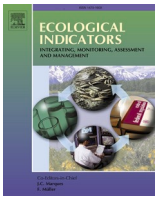
Ecological Indicators 138 (2022) 108827

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind



Original Articles

Using null models and species traits to optimize phytoplankton monitoring: An application across oceans and ecosystems

Leonilde Roselli [a], Stanislao Bevilacqua [b,c,*], Antonio Terlizzi [b,c,d]

[a] Research Institute on Terrestrial Ecosystems, National Research Council, 80131 Napoli, Italy
[b] Department of Life Sciences, University of Trieste, 34127 Trieste, Italy
[c] CoNiSMa, Piazzale Flaminio 9, 00196 Roma, Italy
[d] Stazione Zoologica Anton Dohrn, 80121 Napoli, Italy

ARTICLE INFO

Keywords:
Environmental monitoring and assessment
Multivariate analysis
Spatial patterns
BestAgg
Transitional waters
Marine ecosystems

ABSTRACT

Phytoplankton assemblages are privileged descriptors of the ecological status of marine ecosystems regularly included in routine monitoring programmes. The high spatial and temporal variability of phytoplankton and the intrinsic difficulties of species identifications, however, combine in making reiterate assessments of this component of marine biota particularly demanding. Coarse levels of taxonomic resolution (e.g., genus, family) or morpho-functional categories have been proposed to reduce identification efforts or to ease the analysis of phytoplankton assemblages for monitoring purposes, although with contrasting outcomes. A major issue is that, in the absence of control for the loss of information associated to these alternative approaches, their application may lead to poor representations of genuine spatial and/or temporal patterns of assemblages in relation to natural and anthropogenic sources of variation. We provided a new approach to reduce the efforts required to analyse phytoplankton assemblages that integrate morpho-functional classification of phytoplankton with the use of null models to estimate the consequent loss of information on species-level community patterns. Null models for information loss were built by randomly grouping the original species variables into a progressively decreasing number of groups, in order to identify the minimum number of aggregate variables needed to detect community patters as at species level. Aggregate variables were then defined as morpho-functional groups, by grouping species on the basis of a combination of morpho-functional traits, including general taxonomy, cell size, shape, elongation and complexity. We applied the approach to six case studies investigating the response of phytoplankton assemblages from marine and transitional water ecosystems under different environmental settings in areas spanning the world's ocean, including coral atolls, mangroves, estuaries, coastal lagoons and inlets. The approach allowed obtaining parsimonious sets of morpho-functional groups, which were suitable to detect changes in phytoplankton assemblage structure as at species level in all case studies. Trait-based approaches to phytoplankton research and monitoring are crucial to shed light on processes underlying phytoplankton community assembly and dynamics in the face of global change. In this perspective, our framework incorporates cost-effectiveness, instances from traditional monitoring programmes aiming at the detection of community patterns, and the current need for a deeper understanding of functional responses of phytoplankton to environmental drivers.

1. Introduction

Phytoplankton are a complex group of organisms essential to the functioning of marine ecosystems. They are of basic importance in sustaining marine food webs by providing the largest part of primary productivity of the world's ocean (Uitz et al., 2010), and largely contributing to planetary processes of oxygen production, biogeochemical cycles, climate regulation, and carbon sequestration (Vallina and Simó, 2007; Litchman et al., 2015; Culhane et al., 2018; Zhang et al., 2018). Due to their prominent ecological role, phytoplankton are considered an indispensable target in continuous global monitoring programmes (Batten et al., 2019), and have traditionally been included in routine assessments of environmental agencies worldwide as descriptors of the ecological status of aquatic ecosystems

* Corresponding author at: Dipartimento di Scienze della Vita, Università degli Studi di Trieste, Via Licio Giorgieri, 10 – 34127 Trieste, Italy.
E-mail address: sbevilacqua@units.it (S. Bevilacqua).

https://doi.org/10.1016/j.ecolind.2022.108827
Received 7 February 2022; Received in revised form 29 March 2022; Accepted 30 March 2022
Available online 2 April 2022

(Tweddle et al., 2018; Eriksen et al., 2019).

Monitoring phytoplankton is particularly demanding in terms of time and associated costs because of labour-intensive procedures for sample processing and taxonomic identifications. Phytoplankton assemblages are generally very speciose and exhibit high spatial and temporal heterogeneity (Hutchinson, 1961; Tweddle et al., 2018), thus requiring frequent assessments (typically every two–three months) at multiple sites during monitoring programmes, leading environmental agencies to routinely handle a large number of samples. Preservation does not ensure the integrity of all organisms in the long run, and samples have to be processed as soon as possible to avoid degradation of delicate species (Muñiz et al., 2020), requiring subsampling of water volumes, counting and identifying each cell at species level for many samples at a time. In most cases, fine taxonomic identifications entail a long time to prepare the material for subsequent analysis through traditional electron scanning microscopy and confocal microscopy techniques. *In vivo* observation and preparation of cultures could help the identification process (Tomas, 1997), but further increase time and costs for analysis. In addition, laboratories of environmental agencies deputed for monitoring often lack adequate instruments, and personnel, though carefully trained, has rarely sufficient expertise to conduct species-level identifications for all taxonomic groups (Domingues et al., 2008; Straile et al., 2013).

The limited availability of expert taxonomists and the need for speeding sample processing and species identifications has stimulated continuous efforts to implement rapid and cost-effective methods to monitor phytoplankton. Promising approaches to replace traditional procedures when quantifying the abundance and composition of phytoplankton assemblages exploit measurements of cell pigments (Havskum et al., 2004), or other unconventional techniques, like computer assisted taxonomy (Gaston and O'Neill, 2004) and flow cytometry (Sieracki et al., 1998; Sosik and Olson, 2007), although several issues on their application remain unsolved. For instance, wrong assignments of specimens to taxonomic groups and misinterpretation of suspended particles as phytoplankton cells may severely affect the results of such analyses (Dashkova et al., 2017). The use of sophisticated instruments and screening algorithms complicates these methods, which, moreover, often need complementary evaluations through conventional taxonomic identifications and counting of phytoplankton cells (Alvarez et al., 2014; Coupel et al., 2015).

Attempts to achieve profitable strategies for routine monitoring also explored the potential advantages of using coarse levels of taxonomic resolution in the analysis of phytoplankton assemblages (Cottingham and Carpenter, 1998; Heino and Soininen, 2007; Carneiro et al., 2010). Several studies on fish and aquatic invertebrates demonstrated that identifying organisms to taxonomic levels higher than species, and particularly genus and family-level identification, alleviates costs and time for analysis by reducing the number of operational taxonomic units (i.e. the taxonomic units to be identified) and the difficulties of their identification, while still allowing the detection of community responses as at species level (e.g., Ferraro and Cole, 1995; Thompson et al., 2003; Bates et al., 2007). For many phytoplankton groups, however, the identification of organisms at genus or even at family level is still too challenging and time-consuming to allow substantial benefits in routine monitoring, while coarser taxonomic resolution (e.g., class), although conducive to significant reductions of identification efforts, is often inadequate to capture ecological patterns as at species level (Heino and Soininen, 2007; Carneiro et al., 2010; Gallego et al., 2012). More generally, this approach of taxonomic sufficiency (Ellis, 1985) lacks of sound theoretical basis that could allow modelling information loss at decreasing taxonomic resolution and optimizing trade-offs associated to different taxonomic levels through a formal and transparent procedure (Jones, 2008; Mellin et al., 2011; Bevilacqua et al., 2012). The reduction in the number of operational taxonomic units strongly depends on the particular architecture of the taxonomic hierarchy of the group of organisms under study (Bertrand et al., 2006; Bevilacqua et al., 2009), and analysing higher taxa instead of species, regardless of their ecological importance or actual difficulty in their identification, could often lead to unnecessary loss of ecological information (Terlizzi et al., 2003; Jones, 2008; Groc et al., 2010).

Alternative ways to characterize phytoplankton assemblages rely on classifying species according to their similarity in habitat preference, tolerance and sensitivity to environmental variations (e.g., Reynolds et al., 2002; Padisák et al., 2009), or using morphological, physiological, phenological and behavioural features (traits) which are functionally relevant (Litchman and Klausmeier, 2008). All these functional classification schemes recognize the importance of key morphological traits, and particularly of cell size and shape, as 'master traits' to be considered when defining phytoplankton functional groups (Litchman and Klausmeier, 2008; Weithoff and Beisner, 2019; Ryabov et al., 2021). Approaches strictly relying on morphology have been implemented to obtain a small set of ecologically meaningful operational units of identification (e.g., morpho-functional groups) that, at the same time, can be easily identified and used by researchers, practitioners and environmental managers to investigate phytoplankton assemblages. For example, Kruk et al. (2010) proposed a classification of phytoplankton in seven main categories, namely the 'morphologically based functional groups', obtained by combining several morphological traits (e.g., presence of flagella, presence of siliceous structures, cell volume, surface-volume ratio). Yet, the application of this approach, or even of more complex functional classifications (e.g., Reynolds et al., 2002) are basically thought for freshwater phytoplankton species and, more importantly, may oversimplify the structure of phytoplankton assemblages leading to poor representations of spatial and/or temporal patterns in relation to natural and anthropogenic sources of variation (e.g., Carneiro et al., 2010; Wang et al., 2021).

