Light-dominated selection shaping filamentous cyanobacterial assemblages drives odor problem in a drinking water reservoir

Supplementary Material

Ming Su[1,5] Yiping Zhu[2] Tom Andersen[3] Xianyun Wang[4]
Zhiyong Yu[1,5] Jinping Lu[1,5] Yichao Song[2] Tengxin Cao[1,5]
Jianwei Yu[1,5] Yu Zhang[1,5] Min Yang[1,5,*]

2025-01-09

[1] Key Laboratory of Drinking Water Science and Technology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences

[2] Shanghai Chengtou Raw Water Co. Ltd

[3] Department of Biosciences, University of Oslo

[4] National Engineering Research Center of China (South) for Urban Water

[5] University of Chinese Academy of Sciences

[*] Correspondence: Min Yang <yangmin@rcees.ac.cn>

Supplementary Material

All data analysis and illustration were performed using R 4.0[1]. Data pretreatment and summary were performed using the **dplyr**[2] and **base** packages in R, regression analysis including linear and generalized linear models were performed using the **stats** package[1], generalized additive modelling was performed using the **mgcv** package[3,4] quantile regression analysis was performed using the **quantreg** package[5]; contour figures were created by the **graphics** package[1], other figures were prepared using the **ggplot2** package[6].

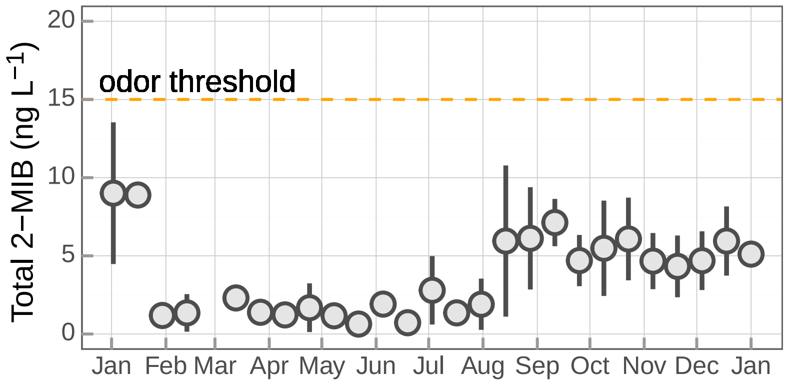

Fig. S1Seasonal dynamics of MIB of Yangtze River water (inlet) from 2011 to 2015"

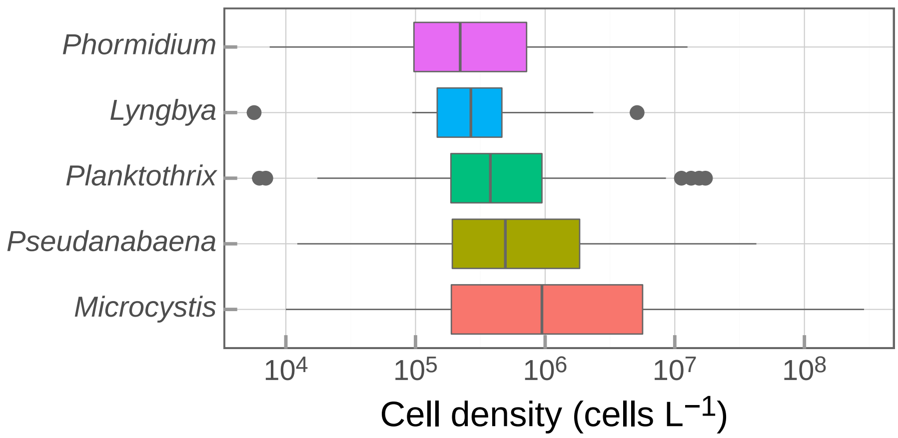

Fig. S2Composition of filamentous cyanobacteria and *Microcystis* in QCS Reservoir

Seasonal and long-term trends of *Planktothrix* and *Pseudanabaena* were evaluated using time series analysis, as illustrated in Table S7, Fig. S6, Table S8 and Fig. S7.

