Diazotrophic cyanobacteria in planktonic food webs

Sanna Suikkanen [a,*], Laura Uusitalo [b], Sirpa Lehtinen [a], Maiju Lehtiniemi [a], Pirkko Kauppila [a], Katja Mäkinen [c], Harri Kuosa [a]

[a] *Finnish Environment Institute, Marine Research Centre, Latokartanonkaari 11, FI-00790 Helsinki, Finland*
[b] *Finnish Environment Institute, Strategic Programme for Environmental Information, Latokartanonkaari 11, FI-00790 Helsinki, Finland*
[c] *Archipelago Research Institute, Biodiversity Unit, University of Turku, Joukahaisenkatu 3-5, FI-02520 Turku, Finland*

ABSTRACT

Blooms of cyanobacteria are recurrent phenomena in coastal estuaries. Their maximum abundance coincides with the productive period of zooplankton and pelagic fish. Experimental studies indicate that diazotrophic, i.e. dinitrogen (N_2)-fixing cyanobacterial (taxonomic order Nostocales) blooms affect zooplankton, as well as other phytoplankton. We used multidecadal monitoring data from one archipelago station (1992–2013) and ten open sea stations (1979–2013) in the Baltic Sea to explore the potential bottom-up connections between diazotrophic and non-diazotrophic cyanobacteria and phyto- and zooplankton in natural plankton communities. Random forest regression, combined with linear regression analysis showed that the biomass of cyanobacteria (both diazotrophic and non-diazotrophic) was barely connected to any of the phytoplankton and zooplankton variables examined. Instead, physico-chemical variables (salinity, temperature, total phosphorus), as well as spatial and temporal variability seemed to have more significant connections to both phytoplankton and zooplankton variables. Zooplankton variables were also connected to the biomass of phytoplankton groups other than cyanobacteria (such as chrysophytes, cryptophytes and prymnesiophytes), and phytoplankton variables had connections with the biomass of different zooplankton groups, especially copepods. Overall, negative relationships between cyanobacteria and other plankton taxa were scarcer than expected based on previous experimental studies.

1. Introduction

Primary production of organic matter is the basis of secondary production in aquatic systems. In the open waters, microscopic phytoplankton is responsible for the majority of primary production with highly variable quantity distribution. However, starting from the classic work by Ryther (1969) it is realized that secondary production up to the level of fish and other top predators is not always directly correlated with primary productivity. Though phytoplankton consists of microscopic cells, which are visible only by their exceptional growth as 'blooms', their communities are a mixture of tens or hundreds of species. These communities include numerous species of prokaryotic cyanobacteria, which are members of the phytoplankton community since they participate in the pelagic primary production similarly to the autotrophic planktonic microalgae. All phytoplankton species have different specific characteristics, which potentially affect trophic transfer efficiency, such as size range, which covers several orders of magnitude from single picoplanktonic (Ø <2 μm) cells to (just) visible colonies (Finkel et al., 2010). In addition to different sizes, phytoplankton species hold a number of functional traits (Barton et al., 2013), including defense systems against herbivory, which may decrease the usability of primary production by zooplankton (Van Donk et al., 2011).

The brackish Baltic Sea is one of the largest single ecosystems globally with recurring massive cyanobacterial blooms (O'Neil et al., 2012). Though diazotrophic, i.e. dinitrogen (N_2) fixing cyanobacterial blooms are a natural phenomenon in the Baltic Sea (Bianchi et al., 2000), the extent and intensity of modern cyanobacterial blooms has been linked to the increasing eutrophication status, the subsequent increased organic matter settling and increased deep-water hypoxia (Backer et al., 2010; Gustafsson et al., 2012; Funkey et al., 2014). At the same time, climate change may favor stronger cyanobacteria blooms that start earlier and last longer in the future climatic conditions (Paerl and Paul, 2012; Hense et al., 2013; Kahru et al., 2016; Olofsson et al., 2020). Indications of an increased total biomass of cyanobacteria in the phytoplankton

community exist in the northern sub-basins of the Baltic Sea (Lehtinen et al., 2016; Kuosa et al., 2017), whereas decreasing trends have been found in the southern Baltic (Wasmund et al., 2011; Griffiths et al., 2020). It is still not clear how the bloom-forming diazotrophic cyanobacteria influence overall phytoplankton community composition, food webs and trophic transfer in the open sea (Fig. 1).

The cyanobacteria community of the Baltic Sea is extremely diverse consisting of well over one hundred taxa (Hällfors, 2004), yet the late summer cyanobacterial communities in the offshore areas of the Bothnian Sea, the Baltic Proper, and the Gulf of Finland are generally dominated by the diazotrophic order Nostocales (Kanoshina et al., 2003; Andersson et al., 2015; Olli et al., 2015). Cyanobacteria produce several toxins with variable effects in aquatic ecosystems (Zanchett and Oliveira-Filho, 2013). The Baltic Sea diazotrophic cyanobacteria have been shown to have both allelopathic (growth-inhibiting) and growth-promoting effects on other groups of phytoplankton in experimental studies (Suikkanen et al., 2004, 2005). Allelopathy may play a role in the competition strategy of cyanobacteria for phosphorus, but its full role is not well understood (Leão et al., 2009). The molecular nitrogen fixed by diazotrophic cyanobacteria is released to the surrounding system, where it might promote the growth of bacteria and concomitant phytoplankton (Capone et al., 1997; Wannicke et al., 2013; Karlson et al., 2015). Both allelopathic and growth-promoting effects towards other systematic groups or individual species may lead to changes in phytoplankton community structure. However, no indications of predominantly positive or negative effects of the diazotrophic cyanobacterial genera *Nodularia* or *Aphanizomenon* on ambient phytoplankton species were found in a species-level analysis of a large multi-decadal monitoring data set from the Baltic Sea (Olli et al., 2015).

Cyanobacteria are one of the groups with potential direct and indirect effects on herbivorous zooplankton, therefore affecting the trophic transfer efficiency (Fig. 1). The organic matter produced by diazotrophic cyanobacteria is an important part of summer energy flow as it is a source of new production (molecular N-uptake) in the midst of production period mainly based on recycled nitrogen. Cyanobacteria may fuel the transfer of energy to higher trophic levels via the microbial loop, but this alternative route is not as efficient as the direct food chain, since it considerably increases food chain length (Berglund et al., 2007). The direct response of zooplankton to cyanobacteria may be species-specific and also depend on e.g. temperature and food availability (Gilbert, 1996a, 1996b). The Baltic planktonic *Nodularia spumigena* is known to produce a hepatotoxin, nodularin (Sivonen et al., 1989), which affects animals with variable efficiency and with both negative and positive effects on e.g. reproduction (Sellner et al., 1994; Koski et al., 1999; Engström et al., 2001; Lehtiniemi et al., 2002; Schmidt et al., 2002; Kozlowsky-Suzuki et al., 2003; Karjalainen et al., 2005; Korpinen et al., 2006; Hogfors et al., 2014). Based on experimental studies, even the toxic strains of *N. spumigena* do not necessarily directly harm e.g. the common copepod *Eurytemora affinis* (Gorokhova and Engström-Öst, 2009), and a mixture of phytoplankton species as a food source may counteract detrimental effects of toxin-producing cyanobacteria on copepods (Kozlowsky-Suzuki et al., 2003; Karjalainen et al., 2007). However, experiments have implied that a large proportion of Nostocales in the phytoplankton community might be harmful for copepod reproduction (Engström-Öst et al., 2015).

In freshwaters, cyanobacteria have been considered as poor food with potential toxicity as in marine waters, even though studies have also shown that cyanobacteria do not necessarily directly disturb herbivores (e.g. Work and Havens, 2003; Chislock et al., 2013; Tõnno et al., 2016). However, zooplankton biomass is found to be proportional to
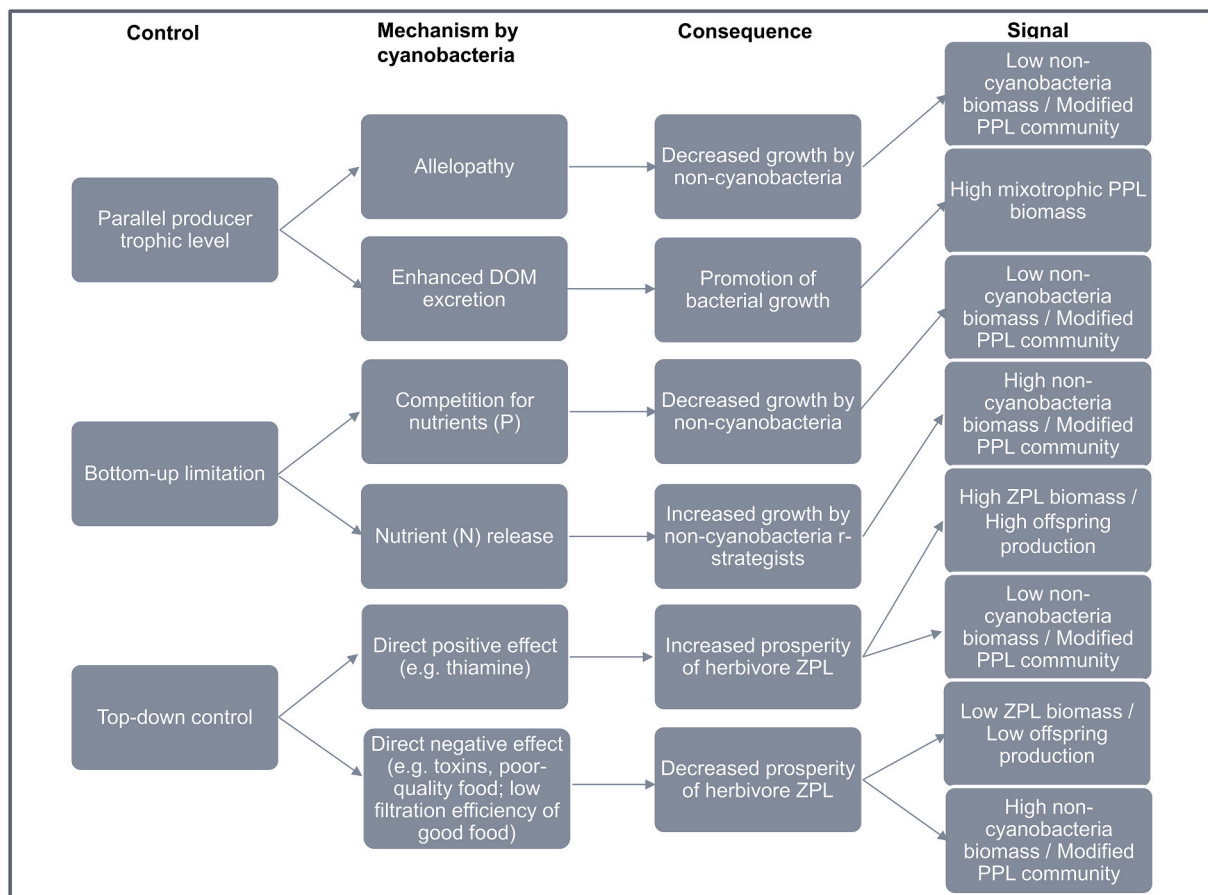