Quantifying the structure of assemblages using any type of species categories implies condensing the information that would have been gained using all separate species variables into a set of 'aggregate' variables, in which species identities are lost and their relative abundances are summed. This necessarily entails a loss of detail which can be more or less prejudicial for the detection of community patterns, depending on the effectiveness of aggregate variables to comprise species with similar ecological responses to the investigated environmental and/or biological drivers and, above all, on the level of 'compression' the set of the original species variables is subjected when it is condensed into a new set of aggregate variables. The higher will be the level of compression, that is the lower will be the number of aggregate variables with respect to the number of the original species variables, the larger will be the ensuing loss of information on the original species-level community patterns (Bevilacqua et al., 2013). Evidence across a wide range of organisms, from terrestrial plants to freshwater and marine invertebrates and vertebrates, demonstrated this inverse relationship between the effectiveness of aggregate variables to reflect species-level community response and the level of compression of species variables (Bevilacqua et al., 2012; Mueller et al., 2013; Rosser, 2017). On this basis, determining to what extent the original species variables can be aggregated before a substantial loss of information occurs represents a crucial step to obtain a set of aggregate variables which are effective surrogates for species in detecting community patterns. Recent developments employed null models of species aggregation to find the 'best practicable aggregation of species' (BestAgg, Bevilacqua et al., 2013; Bevilacqua and Terlizzi, 2016) able to retain sufficient information on species-level community patterns, while reducing as much as possible the number of operational units of identification (i.e., the items to be identified, whatever their type). The procedure assumes that all species are ecologically equivalent and that the loss of information on the original species-level pattern is a pure effect of the aggregation of species variables. In this framework, a null model of information decay at increasing level of compression of species variables is built by randomly aggregating the original species variables into a decreasing number of aggregate variables (Bevilacqua et al., 2013). This allows

identifying the lowest number of aggregate variables needed for quantifying community patterns as at species level, under the null hypothesis that species are virtually interchangeable among groups. This threshold represents an objective reference to guide the experimenter when implementing a set of aggregate variables to be used as surrogates for species: if the number of surrogates is higher than the identified critical limit, then the set of surrogates will be able to reflect consistently the original species-level patterns (with a probability of failure that can be fixed a priori). It is worth noting that this will happen irrespective of the fact that the chosen aggregation criterion actually leads to group ecologically similar species, or that species within surrogates could exhibit incoherent, or even random, ecological responses to environmental changes (Bevilacqua and Terlizzi, 2016). To date, this method has been used to obtain suitable set of aggregate variables in order to quantify spatial and temporal patterns of marine and freshwater macroinvertebrates and fish assemblages in relation to anthropogenic and natural sources of variation (e.g., Milošević et al., 2014; Thiault et al., 2015; Jiang et al., 2017; Bevilacqua et al., 2018), although it is virtually applicable to any type of organisms and environmental settings.

Here, we proposed a new approach to reduce the efforts required to analyse phytoplankton assemblages in reiterated routine assessments by combining the rationale underlying the BestAgg procedure to the easiness of identifications that characterizes classical approaches based on morpho-functional aggregation of species. To demonstrate the generality of the method we used as pilot data a collection of datasets from six case studies on phytoplankton assemblages from marine and transitional water ecosystems worldwide, including coral reefs, mangroves, estuaries, coastal lagoons and inlets. Specifically, (1) we used null models of species aggregation from BestAgg to determine the smallest number of aggregate variables able to reflect the original species-level patterns in phytoplankton assemblages across habitats and marine regions. This threshold served as reference to (2) obtain parsimonious sets of easy-to-identify surrogates for species based on a classification scheme of phytoplankton relying on coarse taxonomy, cell shape and size traits. Finally, (3) we tested the effectiveness of these sets of surrogates to closely reflect species-level patterns of phytoplankton assemblages.

2. Materials and methods

2.1. Study areas and datasets

We analysed six datasets of phytoplankton assemblages from case studies in different habitats and geographic areas (Fig. 1), including transitional water and marine ecosystems. Two of them, SCO1 and SCO2, related to phytoplankton samples collected from brackish basins in the Orkney Islands (NE Atlantic).

Specifically, for SCO1 (73 taxa × 27 samples), phytoplankton assemblages were sampled along a gradient of confinement within a loch extending about 6 km, in order to test for difference in assemblage structure at increasing distance from the sea inlet. Samples were taken at the inlet, in the central and inner parts of the loch, with three sampling sites in each position along the gradient and $n = 3$ replicate samples in each site. SCO2 (116 taxa × 27 samples) referred to phytoplankton assemblages sampled in three small (~1 ha) silled basins 2–3 km apart, with three sampling sites in each basin and $n = 3$ replicates per site. In this case, the aim was to assess spatial variations in phytoplankton assemblages among basins. The MED dataset (128 taxa × 81 samples) included data from three Mediterranean brackish coastal lakes, one in the North Ionian Sea (Korission lake, Kerkira, Greece) and two in the Levantine (Akgöl and Paradeniz lakes, Turkey), and focused on assessing spatial variations in phytoplankton assemblages among lakes and increasing confinement within lakes. In each lake, three sites were sampled near the connection with the open sea, and in the inner and central parts of the lakes, with $n = 3$ replicates in each site. The BRA dataset (123 taxa × 81 samples) concerned a case study where changes in phytoplankton assemblages along the estuarine gradient, from the

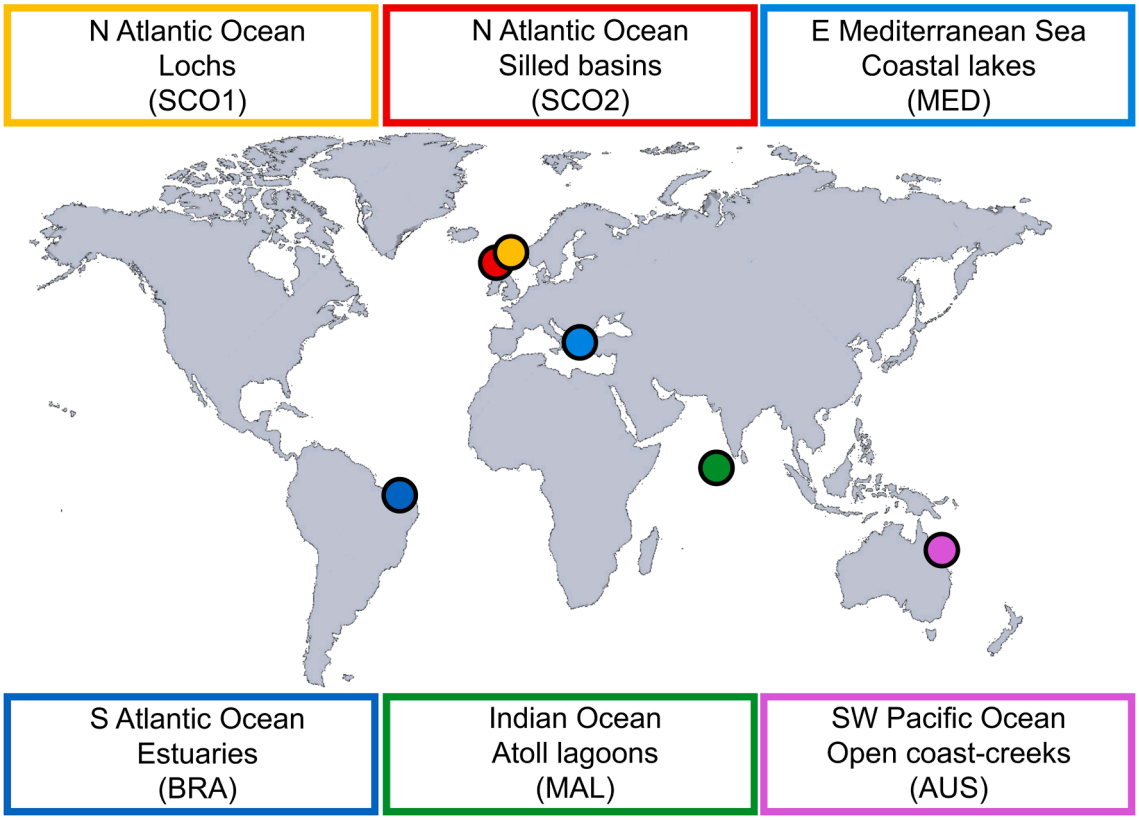


Fig. 1. Global distribution and type of study areas.

mouth to the interior part, were compared among different estuaries located along the northern coast of Brazil (Rio Grande do Norte, SW Atlantic Ocean). Three estuaries were sampled, each of them at three positions along the estuary from the mouth to the interior part, with three sampling sites at each position and $n = 3$ replicate samples in each site. In the MAL case study (97 taxa × 54 samples), phytoplankton assemblages from coral atolls of different size were compared. Phytoplankton were sampled in the inner lagoon of two atolls of small (~1 ha), medium (~50 ha), and large (~150 ha) size located in the Maldives archipelago (Indian Ocean). Within each lagoon, three random sites were sampled, with $n = 3$ replicate samples in each site. Finally, in the last case study AUS (132 taxa × 27 samples), phytoplankton from sandy coast, creek and mangrove areas (Magnetic Island, SW Pacific) were compared to test for difference in assemblage structure at varying habitat type. Three random sites were sampled in each habitat, with $n = 3$ replicate samples in each site. Sampling design for all case studies were summarized in Table 1. All data are available at LifeWatch ERIC (https://www.lifewatch.eu), the e-Science Infrastructure for Biodiversity and Ecosystem Research (see Appendix A in supplementary material for dataset DOI).

For all case studies, phytoplankton samples were collected through 6 μm mesh plankton net equipped with a flow meter to determine the filtered water volume. Samples were preserved with Lugol (15 mL/L of sample). Phytoplankton were examined following Utermöhl's method (1958). Phytoplankton were analysed by inverted microscope (Nikon T300E) connected to a video-interactive image analysis system (L.U.C.I. A Version 4.8, Laboratory Imaging). Taxonomic identification, counting, and morphological traits' quantification were performed on 400 phytoplankton cells for each sample. Phytoplankton were identified to species or genus level referring to different sources of taxonomic keys since the phytoplankton community in most cases has not been previously investigated (see Roselli et al., 2017 for further details).