We select water temperature (temp), pre-week photosynthetically active radiation (weekPAR), total nitrogen (TN), nitrate (NO3), total phosphate (TP), ammonia (NH4), wind speed (wind) and maximum daily air temperature (maxtemp) as the potential predictors for abundances of *Planktothrix* (Table S9) and *Pseudanabaena* (Table S10). Linear model ($Y = \sum \beta_i X_i + \beta_0 + \epsilon, \epsilon \in N(0,1), i = 1,2,...$) between these predictors (X_i, i = 1, 2, …) and logarithm transformed abundances ($Y = log_{10}(1 + N)$).

We performed Backward Stepwise Regressions to identify the significant variables responsible for the abundances of *Planktothrix* and *Pseudanabaena*. The regression started with a model that contains all variables, and then removing the least significant variables one by one, until a pre-specified stopping rule (here we use AIC rule) is reached.

Below is the Backward Stepwise Regressions of linear model for *Planktothrix*, the water temperature (temp), pre-week PAR (weekPAR), total nitrogen (TN) and wind speed (wind) were considered as effect predictors for *Planktothrix*.

Below is the Backward Stepwise Regressions of linear model for *Pseudanabaena*, the water temperature (temp), pre-week PAR (weekPAR), total nitrogen (TN), ammonia (NH4) and maximum daily air temperature (maxtemp) were considered as effect predictors for *Pseudanabaena*.

Backward Stepwise Regression approaches were performed for the two linear models to find out the optimum selections of predictors, as summarize in Table S11 (*Planktothrix*) and Table S12 (*Pseudanabaena*), respectively. Water temperature (temp), pre-week PAR (weekPAR), total nitrogen (TN), wind speed (wind) and maximum daily air temperature (maxtemp) were sorted out.

According to the Backward Stepwise Regressions, water temperature (temp), pre-week PAR (weekPAR), total nitrogen (TN), total phosphate (TP), ammonia (NH4) and wind speed (wind) were selected as the valid predictors and used for following general additive models (GAMs). Variance inflation factors (VIF) were calculated and validated the rationality (<5) for these predictors , as summarized in Table S13.

Correlation coefficients among these predictors were computed by Pearson Method, as illustrated in Fig. S8. The results were used for the optimization of non-parameter smooth functions (smoother) in following GAMs for *Planktothrix* and *Pseudanabaena*. Specifically, high correlation coefficient between water temperature (temp) and pre-week PAR (PAR) suggests they are interacting predictors; similarly, total nitrogen (TN) and ammonia (NH4) are interacting predictors.

Iterations between water temperature and pre-week PAR (t2(temp, weekPAR)), and between total nitrogen and ammonia (t2(TN, NH4)) were both evaluated by the tensor products ($f_1(x_1) \otimes f_2(x_2)$).

The results of "GAM1" suggest water temperature (temp), pre-week PAR (weekPAR), total nitrogen (TN) and ammonia (NH4) are most important predictors for *Planktothrix* abundance.

The results of "GAM2" suggest water temperature (temp), pre-week PAR (weekPAR), total phosphrus (TP) and wind speed (wind) are most important predictors for *Pseudanabaena* abundance.

According to “GAM1”, the predictors of *Planktothrix* abundances were further optimized to 4 factors, which are water temperature (temp), pre-week PAR (weekPAR), total nitrogen (TN) and ammonia (NH4). The summary of optimized *Planktothrix* GAM (GAM3) results is shown in Table S16.

According to “GAM2”, the predictors of *Pseudanabaena* abundances were further optimized to 4 factors, which are water temperature (temp), pre-week PAR (weekPAR), total phosphorus (TP) and wind speed (wind). The summary of optimized *Pseudanabaena* GAM (GAM4) results is shown in Table S17.

R Language demonstration code

```
# demo data frame
require(lubridate)
require(mgcv)
modeldf <- data.frame(date = ymd("2020-01-01") + 0:365,
                      y = rnorm(366))
# seasonal pattern: week number
modeldf$x1 <- week(modeldf$date)
# long-term pattern: decimal number
modeldf$x2 <- year(modeldf$date) + yday(modeldf$date) / 366
# demo gam model
m <- gamm(y ~ s(x1, bs = "cc") + s(x2), data = modeldf)
```

1. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2020).

