities, nitrate and ammonium conditions. *Eur. J.*
1017 *Phycol*. 35: 173–182. https://doi.org/10.1080/09670260010001735761

1018 Underwood, G.J.C., Perkins, R.G., Consalvey, M.C., Hanlon, A.R.M., Oxborough, K., Baker, N.R.
1019 & Paterson, D.M. 2005. Patterns in microphytobenthic primary productivity: Species-specific
1020 variation in migratory rhythms and photosynthetic efficiency in mixed-species biofilms.
1021 *Limnol. Oceanogr*. 50: 755–767. https://doi.org/10.4319/lo.2005.50.3.0755.

1022 Vasselon, V., Rimet, F., Tapolczai, K. & Bouchez, A. 2017. Assessing ecological status with
1023 diatoms DNA metabarcoding: scaling-up on aWFD monitoring network (Mayotte island,
1024 France). *Ecol. Indic.* 82: 1–12. https://doi.org/10.1016/j.ecolind.2017.06.024

1025 Vasselon, V., Bouchez, A., Rimet, F., Jacquet, S., Trobajo, R., Corniquel, M., Tapolczai, K. &
1026 Domaizon, I. 2018. Avoiding quantification bias in metabarcoding: application of a cell
1027 biovolume correction factor in diatom molecular biomonitoring. *Methods Ecol. Evol.* 9:
1028 1060–1069. https://doi.org/10.1111/2041-210X.12960

1029 Virta, L., Gammal, J., Järnström, M., Bernard, G., Soininen, J., Norkko, J. & Norkko, A. 2019. The
1030 diversity of benthic diatoms affects ecosystem productivity in heterogeneous coastal
1031 environments. *Ecology* 100: e02765. https://doi.org/10.1002/ecy.2765

1032 Wang, Q., Garrity, G.M., Tiedje, J.M. & Cole, J.R. 2007. Naïve Bayesian classifier for rapid
1033 assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol*.
1034 73: 5261–5267. https://doi.org/10.1128/AEM.00062-07

1035 Wetherbee, R., Bringloe, T.T., Costa, J.F., van de Meene, A., Andersen, R.A. & Verbruggen, H.
1036 2021. New pelagophytes show a novel mode of algal colony development and reveal a
1037 perforated theca that may define the class. *J. Phycol*. 57: 396–411.
1038 https://doi.org/10.1111/jpy.13074

1039 Witak, M., Pędziński, J., Olwa, S. & Hetko, D. (2020). Biodiversity of benthic diatom flora in the
1040 coastal zone of Puck Bay (southern Baltic Sea): a case study of the Hel Peninsula. *Oceanol.*
1041 *Hydrobiol. Studies* 49: 304–318. https://doi.org/10.1515/ohs-2020-0027

1042 Yamada, N., Bolton, J.J., Trobajo, R., Mann, D.G., Dąbek, P., Witkowski, A., Onuma, R.,
1043 Horiguchi, T. & Kroth, P.G. 2019. Discovery of a kleptoplastic 'dinotom' dinoflagellate and
1044 the unique nuclear dynamics of converting kleptoplastids to permanent plastids. *Sci. Rep*. 9:
1045 10474. https://doi.org/10.1038/s41598-019-46852-y

1046 Yamada, N., Sakai, H., Onuma, R., Kroth, P.G. & Horiguchi, T. 2020. Five non-motile dinotom
1047 dinoflagellates of the genus *Dinothrix*. *Front. Plant Sci*. 11: 1764.

1048 You, X., Luo, Z., Su, Y., Gu, L. & Gu, H. 2015. *Peridiniopsis jiulongensis*, a new freshwater
1049 dinoflagellate with a diatom endosymbiont from China. *Nova Hedwigia* 101: 313–326.
1050 https://doi.org/10.1127/nova_hedwigia/2015/0272

1051 Zgrundo, A., Lemke, P., Pniewski, F., Cox, E.J., Latala, A. 2013. Morphological and molecular
1052 phylogenetic studies on *Fistulifera saprophila*. *Diat. Res*. 28: 431–443. https://
1053 doi.org/10.1080/0269249X.2013.833136.