2.2. Null models for species aggregation

Using operational units of identification alternative to species (e.g., taxa of higher Linnaean ranks, morphological or functional groups, or any other species grouping) is equivalent to compress the multivariate information on assemblage structure, that would be obtained from S species variables, into G aggregate variables, with $S > G$. If the complexity and variability of ecological responses of species within the G groups, irrespective of the criterion used for species aggregation, may often lead such groups to be indistinguishable from random subsets of the original S species, then the loss of information on species-level community patterns will mostly depend on the aggregation *per se*, rather than on the way in which the aggregation is done. Under this assumption, the smaller will be G, the higher will be the level of compression of the original species variables, and the higher will be the loss of information on the original species-level patterns. The level of compression can be simply defined as $\varphi = G/S$, whereas the loss of information can be expressed as the Spearman's rank matrix correlation ρbetween the original dissimilarity data matrix at species level and the dissimilarity matrix based on the aggregate variables (for any dissimilarity or similarity measure of interest) (Bevilacqua et al., 2012; 2013). The construction of null models of species aggregation from the BestAgg procedure allows determining the minimum number of aggregate variables (G_{min}), below which the probability to fail in detecting the original species-level patterns is higher than a tolerable level (β) fixed a priori.

The original S species variables in the data matrix are progressively compressed (i.e., grouped and their abundances summed) at random into a decreasing number of G aggregate variables, simulating increasing compression levels (i.e., decreasing value of φ). The randomization procedure concerns the identity of species and their distribution within the G groups, so that not only the S species are randomly assigned to the G groups but also their number within each group is determined at random. For each aggregation step (i.e., for each simulated value of φ), 1,000 such randomizations were performed obtaining 1,000 aggregated data matrices. For each randomization, the Spearman's rank correlation ρ between the original species matrix and the aggregated matrix was calculated, and a distance-based permutational multivariate analysis of variance (PERMANOVA; Anderson, 2001) was performed based on each randomly aggregated matrix to test for the term of interest in the analysis (e.g., the effect of the investigated source of variation on multivariate assemblage structure). The lowest practicable aggregation level was then determined by the lowest value of G (i.e., G_{min}) allowing the

Table 1
Dataset information and design for analysis.

Dataset	Environmental settings	Source of variation	Design for the analysis	Term of interest
SCO1	Temperate cold Transitional waters Brackish loch	Confinement gradient	Confinement [Co], 3 levels (sea inlet, central and interior part of the loch), fixed; Station [St(Co)], 3 levels, random, nested in Co; $n = 3$	Co
SCO2	Temperate cold Transitional waters Silled brackish basins	Variability among water bodies	Basin [Ba], 3 levels (basin 1, basin 2, basin 3), random; Station [St(Ba)], 3 levels, random, nested in Ba; $n = 3$	Ba
MED	Temperate warm Transitional waters Brackish coastal lakes	Confinement gradient and variability among water bodies	Lake [La], 3 levels (lake 1, lake 2, lake 3), random; Confinement [Co] 3 levels (sea inlet, central and interior part of lakes), fixed, crossed; Station [St(La × Co)], 3 levels, random, nested in La × Co; $n = 3$	La × Co interaction
BRA	Tropical Transitional waters Estuaries	Estuarine gradient and variability among estuaries	Estuary [Es], 3 levels (estuary 1, estuary 2, estuary 3), random; Position [Po], 3 levels (mouth, central and interior part of estuaries), fixed, crossed; Station [St(Es × Po)], 3 levels, random, nested in Es × Po, $n = 3$	Es × Po interaction
MAL	Tropical Marine Atoll lagoons (coral reefs)	Lagoon size	Size [Sz], 3 levels (small, medium, large), fixed; Lagoon [La(Sz)], 2 levels, random, nested in Sz, Station [St(La (Sz))], 3 levels random, nested in Sz, $n = 3$	Sz
AUS	Tropical Marine-brackish Open coast/ creek (mangroves)	Changes among habitats	Habitat [Ha], 3 levels (sand, creek, mangroves), fixed, Station [St (Ha)], 3 levels, random, nested in Ha, $n = 3$	Ha

95% of PERMANOVAs tests to give a result consistent with (i.e., the *p*-values of the tests for the term of interest are in the same range of, or lower than) that obtained at species level. In other words, the null hypothesis that G_{min} is not sufficient to allow consistent results with species-level analysis can be rejected with a probability of Type-I error of $P < 0.05$, under the assumption that aggregate variables are random subsets of the original species variables. Note that this is a threshold value, and therefore the number of aggregate variables to be used can be even higher than G_{min}. Full details on the BestAgg procedure and null model construction for main experimental designs (including asymmetrical designs) with nested and crossed factors, along with the R code to run the procedure, are provided in Bevilacqua et al. (2013), Bevilacqua and Terlizzi (2016).

Following the above procedure, we obtained six separate null models of species aggregation, one for each dataset, in order to determine G_{min} for each case study. In all cases, to simulate decreasing φ values, we considered a stepwise reduction of 10 variables at a time, starting from the total number of species found in the study. For example, for SCO1, the original 73 species were progressively aggregated into 70, 60, 50, 40, 30, 20, and 10 groups, obtaining 1,000 randomly aggregated data matrices for each step. Stepwise reductions of approximately 10% of *S* allow defining a representative set of simulated decreasing φ values for a wide range of *S* (Bevilacqua et al., 2013). For each null model, the relationship between the level of compression of the original variables and the loss of information was checked by fitting a linear regression of ρ values from random aggregations against the corresponding ln(φ). For all the regression analyses, the intercept of linear models was fixed to 1, since the maximum Spearman's correlation between matrices is obtained when the number of aggregate *G* variables is exactly equal to the number of the original *S* species variables, that is when $\varphi = 1$ and, therefore, $\ln(\varphi) = 0$. PERMANOVAs testing for the term of interest in the analysis in all models were based on Bray-Curtis dissimilarities of untransformed abundance data, with 5,000 permutations. The design for analysis and the terms of interest in the different case studies are summarized in Table 1. All analyses were performed using R (R Core Team, 2020).

2.3. Morpho-functional classification of phytoplankton species

Once the G_{min} has been determined, any aggregation criterion of species can be used in order to obtain the set of aggregate variables, having in mind that if their number is lower than G_{min} then the ability to detect species-level patterns is very likely to be compromised. Clearly, the aggregation criterion should allow obtaining ecologically meaningful aggregate variables while increasing the easiness of their identification. Commonly, aggregation criteria are calibrated upon the study aim, considering the group of organisms under study, the environmental context, the ecological relevance of species, the available taxonomic expertise and, obviously, the study-specific G_{min} (Bevilacqua et al., 2013). However, since our study focused exclusively on phytoplankton and our aim is to demonstrate the generality of the approach, we preferred not to obtain study-specific sets of surrogates. Instead, we proposed a general aggregation scheme that, while being suitable to respect the threshold of G_{min} in all the six case studies, can be of general application to different environmental contexts and geographic areas. Our scheme included taxonomic, morphological (i.e., cell shape, elongation, and complexity) and morphometric (i.e., typical cell size) criteria, which were combined to achieve a hierarchical classification of phytoplankton species.

As first, the aggregation scheme distinguished phytoplankton into three main taxonomic groups, which are easy to identify while being informative of the primary ecological features of phytoplankton (e.g., Mutshinda et al., 2016; Wasmund et al., 2017), and namely *Diatoms*, *Dinoflagellates* and *Other* (e.g., Cyanobacteria, Chlorophyta) phytoplankton. Then, phytoplankton were further classified based on maximum linear dimension (MLD) and cell shape. Cell size (MLD) classes were *small* (<10 μm), *medium* (10 μm-20 μm), *large* (20 μm-100 μm), and *very large* (>100 μm), whereas all taxa found in the whole dataset were ascribed to 19 main three-dimensional shapes, whether simple (i.e., cone, pyramid, cube, sphere, ellipsoid, half sphere, parallelepiped, prism with parallelogram base, prism with triangular base, prism with elliptic base, prism sickle-shaped, cylinder, prolate spheroid, cymbelloid, gomphonemoid) or composite (i.e., double ellipsoid, double cone, double truncated cone, half sphere or half ellipsoid + cone). In addition, *elongation* (i.e., cell shape stretched in one dimension) and *complexity* (i.e., the presence of elements increasing the complexity of the main shape, such as flagella, spines, additional shapes to the main body) were also considered as classification criteria. Finally, the genera *Chaetoceros* and *Pseudo-nitzschia*, which are relatively easy to determine and are of particular ecological interest because of their harmfulness, were considered separately.

Size and shape are strongly correlated to nutrient uptake, light use, growth, grazer resistance and anthropogenic disturbance (Rojo and Rodríguez, 1994; Mouillot et al., 2006; Naselli-Flores et al., 2007; Litchman and Klausmeier, 2008; Lugoli et al., 2012). Such traits virtually convey useful information on the ecological role of morphologically different groups of organisms for the functioning of aquatic ecosystems, including energy and matter flow (Weithoff and Beisner, 2019). Therefore, cell size, shape, elongation and complexity were selected due to their widely recognized relations with environmental and biological drivers of phytoplanktonic community changes and, ultimately, to their role as determinants of fitness for phytoplankton (Ryabov et al., 2021).