2. Wickham, H., François, R., Henry, L. & Müller, K. *dplyr: A grammar of data manipulation*. (2018).

3. Wood, S. N. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association* **99**, 673–686 (2004).

4. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (b)* **73**, 3–36 (2011).

5. Koenker, R. *Quantreg: Quantile regression*. (2017).

6. Wickham, H. *ggplot2: Elegant graphics for data analysis*. (Springer-verlag New York, 2016).

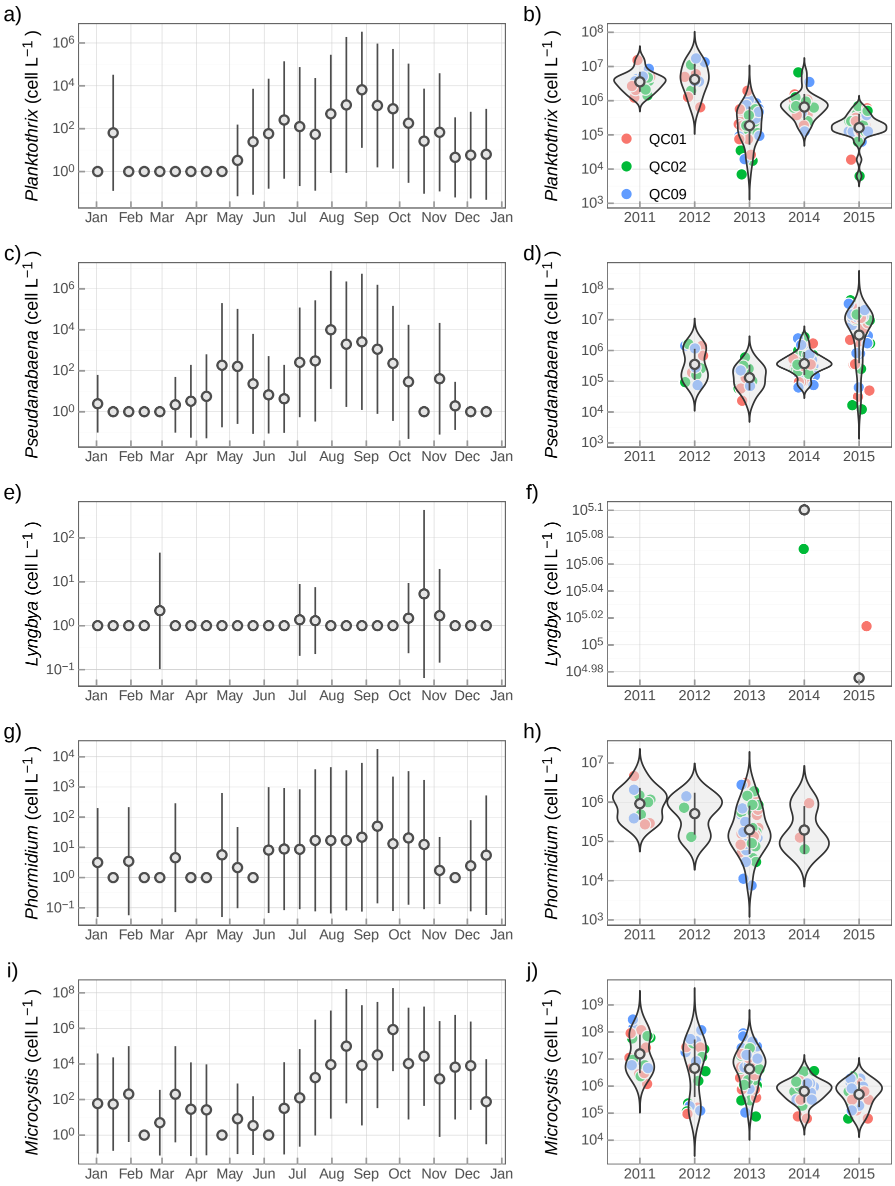

Fig. S3Seasonal dynamics and long-term trends of *Planktothrix*, *Pseudanabaena*, *Lyngbya*, *Phormidium* and *Microcystis* in QCS Reservoir