Fig. 1. Potential effects of diazotrophic cyanobacteria in the ecosystem of the open Baltic Sea.

total phytoplankton biomass only when large cyanobacteria are removed from the edible fraction of phytoplankton (Heathcote et al., 2016). In high-salinity marine systems, the biomass of cyanobacteria colonies is less significant than in fresh and brackish waters, apart from the genus *Trichodesmium*, which also largely remains ungrazed (Capone et al., 1997).

In an experimental study with freshwater taxa, increasing frequency and intensity of cyanobacteria have been shown to increase small-bodied zooplankters in the community (Jiang et al., 2014). In the Baltic Sea, a shift in the plankton community to smaller-sized organisms was attributed to complex interactions between warming, eutrophication and increased top-down pressure, and the resulting trophic cascades (Suikkanen et al., 2013). Zooplankton community structure, including mean size of zooplankters in the community, has been utilized in ecological indicator development and respective environmental status assessments to indicate both eutrophication and fish feeding conditions in the Baltic Sea (Gorokhova et al., 2016). Within that indicator approach, abundant zooplankton with high mean individual size represents both high grazing potential of zooplankton as well as favorable feeding conditions for fish (Gorokhova et al., 2016). Thus, a link between the abundance of cyanobacteria and zooplankton mean size would expand the potential links of this indicator to other anthropogenic effects in the sea i.e. cyanobacterial blooms.

In this article, we analyzed Baltic Sea multidecadal monitoring data to find indications of direct and indirect bottom-up connections of diazotrophic vs. non-diazotrophic cyanobacteria with zooplankton communities. In addition, we studied if there are connections of diazotrophic vs. non-diazotrophic cyanobacteria biomass with the other phytoplankton groups. Our aim was to test whether results obtained by experimental studies, concentrating especially on the effects of cyanobacteria on zooplankton, can be tracked down from field data as well. In most cases, however, the exact underlying mechanisms are difficult to pinpoint using long-term monitoring data. As the study only focused on investigating potential bottom-up effects of cyanobacteria on zooplankton, as well as on studying effects on concurrent phytoplankton community, the higher trophic levels or top-down effects were not included. Nevertheless, it is known that both top-down and bottom-up forces on zooplankton are important and that their relative strength can vary (Casini et al., 2009). Effects of top-down control by fish predation on zooplankton and plankton-fish interactions have already been assessed in our study area by e.g. Flinkman et al. (1998), Rönkkönen et al. (2004), Peltonen et al. (2007) and Kuosa et al. (2017).

Our hypotheses were that (1) diazotrophic cyanobacteria (class Nostocales) affect the concurrent phytoplankton community either by harming or benefiting taxa and simultaneously decreasing phytoplankton diversity, (2) potentially mixotrophic phytoplankton benefit from diazotrophic cyanobacteria by the promoted productivity of their food items, i.e. bacteria which are supported by the organic carbon release from the decaying cyanobacterial filaments, and (3) concurrent diazotrophic cyanobacterial biomass shows indications of affecting the zooplankton community either by harming or benefiting taxa, decreasing zooplankton diversity and their mean size. To reach this goal we statistically examined data with concurrent phyto- and zooplankton sampling for indications of effects by biomass of cyanobacteria on autotrophic, mixotrophic and total phytoplankton biomass and class-level phytoplankton kurtosis (a measure of peakedness of biomass distribution), as well as effects of cyanobacteria and concomitant phytoplankton on zooplankton mean size, production, kurtosis, and biomass of dominant zooplankton taxa. Effects of cyanobacteria on co-occurring phytoplankton were not studied on species level, since it has been recently done by Olli et al. (2015) using the Baltic Sea monitoring data.

Kurtosis was used as the diversity measure, since instead of species-level examination, we studied the possible effects of cyanobacteria on class-level phytoplankton diversity. The number of classes in samples is usually very constant within a study area, i.e. all classes are usually present in the samples even though their biomasses and biomass shares

vary (e.g. Andersson et al., 2017, and references therein) and thus utilizing class number as a diversity measure is not informative. Low kurtosis values reflect an even distribution of biomass and a high diversity, and increasing kurtosis reflects decreasing diversity (Gross et al., 2017). In addition, effects by physico-chemical variables on both phytoplankton and zooplankton variables were tested.

2. Material and methods

2.1. Study area

Our data originate from Finnish national monitoring stations in the northern and central Baltic Sea (Fig. 2). It includes 10 open sea stations that were sampled annually from 1979 to 2013, and one monitoring station in the SW middle archipelago, sampled ca. monthly during the open-water season from 1992 to 2013. The bottom depth of all stations is more than 30 m. Data from the northernmost sub-basin of the Baltic Sea, Bothnian Bay, were excluded from the analysis because the area is phosphorus-limited (Tamminen and Andersen, 2007), resulting in differences in plankton community structure compared with the more southern areas, e.g. diazotrophic cyanobacteria are not abundant in the open sea areas of the Bothnian Bay.

2.2. Data sets and variables

We used late summer (July–September) data of phytoplankton, zooplankton and hydrography, since cyanobacteria are most common in the study area during that period, and thus also the coupling between cyanobacteria and other phytoplankton as well as zooplankton is considered to be highest during this period. The samples were taken between 4 July and 28 September, but the majority (79%) of the samplings took place in August (8% were taken in in July and 13% in
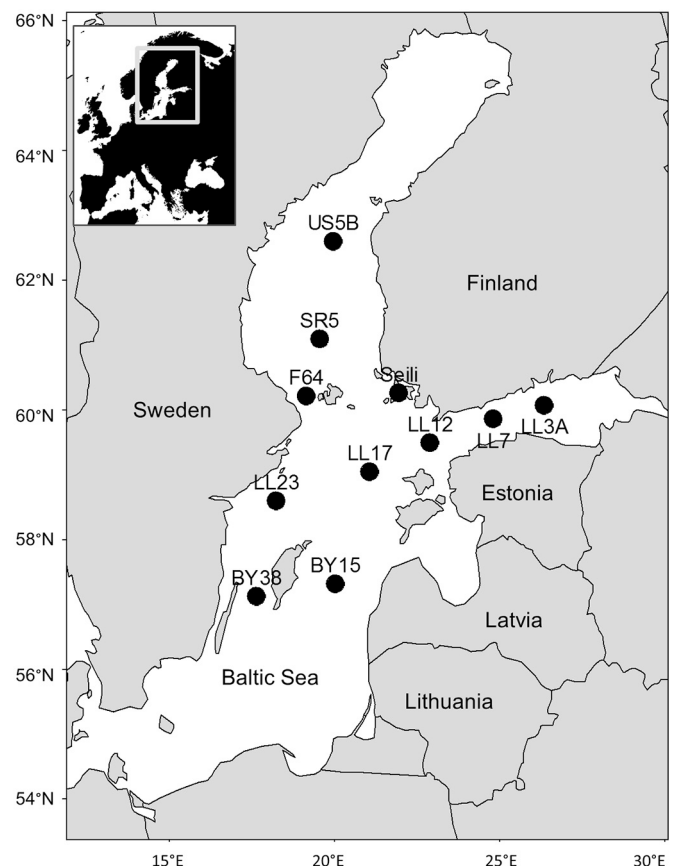

Fig. 2. Map of sampling stations.

September), the median sampling date being 18 August. There was no long-term trend in sampling dates over the study period. The risk born by static sampling in relation to potentially changed bloom phenology (Kahru et al., 2016) was not considered significant, as the purpose of the study was to examine potential interactions between cyanobacteria and the rest of the plankton community, no matter the seasonal succession phase of the phytoplankton community.