2.4. Comparison with species-level patterns

For each dataset, the original *S* species (or taxa) variables were aggregated (i.e., grouped and their abundances summed), simulating the identification of phytoplankton using the morpho-functional groups from the proposed aggregation scheme. The Spearman's rank matrix correlation ρ between the original dissimilarity data matrix and the dissimilarity matrix based on the morpho-functional groups was then calculated. Since it is assumed that the aggregate variables are random subsets of the original *S* variables, the information on the original community patterns retained in the aggregated data matrix (expressed as ρ) should fall within, or even above (if the aggregate variables capture more information than what expected to occur by chance) random expectations from the null model. Correlation ρ values below expectations would indicate, instead, a set of aggregate variables that is likely unsuitable (Bevilacqua et al., 2015). To check for this, a randomization test was built following the same procedure underlying the construction of the general null model. For each dataset, the original *S* variables were randomly aggregated into the morpho-functional groups. Random aggregations were repeated 1000 times. Correlation ρ values between the original matrix and each randomly aggregated matrix were then calculated obtaining a frequency distribution of ρ values. The true correlation value, i.e. the correlation between the original matrix and the matrix based on true morpho-functional groups, was tested against this random frequency distribution. PERMANOVA was also performed based on randomly aggregated matrices, and the percentage of tests for the term of interest consistent with those obtained using the original variables, represented the probability of Type-I error associated to the specific set of morpho-functional groups in each case study.

The effectiveness of the approach to provide sets of operational units of identification able to reflect species-level community patterns was checked by comparing the outcomes of PERMANOVA results based on the original species variables with those obtained using the morpho-functional groups. Results of multivariate tests using the morpho-functional groups were considered as consistent with those obtained analysing the original variables when *p*-values of tests based on morpho-functional groups were in the same range (or lower) than those resulting from tests on the original variables. Moreover, when the analysis involved fixed factors with more than two levels, we considered results

consistent if also the pair-wise *post-hoc* comparisons returned the same patterns of difference among the factor's levels. To check whether the use of morpho-functional groups allowed depicting multivariate patterns as at species level, a non-metric multidimensional scaling ordination (nMDS) of centroids for the term of interest in the analysis was performed in both cases. Centroids were obtained calculating principal coordinates (PCO) on the basis of the Bray–Curtis dissimilarity matrix among all pairs of samples. Configurations of centroids in the multivariate space based on morpho-functional groups and the original species variables were then compared through a Procrustean randomization test (PROTEST; Jackson, 1995). The m^2 statistic in the analysis provide a measure of fit between configurations of points (i.e., the lower is m^2, the higher is the concordance between the two configurations), which is then compared against a random distribution of m^2 values to determine whether the original values is significantly smaller than what expected by chance. All analyses were performed using R (R Core Team, 2020).

3. Results

The Spearman's correlation ρ between randomly aggregated matrices and the original species matrix decreased at decreasing ϕ (with $\phi = G/S$, that is the level of compression of original S species variable into the G aggregate variables) following a semi-log model (Table 2), indicating that the information on species-level patterns was progressively lost as the number of aggregate variables decreased (Appendix B in supplementary material, Fig. B1). The minimum number of aggregate variables G_{min}, allowing at least 95% of multivariate tests for the term of interest to give results consistent with those obtained at species level ranged between 10 and 30, depending on the case study (Table 2).

The 384 phytoplankton taxa (63% identified at species level, 27% at genus level, and the remaining 10% at higher taxonomic level) found across the study areas (Appendix A in supplementary material) were classified in a total of 82 morpho-functional groups (Fig. 2, Appendix C in supplementary material).

For each case study, the ρ value between the original matrix and the matrix based on morpho-functional groups fell within the distribution of random values generated for an equal set of aggregate variables except for MED (Fig. 3). In this latter case, the correlation value was significantly higher than random expectations ($P = 0.002$), indicating that the morpho-functional groups allowed retaining more information on species-level patterns than what expected to occur by chance.

For all case studies, PERMANOVA on species data highlighted significant effects of the investigated sources of variation (Table 3). For SCO1, the analysis detected significant changes in phytoplankton assemblages at increasing distance from the sea inlet. In SCO2, a significant variability in phytoplankton among the three silled basins was found. For Mediterranean coastal lakes (MED), phytoplankton assemblages differed at increasing level of confinement, although this pattern was inconsistent among the three lakes. In the BRA case study, phytoplankton assemblages changed along the estuarine gradient in E1 and E2, but not in E3. In MAL, medium and large coral atolls hosted comparable phytoplankton assemblages, which however, were significantly different from those found at the small atolls. Finally, phytoplankton from mangroves were different from those found in correspondence of sandy areas and creeks in the AUS case study. Results of PERMANOVA based on morpho-functional groups tightly reflected those obtained using the original species variables (Table 3), with a probability of failure in detecting community patterns as at species level of $P < 0.05$ in all cases.

Multivariate patterns of changes in phytoplankton assemblages were clearly depicted by the nMDS ordination plots, irrespective of using species or morpho-functional groups (Fig. 4). The values of m^2 from PROTEST were generally very low, indicating that ordination plots were almost interchangeable between original and morpho-functional groups, except for SCO1 and MED, which showed higher values (Table 4). However, correlation values of centroids' configurations in the multivariate space when comparing species-level ordinations and those obtained using morpho-functional groups were very high in all cases (Table 4).

Table 2
Summary of results of regression analysis of ρ values (Spearman's correlation) between the species-level matrix and each randomly aggregated matrix against the natural logarithm of the level of variable aggregation φ (=$G/S$where S is the number of species variable and G the number of aggregate variables). The adjusted R^2 and slope are reported for each case study; regression plots and linear models are provided in Fig. B1 in Appendix B. The lowest φ (φ_{min}) from null models, which indicate the maximum possible aggregation of the original species variables that still allowed multivariate tests for the term of interest to give comparable results to those obtained using species, is reported along with the corresponding G_{min} (i.e., the minimum number of aggregate variables G that should be used). All Adjusted R^2 and slopes were significant at $P < 0.001$.

Dataset	Adj. R^2	Slope	φ_{min}	G_{min}
SCO1	0.48	0.10	0.14	10
SCO2	0.39	0.12	0.09	10
MED	0.65	0.27	0.08	10
BRA	0.45	0.10	0.08	30
MAL	0.94	0.08	0.31	30
AUS	0.88	0.14	0.08	20

4. Discussion

The impact of global change on seas and oceans will likely involve a profound alteration of phytoplankton communities (Van de Waal and Litchman, 2020; Henson et al., 2021), with cascading effects affecting the whole marine system, from the surface until the deep-sea environments (Nomaki et al., 2021). Trait-based perspectives to phytoplankton research will be critical to broaden our understanding of phytoplankton community assembly and dynamics under such changes and the ensuing effects on the functioning of marine ecosystems (Litchman et al., 2007; Litchman et al., 2015; Weithoff and Beisner, 2019). However, comprehensive investigations of functional aspects often require identifying organisms at the level of species and a deep knowledge of their autecology, which is still far from being adequate for phytoplankton (Litchman and Klausmeier, 2008), making these analyses overly demanding and impractical especially in reiterate assessments (Weithoff, 2003; Kruk et al., 2010).

A suite of ecological responses of phytoplankton species to environmental and biological variables, such as light and nutrient availability, temperature, salinity, or zooplankton grazing are, nevertheless, strongly related to specific and recognizable morphological attributes of phytoplankton and particularly to cell size and shape (e.g., Litchman and Klausmeier, 2008; Morán et al., 2010; Naselli-Flores and Barone, 2011; Svensson et al., 2014; Roselli et al., 2015; Cloern, 2018; Ryabov et al., 2021). For instance, the efficiency in nutrient uptake and photosynthesis are directly correlated to the surface/volume ratio of phytoplankton cells, so that under conditions of limited nutrient supply and light, small-sized species rely on a better nutrient flux per unit volume and higher photosynthetic rates with respect to larger ones (Morabito et al., 2007; Karp-Boss and Boss, 2016). In contrast, large species are advantaged in well illuminated and nutrient rich environments (Naselli-Flores et al., 2007; Stanca et al., 2013), where resources are not limiting and large size may confer higher plasticity to other selective factors, such as grazing pressure (Sunda and Hardison, 2010; Cloern, 2018). Cell shape also play a crucial role in determining the efficiency of resource exploitation in phytoplankton. If compared with rounded, compact forms, particular cell shapes (e.g., cylindric, prolate, oblong) or elongated forms allow increasing the surface/volume ratio maximizing nutrient uptake and enhancing chloroplast packing (O'Farrell et al., 2007; Naselli-Flores and Barone, 2011). The complexity of cell shape, which increase with the presence of appendices, flagella, spines, and defensive structures, further improves the nutrient absorption and light