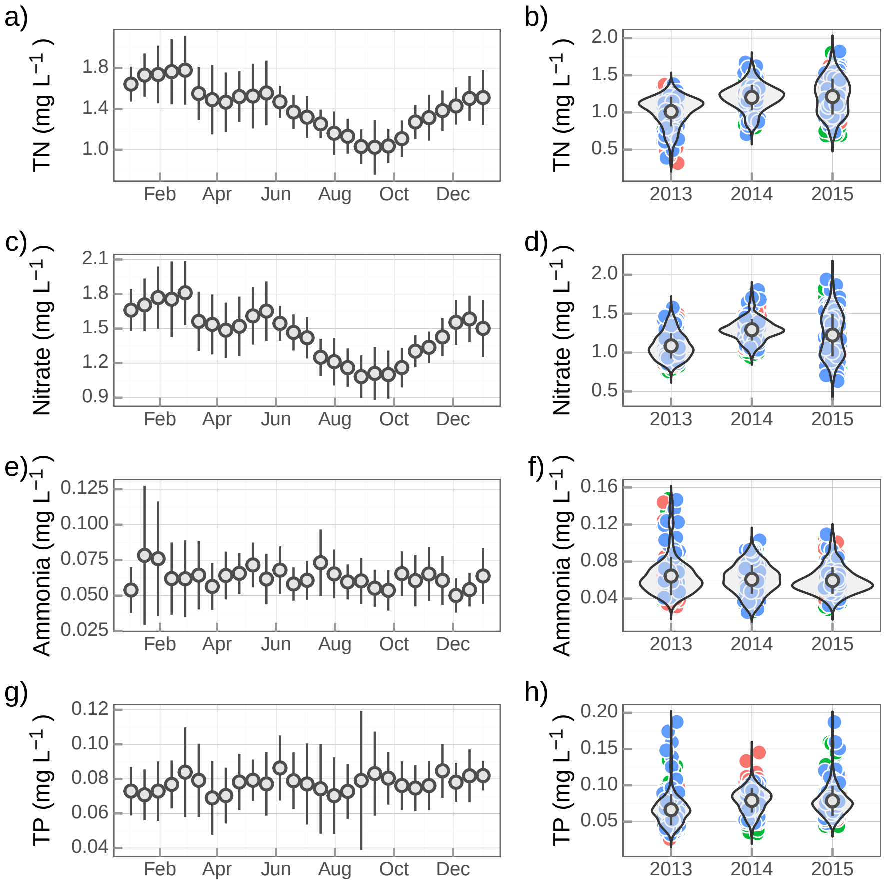

Fig. S4Seasonal dynamics and long-term trends of nutrients (total nitrogen, nitrate, ammonia and total phosphorus) in QCS Reservoir