1054 Zimmermann, J., Glöckner, G., Jahn, R., Enke, N. & Gemeinholzer, B. 2015. Metabarcoding vs.
1055 morphological identification to assess diatom diversity in environmental studies. *Mol. Ecol.*
1056 *Resour.* 15: 526–542. https://doi.org/10.1111/1755-0998.12336.

1057

Figures

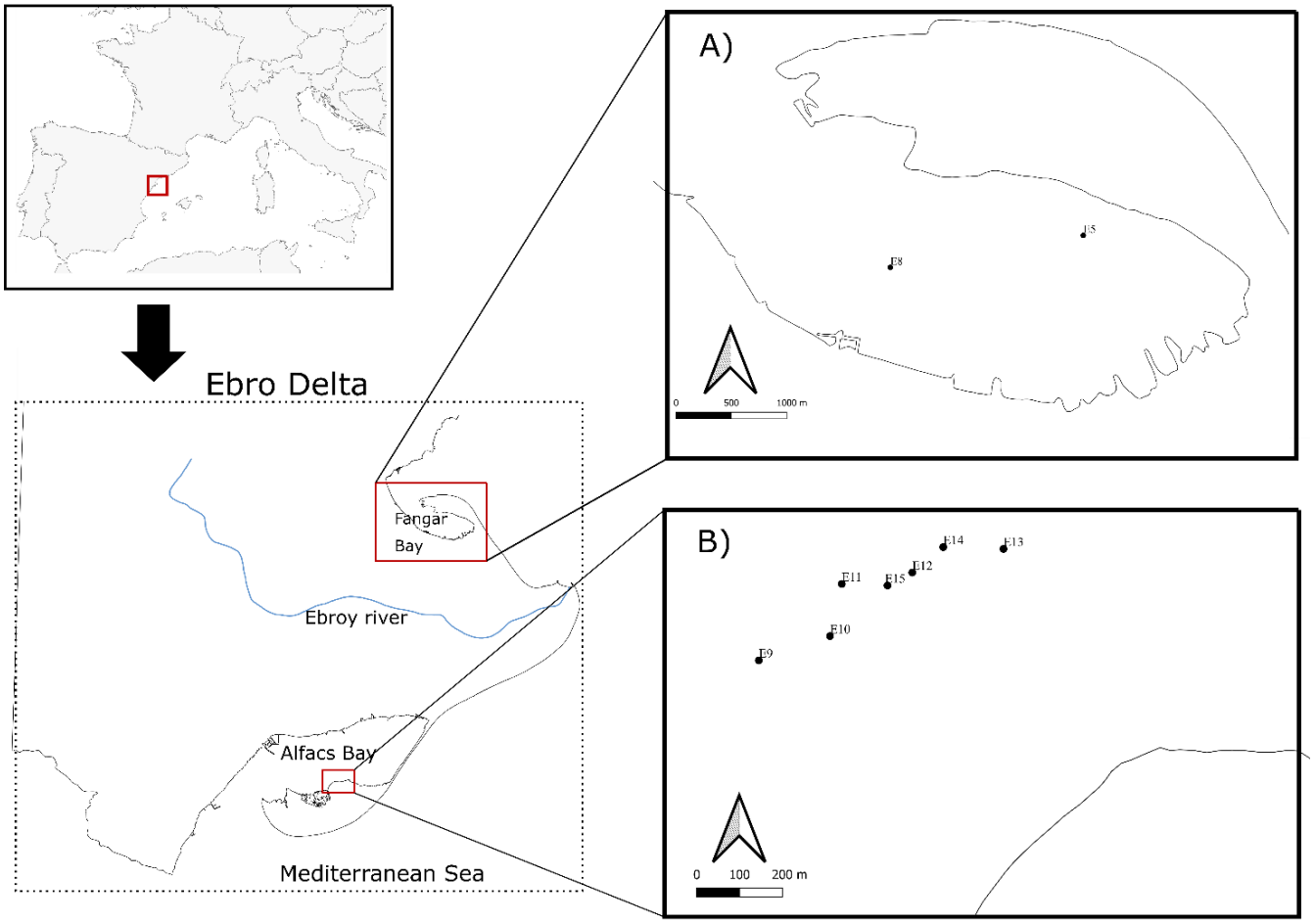