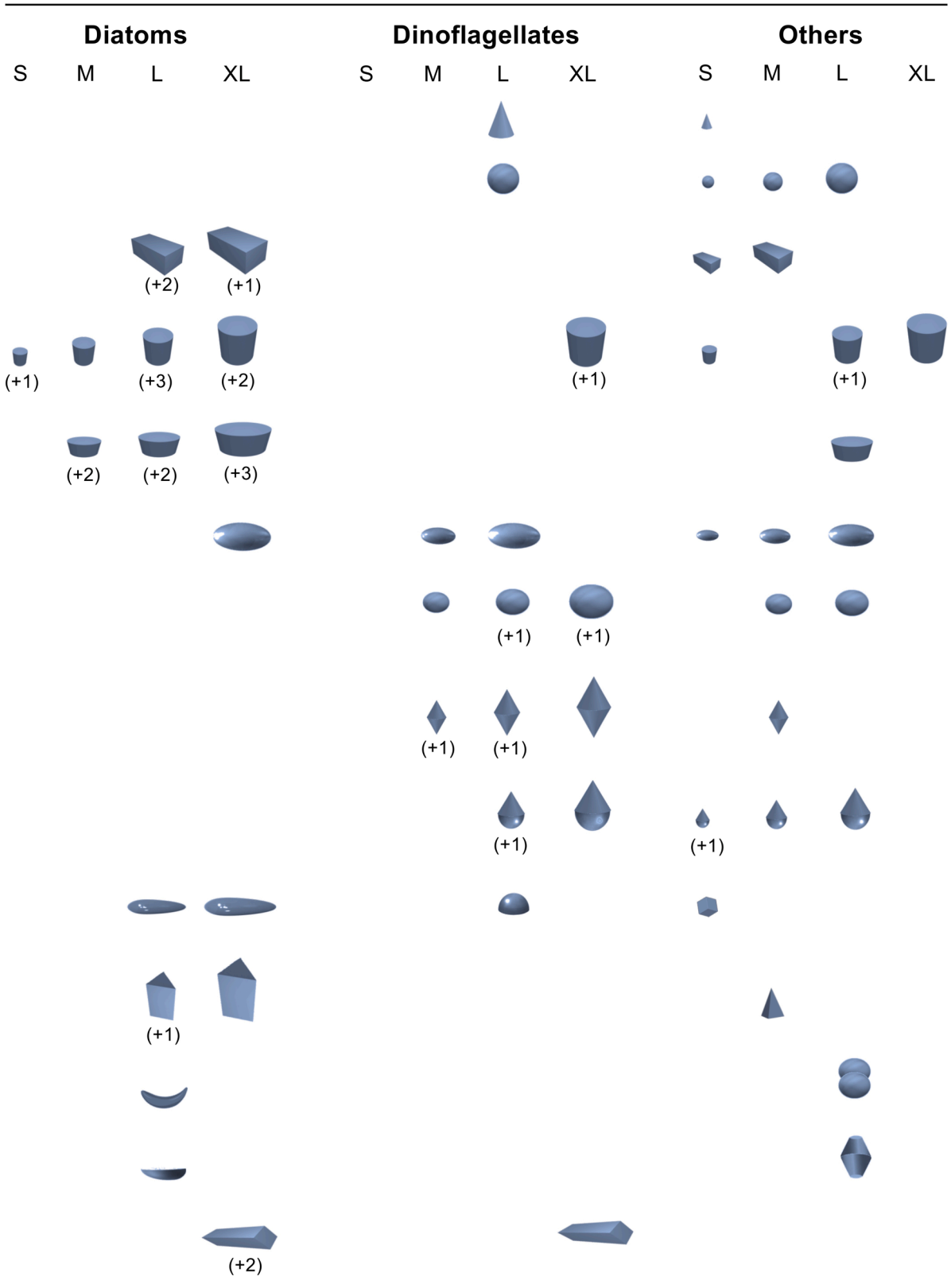


Fig. 2. Schematic representation of the 82 morpho-functional groups of phytoplankton obtained as a combination of the five classification criteria: general taxonomy, cell size, shape, elongation and complexity (see text for further details). Only groups combining general taxonomy (Diatoms, Dinoflagellates, Other), cell size (S, M, L, XL, respectively MLD $<$ 10 μm, 10–20 μm, 20–100 μm, $>$100 μm) and shape are showed. Numbers in brackets indicate groups not showed for that combination of traits, referring to additional grouping of elongated and/or complex forms. The full list of groups is reported in Appendix C in the supplementary material.

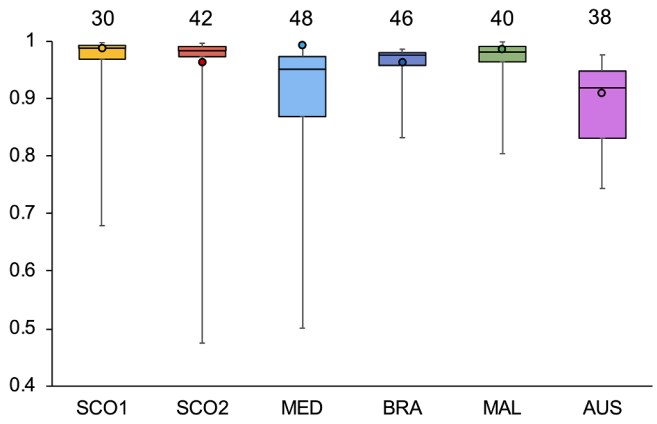


Fig. 3. Box plots of distribution ($n = 1{,}000$) of ρ correlation values between the species-level matrix and matrices in which species were randomly aggregated into a number of variables equal to the number of morpho-functional groups identified for each specific case study. Circles indicate the correlation value between the species-level matrix and the matrix aggregated using the true morpho-functional groups. Whiskers correspond to the 1th and 4th quartile, whereas boxes indicate the median, the 2th and the 3th quartile.

Table 3
Summary of PERMANOVA results using the original number of variables (species) and the aggregate variables (morpho-functional groups) for the term of interest in the analysis for each case study (see Table 1). SCO1: changes along the confinement gradient (Co) from the sea inlet to the interior of the loch (I = sea inlet, II = intermediate, III = interior); SCO2: variation among basins (Ba); MED: patterns of confinement (Co, I = sea inlet, II = intermediate, III = interior) among coastal lakes (La); BRA: differences along the estuarine gradient (Po, I = mouth, I = intermediate, III = interior) among estuaries (Es); MAL: differences among atolls of different size (Sz, S = small, M = medium, L = large); AUS: differences among habitats (Ha, Sa = sand, Ma = mangroves, Cr = creek). **= $P < 0.01$, ***= $P < 0.001$.

Dataset	Number of species	Number of morpho-functional groups	Reduction of operational units	PERMANOVA tests for the term of interest in the analysis	
				Original species variables	Morpho-functional groups
SCO1	73	30	59%	Co*** (I ≠ II ≠ III)	Co*** (I ≠ II ≠ III)
SCO2	116	42	64%	Ba***	Ba***
MED	128	48	63%	La × Co*** L1 (I = II ≠ III) L2 (I ≠ II = III) L3 (I ≠ III ≠ III)	La × Co*** L1 (I = II ≠ III) L2 (I ≠ II = III) L3 (I ≠ III ≠ III)
BRA	123	46	63%	Es × Po*** E1 (I ≠ II = III) E2 (I = II ≠ III) E3 (I = II = III)	Es × Po*** E1 (I ≠ II = III) E2 (I = II ≠ III) E3 (I = II = III)
MAL	97	40	59%	Sz** (S ≠ M = L)	Sz** (S ≠ M = L)
AUS	132	38	71%	Ha** (Ma ≠ Sa = Cr)	Ha** (Ma ≠ Sa = Cr)

harvesting, contrasts cell sinking, and increase the resistance to grazing (Padisák et al., 2003; Stanca et al., 2013; Pančíc and Kiørboe, 2018; Durante et al., 2019).

The use of easily decipherable morphological characteristics of phytoplankton species as traits underlying their functional aspects, therefore, could help extending the application of functional trait-based approaches also to routine phytoplankton monitoring and assessment (Reynolds et al., 2002; Kruk et al., 2010). Yet, when the aim is to analyse community composition using ecologically meaningful and cost-effective operational units as an alternative to labour-intensive fine taxonomic identifications, it should be considered that such operational units, if not designed to serve as surrogates of species-level responses, may not reflect spatial and temporal patterns as outlined by analyses based on multi-species abundances (Bevilacqua et al., 2021). One reason for this may simply relies on the fact that morpho-functional and taxonomic composition of assemblages focus on different aspects of phytoplankton diversity, so that diversity patterns in the two cases may not necessarily overlap (Mutshinda et al., 2016; Graco-Roza et al., 2021). Moreover, as for other organisms (Bevilacqua et al., 2012), our findings from different habitats and marine regions highlighted an intimate relationship between the loss of information on species-level patterns and the level of aggregation of species also for phytoplankton. Hence, grouping phytoplankton species using a priori defined categories, such as taxonomic surrogates (or morpho-functional groups), without controlling for the effect of variable aggregation, may lead to compromise the recognition of genuine community patterns (Carneiro et al., 2010; Gallego et al., 2012; Machado et al., 2015).

If compared to more classical methods, like those involving the use of coarse levels of taxonomic resolution (e.g., genus, family), this new approach can be more profitable in terms of time saved during sample processing, assuming that the time required to identify organisms is proportional to the number of categories in which they have to be classified (Ferraro and Cole, 1995; Thompson et al., 2003). For identifications based on morpho-functional groups the number of categories was about 60% lower than for species-level identifications in all case studies (Table 3), while still allowing the detection of community patterns. Instead, the identification of phytoplankton at genus and family-level would lead, on average, to reduce the number of categories of only 30% and 44% respectively (Appendix B in supplementary material, Table B1). Correlation values between the original species-level matrix and matrices at genus or family level were not significantly higher than what expected to occur by chance (Table B1). Such findings confirmed what has been found for other groups of organisms, from terrestrial plants to marine invertebrates, reinforcing the idea that higher taxa can be viewed as random group of species not conveying consistent ecological responses to natural and human-driven environmental changes (Bevilacqua et al., 2012). Indeed, phylogenetic or taxonomic relatedness and ecological similarity of species are often unrelated for many taxa, or relationships may be limited only to some lineages or species traits (Losos, 2008; Crisp and Cook, 2012), and the responses of relatives to environmental variations may not converge showing, instead, neutral dynamics (Carranza et al., 2011; Bevilacqua et al., 2013; Siqueira et al., 2012). In half of cases, correlations fell below the 95% CI from random expectations or very close to the lower bound (Table B1), indicating that higher taxa of phytoplankton may often tend to include species having opposite patterns of abundance in relation to the investigated environmental gradient. For morpho-functional groups correlation values never fell below the 95% CI, and for the MED case study the value was significantly higher than random expectations. Thus, the morpho-functional approach seems having the potential to be more effective than taxonomic aggregation in grouping species with common ecological features. However, in most cases the information provided by morpho-functional groups did not correlate with species-level information better than what occurred by chance, suggesting that categorizing phytoplankton species according to shared cell traits may not ensure obtaining ecologically coherent groups. Modelling species aggregation