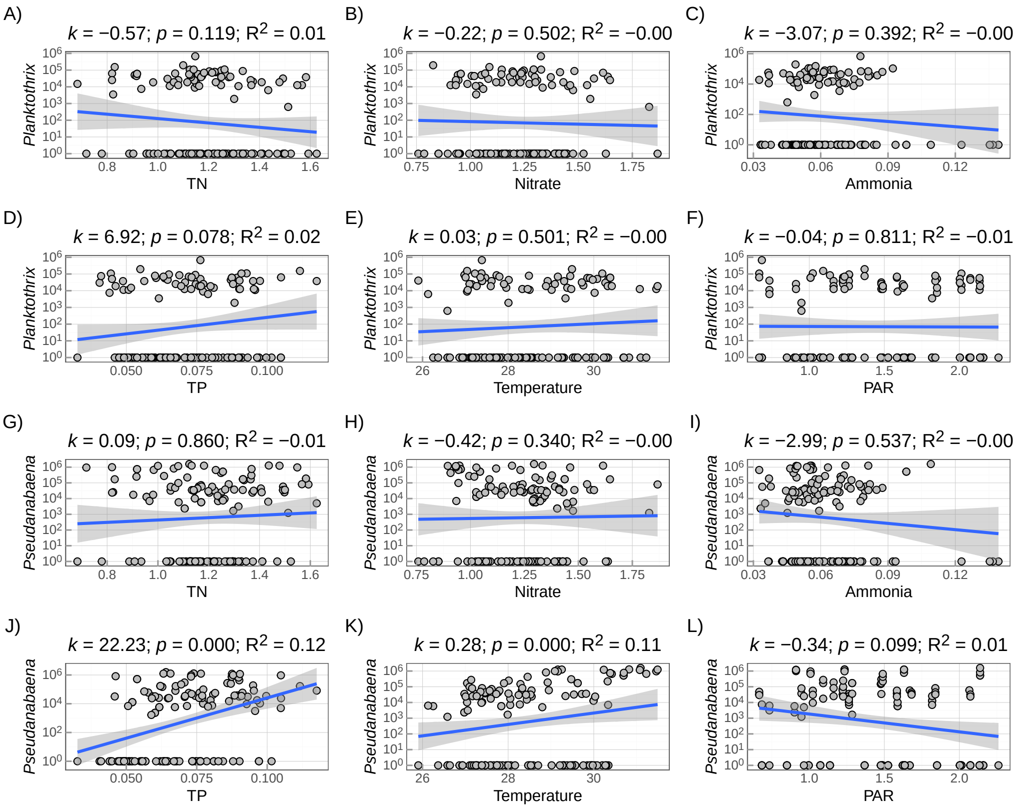

Fig. S5Correlation between factors and filamentous abundances

Year	Season	Min	Max	Median	IQR	Mean	SD	SE	CI
2013	Spring	0.037	0.078	0.055	0.013	0.055	0.010	0.002	0.004
2013	Summer	0.033	0.139	0.058	0.018	0.063	0.023	0.003	0.006
2013	Autumn	0.035	0.127	0.067	0.027	0.065	0.021	0.004	0.007
2013	Winter	0.028	0.201	0.053	0.089	0.088	0.064	0.021	0.050
2014	Spring	0.045	0.107	0.071	0.019	0.074	0.016	0.002	0.005
2014	Summer	0.032	0.093	0.059	0.016	0.061	0.013	0.001	0.003
2014	Autumn	0.024	0.105	0.044	0.027	0.051	0.019	0.003	0.007
2014	Winter	0.034	0.183	0.067	0.024	0.078	0.034	0.007	0.015
2015	Spring	0.034	0.104	0.069	0.019	0.068	0.016	0.003	0.006
2015	Summer	0.033	0.109	0.058	0.022	0.062	0.016	0.002	0.005
2015	Autumn	0.039	0.077	0.054	0.008	0.056	0.011	0.003	0.007
2015	Winter	0.022	0.081	0.053	0.028	0.052	0.022	0.009	0.023

Table S1Descriptive statistics of ammonia nitrogen in QCS Reservoir (unit: mg L^{-1})

Year	Season	Min	Max	Median	IQR	Mean	SD	SE	CI
2013	Spring	1.213	1.702	1.415	0.179	1.431	0.118	0.024	0.05
2013	Summer	0.759	1.665	1.110	0.269	1.137	0.194	0.025	0.05
2013	Autumn	0.843	1.71	1.029	0.445	1.140	0.281	0.047	0.095
2013	Winter	1.390	1.828	1.557	0.167	1.600	0.136	0.045	0.105
2014	Spring	1.578	2.162	1.796	0.121	1.800	0.130	0.019	0.038
2014	Summer	0.944	1.737	1.293	0.176	1.306	0.163	0.019	0.038
2014	Autumn	1.159	1.707	1.309	0.119	1.352	0.129	0.023	0.047
2014	Winter	1.561	2.362	2.008	0.196	2.008	0.189	0.039	0.082
2015	Spring	1.157	1.793	1.485	0.271	1.474	0.186	0.033	0.068
2015	Summer	0.895	1.867	1.411	0.414	1.333	0.264	0.037	0.074
2015	Autumn	0.729	1.365	1.145	0.234	1.092	0.189	0.052	0.114
2015	Winter	1.595	1.774	1.708	0.131	1.694	0.078	0.032	0.082

Table S2Descriptive statistics of nitrate nitrogen in QCS Reservoir (unit: mg L^{-1})