The different types of data (phytoplankton, zooplankton, hydrography) were combined for each sampling station and time. If sampling dates differed for the data types, the data were matched so that there was a maximum of 14 days between the different types of data. In practice, hydrography data were not always available for the date when a plankton sample was taken. In such cases, the closest date where hydrography data were available was selected, and those data were used as a substitute. The selection was done so that a maximum difference between the dates was at most 14 days. In 259 out of 297 cases (87%), all three types of data were from the same date. For the analyses, data from all stations were pooled. The combination of all data was done despite the slightly differing zooplankton communities between the Baltic Proper (with a lower share of cladocerans and a higher share of marine copepods) and the Gulf of Finland, including the archipelago station Seili, as the aim was to search for more general indications of effects of cyanobacteria on phytoplankton and zooplankton communities in the field data, even visible over slightly different communities or areas.

Phytoplankton was sampled, fixed, stored, and counted following the guidelines described in the latest version of the Manual for Marine Monitoring in the COMBINE program of the Helsinki Commission (HELCOM) (current version: HELCOM, 2017). In the open sea stations, phytoplankton was analyzed from pooled 0–10 m surface water samples. In the archipelago station, pooled samples were taken from the surface down to twice the Secchi depth (max 10 m) on each sampling occasion. Phytoplankton samples were fixed with acid Lugol's solution, counted using inverted light microscopy (Utermöhl, 1958), and species-specific biomasses were calculated according to Olenina et al. (2006). Species nomenclature follows the HELCOM PEG list version 2012 (available at http://ices.dk/marine-data/Documents/ENV/PEG_BVOL.zip). Biomasses were reported as µg L^{-1} wet weight.

The following phytoplankton variables were used in the statistical analysis: biomasses of the diazotrophic Nostocales, non-diazotrophic cyanobacteria, division Chlorophyta (including classes Prasinophyceae, Charophyceae and Chlorophyceae), classes Chrysophyceae, Cryptophyceae, Diatomophyceae, Dinophyceae, Euglenophyceae, Prymnesiophyceae, autotrophs excluding cyanobacteria, mixotrophs, total phytoplankton (autotrophs and mixotrophs) excluding cyanobacteria, and community class-level kurtosis excluding cyanobacteria, unidentified taxa and the autotrophic ciliate *Mesodinium rubrum* (only counted since 1986). Heterotrophic taxa (according to the HELCOM PEG list version 2012) were excluded from the phytoplankton data set. Kurtosis is a measure of diversity and represents the relative peakedness of the taxon-specific biomass distribution and the heaviness of its tails. High kurtosis values characterize peaked biomass distributions, indicating a high number of taxa with similar mean biomass values, and therefore reflect a low diversity. Low kurtosis values reflect an even distribution of biomass and a high diversity (Gross et al., 2017). Class-level phytoplankton data were used as species-level analysis of the effects of cyanobacteria on phytoplankton was already published by Olli et al. (2015). In addition, class-level phytoplankton data have been used in phytoplankton-based indicators developed for marine status assessments (Wasmund et al., 2017).

In the open sea stations, zooplankton samples were taken from the whole water column (sampling depth varying from ca. 0–200 m to 0–50 m) using vertical tows of a WP-2 net (mesh 100 µm), equipped with a flow meter, and fixed with formaldehyde (HELCOM, 2017). In the archipelago station, the samples were taken from 0 to 25 m depth (total depth 50 m) using a 150 µm-mesh net and fixed as described above. The samples were counted with an inverted microscope to the most accurate

taxonomic level possible (HELCOM, 1988). Zooplankton abundance was calculated per volume (m^3) of water. Biomass was computed using species-, sex-, and state-specific individual biomass values (Hernroth, 1985).

Zooplankton variables used in the statistical analysis included zooplankton mean size, copepod nauplius-to-female ratio, zooplankton community species-level kurtosis, and biomasses of rotifers (Rotatoria), cladocerans *Bosmina (Eubosmina) coregoni maritima* and *Cercopagis pengoi*, copepods *Acartia* spp., *Eurytemora* spp., *Limnocalanus macrurus* and *Pseudocalanus elongatus*, and the combined biomass of all copepod nauplii. Mean size of zooplankton (µg individual^{-1}) was calculated by dividing total zooplankton biomass by total zooplankton abundance (McCune and Grace, 2002). Copepod nauplius-to-female ratio was used to indicate the reproduction potential of copepods.

Data of the mean surface water (0–10 m) temperature, salinity, total phosphorus and chlorophyll *a* concentration were additionally used in the analysis. Surface layer data were chosen as the phytoplankton data are from that layer only. In the open sea stations, temperature and salinity were measured using an SBE 911plus CTD system. In the archipelago station, this data originate from discrete water samples collected using a Limnos water sampler. The total phosphorus concentration was measured spectrophotometrically (Grasshoff et al., 1999; Koroleff, 1979), with a detection limit of 0.01 µM. Chlorophyll *a* concentration was measured with a spectrofluorometer according to the standard (HELCOM, 1988).

2.3. Analyses

13 response variables were investigated by random forest regression, to identify the most important explaining variables predicting their value (Table 1). The potential explaining variables were pre-selected in the following manner: Phytoplankton variables were explained using physico-chemical and zooplankton variables, biomass of Nostocales and other cyanobacteria, year, latitude, and longitude of the sampling station. Zooplankton variables were predicted using physico-chemical and phytoplankton variables, sampling year, latitude and longitude. The zooplankton species *Cercopagis pengoi* and *Limnocalanus macrurus* were not evaluated as response variables because the former is a predatory species and the latter occurs in deeper water layers than cyanobacteria, and both are thus less likely to be influenced by cyanobacteria.

A separate random forest regression model (Breiman, 2001) was fit for each response variable. Random forest is a relatively simple, yet powerful method for finding the best predictors and using them to predict the value of the response variable. Random forest regression is based on a large number of regression trees, each of which is allowed to evaluate a subset of the predictors; this helps avoid the problem of high correlation between the trees and makes the result more reliable. A regression tree splits the data into distinct sub-parts through searching repeatedly for the best predictor variable and the best cut point that minimizes the variance within the two sides of the split (James et al., 2013), and finally predicts the mean value of observations in the region where the new observation falls.

Using a large number of predictors may lead to overfitting, i.e. reducing the error in the data set that is used for fitting the model, but actually increasing the error when predicting new data points. Random forests avoid overfitting through using a subset of data for each tree and using the rest of the data (out-of-bag) to evaluate the prediction accuracy of the tree. This way, the goodness of fit of the model can be estimated using data that has not been used for fitting the model, providing a fair estimate of the prediction accuracy. The optimal number of explaining variables, as well as the most important explaining variables, were found using the out-of-box sample and root mean squared error (RMSE) metric as the measure of the goodness of fit. In addition, the coefficients of determination (R^2) of the models were computed.

Random forest regression was run using 100,000 trees for each response variable. The high number of trees was needed to ensure the

Table 1

Results of the random forest and linear regression analyses.

	Au bm	Mx bm	Tot bm	PPL kurtosis	ZPL size	N:F ratio	ZPL kurtosis	Rotatoria	*Bosmina*	Nauplii	*Acartia*	*Eurytemora*	*Pseudocalanus*
Year	2	3		1	2		1						4
Latitude	12	8				1	9		1	4	3		1
Longitude	15						3		3	2	1	1	3
Temperature	1		1				8		2				
Salinity	7	1	6			1	2			5	3	6	2
Total phosphorus	3		2				7				2		8
Nostocales	16												
Other cyanobacteria	6		7							1			
Chlorophyta													
Chrysophyceae							6						5
Cryptophyceae						2	4		4		5	5	
Diatomophyceae													
Dinophyceae													6
Euglenophyceae													
Prymnesiophyceae										3		2	9
Autotroph biomass							5						
Mixotroph biomass												4	7
Total biomass													
Chlorophyll *a*								1					
PPL curtosis													
ZPL size	8		12										
N:F ratio	11		10										
ZPL kurtosis													
Rotatoria	9		13										
Bosmina spp.	14												
Cercopagis spp.													
Nauplii	5		5										
Acartia spp.	13		9										
Eurytemora spp.	17	2											
Limnocalanus spp.	10		11										
Pseudocalanus spp.	4		4										

Rows = explaining variables, columns = response variables. All explaining variables considered for each response variable are marked in grey. White cells indicate relationships that were not investigated because the variables are interdependent. Significant explaining variables based on random forest analysis are marked with numbers describing their decreasing order of importance. Significant linear regressions ($p < 0.05$) are marked with red (positive regression) or blue (negative regression). Au bm = autotrophic phytoplankton biomass, Mx bm = mixotrophic phytoplankton biomass, Tot bm = total phytoplankton biomass.

stability of the results. The computations were implemented in R environment (R Core Team, 2018) using the *caret* package (Kuhn, 2008); R markdown files (Xie, 2014, 2015, 2018) containing the scripts and the results are available in https://github.com/luusitalo/ZplCyanoRF.

Random forest regression is not limited to finding linear or other functional form correlations between the variables; due to iteratively splitting the space in order to find subspaces with high explanation power, it can find highly non-linear data structures that are missed by traditional parametric methods but which can be used to predict the value of the response variable with high accuracy. Therefore, the variables identified as important explaining variables may not be those that are identified through e.g. regression analysis. Also, it is not in general possible to say whether the relationship between the explaining and response variable is positive or negative, as this may vary in different subspaces, and depending on other explaining variables. The random forest method has been used earlier as a machine learning technique (e. g. García-Comas et al., 2011; Verikas et al., 2012; Thomas et al., 2018),

but also in freshwater plankton studies (Kruk and Segura, 2012; Pomati et al., 2020), and marine macrofaunal studies (Wei et al., 2012). Concerning ecological studies focusing on the Baltic Sea ecosystem, the random forest method has been used at least in studying spatial prediction of fish diversity (Smoliński and Radtke, 2016) and the correlation between environmental factors and genetic differentiation of mussel populations (Kijewski et al., 2019). Therefore, the random forest method has been found suitable for several types of community studies such as the present one.