Fig 1. Location of Ebro Delta (NE of Spain) and samples sites of Fangar (A) and Alfacs (B) bays. 2 biofilms samples (E5 and E8) were taken from the surface of *Crassostrea gigas* individuals located in the Fangar Bay (A). 7 biofilms samples were taken from the surface of *Pinna nobilis* (E9, E11 and E12), *Cymodocea nodosa* (E14) and *Caulerpa prolifera* (E15) individuals located in the Alfacs Bay (B). The eediment samples (E10 and E13) were taken from the sediment adjacent to specimens of *P.nobilis* located in the Alfacs Bay (B).

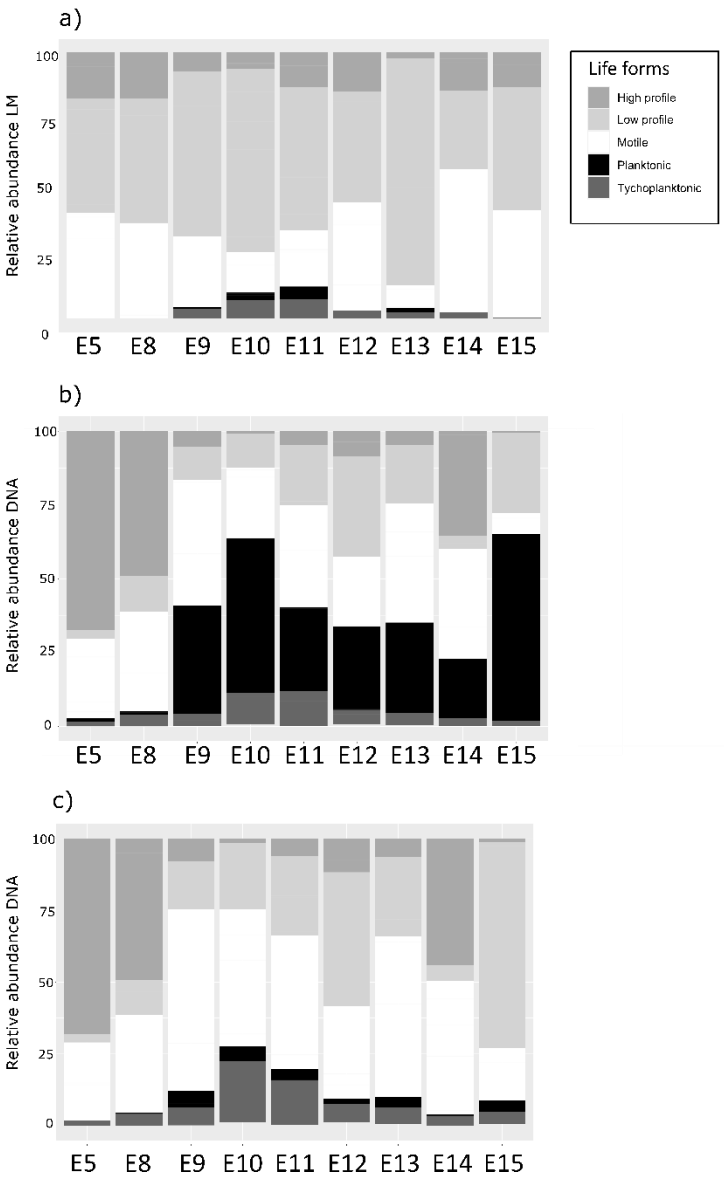
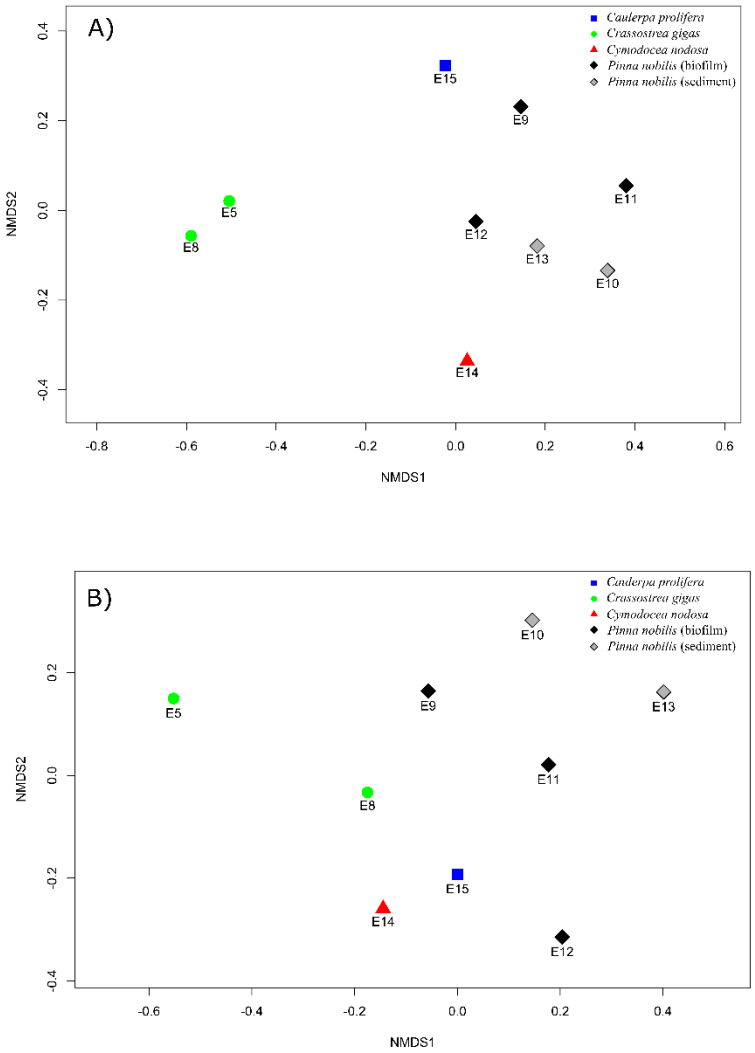