Year	Season	Min	Max	Median	IQR	Mean	SD	SE	CI
2013	Spring	1.008	1.469	1.326	0.156	1.313	0.113	0.023	0.048
2013	Summer	0.683	1.411	1.126	0.150	1.094	0.153	0.020	0.040
2013	Autumn	0.319	1.592	1.006	0.583	1.052	0.351	0.059	0.119
2013	Winter	1.367	1.735	1.543	0.161	1.542	0.132	0.044	0.101
2014	Spring	1.478	2.265	1.808	0.217	1.795	0.164	0.024	0.048
2014	Summer	0.808	1.705	1.237	0.192	1.228	0.195	0.023	0.045
2014	Autumn	0.959	1.601	1.236	0.125	1.250	0.142	0.025	0.052
2014	Winter	1.361	2.429	2.073	0.337	1.974	0.259	0.054	0.112
2015	Spring	1.234	1.875	1.419	0.191	1.450	0.154	0.028	0.056
2015	Summer	0.718	1.654	1.317	0.355	1.292	0.223	0.031	0.063
2015	Autumn	0.726	1.348	1.013	0.135	1.017	0.156	0.043	0.094
2015	Winter	1.348	1.75	1.59	0.279	1.581	0.177	0.072	0.186

Table S3Descriptive statistics of total nitrogen (TN) in QCS Reservoir (unit: mg L^{-1})

Year	Season	Min	Max	Median	IQR	Mean	SD	SE	CI
2013	Spring	0.053	0.095	0.077	0.016	0.078	0.012	0.003	0.005
2013	Summer	0.033	0.25	0.064	0.023	0.07	0.036	0.005	0.009
2013	Autumn	0.046	0.121	0.070	0.017	0.072	0.015	0.003	0.005
2013	Winter	0.055	0.095	0.060	0.018	0.067	0.013	0.004	0.010
2014	Spring	0.050	0.133	0.089	0.022	0.089	0.019	0.003	0.005
2014	Summer	0.047	0.118	0.081	0.024	0.079	0.017	0.002	0.004
2014	Autumn	0.056	0.126	0.084	0.011	0.088	0.013	0.002	0.005
2014	Winter	0.058	0.133	0.092	0.037	0.091	0.021	0.004	0.009
2015	Spring	0.042	0.115	0.073	0.023	0.073	0.019	0.003	0.007
2015	Summer	0.046	0.222	0.077	0.022	0.083	0.027	0.004	0.007
2015	Autumn	0.065	0.092	0.072	0.012	0.075	0.008	0.002	0.005
2015	Winter	0.069	0.096	0.075	0.006	0.078	0.010	0.004	0.010

Table S4Descriptive statistics of total phosphrus (TP) in QCS Reservoir (unit: mg L^{-1})

Year	Season	Min	Max	Median	IQR	Mean	SD	SE	CI
2013	Spring	11.208	24.642	19.468	5.594	19.475	4.088	0.834	1.726
2013	Summer	24.563	30.343	28.701	1.977	28.348	1.609	0.208	0.416
2013	Autumn	11.564	27.028	21.502	7.032	20.788	5.019	0.836	1.698
2013	Winter	3.927	9.514	5.846	4.501	6.853	2.403	0.801	1.847
2014	Spring	12.100	25.821	20.99	6.767	20.322	4.091	0.591	1.188
2014	Summer	25.688	28.549	27.386	0.832	27.367	0.691	0.08	0.159
2014	Autumn	12.874	26.943	21.812	7.086	21.447	4.271	0.767	1.567
2014	Winter	7.19	11.001	8.403	1.759	8.637	1.168	0.244	0.505
2015	Spring	11.17	25.049	22.13	3.92	20.856	4.131	0.742	1.515
2015	Summer	24.69	31.491	28.798	2.92	28.475	1.953	0.273	0.549
2015	Autumn	16.102	28.22	26.413	5.503	23.534	4.587	1.272	2.772
2015	Winter	7.861	8.981	8.092	0.554	8.278	0.447	0.182	0.469

Table S5Descriptive statistics of water temperature in QCS Reservoir (unit: °C)