In order to help the intuitive interpretation of the results and shed additional light on the type of interaction between these variables, a simple linear regression was run between the response variable and those explaining variables that were deemed relevant for the response variable in the random forest analysis. In some cases, different variables were found important linear regression and random forest regression (Table 1). This is likely caused by nonlinear interactions between the response variable and the explanatory variables. It is also possible that

some of the explanatory variables are good predictors together, but less so individually.

Long-term trends in autotrophic, mixotrophic and total phytoplankton biomass, phytoplankton kurtosis, zooplankton mean size, nauplius-to-female ratio, and kurtosis were analyzed using generalized additive models (GAM) (R package *mgcv*, Wood, 2014). For this, the data were first centered for sea areas by subtracting the sea area mean from each value, and finally data from all sea areas were pooled for analyzing time series over the entire study area. Curves estimated with GAM and their 95% confidence intervals were plotted on the data for visualizing the significant long-term variation in the study area (Fig. 3). Long-term trends for other considered variables have been published elsewhere (Suikkanen et al., 2013; Lehtinen et al., 2016; Kuosa et al., 2017).

3. Results

In our data, the mean proportion of cyanobacteria biomass of the total phytoplankton biomass varied from 20% at Seili (Archipelago Sea) to 43% at LL12 (Northern Baltic Proper). The total cyanobacteria biomass included diazotrophic Nostocales (on average 90%) and non-diazotrophic cyanobacteria (10%). These shares varied from an average of 80% Nostocales and 20% non-diazotrophic cyanobacteria at Seili to 95% Nostocales and 5% non-diazotrophic cyanobacteria at F64 (Åland Sea). Further, Nostocales biomass was composed of the genera *Aphanizomenon* (78%), *Nodularia* (17%) and *Dolichospermum* (5%). The biomass of non-diazotrophic cyanobacteria was formed by species of the orders Chroococcales (77%) and Oscillatoriales (23%). Based on the average biomass, the most common genera of Chroococcales were *Snowella* (48% of total Chroococcales biomass), *Lemmermanniella* (8%) and *Aphanocapsa* (7%), whereas *Pseudanabaena* was the main genus within Oscillatoriales (67% of total Oscillatoriales biomass).

Results of the random forest and linear regression analyses are shown in Table 1 and in https://github.com/luusitalo/ZplCyanoRF. In Table 1, significant explaining variables for each response variable, based on random forest analysis were marked with numbers describing their order of importance, and significant linear regressions ($p < 0.05$) were marked with red (positive regression) or blue (negative regression) color.

Based on random forest analysis, Nostocales biomass, as well as the biomass of other non-diazotrophic cyanobacteria, explained the concomitant autotrophic biomass (excluding cyanobacteria) (Table 1). However, biomass of Nostocales was only 16th in the order of importance in explaining the concomitant autotrophic biomass, while the other cyanobacteria were 6th in the same order of importance (Table 1). In addition, random forest analysis indicated that biomass of other cyanobacteria was significant in explaining total phytoplankton biomass (excluding cyanobacteria), while biomass of Nostocales was not (Table 1). Biomass of mixotrophic phytoplankton or phytoplankton

kurtosis were not explained by the biomass of Nostocales or by the biomass of other cyanobacteria based on random forest analysis (Table 1). Linear regression analyses did not indicate any significant relations between biomass of Nostocales or other cyanobacteria and concomitant autotrophic biomass, mixotrophic biomass, total biomass, or kurtosis of phytoplankton (Table 1).

In the random forest analysis, the autotrophic biomass was significantly related to all but two explaining variables (zooplankton kurtosis and biomass of *Cercopagis pengoi*), but linear regressions were significant only for temperature, copepod nauplii and *Pseudocalanus* (negative), as well as total phosphorus and *Acartia* spp. (positive). Similarly, the total phytoplankton biomass was significantly related to 13 of the 19 investigated variables in random forest analysis, and showed significant linear regression with temperature, copepod nauplius-to-female ratio, copepod nauplii, *Limnocalanus* and *Pseudocalanus* (negative), as well as total phosphorus and *Acartia* spp. (positive). Mixotrophic biomass was significantly connected to salinity and *Eurytemora* spp. in random forest analysis, but only the latter connection was significant as linear regression (positive). Phytoplankton kurtosis was related to sampling year in random forest, but it had no significant linear regressions. There was also a significant decreasing trend of phytoplankton kurtosis between 1979 and 2013 ($p = 0.030$, Fig. 3). None of the other phytoplankton variables examined (autotrophic, mixotrophic and total phytoplankton biomass) showed significant long-term trends in the study period ($p < 0.05$).

Based on both random forest and linear regression analyses, neither biomass of Nostocales nor the biomass of other cyanobacteria explained mean size of zooplankton, copepod nauplius:female ratio, or zooplankton kurtosis (Table 1). Only the biomass of the cladoceran *Bosmina coregoni maritima* was connected to biomass of non-diazotrophic cyanobacteria based on random forest analysis, but not based on linear regression analysis (Table 1). Significant relationships were found between zooplankton size and sampling year and salinity in random forest analysis. Of these, there was also a significant negative linear regression between zooplankton size and sampling year. The long-term trend analysis additionally confirmed the significant decreasing trend in zooplankton mean size in the study area between 1979 and 2013 ($p = 0.029$, Fig. 3). Copepod nauplius:female ratio was significantly connected to latitude and biomass of Cryptophyceae according to random forest analysis, but these connections were not significant according to the linear regression analysis. There was also not a significant long-term trend in the nauplius:female ratio ($p = 0.325$). Zooplankton kurtosis was linked to sampling year, salinity, longitude, biomass of Cryptophyceae, Chrysophyceae and total autotrophs, total phosphorus, temperature and latitude in random forest regression, but significant linear regressions were only shown by year, temperature, longitude, total phosphorus (positive) and salinity (negative). There was also a significant increasing trend of zooplankton kurtosis between 1979 and
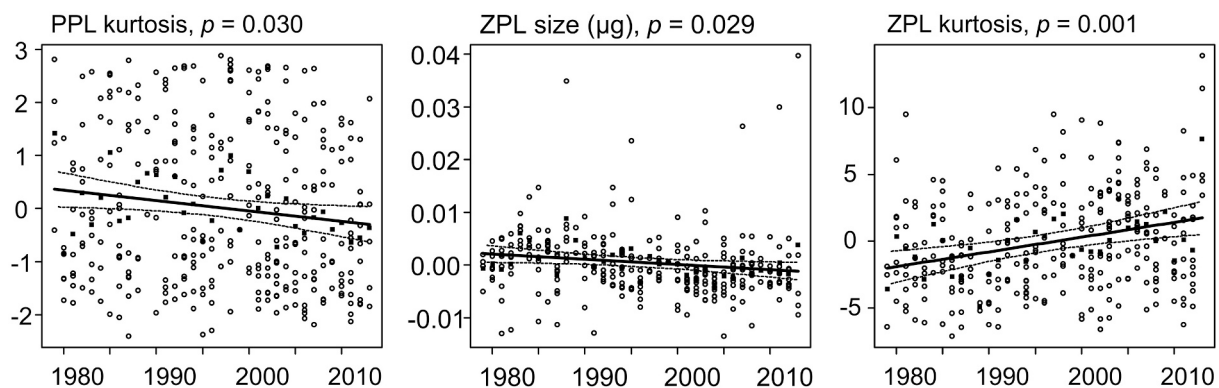

Fig. 3. Significant ($p < 0.05$) long-term trends in zooplankton size, zooplankton kurtosis and phytoplankton kurtosis in the study area. A GAM curve (solid line) is plotted with a 95% confidence interval (dashed line). Raw data are plotted as open circles and annual averages as filled circles.

2013 ($p = 0.001$, Fig. 3).

For the individual zooplankton taxa, there was a significant connection between the biomass of Rotatoria and chlorophyll *a* concentration in random forest analysis, but not a significant linear regression. The biomass of *Bosmina coregoni maritima* was connected to the biomass of non-diazotrophic cyanobacteria, temperature and longitude in random forest. There was also a significant positive linear regression between the biomass of *B. coregoni maritima* and both temperature and longitude. Biomass of copepod nauplii was related to latitude, longitude, biomass of Prymnesiophyceae and Cryptophyceae, as well as salinity in random forest. Of these, the relationships with latitude (negative), longitude (positive) and Cryptophyceae (negative) were significant based on linear regression analysis. *Acartia* spp. had significant random forest connections with longitude, total phosphorus concentration, salinity, latitude, and the biomass of Cryptophyceae, linear regressions being significant and positive with all but latitude. Biomass of *Eurytemora* spp. was significantly connected to longitude, Prymnesiophyceae, latitude, mixotroph biomass, Cryptophyceae and salinity, all connections being significant as linear regressions as well, and positive except for the regression with salinity. *Pseudocalanus elongatus* was related to latitude, salinity, longitude, year, Chrysophyceae, Dinophyceae, mixotroph biomass, total phosphorus and Prymnesiophyceae in random forest, and all but Dinophyceae and total phosphorus also had a significant regression with *P. elongatus* biomass. Of the significant regressions, all were negative except for salinity.