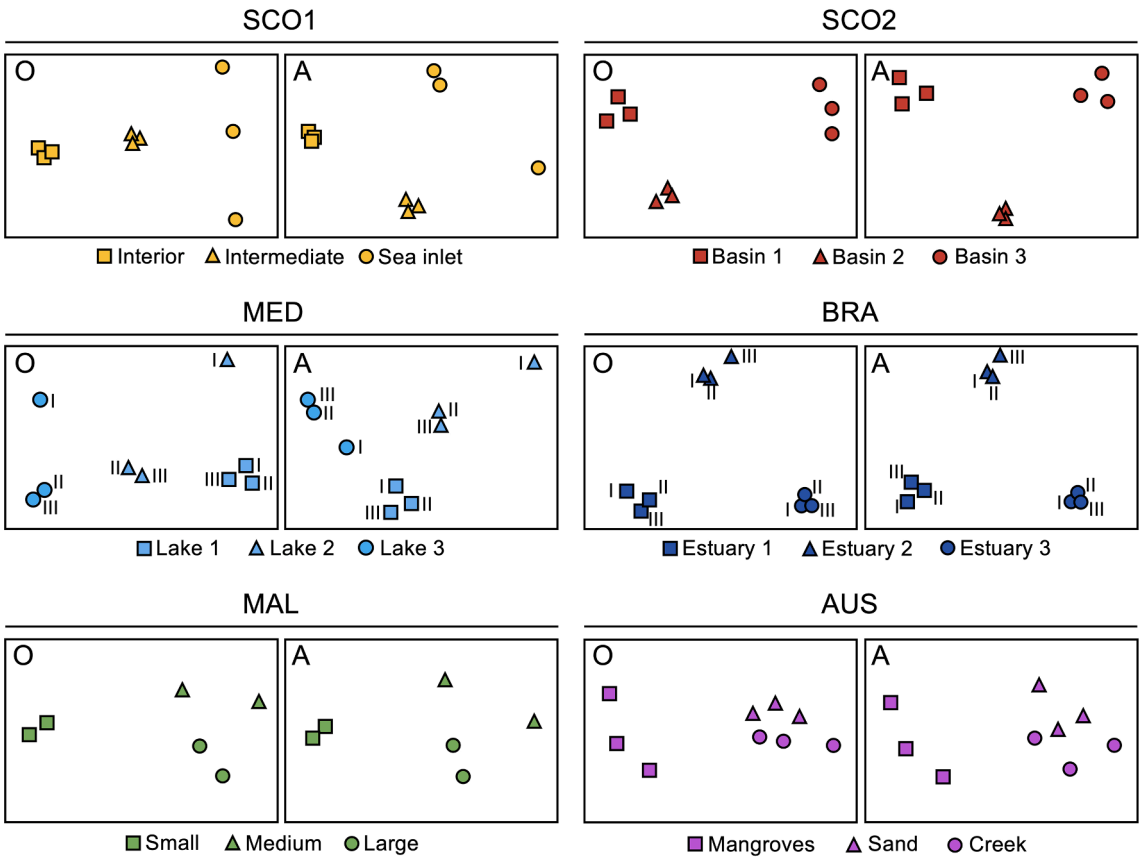


Fig. 4. Non-metric multidimensional scaling ordinations (nMDS) of relevant centroids for each case study (see also Table 3) based on Bray–Curtis dissimilarity for species-level data (O), and data aggregated using the morpho-functional groups (A). SCO1: centroids of stations along the confinement gradient from the sea inlet to the interior of the loch; SCO2: centroids of stations in the three basins; MED: centroids of confinement positions in the three coastal lakes (I = sea inlet, II = intermediate, III = interior); BRA: centroids of positions along the estuarine gradient in the three estuaries (I = sea inlet, II = intermediate, III = interior); MAL: centroids of atolls of different size; AUS: centroids of stations of the different habitats. In all cases, stress value was $\ll 0.1$.

Table 4
Results of PROTEST comparing the nMDS ordination plots based on the original species variables and morpho-functional groups for each case study.

	SCO1	SCO2	MED	BRA	MAL	AUS
m^2	0.248	0.091	0.484	0.006	0.025	0.050
Pearson's correlation	0.867**	0.954**	0.961**	0.997**	0.987**	0.975**

based on functional traits assumes intraspecific trait invariance, which could not be always the case (Yoshida et al., 2003; Merico et al., 2014). More generally, the complex interplay among niche preferences, competition, and predation, may lead phytoplankton species within functional groups to respond to environmental variations inconsistently (Segura et al., 2011; Mutshinda et al., 2016), and a deeper understanding of trade-off among different and often contrasting trait-mediated competitive abilities is needed for improving trait-based assessments (Litchman and Klausmeier, 2008).

The approach to integrate estimates of information loss and morpho-functional classification of phytoplankton allowed obtaining parsimonious sets of morpho-functional operational units, which were suitable to detect changes in phytoplankton assemblage structure as at species level in several areas spanning the world's ocean. Considering the full list of species found across the investigated geographic areas, the original 384 taxa can be condensed into 82 effective morpho-functional groups with a decrease of about 80% of units of identification across studies. At the scale of single study areas, about 40 morpho-functional groups, on average, were sufficient to reflect species-level patterns, with an average reduction of units of identification $>$ 60%. This reduction could have been even higher, since null model predictions for G_{min} ranged between 10 and 30, depending on the study case. This does not mean that random grouping is more effective than morpho-functional groups, but simply that our classification scheme has led to a number of aggregate variables higher than the predicted G_{min}, which represent a threshold below which further aggregation are prejudicial. It is worth noting, also, that we do not intend to provide a static morpho-functional classification of phytoplankton, nor an exhaustive list of morpho-functional groups. Our aim was, rather, to implement a standardized process to optimize and facilitate the analysis of phytoplankton assemblages in routine monitoring and assessments, while controlling for the consequent loss of information. Although representative of different marine habitats and biogeographic regions, our study is clearly not comprehensive of the whole phytoplankton diversity, and further morpho-functional groups based on the selected criteria have probably to be defined for classifying phytoplankton species not included in our list, and/or additional criteria could be incorporated to refine the morpho-functional classification in other habitats or under different environmental settings. However, since phytoplankton are primary biological quality elements for coastal and marine ecosystems (e.g., the European Water Framework Directive; EC, 2000), data on phytoplankton assemblages can be largely available for many marine water

bodies, serving as baseline to define the set of morpho-functional groups needed to reflect community patterns as at species-level. The application of our approach, eventually complemented by periodic species-level surveys to check and refine the effectiveness of morpho-functional units of identification, could represent a valid alternative to more traditional and expensive procedures, at least in routine monitoring and assessments of phytoplankton. A major strength of this procedure is that the use of null models of species aggregation confers high robustness to the confounding effects of natural variability of assemblages in space and time (Bevilacqua et al. 2015, Thiault et al. 2015), and allows controlling for uncertainty in the application of the selected set of aggregate variables. Also, by identifying the number of morpho-functional groups, the use of this procedure combined to automated morpho-functional classification, through imaging-in-flow cytometry techniques (Sosik and Olson, 2007; Thomas et al., 2018), can be envisaged.

Standardized methods relying on easy-to-identify operational units can reduce identification errors which are likely to occur in taxonomic analyses of difficult taxa or due to the heterogeneity in taxonomic expertise of practitioners and researchers (Straile et al., 2013; Muñiz et al., 2020), reduce identification efforts in terms of time spent for sample processing and resources, which in turn can be invested in extending monitoring programmes in space and time (Ellingsen et al., 2017), and allow comparisons of community patterns among different geographic areas or periods (Bevilacqua et al., 2018). In this view, our approach condensed into a single framework cost-effectiveness, instances from conventional monitoring programmes focusing on detecting patterns of spatial (and/or temporal) variations in the structure of phytoplankton assemblages, and the current need for further insights from a functional perspective (Edwards et al., 2013; Litchman et al., 2012).

CRediT authorship contribution statement

Leonilde Roselli: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. **Stanislao Bevilacqua:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Antonio Terlizzi:** Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank E. Stanca for assistance in the identification of phytoplankton, and recognize the invaluable support of A. Basset. This research was supported by “POR PUGLIA Progetto Strategico 2009–2012”.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2022.108827.