Fig 2. Relative abundance comparison of diatom growth forms between LM (a), DNA metabarcoding inventories (b) and DNA metabarcoding without considering the planktonic species *Thalassiosira profunda* (c).



1070

1071 Figure 3. Non-metric multidimensional scaling of the Bray-Curtis dissimilarity calculated on diatom ASVs
1072 relative abundance (A; stress=0.09) and relative abundance of diatom taxa identified by LM (B; stress=0.13).

1073

1074

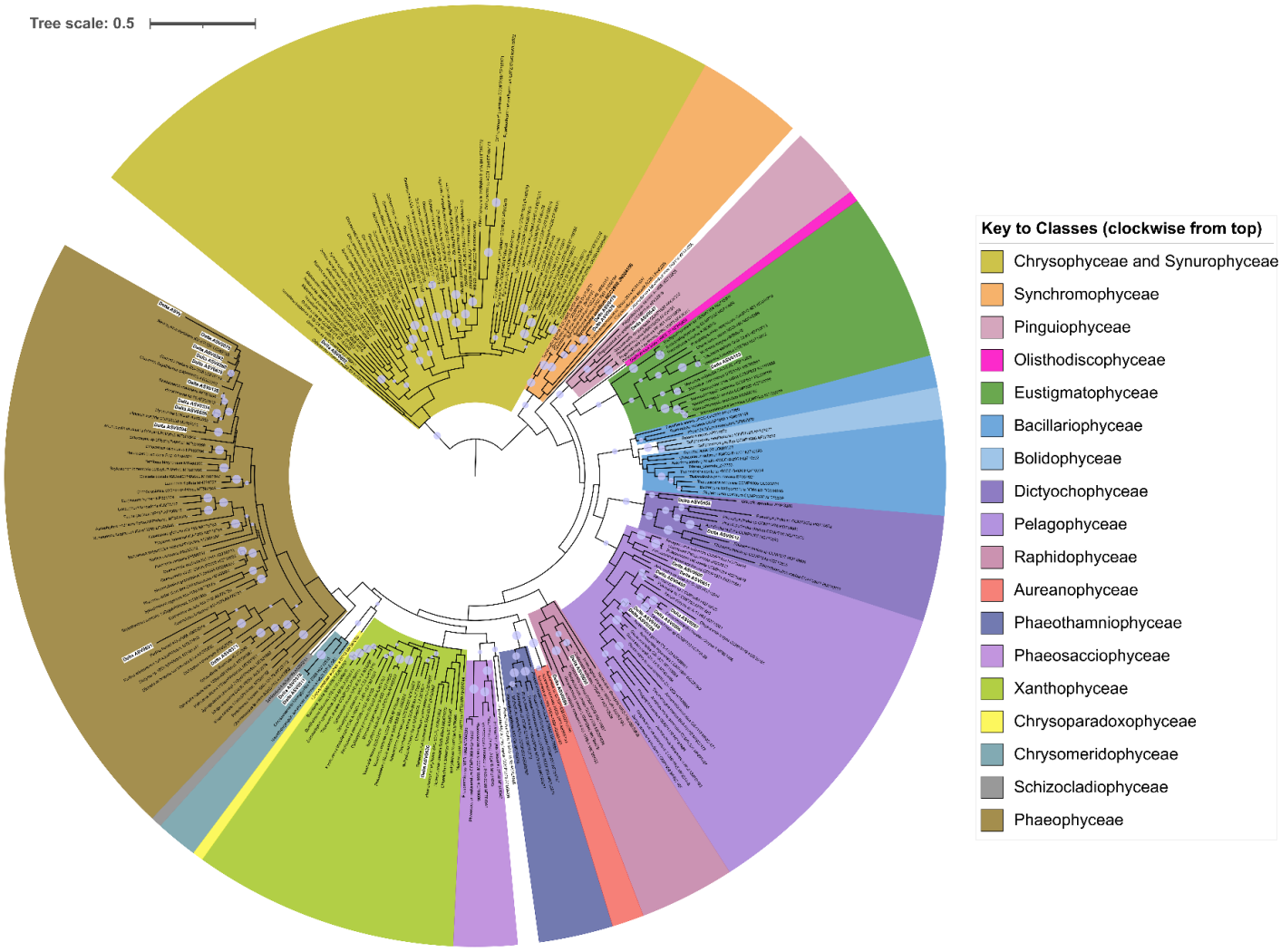


Fig 4. Maximum likelihood phylogenetic tree based on non-diatom ASVs related to different heterekont classes. *RbcL* representative sequences included in the tree were extracted from Graf et al. (2020), Watherbee et al. (2021) and GenBank database. The tree was built using using raxmlGUI on an alignment partitioned by codon position and setting the GRT-Gamma model with 1000 replicates for the bootstrap analyses. The tree was drawn using iTOL. Bootstrap support values from 70 to 100 are represented. ASVs from *rbcL* metabarcoding are highlighted by white boxes.

Tables

Table 1: Physico-chemical measurements registered in the different sampling sites at the time and date of sampling. Note that there are samples (E9 & E10; E12-E15) with the same physico-chemical data; this is because these samples were collected on the same date from the same small area, being separated from each other by distances in the order of 10s of meters.