4. Discussion

Based on the current, >30-year monitoring data, potential direct and indirect bottom-up relationships between cyanobacteria and zooplankton, as well as between cyanobacteria and other phytoplankton taxa summarized in Fig. 1 did rarely manifest in the statistical analysis. A recurring challenge with environmental data analysis is that even a relatively rich data are often sparse when compared to the spatial, temporal, and structural complexity of the system. Different data analyses, lab experiments, and theoretical considerations all provide a different lens and a different viewpoint to the system, and through combining the knowledge created from all of these angles, a better picture emerges. The aim of this work was to provide one such lens and viewpoint to add to the other studies, from the perspective of monitoring data. However, the information gained from this analysis must be evaluated together with the information gained from earlier studies taking different viewpoints.

Random forest analysis revealed a connection between Nostocales biomass and the biomass of the autotrophic phytoplankton community (all cyanobacteria excluded). This connection may be explained by Nostocales and other autotrophic phytoplankton being favored by the same environmental conditions including e.g. temperature, light, mixing, and nutrient availability. In addition, since growth of autotrophic phytoplankton communities is usually N-limited during summer in the study area, the results may also indicate that the atmospheric N_2 fixed by diazotrophic cyanobacteria was transferred through the food web to coexisting primary producers and thus supported the increase of also other autotrophic phytoplankton biomass. Evidence for such transfer of the N_2 fixed by diazotrophic cyanobacteria has been found empirically using e.g. ^{15}N tracers (e.g. Ohlendieck et al., 2000; Stal et al., 2003; Ploug et al., 2011). N_2-fixation by diazotrophic cyanobacteria probably plays a significant role in shaping the plankton communities of the Baltic Sea: it has been estimated that 14.3 to 279 mmol N m^{-2} is fixed in the Baltic Proper annually (Karlson et al., 2015 and references therein), and that up to 90% of the fixed N_2 is transferred through the microbial loop from picoplankton towards the mesozooplankton community (e.g. Mulholland and Capone, 2001; Ploug et al., 2010, 2011; Wannicke et al., 2013). Even non-diazotrophic cyanobacteria were significantly connected to the biomass of autotrophs and the total phytoplankton community (all cyanobacteria excluded). The reason for this may be that

non-diazotrophic cyanobacteria are also limited by available dissolved nutrients and benefit from N_2-fixation by diazotrophic cyanobacteria similarly to the other autotrophic phytoplankton.

In our study, we did not investigate the entire food web or top-down effects on either phytoplankton or zooplankton. However, it is known that phytoplankton communities are directly affected by zooplankton grazing as well as by indirect top-down trophic cascading effects caused by changes in feeding pressure of predatory zooplankton on herbivorous zooplankton and microzooplankton (Kivi et al., 1996; Sommer, 2008; Lehtinen et al., 2010). Since our results did not show similar connections between Nostocales biomass and zooplankton as between Nostocales biomass and other phytoplankton, it can be inferred that there were no indications of Nostocales biomass being connected to potential changes in the top-down pressure by zooplankton or by trophic cascading effects via microzooplankton on phytoplankton.

Based on our results, the increased biomass of Nostocales cyanobacteria did not change the concurrent phytoplankton community either by harming or benefiting any of the class-level phytoplankton taxa. Thus, our class-level result supports the earlier findings of a species-level study showing that diazotrophic cyanobacteria do not have predominantly positive or negative effects on ambient phytoplankton species (Olli et al., 2015). Concerning multi-decadal changes, phytoplankton class-level kurtosis significantly decreased in the study period, indicating an increased diversity of phytoplankton. The result is in line with earlier results by Olli et al. (2014), who found an increase in Baltic Sea phytoplankton genus richness during 1966–2008. Olli et al. (2014) concluded that it potentially reflects a delayed long-term response to the anthropogenic fertilization.

Our hypothesis that potentially mixotrophic phytoplankton would benefit from the increased biomass of Nostocales cyanobacteria was not supported either. It was hypothesized that increasing cyanobacterial biomass increases the productivity of bacteria due to organic carbon release, which would benefit bacterivorous mixotrophic taxa (Rolff, 2000; Bunse et al., 2019). Mixotrophic phytoplankton biomass was significantly connected to salinity and *Eurytemora* spp. in random forest analysis, and there was a significant (positive) regression between the biomass of mixotrophs and *Eurytemora* spp. One explanation to this may be a top-down trophic cascading effect. *E. affinis* has been found to clear and ingest ciliate prey at a higher rate than phytoplankton in experimental studies (Merrell and Stoecker, 1998), and when *Eurytemora* spp. feed effectively on ciliates, they decrease the grazing pressure of ciliates on nanoflagellates, in which mixotrophy is a very common trait in the study area. Another potential reason for the positive connection between *Eurytemora* spp. and mixotrophic phytoplankton could be that mixotrophs can utilize bacteria as food (Kuuppo-Leinikki, 1990), and bacteria are known to benefit nutritionally from the sloppy feeding of mesozooplankton (Vargas et al., 2007).

Based on the random forest analysis, autotrophic phytoplankton biomass was significantly related to all but two explaining variables. Similarly, total phytoplankton biomass was significantly related to 13 of the 19 investigated variables in random forest analysis. However, in linear regressions both autotrophic and total phytoplankton biomasses were significantly related only to temperature, copepod nauplii, *Pseudocalanus,* total phosphorus and *Acartia* spp., and total phytoplankton biomass to copepod nauplius-to-female ratio and *Limnocalanus*. Based on our data, phytoplankton class-level or total biomasses were positively or not at all related to the dominating zooplankton of the area, e.g. cladocerans, *Acartia* spp. and *Eurytemora* spp. Thus, the results indicate that the direct top-down pressure of zooplankton on phytoplankton was not very high, but on the other hand, there may have been positive bottom-up effects by phytoplankton on *Acartia* spp. (cryptophytes) and *Eurytemora* spp. (cryptophytes and prymnesiophytes). In addition, a top-down trophic cascading effect could be one explanation for the positive relation, i.e. copepods may have grazed ciliates and thus relieved these nanoflagellate-sized phytoplankton from ciliate crazing pressure. These results demonstrate that the random forest analysis, not being

restricted to linear (or transformed linear) relationships, can help identify potentially important, non-linear interactions that are not revealed through regression techniques such as principal component analysis, which assume linearity.

Zooplankton variables were connected to the biomass of phytoplankton groups other than cyanobacteria (such as chrysophytes, cryptophytes and prymnesiophytes). The significant connections between the biomass of rotifers and chlorophyll *a* and between the biomass of the cladoceran *Bosmina coregoni maritima* and biomass of non-diazotrophic cyanobacteria and temperature seem to well support the knowledge that these species with short generation times and capable of rapid reproduction under optimal conditions (warm water and lot of food indicated by e.g. chlorophyll *a*) thrive in eutrophic ecosystems (Orcutt and Pace, 1984; Heerkloss et al., 1991).

During the study period, zooplankton kurtosis increased, but it was not significantly connected to cyanobacteria biomass. The mean size of zooplankton significantly decreased in the study period, and based on the random forest analysis, the decrease was connected to salinity. This is due to the decrease of large marine copepods with the decreasing salinities (Mäkinen et al., 2017). The increase of smaller-sized zooplankters has been earlier connected with e.g. warming and eutrophication, and increasing cyanobacteria (Suikkanen et al., 2013; Jiang et al., 2014). The increase of small-sized zooplankton species is linked to stimulation of the microbial loop by increased nutrient intake through cyanobacterial exudates, resulting in an increase of bacterio- and nanoplankton prey (Motwani and Gorokhova, 2013). However, our monitoring data did not show a direct connection, and the effect of salinity should be analyzed further, potentially finding more evidence on riverine inputs of organic matter and its effect on bacterial productivity. By using a monitoring data set that also includes fish data the relative strength of top-down effects caused by fish predation vs. bottom-up effects caused by cyanobacteria on zooplankton could be further investigated.

Engström-Öst et al. (2015) found that the reproductive output and population dynamics of *Acartia* may be negatively affected by *N. spumigena*. Our data were strongly dominated by *Aphanizomenon* sp., which complicates observation of potential effects of other diazotrophic species on copepod reproduction. Copepod nauplius:female ratio was not connected to diazotrophic cyanobacteria in the random forest analysis but to latitude and biomass of Cryptophyceae. However, these connections were likely non-linear and difficult to interpret. E.g. compound-specific isotopes might be utilized to shed light on the potentially significant, non-linear food web interactions (Ek et al., 2018).

In experimental studies, negative effects of cyanobacteria on phyto- and zooplankton have been found (summarized e.g. for the Baltic Sea by Karjalainen et al., 2007), but such effects were not found in the monitoring data by Olli et al. (2015) or us. Experimental studies may not reliably represent the natural situation where e.g. zooplankton can avoid cyanobacteria blooms to some extent, at least by selecting their food items or actively moving in the water mass. Experimental studies reveal relevant potential relationships, as direct causes and effects cannot be easily extracted from monitoring data. Monitoring data include all potential relationships at the same time, and thus only the strongest may be revealed with statistical analysis. However, it can be concluded from our data that, considering the planktonic system of the northern Baltic Sea, cyanobacterial blooms do not have a significant effect on either co-occurring phyto- or zooplankton communities. More information on bacterial productivity and other parts of the microbial loop must be collected to reveal food web and trophic transfer effects in situ. These data sets can be smaller in spatial extent but should be intensively sampled to follow the rapidly variating processes.