References

Alvarez, E., Moyano, M., López-Urrutia, Á., Nogueira, E., Scharek, R., 2014. Routine determination of plankton community composition and size structure: a comparison between FlowCAM and light microscopy. J. Plankton Res. 36, 170–184.
Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 26 (1), 32–46.
Bates, C.R., Scott, G., Tobin, M., Thompson, R., 2007. Weighing the costs and benefits of reduced sampling resolution in biomonitoring studies: perspectives from the temperate rocky intertidal. Biol. Conserv. 137 (4), 617–625.
Batten, S.D., Abu-Alhaija, R., Chiba, S., Edwards, M., Graham, G., Jyothibabu, R., Kitchener, J.A., Koubbi, P., McQuatters-Gollop, A., Muxagata, E., Ostle, C., Richardson, A.J., Robinson, K.V., Takahashi, K.T., Verheye, H.M., Wilson, W., 2019. A global plankton diversity monitoring program. Front. Mar. Sci. 6, 321.
Bertrand, Y., Pleijel, F., Rouse, G.W., 2006. Taxonomic surrogacy in biodiversity assessments, and the meaning of Linnaean ranks. System. Biodivers. 4 (2), 149–159.
Bevilacqua, S., Terlizzi, A., 2016. Species surrogacy in environmental impact assessment and monitoring: extending the BestAgg approach to asymmetrical designs. Mar. Ecol. Progr. Ser. 547, 19–32.
Bevilacqua, S., Fraschetti, S., Musco, L., Terlizzi, A., 2009. Taxonomic sufficiency in the detection of natural and human-induced changes in marine assemblages: a comparison of habitats and taxonomic groups. Mar. Pollut. Bull. 58 (12), 1850–1859.
Bevilacqua, S., Terlizzi, A., Fraschetti, S., Claudet, J., Boero, F., 2012. Taxonomic relatedness does not matter for species surrogacy in the assessment of community responses to environmental drivers. J. Appl. Ecol. 49, 357–366.
Bevilacqua, S., Claudet, J., Terlizzi, A., 2013. Best practicable aggregation of species: a step forward for species surrogacy in environmental assessment and monitoring. Ecol. Evol. 3 (11), 3780–3793.
Bevilacqua, S., Terlizzi, A., Mistri, M., Munari, C., 2015. New frameworks for species surrogacy in monitoring highly variable coastal ecosystems: applying the BestAgg approach to Mediterranean coastal lagoons. Ecol. Indic. 52, 207–218.
Bevilacqua, S., Mistri, M., Terlizzi, T., Munari, C., 2018. Assessing the effectiveness of surrogates for species over time: evidence from decadal monitoring of a Mediterranean transitional water ecosystem. Mar. Pollut. Bull. 131, 507–514.
Bevilacqua, S., Anderson, M.J., Ugland, K.I., Somerfield, P.J., Terlizzi, A., 2021. The use of taxonomic relationships among species in applied ecological research: baseline, steps forward and future challenges. Austral Ecol. 46 (6), 950–964.
Carneiro, F.M., Bini, L.M., Rodrigues, L.C., 2010. Influence of taxonomic and numerical resolution on the analysis of temporal changes in phytoplankton communities. Ecol. Indic. 10 (2), 249–255.
Carranza, A., Defeo, O., Arim, M., 2011. Taxonomic relatedness and spatial structure of a shelf benthic gastropod assemblage. Divers. Distrib. 17, 25–34.
Cloern, J.E., 2018. Why large cells dominate estuarine phytoplankton. Limnol. Oceanogr. 63 (S1), S392–S409.
Cottingham, K.L., Carpenter, S.R., 1998. Population, community, and ecosystem variates as ecological indicators: phytoplankton responses to whole-lake enrichment. Ecol. Appl. 8 (2), 508–530.
Coupel, P., Matsuoka, A., Ruiz-Pino, D., Gosselin, M., Marie, D., Tremblay, J.-É., Babin, M., 2015. Pigment signatures of phytoplankton communities in the Beaufort Sea. Biogeosciences 12, 991–1006.
Crisp, M.D., Cook, L.G., 2012. Phylogenetic niche conservatism: what are the underlying evolutionary and ecological causes? New Phytol. 196 (3), 681–694.
Culhane, F.E., Frid, C.L.J., Royo Gelabert, E., White, L., Robinson, L.A., 2018. Linking marine ecosystems with the services they supply: what are the relevant service providing units? Ecol. Appl. 28 (7), 1740–1751.
Dashkova, V., Malashenkov, D., Poulton, N., Vorobjev, I., Barteneva, N.S., 2017. Imaging flow cytometry for phytoplankton analysis. Methods 112, 188–200.
Domingues, R.B., Barbosa, A., Galvão, H., 2008. Constraints on the use of phytoplankton as a biological quality element within the Water Framework Directive in Portuguese waters. Mar. Pollut. Bull. 56 (8), 1389–1395.
Durante, G., Basset, A., Stanca, E., Roselli, L., 2019. Allometric scaling and morphological variation in sinking rate of phytoplankton. J. Phycol. 55 (6), 1386–1393.
EC, 2000. Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for community action in the field of water policy.
Edwards, K.F., Litchman, E., Klausmeier, C.A., Elser, J., 2013. Functional traits explain phytoplankton community structure and seasonal dynamics in a marine ecosystem. Ecol. Lett. 16 (1), 56–63.
Ellingsen, K.E., Yoccoz, N.G., Tveraa, T., Hewitt, J.E., Thrush, S.F., 2017. Long-term environmental monitoring for assessment of change: measurement inconsistencies over time and potential solutions. Environ. Monit. Assess. 189, 595.
Ellis, D., 1985. Taxonomic sufficiency in pollution assessment. Mar. Pollut. Bull. 16 (12), 459.
Eriksen, R.S., Davies, C.H., Bonham, P., Coman, F.E., Edgar, S., McEnnulty, F.R., McLeod, D., Miller, M.J., Rochester, W., Slotwinski, A., Tonks, M.L., Uribe-Palomino, J., Richardson, A.J., 2019. Australia's long-term plankton observations: the integrated marine observing system National Reference Station Network. Front. Mar. Sci. 6, 161.
Ferraro, S.P., Cole, F.A., 1995. Taxonomic level sufficient for assessing pollutionimpacts on the southern Californian Bight macrobenthos – revisited. Environ. Toxicol. Chem. 14, 1031–1040.
Gallego, I., Davidson, T.A., Jeppesen, E., Pérez-Martínez, C., Sánchez-Castillo, P., Juan, M., Fuentes-Rodríguez, F., León, D., Peñalver, P., Toja, J., Casas, J.J., 2012. Taxonomic or ecological approaches? Searching for phytoplankton surrogates in the determination of richness and assemblage composition in ponds. Ecol. Indic. 18, 575–585.
Godfray, H.C.J., Knapp, S., Gaston, K.J., O'Neill, M.A., 2004. Automated species identification: why not? Phil. Trans. R. Soc. Lond. B 359 (1444), 655–667.
Graco-Roza, C., Soininen, J., Corrêa, G., Pacheco, F.S., Miranda, M., Domingos, P., Marinho, M.M., 2021. Functional rather than taxonomic diversity reveals changes in the phytoplankton community of a large dammed driver. Ecol. Indic. 121, 107048.
Groc, S., Delabie, J.H.C., Longino, J.T., Orivel, J., Majer, J.D., Vasconcelos, H.L., Dejean, A., 2010. A new method based on taxonomic sufficiency to simplify studies on Neotropical ant assemblages. Biol. Conserv. 143 (11), 2832–2839.