Sample	Sampling date	Sampling time	Water temperature (ºC)	Salinity (g/L)	pH	Dissolved oxygen (mg/L)	% Dissolved oxygen (mg/L)
E5 - *Crassostrea gigas*	03/03/2020	12:35	11	35.87	7.95	7.49	85.1
E8 - *Crassostrea gigas*	10/03/2020	13:00	12.3	36.99	8.05	6.97	82.2
E9 - *Pinna nobilis* biofilm E10 - *Pinna nobilis* sediment	12/03/2020	12:15	14.9	36.5	8.11	7.79	96.7
E11 - *Pinna nobilis* biofilm	12/03/2020	14:30	16.6	37.27	8.24	8.65	111.6
E12 - *Pinna nobilis* biofilm E13 - *Pinna nobilis* sediment E14 - *Cymodocea nodosa* E15 - *Caulerpa prolifera*	12/03/2020	15:50	15.4	36.24	8.14	7.72	97.6

1090 Table 2. Comparison of taxa richness and Shannon diversity index values obtained for the LM and DNA
1091 metabarcoding methods.

1092

	Microscopy		DNA metabarcoding	
Sample	**Taxa richness**	**Shannon index**	**Taxa richness**	**Shannon index**
E5 - *Crassostrea gigas*	69	3.50	37	2.15
E8 - *Crassostrea gigas*	48	2.90	52	2.92
E9 - *Pinna nobilis* biofilm	44	2.81	37	2.43
E10 - *Pinna nobilis* sediment	75	3.73	47	2.23
E11 - *Pinna nobilis* biofilm	71	3.74	62	3.04
E12 - *Pinna nobilis* biofilm	67	3.46	55	3.04
E13 - *Pinna nobilis* sediment	72	3.50	57	2.88
E14 - *Cymodocea nodosa*	40	2.61	35	2.21
E15 - *Caulerpa prolifera*	44	3.34	25	1.59

1093

1094

Table 3. SIMPER analyses showing taxa contribution to the total dissimilarities between DNA metabarcoding and LM methods. Only the first thirty taxa with the greatest contribution to dissimilarities are shown. It is also indicated the taxa for which there are or not representative sequences in the reference library Diat.barcode v9.

Taxon	Relative abundance DNA metabarcoding	Relative abundance LM	Contribution to dissimilarities (%)	Cumulative Contribution to dissimilarities (%)	Availability of a representative sequence in Diat.barcode v9
Thalassiosira profunda	27.36	0	14.37	14.37	yes
Navicula sp.4	0	9.3	4.88	19.25	no
Amphora helenensis	1.74	9.08	4.27	23.52	yes
Achnanthes longipes	7.64	0.45	4.15	27.67	yes
Berkeleya fennica	7.17	2.02	3.51	31.18	yes
Nanofrustulum shiloi	5.17	2.61	2.97	34.15	yes
Navicula sp.	4.47	0	2.35	36.50	yes
Amphora cf helenensis	0	4.17	2.19	38.69	yes
Nitzschia spathulata	3.57	0	1.87	40.56	no
Cocconeis scutellum v. posidoniae	0	3.47	1.82	42.38	no
Navicula normaloides	0	2.89	1.52	43.90	yes
Haslea howeana	2.79	0	1.47	45.37	no
Cyclotella choctawhatcheeana	0.11	2.77	1.43	46.80	yes
Navicula normalis	0	2.69	1.41	48.21	no
Cocconeis scutellum	0	2.56	1.35	49.56	no
Cyclotella sp.	2.15	0	1.13	50.68	yes
Seminavis robusta	1.57	0.92	1.11	51.79	yes
Nitzschia sp.	2.02	0	1.06	52.86	yes
Pleurosigma sp.	1.94	0	1.02	53.88	yes
Halamphora sp.2	0	1.81	0.95	54.83	yes
Navicula perminuta	1.75	0.53	0.92	55.75	no
Serratifera sp.3	0	1.71	0.90	56.65	no
Plagiogramma minus	0	1.71	0.90	57.55	no
Seminavis cf. robusta	1.69	0	0.89	58.43	yes
Navicula subagnita	0	1.68	0.88	59.31	no
Pteroncola marina	0	1.43	0.75	60.06	yes
Craspedostauros constricta	1.42	0	0.74	60.81	yes
Psammodictyon sp.	1.39	0	0.73	61.54	no
Mastogloia crucicula	0	1.36	0.72	62.26	yes
Halamphora sp.	1.36	0	0.71	62.97	no