Year	Season	Min	Max	Median	IQR	Mean	SD	SE	CI
2013	Spring	0.815	1.879	1.273	0.410	1.322	0.314	0.064	0.133
2013	Summer	0.901	2.261	1.638	0.810	1.621	0.416	0.054	0.107
2013	Autumn	0.701	1.468	1.152	0.409	1.110	0.247	0.041	0.084
2013	Winter	0.483	1.311	0.798	0.139	0.873	0.298	0.099	0.229
2014	Spring	0.854	1.977	1.552	0.395	1.528	0.284	0.041	0.083
2014	Summer	0.666	1.840	1.226	0.453	1.235	0.316	0.036	0.073
2014	Autumn	0.643	1.487	0.957	0.368	0.999	0.254	0.046	0.093
2014	Winter	0.461	1.254	0.938	0.241	0.857	0.248	0.052	0.107
2015	Spring	0.870	2.033	1.462	0.572	1.418	0.355	0.064	0.130
2015	Summer	0.734	2.138	1.235	0.482	1.293	0.352	0.049	0.099
2015	Autumn	0.638	1.460	1.253	0.185	1.195	0.329	0.091	0.199
2015	Winter	0.415	0.994	0.598	0.297	0.700	0.239	0.097	0.251

Table S6Descriptive statistics of pre-week PAR in QCS Reservoir (unit: mol m^{-2} d^{-1})

Term	EDF	Ref.DF	Statistic	P.Value
s(nweek)	5.293509	8.000000	21.395662	0.000000
s(ndate)	1.757591	1.757591	3.755324	0.091522

Table S7Summary of time series analysis for *Planktothrix*; s(nweek) denotes the seasonal pattern, and s(ndate) denotes the long-term pattern

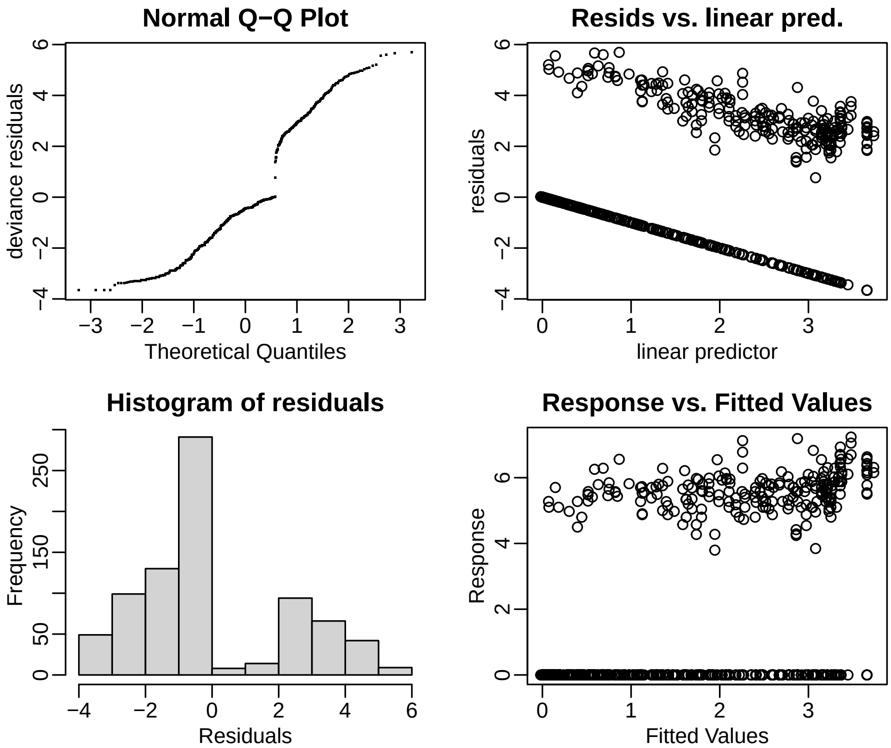

Fig. S6Time series analysis of *Planktothrix* based on seasonal and long-term trend smooth functions

Term	EDF	Ref.DF	Statistic	P.Value
s(nweek)	6.246786	8.000000	11.302972	0.000000
s(ndate)	1.943873	1.943873	82.456653	0.000000

Table S8Summary of time series analysis for *Pseudanabaena*; s(nweek) denotes the seasonal pattern, and s(ndate) denotes the long-term pattern