Havskum, H., Schlüter, L., Scharek, R., Berdalet, E., Jacquet, S., 2004. Routine quantification of phytoplankton groups – microscopy or pigment analyses? Mar. Ecol. Prog. Ser. 273, 31–42.
Heino, J., Soininen, J., 2007. Are higher taxa adequate surrogates for species level assemblage patterns and species richness in stream organisms? Biol. Conserv. 137 (1), 78–89.
Henson, S.A., Cael, B.B., Allen, S.R., Dutkiewicz, S., 2021. Future phytoplankton diversity in a changing climate. Nature Comm. 12, 5372.
Hutchinson, G.E., 1961. The paradox of the plankton. Am. Nat. 95 (882), 137–145.
Jackson, D.A., 1995. PROTEST: a PROcrustean randomization TEST of community environment concordance. Ecoscience 2 (3), 297–303.
Jiang, X., Song, Z., Xiong, J., Proctor, H., Xie, Z., 2017. Different surrogacy approaches for stream macroinvertebrates in discriminating human disturbances in Central China. Ecol. Indic. 75, 182–191.
Jones, F.C., 2008. Taxonomic sufficiency: the influence of taxonomic resolution on freshwater bioassessments using benthic macroinvertebrates. Environ. Rev. 16 (NA), 45–69.
Karp-Boss, L., Boss, E., 2016. The elongated, the squat and the spherical: Selective pressures for phytoplankton shape. In: Glibert, P.M., Kana, T.M. (Eds.), Aquatic Microbial Ecology and Biogeochemistry: A Dual Perspective. Springer International Publication, Geneva, Switzerland, pp. 25–34.
Kruk, C., Huszar, V.L.M., Peeters, E.T.H.M., Bonilla, S., Costa, L., Lürling, M., Reynolds, C.S., Scheffer, M., 2010. A morphological classification capturing functional variation in phytoplankton. Freshw. Biol. 55, 614–627.
Litchman, E., Klausmeier, C.A., 2008. Trait-based community ecology of phytoplankton. Annu. Rev. Ecol. Evol. Syst. 39 (1), 615–639.
Litchman, E., Klausmeier, C.A., Schofield, O.M., Falkowski, P.G., 2007. The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level. Ecol. Lett. 10 (12), 1170–1181.
Litchman, E., Edwards, K.F., Klausmeier, C.A., Thomas, M.K., 2012. Phytoplankton niches, traits and eco-evolutionary responses to global environmental change. Mar. Ecol. Progr. Ser. 470, 235–248.
Litchman, E., Tezanos Pinto, P., Edwards, K.F., Klausmeier, C.A., Kremer, C.T., Thomas, M.K., Austin, A., 2015. Global biogeochemical impacts of phytoplankton: a trait-based perspective. J. Ecol. 103 (6), 1384–1396.
Losos, J.B., 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. Ecol. Lett. 11, 995–1007.
Lugoli, F., Garmendia, M., Lehtinen, S., Kauppila, P., Moncheva, S., Revilla, M., Roselli, L., Slabakova, N., Valencia, V., Dromph, K.M., Basset, A., 2012. Application of a new multi-metric phytoplankton index to the assessment of ecological status in marine and transitional waters. Ecol. Indic. 23, 338–355.
Machado, K.B., Borges, P.P., Carneiro, F.M., Santana, J.F., Vieira, L.C.G., Huszar, V.L.M., Nabout, J.C., 2015. Using lower taxonomic resolution and ecological approches as a surrogate for plankton species. Hydrobiologia 743, 255–267.
Mellin, C., Delean, S., Caley, J., Edgar, G., Meekan, M., Pitcher, R., Przeslawski, R., Williams, A., Bradshaw, C., 2011. Effectiveness of biological surrogates for predicting patterns of marine biodiversity: a global meta-analysis. PLoS One 6, e20141.
Merico, A., Brandt, G., Smith, S.L., Oliver, M., 2014. Sustaining diversity in trait-based models of phytoplankton communities. Front. Ecol. Evol. 2, 59.
Milošević, D., Stojković, M., Čerba, D., Petrović, A., Paunović, M., Simić, V., 2014. Different aggregation approaches in the chironomid community and the threshold of acceptable information loss. Hydrobiologia 727 (1), 35–50.
Morabito, G., Oggioni, A., Caravati, E., Panzani, P., 2007. Seasonal morphological plasticity of phytoplankton in Lago Maggiore (N. Italy). Hydrobiologia 578 (1), 47–57.
Morán, X.A., López-Urrutia, Á., Calvo-Díaz, A., Li, W.K.W., 2010. Increasing importance of small phytoplankton in a warmer ocean. Global Change Biol. 16, 1137–1144.
Mouillot, D., Spatharis, S., Reizopoulou, S., Laugier, T., Sabetta, L., Basset, A., Do Chi, T., 2006. Alternatives to taxonomic-based approaches to assess changes in transitional water communities. Aquat. Conserv. Mar. Freshw. Syst. 16 (5), 469–482.
Mueller, M., Pander, J., Geist, J., 2013. Taxonomic sufficiency in freshwater ecosystems: effects of taxonomic resolution, functional traits, and data transformation. Freshw. Sci. 32 (3), 762–778.
Muñiz, O., Rodríguez, J.G., Revilla, M., Laza-Martínez, A., Seoane, S., Franco, J., 2020. Inhomogeneity detection in phytoplankton time series using multivariate analyses. Oceanologia 62 (3), 243–254.
Mutshinda, C.M., Finkel, Z.V., Widdicombe, C.E., Irwin, A.J., Norden, N., 2016. Ecological equivalence of species within phytoplankton functional groups. Funct. Ecol. 30 (10), 1714–1722.
Naselli-Flores, L., Barone, R., 2011. Invited review - fight on plankton! or, phytoplankton shape and size as adaptive tools to get ahead in the struggle for life. Cryptogam. Algol. 32 (2), 157–204.
Naselli-Flores, L., Padisák, J., Albay, M., 2007. Shape and size in phytoplankton ecology: do they matter? Hydrobiologia 578 (1), 157–161.
Nomaki, H., Rastelli, E., Ogawa, N.O., Matsui, Y., Tsuchiya, M., Manea, E., Corinaldesi, C., Hirai, M., Ohkouchi, N., Danovaro, R., Nunoura, T., Amaro, T., 2021. In situ experimental evidences for responses of abyssal benthic biota to shifts in phytodetritus compositions linked to global climate change. Global Change Biol. 27 (23), 6139–6155.
O'Farrell, I., de Tezanos Pinto, P., Izaguirre, I., 2007. Phytoplankton morphological response to the underwater light conditions in a vegetated wetland. Hydrobiologia 578 (1), 65–77.
Padisák, J., Soróczki-Pintér, È., Rezner, Z., 2003. Sinking properties of some phytoplankton shapes and the relation of form resistance to morphological diversity of plankton – an experimental study. Hydrobiologia 500, 243–257.
Padisák, J., Crossetti, L.O., Naselli-Flores, L., 2009. Use and misuse in the application of the phytoplankton functional classification: a critical review with updates. Hydrobiologia 621 (1), 1–19.
R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Reynolds, C., Huszar, V., Kruk, C., Naselli-Flores, L., Melo, S., 2002. Towards a functional classification of the freshwater phytoplankton. J. Plankton Res. 24, 417–428.
Roselli, L., Basset, A., 2015. Decoding size distribution patterns in marine and transitional water phytoplankton: from community to species level. PLoS One 10, e0127193.
Roselli, L., Litchman, E., Stanca, E., Cozzoli, F., Basset, A., 2017. Individual trait variation in phytoplankton communities across multiple spatial scales. J. Plankton Res. 39, 577–588.
Rosser, N., 2017. Shortcuts in biodiversity research: What determines the performance of higher taxa as surrogates for species? Ecol. Evol. 7 (8), 2595–2603.
Ryabov, A., Kerimoglu, O., Litchman, E., Olenina, I., Roselli, L., Basset, A., Stanca, E., Blasius, B., Chase, J., 2021. Shape matters: the relationship between cell geometry and diversity in phytoplankton. Ecol. Lett. 24 (4), 847–861.
Segura, A.M., Calliari, D., Kruk, C., Conde, D., Bonilla, S., Fort, H., 2011. Emergent neutrality drives phytoplankton species coexistence. Proc. R. Soc. B 278 (1716), 2355–2361.
Sieracki, C.K., Sieracki, M.E., Yentsch, C.S., 1998. An imaging-in-flow system for automated analysis of marine microplankton. Mar. Ecol. Progr. Ser. 168, 285–296.
Siqueira, T., Bini, L.M., Roque, F.O., Cottenie, K., 2012. A metacommunity framework for enhancing the effectiveness of biological monitoring strategies. PLoS One 7, e43626.
Sosik, H.M., Olson, R.J., 2007. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. Limnol. Oceanogr. Methods 5 (6), 204–216.
Stanca, E., Cellamare, M., Basset, A., 2013. Geometric shape as a trait to study phytoplankton distributions in aquatic ecosystems. Hydrobiologia 701 (1), 99–116.
Straile, D., Jochimsen, M.C., Kümmerlin, R., 2013. The use of long-term monitoring data for studies of planktonic diversity: a cautionary tale from two Swiss lakes. Freshw. Biol. 58, 1292–1301.
Sunda, W.G., Hardison, D.R., 2010. Evolutionary tradeoffs among nutrient acquisition, cell size, and grazing defense in marine phytoplankton promote ecosystem stability. Mar. Ecol. Progr. Ser. 401, 63–76.
Svensson, F., Norberg, J, Snoeijs, P., 2014. Diatom Cell Size, Coloniality and Motility: Trade-Offs between Temperature, Salinity and Nutrient Supply with Climate Change. PLoS ONE 9, e109993.
Terlizzi, A., Bevilacqua, S., Fraschetti, S., Boero, F., 2003. Taxonomic sufficiency and the increasing insufficiency of taxonomic expertise. Mar. Pollut. Bull. 46 (5), 556–561.
Thiault, L., Bevilacqua, S., Terlizzi, A., Claudet, J., 2015. Taxonomic relatedness does not reflect coherent ecological response of fish to protection. Biol. Conserv. 190, 98–106.
Thomas, M.K., Fontana, S., Reyes, M., Pomati, F., 2018. Quantifying cell densities and biovolumes of phytoplankton communities and functional groups using scanning flow cytometry, machine learning and unsupervised clustering. PLoS ONE 13, e0196225.
Thompson, B.W., Riddle, M.J., Stark, J.S., 2003. Cost-efficient methods for marine pollution monitoring at Casey Station, East Antarctica: the choice of sieve mesh-size and taxonomic resolution. Mar. Pollut. Bull. 46 (2), 232–243.
Tomas, C.R., 1997. Identifying Marine Phytoplankton. Academic Press, San Diego, p. 858.
Tweddle, J.F., Gubbins, M., Scott, B.E., 2018. Should phytoplankton be a key consideration for marine management? Mar. Policy 97, 1–9.
Uitz, J., Claustre, H., Gentili, B., Stramski, D., 2010. Phytoplankton class-specific primary production in the world's oceans: seasonal and interannual variability from satellite observations. Global Biogeochem. Cycles 24 (3), n/a–n/a.
Utermöhl, H., 1958. Zur Vervollkommnung der quantitativen Phytoplankton-methodik. Mitt. Int. Ver. Theoret. Angew. Limnol. 9 (1), 1–38.
Vallina, S.M., Simó, R., 2007. Strong relationship between DMS and the solar radiation dose over the global surface ocean. Science 315 (5811), 506–508.
Van de Waal, D.B., Litchman, E., 2020. Multiple global change stressor effects on phytoplankton nutrient acquisition in a future ocean. Phil. Trans. R. Soc. Lond. B 375 (1798), 20190706.
Wang, C., Jia, H., Wei, J., Yang, W., Gao, Y., Liu, Q., Ge, D., Wu, N., 2021. Phytoplankton functional groups as ecological indicators in a subtropical estuarine river delta system. Ecol. Indic. 126, 107651.
Wasmund, N., Kownacka, J., Göbel, J., Jaanus, A., Johansen, M., Jurgensone, I., Lehtinen, S., Powilleit, M., 2017. The diatom/dinoflagellate index as an indicator of ecosystem changes in the Baltic Sea 1. Principle and handling instruction. Front. Mar. Sci. 4, 22.
Weithoff, G., 2003. The concepts of 'plant functional types' and 'functional diversity' in lake phytoplankton – a new understanding of phytoplankton ecology? Freshw. Biol. 48, 1669–1675.
Weithoff, G., Beisner, B.E., 2019. Measures and approaches in trait-based phytoplankton community ecology – from freshwater to marine ecosystems. Front. Mar. Sci. 6, 40.
Yoshida, T., Jones, L.E., Ellner, S.P., Fussmann, G.F., Hairston, N.G., 2003. Rapid evolution drives ecological dynamics in a predator-prey system. Nature 424 (6946), 303–306.
Zhang, C., Dang, H., Azam, F., Benner, R., Legendre, L., Passow, U., Polimene, L., Robinson, C., Suttle, C.A., Jiao, N., 2018. Evolving paradigms in biological carbon cycling in the ocean. Natl. Sci. Rev. 5, 481–499.