Declaration of Competing Interest

There are no conflicts of interest concerning our article.

References

Andersson, A., Höglander, H., Karlsson, C., Huseby, S., 2015. Key role of phosphorus and nitrogen in regulating cyanobacterial community composition in the northern Baltic Sea. Estuar. Coast. Shelf Sci. 164, 161–171. https://doi.org/10.1016/j.ecss.2015.07.013.

Andersson, A., Tamminen, T., Lehtinen, S., Jürgens, K., Labrenz, M., Viitasalo, M., 2017. The pelagic food web. In: Snoeijs-Leijonmalm, P., Schubert, H., Radziejewska, T. (Eds.), Biological Oceanography of the Baltic Sea. Springer Nature, The Netherlands, pp. 281–332.

Backer, H., Leppänen, J.-M., Brusendorff, A.C., Forsius, K., Stankiewicz, M., Mehtonen, J., Pyhälä, M., Laamanen, M., Paulomäki, H., Vlasov, N., Haaranen, T., 2010. HELCOM Baltic Sea Action Plan – a regional programme of measures for the marine environment based on the Ecosystem Approach. Mar. Pollut. Bull. 60, 642–649. https://doi.org/10.1016/j.marpolbul.2009.11.016.

Barton, A.D., Pershing, A.J., Litchman, E., Record, N.R., Edwards, K.F., Finkel, Z.V., Kiørboe, T., Ward, B.A., 2013. The biogeography of marine plankton traits. Ecol. Lett. 16, 522–534. https://doi.org/10.1111/ele.12063.

Berglund, J., Müren, U., Båmstedt, U., Andersson, A., 2007. Efficiency of a phytoplankton-based and a bacteria-based food web in a pelagic marine system. Limnol. Oceanogr. 52, 121–131. https://doi.org/10.4319/lo.2007.52.1.0121.

Bianchi, T.S., Engelhaupt, E., Westman, P., Andrén, T., Rolff, C., Elmgren, R., 2000. Cyanobacterial blooms in the Baltic Sea: natural or human-induced? Limnol. Oceanogr. 45, 716–726. https://doi.org/10.4319/lo.2000.45.3.0716.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Bunse, C., Israelsson, S., Baltar, F., Bertos-Fortis, M., Fridolfsson, E., Legrand, C., Lindehoff, E., Lindh, M.V., Martinez-Garcia, S., Pinhassi, J., 2019. High frequency multi-year variability in Baltic Sea microbial plankton stocks and activities. Front. Microbiol. 9, 3296. https://doi.org/10.3389/fmicb.2018.03296.

Capone, D.G., Zehr, J.P., Paerl, H.W., Bergman, B., Carpenter, E.J., 1997. *Trichodesmium*, a globally significant marine cyanobacterium. Science 276, 1221–1229. https://doi.org/10.1126/science.276.5316.1221.

Casini, M., Hjelm, J., Molinero, J.C., Lövgren, J., Cardinale, M., Bartolino, V., Belgrano, A., Kornilovs, G., 2009. Trophic cascades promote threshold-like shifts in pelagic marine ecosystems. Proc. Natl. Acad. Sci. U. S. A. 106, 197–202. https://doi.org/10.1073/pnas.0806649105.

Chislock, M.F., Sarnelle, O., Olsen, B.K., Doster, E., Wilson, A.E., 2013. Large effects of consumer offense on ecosystem structure and function. Ecology 94, 2375–2380. https://doi.org/10.1890/13-0320.1.

Ek, C., Holmstrand, H., Mustajärvi, L., Garbaras, A., Barisevičiu̅te, R., Šapolaitè, J., Sobek, A., Gorokhova, E., Karlson, A.M.L., 2018. Using compound-specific and bulk stable isotope analysis for trophic positioning of bivalves in contaminated Baltic Sea sediments. Environ. Sci. Technol. 52, 4861–4868. https://doi.org/10.1021/acs.est.7b05782.

Engström, J., Viherluoto, M., Viitasalo, M., 2001. Effects of toxic and non-toxic cyanobacteria on grazing, zooplanktivory and survival of the mysid shrimp *Mysis mixta*. J. Exp. Mar. Biol. Ecol. 257, 269–280. https://doi.org/10.1016/S0022-0981(00)00339-7.

Engström-Öst, J., Brutemark, A., Vehmaa, A., Motwani, N.H., Katajisto, T., 2015. Consequences of a cyanobacteria bloom for copepod reproduction, mortality and sex ratio. J. Plankton Res. 37, 388–398. https://doi.org/10.1093/plankt/fbv004.

Finkel, Z.V., Beardall, J., Flynn, K.J., Quigg, A., Alwyn, T., Rees, V., Raven, J.A., 2010. Phytoplankton in a changing world: cell size and elemental stoichiometry. J. Plankton Res. 32, 119–137. https://doi.org/10.1093/plankt/fbp098.

Flinkman, J., Aro, E., Vuorinen, I., Viitasalo, M., 1998. Changes in northern Baltic zooplankton and herring nutrition from 1980s to 1990s: top-down and bottom-up processes at work. Mar. Ecol. Prog. Ser. 165, 127–136. https://doi.org/10.3354/meps165127.

Funkey, C.P., Conley, D.J., Reuss, N.S., Humborg, C., Jilbert, T., Slomp, C.P., 2014. Hypoxia sustains cyanobacteria blooms in the Baltic Sea. Environ. Sci. Technol. 48, 2598–2602. https://doi.org/10.1021/es404395a.

García-Comas, C., Stemmann, L., Ibanez, F., Berline, L., Mazzocchi, M.G., Gasparini, S., Picheral, M., Gorsky, G., 2011. Zooplankton long-term changes in the NW Mediterranean Sea: decadal periodicity forced by winter hydrographic conditions related to large-scale atmospheric changes? J. Mar. Syst. 87, 216–226. https://doi.org/10.1016/j.jmarsys.2011.04.003.

Gilbert, J.J., 1996a. Effect of temperature on the response of planktonic rotifers to a toxic cyanobacterium. Ecology 77, 1174–1180. https://doi.org/10.2307/2265586.

Gilbert, J.J., 1996b. Effect of food availability on the response of planktonic rotifers to a toxic strain of the cyanobacterium *Anabaena flos-aquae*. Limnol. Oceanogr. 41, 1565–1572. https://doi.org/10.4319/lo.1996.41.7.1565.

Gorokhova, E., Engström-Öst, J., 2009. Toxin concentration in *Nodularia spumigena* is modulated by mesozooplankton grazers. J. Plankton Res. 31, 1235–1247. https://doi.org/10.1093/plankt/fbp060.

Gorokhova, E., Lehtiniemi, M., Postel, L., Rubene, G., Amid, C., Lesutiene, J., Uusitalo, L., Strake, S., Demereckiene, N., 2016. Indicator properties of Baltic zooplankton for classification of environmental status within Marine Strategy Framework Directive. PLoS One 11 (7), e0158326. https://doi.org/10.1371/journal.pone.0158326.

Grasshoff, K., Kremling, K., Ehrhardt, M., 1999. Methods of Seawater Analysis. Wiley-VCH Verlag, Weinheim.

Griffiths, J.R., Lehtinen, S., Suikkanen, S., Winder, M., 2020. Limited evidence for common interannual trends in Baltic Sea summer phytoplankton biomass. PLoS One 15 (4), e0231690. https://doi.org/10.1371/journal.pone.0231690.

Gross, N., Bagousse-Pinguet, Y., Liancourt, P., Berdugo, M., Gotelli, N.J., Maestre, F.T., 2017. Functional trait diversity maximizes ecosystem multifunctionality. Nat. Ecol. Evol. 1, 0132. https://doi.org/10.1038/s41559-017-0132.

Gustafsson, B.G., Schenk, F., Blenckner, T., Eilola, K., Meier, H.E.M., Müller-Karulis, B., Neumann, T., Ruoho-Airola, T., Savchuk, O.P., Zorita, E., 2012. Reconstructing the development of Baltic Sea eutrophication 1850–2006. Ambio 41, 534–548. https://doi.org/10.1007/s13280-012-0318-x.

Hällfors, G., 2004. Checklist of Baltic Sea phytoplankton species (including some heterotrophic protistan groups). In: Baltic Sea Environment Proceedings, vol. 95.

Heathcote, A.J., Filstrup, C.T., Kendall, D., Downing, J.A., 2016. Biomass pyramids in lake plankton: influence of cyanobacteria size and abundance. Inland Waters 6, 250–257. https://doi.org/10.5268/IW-6.2.941.

Heerkloss, R., Schnese, W., Adamkiewicz-Chojnacka, B., 1991. Seasonal variation in the biomass of zooplankton in two shallow coastal water inlets differing in their stage of eutrophication. Int. Rev. Hydrobiol. 76, 397–404. https://doi.org/10.1002/iroh.19910760312.

HELCOM, 1988. Guidelines for the Baltic monitoring programme for the third stage. Part D. Biological determinands. In: Baltic Sea Environment Proceedings 27D.

HELCOM, 2017. Manual for Marine Monitoring in the COMBINE Programme of HELCOM. http://www.helcom.fi/action-areas/monitoring-and-assessment/manuals-and-guidelines/combine-manual (accessed 7 November 2017).

Hense, I., Meier, H.E.M., Sonntag, S., 2013. Projected climate change impact on Baltic Sea cyanobacteria. Clim. Chang. 119, 391–406. https://doi.org/10.1007/s10584-013-0702-y.

Hernroth, L., 1985. Recommendations on methods for marine biological studies in the Baltic Sea. In: Mesozooplankton Biomass Assessment, vol. 10. Baltic Marine Biological Publications.

Hogfors, H., Motwani, N.H., Hajdu, S., El-Shehawy, R., Holmborn, T., Vehmaa, A., Engström-Öst, J., Brutemark, A., Gorokhova, E., 2014. Bloom-forming cyanobacteria support copepod reproduction and development in the Baltic Sea. PLoS One 9 (11), e112692. https://doi.org/10.1371/journal.pone.0112692.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: With Applications in R. Springer Texts in Statistics, New York.

Jiang, X., Yang, W., Zhang, L., Chen, L., Niu, Y., 2014. Predation and cyanobacteria jointly facilitate competitive dominance of small-bodied cladocerans. J. Plankton Res. 36, 956–965. https://doi.org/10.1093/plankt/fbu032.

Kahru, M., Elmgren, R., Savchuk, O.P., 2016. Changing seasonality of the Baltic Sea. Biogeosciences 13, 1009–1018. https://doi.org/10.5194/bg-13-1009-2016.

Kanoshina, I., Lips, U., Leppänen, J.-M., 2003. The influence of weather conditions (temperature and wind) on cyanobacterial bloom development in the Gulf of Finland (Baltic Sea). Harmful Algae 2, 29–41. https://doi.org/10.1016/S1568-9883(02)00085-9.

Karjalainen, M., Reinikainen, M., Spoof, L., Meriluoto, J.A.O., Sivonen, K., Viitasalo, M., 2005. Transfer of cyanobacterial toxins from zooplankton to planktivores: consequences for pike larvae and mysid shrimps. Environ. Toxicol. 20, 354–362. https://doi.org/10.1002/tox.20112.

Karjalainen, M., Engström-Öst, J., Korpinen, S., Peltonen, H., Pääkkönen, J.-P., Rönkkönen, S., Suikkanen, S., Viitasalo, M., 2007. Ecosystem consequences of cyanobacteria in the Northern Baltic Sea. Ambio 36, 195–202. https://doi.org/10.1579/0044-7447(2007)36[195:ECOCIT]2.0.CO;2.

Karlson, A.M.L., Duberg, J., Motwani, N.H., Hogfors, H., Klawonn, I., Ploug, H., Svedén, J.B., Garbaras, A., Sundelin, B., Hajdu, S., Larsson, U., Elmgren, R., Gorokhova, E., 2015. Nitrogen fixation by cyanobacteria stimulates production in Baltic food webs. Ambio 44, 413–426. https://doi.org/10.1007/s13280-015-0660-x.

Kijewski, T., Zbawicka, M., Strand, J., Kautsky, H., Kotta, J., Rätsep, M., Wenne, R., 2019. Random forest assessment of correlation between environmental factors and genetic differentiation of populations: case of marine mussels *Mytilus*. Oceanologia 61, 131–142. https://doi.org/10.1016/j.oceano.2018.08.002.

Kivi, K., Kuosa, H., Tanskanen, S., 1996. An experimental study on the role of crustacean and microprotozoan grazers in the planktonic food web. Mar. Ecol. Prog. Ser. 136, 59–68. https://doi.org/10.3354/meps136059.

Koroleff, F., 1979. The General Chemical Analysis Methods of Sea Water. Institute of Marine Research, Finland. Meri 7 (in Finnish).

Korpinen, S., Karjalainen, M., Viitasalo, M., 2006. Effects of cyanobacteria on survival and reproduction of the littoral crustacean *Gammarus zaddachi* (Amphipoda). Hydrobiologia 559, 285–295. https://doi.org/10.1007/s10750-005-1172-7.

Koski, M., Engström, J., Viitasalo, M., 1999. Reproduction and survival of the calanoid copepod *Eurytemora affinis* fed with toxic and non-toxic cyanobacteria. Mar. Ecol. Prog. Ser. 186, 187–197. https://doi.org/10.3354/meps186187.

Kozlowsky-Suzuki, B., Karjalainen, M., Lehtiniemi, M., Engström-Öst, J., Koski, M., Carlsson, P., 2003. Feeding, reproduction and toxin accumulation by the copepods *Acartia bifilosa* and *Eurytemora affinis* in the presence of the toxic cyanobacterium *Nodularia spumigena*. Mar. Ecol. Prog. Ser. 249, 237–249. https://doi.org/10.3354/meps249237.

Kruk, C., Segura, A.M., 2012. The habitat template of phytoplankton morphology-based functional groups. Hydrobiologia 698, 191–202. https://doi.org/10.1007/978-94-007-5790-5_15.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28 (5).

Kuosa, H., Fleming-Lehtinen, V., Lehtinen, S., Lehtiniemi, M., Nygård, H., Raateoja, M., Raitaniemi, J., Tuimala, J., Uusitalo, L., Suikkanen, S., 2017. A retrospective view of the development of the Gulf of Bothnia ecosystem. J. Mar. Syst. 167, 78–92. https://doi.org/10.1016/j.jmarsys.2016.11.020.

Kuuppo-Leinikki, P., 1990. Protozoan grazing on planktonic bacteria and its impact on bacterial population. Mar. Ecol. Prog. Ser. 63, 227–238.

Leão, P.N., Vasconcelos, M.T., Vasconcelos, V.M., 2009. Allelopathy in freshwater cyanobacteria. Crit. Rev. Microbiol. 35, 271–282. https://doi.org/10.3109/10408410902823705.

Lehtinen, S., Sopanen, S., Tamminen, T., 2010. Community-level and species-specific responses of coastal phytoplankton to the presence of two mesozooplanktonic grazers. Mar. Biol. 157, 1555–1565. https://doi.org/10.1007/s00227-010-1429-1.

Lehtinen, S., Suikkanen, S., Hällfors, H., Kauppila, P., Lehtiniemi, M., Tuimala, J., Uusitalo, L., Kuosa, H., 2016. Approach for supporting food web assessments with multi-decadal phytoplankton community analyses – case Baltic Sea. Front. Mar. Sci. 3, 220. https://doi.org/10.3389/fmars.2016.00220.

Lehtiniemi, M., Engström-Öst, J., Karjalainen, M., Kozlowsky-Suzuki, B., Viitasalo, M., 2002. Fate of cyanobacterial toxins in the pelagic food web: transfer to copepods or to faecal pellets? Mar. Ecol. Prog. Ser. 241, 13–21. https://doi.org/10.3354/meps241013.

Mäkinen, K., Vuorinen, I., Hänninen, J., 2017. Climate-induced hydrography change favours small-bodied zooplankton in a coastal ecosystem. Hydrobiologia 792, 83–96. https://doi.org/10.1007/s10750-016-3046-6.

McCune, B., Grace, J.B., 2002. Analysis of Ecological Communities. MjM Software Design, Oregon, USA.

Merrell, J.R., Stoecker, D.K., 1998. Differential grazing on protozoan microplankton by developmental stages of the calanoid copepod *Eurytemora affinis* Poppe. J. Plankton Res. 20, 289–304. https://doi.org/10.1093/plankt/20.2.289.

Motwani, N.H., Gorokhova, E., 2013. Metazooplankton grazing on picocyanobacteria in the Baltic Sea as inferred from molecular diet analysis. PLoS One 8, e79230. https://doi.org/10.1371/journal.pone.0079230.

Mulholland, M.R., Capone, D.G., 2001. Stoichiometry of nitrogen and carbon utilization in cultured population of *Trichodesmium* IMS101: implications for growth. Limnol. Oceanogr. 46, 436–443. https://doi.org/10.4319/lo.2001.46.2.0436.

Ohlendieck, U., Stuhr, A., Siegmund, H., 2000. Nitrogen fixation by diazotrophic cyanobacteria in the Baltic Sea and transfer of the newly fixed nitrogen to picoplankton organisms. J. Mar. Syst. 25, 213–219. https://doi.org/10.1016/S0924-7963(00)00016-6.

Olenina, I., Hajdu, S., Edler, L., Andersson, A., Wasmund, N., Busch, S., Göbel, J., Gromisz, S., Huseby, S., Huttunen, M., Jaanus, A., Kokkonen, P., Ledaine, I., Niemkiewicz, E., 2006. Biovolumes and size-classes of phytoplankton in the Baltic Sea. In: Baltic Sea Environment Proceedings, vol. 106.

Olli, K., Ptacnik, R., Andersen, T., Trikk, O., Klais, R., Lehtinen, S., Tamminen, T., 2014. Against the tide: recent diversity increase enhances resource use in a coastal ecosystem. Limnol. Oceanogr. 59, 267–274. https://doi.org/10.4319/lo.2014.59.1.0267.

Olli, K., Klais, R., Tamminen, T., 2015. Rehabilitating the cyanobacteria – niche partitioning, resource use efficiency and phytoplankton community structure during diazotrophic cyanobacterial blooms. J. Ecol. 103, 1153–1164. https://doi.org/10.1111/1365-2745.12437.

Olofsson, M., Suikkanen, S., Kobos, J., Wasmund, N., Karlson, B., 2020. Basin-specific changes in filamentous cyanobacteria community composition across four decades in the Baltic Sea. Harmful Algae 91, 101685. https://doi.org/10.1016/j.hal.2019.101685.

O'Neil, J.M., Davis, T.W., Burford, M.A., Gobler, C.J., 2012. The rise of harmful cyanobacteria blooms: the potential roles of eutrophication and climate change. Harmful Algae 14, 313–334. https://doi.org/10.1016/j.hal.2011.10.027.

Orcutt, J.D., Pace, M.L., 1984. Seasonal dynamics of rotifer and crustacean zooplankton populations in a eutrophic, monomictic lake with a note on rotifer sampling techniques. Hydrobiologia 119, 73–80. https://doi.org/10.1007/BF00016866.

Paerl, H.W., Paul, V.J., 2012. Climate change: links to global expansion of harmful cyanobacteria. Water Res. 46, 1349–1363. https://doi.org/10.1016/j.watres.2011.08.002.

Peltonen, H., Luoto, M., Pääkkönen, J.-P., Karjalainen, M., Tuomaala, A., Pönni, J., Viitasalo, M., 2007. Pelagic fish abundance in relation to regional environmental variation in the Gulf of Finland, northern Baltic Sea. ICES J. Mar. Sci. 64, 487–495.

Ploug, H., Musat, N., Adam, B., Moraru, C.L., Lavik, G., Vagner, T., Bergman, B., Kuypers, M.M.M., 2010. Carbon and nitrogen fluxes associated with the

cyanobacterium *Aphanizomenon* sp. in the Baltic Sea. ISME J. 4, 1215–1223. https://doi.org/10.1038/ismej.2010.53.

Ploug, H., Adam, B., Musat, N., Kalvelage, T., Lavik, G., Wolf-Gladrow, D., Kuypers, M.M.M., 2011. Carbon, nitrogen and O_2 fluxes associated with the cyanobacterium *Nodularia spumigena* in the Baltic Sea. ISME J. 5, 154–1558. https://doi.org/10.1038/ismej.2011.20.

Pomati, F., Shurin, J.B., Andersen, K.H., Tellenbach, C., Barton, A.D., 2020. Interacting temperature, nutrients and zooplankton grazing control phytoplankton size-abundance relationships in eight Swiss lakes. Front. Microbiol. 10, 3155. https://doi.org/10.3389/fmicb.2019.03155.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rolff, C., 2000. Seasonal variation in ^{13}C and ^{15}N of size-fractionated plankton at a coastal station in the northern Baltic proper. Mar. Ecol. Prog. Ser. 203, 47–65. https://doi.org/10.3354/meps203047.

Rönkkönen, S., Ojaveer, E., Raid, T., Viitasalo, M., 2004. Long-term changes in Baltic herring (*Clupea harengus membras*) growth in the Gulf of Finland. Can. J. Fish. Aquat. Sci. 61, 219–229. https://doi.org/10.1139/f03-167.

Ryther, J.H., 1969. Photosynthesis and fish production in the sea. Science 166, 72–76.

Schmidt, K., Koski, M., Engström-Öst, J., Atkinson, A., 2002. Development of Baltic Sea zooplankton in the presence of a toxic cyanobacterium: a mesocosm approach. J. Plankton Res. 24, 979–992. https://doi.org/10.1093/plankt/24.10.979.

Sellner, K.G., Olson, M.M., Kononen, K., 1994. Copepod grazing in a summer cyanobacteria bloom in the Gulf of Finland. Hydrobiologia 292 (293), 249–254.

Sivonen, K., Kononen, K., Carmichael, W.W., Dahlem, A.M., Rinehart, K.L., Kiviranta, J., Niemelä, S.I., 1989. Occurrence of the hepatotoxic cyanobacterium *Nodularia spumigena* in the Baltic Sea and structure of the toxin. Appl. Environ. Microbiol. 55, 1990–1995.

Smoliński, S., Radtke, K., 2016. Spatial prediction of demersal fish diversity in the Baltic Sea: comparison of machine learning and regression-based techniques. ICES J. Mar. Sci. 74, 102–111. https://doi.org/10.1093/icesjms/fsw136.

Sommer, U., 2008. Trophic cascades in marine and freshwater plankton. Int. Rev. Hydrobiol. 93, 506–516. https://doi.org/10.1002/iroh.200711039.

Stal, L.J., Albertano, P., Bergman, B., von Bröckel, K., Gallone, J.R., Hayes, P.K., Sivonen, K., Walsby, A.E., 2003. BASIC: Baltic Sea cyanobacteria. An investigation of the structure and dynamics of water blooms of cyanobacteria in the Baltic Sea - responses to a changing environment. Cont. Shelf Res. 23, 1695–1714. https://doi.org/10.1016/j.csr.2003.06.001.

Suikkanen, S., Fistarol, G.O., Granéli, E., 2004. Allelopathic effects of the Baltic cyanobacteria *Nodularia spumigena, Aphanizomenon flos-aquae* and *Anabaena lemmermannii* on algal monocultures. J. Exp. Mar. Biol. Ecol. 308, 85–101. https://doi.org/10.1016/j.jembe.2004.02.012.

Suikkanen, S., Fistarol, G.O., Granéli, E., 2005. Effects of cyanobacterial allelochemicals on a natural plankton community. Mar. Ecol. Prog. Ser. 287, 1–9. https://doi.org/10.3354/meps287001.

Suikkanen, S., Pulina, S., Engström-Öst, J., Lehtiniemi, M., Lehtinen, S., Brutemark, A., 2013. Climate change and eutrophication induced shifts in northern summer plankton communities. PLoS One 8, e66475. https://doi.org/10.1371/journal.pone.0066475.

Tamminen, T., Andersen, T., 2007. Seasonal phytoplankton nutrient limitation patterns as revealed by bioassays over Baltic Sea gradients of salinity and eutrophication. Mar. Ecol. Prog. Ser. 340, 121–138. https://doi.org/10.3354/meps340121.

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 (5), e0196225. https://doi.org/10.1371/journal.pone.0196225.

Tõnno, I., Agasild, H., Kõiv, T., Freiberg, R., Nõges, P., Nõges, T., 2016. Algal diet of small-bodied crustacean zooplankton in a cyanobacteria-dominated eutrophic lake. PLoS One 11 (4), e0154526. https://doi.org/10.1371/journal.pone.0154526.

Utermöhl, H., 1958. Zur Vervollkommung der quantitativen Phytoplankton-Methodik. Mitt. Int. Ver. Limnol. 9, 1–38.

Van Donk, E., Ianora, A., Vos, M., 2011. Induced defences in marine and freshwater phytoplankton: a review. Hydrobiologia 668, 3–19. https://doi.org/10.1007/s10750-010-0395-4.

Vargas, C.A., Cuevas, L.A., Gonzáles, H.E., Daneri, G., 2007. Bacterial growth response to copepod grazing in aquatic ecosystems. J. Mar. Biol. Assoc. UK 87, 667–674. https://doi.org/10.1017/s0025315407056275.

Verikas, A., Gelzinis, A., Bacauskiene, M., Olenina, I., Olenin, S., Vaiciukynas, E., 2012. Phase congruency-based detection of circular objects applied to analysis of phytoplankton images. Pattern Recogn. 45, 1659–1670. https://doi.org/10.1016/j.patcog.2011.10.019.

Wannicke, N., Korth, F., Liskow, I., Voss, M., 2013. Incorporation of diazotrophic fixed N_2 by mesozooplankton – case studies in the southern Baltic Sea. J. Mar. Syst. 117–118, 1–13. https://doi.org/10.1016/j.jmarsys.2013.03.005.

Wasmund, N., Tuimala, J., Suikkanen, S., Vandepitte, L., Kraberg, A., 2011. Long-term trends in phytoplankton composition in the western and Central Baltic Sea. J. Mar. Syst. 87, 145–159. https://doi.org/10.1016/j.jmarsys.2011.03.010.

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. https://doi.org/10.3389/fmars.2017.00022.

Wei, C.-L., Rowe, G.T., Escobar-Briones, E., Nunnally, C., Soliman, Y., Ellis, N., 2012. Standing stocks and body size of deep-sea macrofauna: predicting the baseline of 2010 Deepwater Horizon oil spill in the northern Gulf of Mexico. Deep Sea Res. I 69, 82–99. https://doi.org/10.1016/j.dsr.2012.07.008.

Wood, S.N., 2014. mgcv: GAMs with GCV/AIC/REML Smoothness Estimation and GAMMs by PQL. R package version 1.8-2. http://CRAN.R-project.org/package=mgcv.

Work, K.A., Havens, K.E., 2003. Zooplankton grazing on bacteria and cyanobacteria in a eutrophic lake. J. Plankton Res. 25, 1301–1307. https://doi.org/10.1093/plankt/fbg092.

Xie, Y., 2014. knitr: a comprehensive tool for reproducible research in R. In: Stodden, V., Leisch, F., Peng, R.D. (Eds.), Implementing Reproducible Computational Research. Chapman and Hall/CRC, Boca Raton.

Xie, Y., 2015. Dynamic Documents with R and knitr, 2nd ed. Chapman and Hall/CRC, Boca Raton.

Xie, Y., 2018. knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.20. https://cran.r-project.org/web/packages/knitr/index.html.

Zanchett, G., Oliveira-Filho, E.C., 2013. Cyanobacteria and cyanotoxins: from impacts on aquatic ecosystems and human health to anticarcinogenic effects. Toxins 5, 1896–1917. https://doi.org/10.3390/toxins5101896.