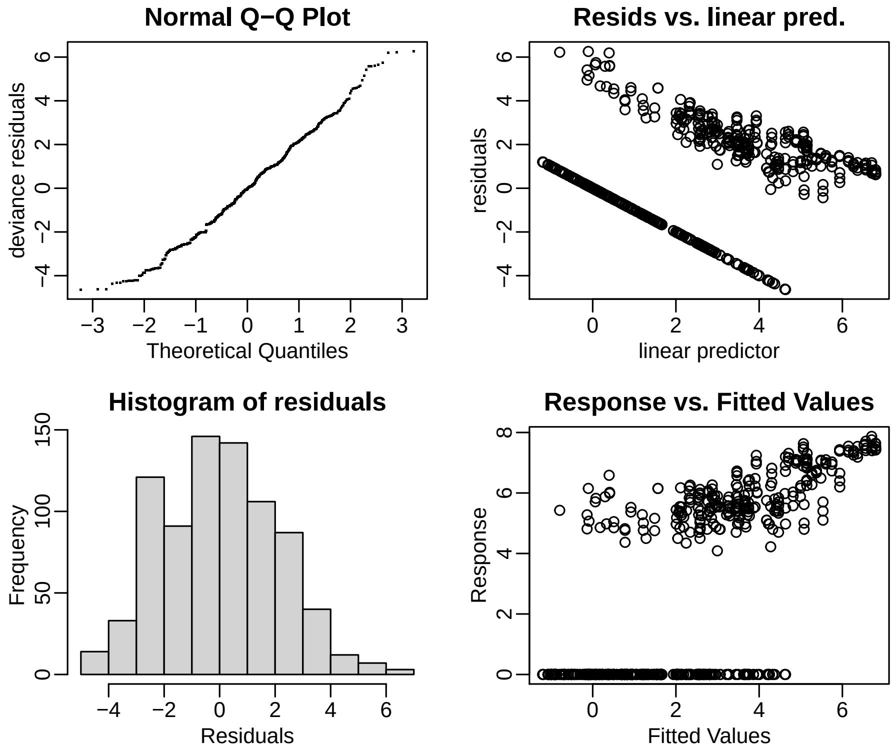

Fig. S7Time series analysis for *Pseudanabaena* based on seasonal and long-term trend smooth functions

Term	Estimate	Std.Error	Statistic	P.Value
(Intercept)	0.055381	0.204114	0.271325	0.786326
temp	0.020686	0.006190	3.342053	0.000936
weekPAR	-0.043151	0.062931	-0.685693	0.493431
TN	-0.125703	0.134448	-0.934960	0.350554
NO3	-0.110292	0.143055	-0.770975	0.441323
TP	0.915797	1.105912	0.828092	0.408270
NH4	-0.721135	0.911296	-0.791330	0.429371
wind	0.085435	0.042653	2.003032	0.046064
maxtemp	-0.003568	0.005304	-0.672707	0.501646

Table S9Correlation analysis between *Planktothrix* and environmental factors (named as LM1)

Term	Estimate	Std.Error	Statistic	P.Value
(Intercept)	-0.249268	0.208706	-1.194353	0.233274
temp	0.041253	0.006329	6.518053	2.97e-10
weekPAR	-0.119826	0.064347	-1.862196	0.063543
TN	0.189667	0.137473	1.379667	0.168707
NO3	-0.068600	0.146274	-0.468984	0.639418
TP	0.202906	1.130792	0.179437	0.857715
NH4	-0.737178	0.931798	-0.791136	0.429484
wind	-0.034087	0.043612	-0.781579	0.435072
maxtemp	-0.012824	0.005423	-2.364817	0.018669

Table S10Correlation analysis between *Pseudanabaena* and environmental factors (named as LM2)

Term	Estimate	Std.Error	Statistic	P.Value
(Intercept)	0.029206	0.183201	0.159421	0.873442
temp	0.018787	0.003960	4.744185	3.21e-06
weekPAR	-0.066606	0.057760	-1.153142	0.249749
TN	-0.194775	0.073604	-2.646272	0.008558
wind	0.075903	0.040487	1.874726	0.061780

Table S11Backward Stepwise Regression of linear models for *Planktothrix*

Term	Estimate	Std.Error	Statistic	P.Value
(Intercept)	-0.351	0.162542	-2.159404	0.031596
temp	0.043007	0.005947	7.231744	3.84e-12
weekPAR	-0.118911	0.063542	-1.871372	0.062248
TN	0.142841	0.078959	1.809052	0.071424
NH4	-0.888384	0.900174	-0.986903	0.324470
maxtemp	-0.013610	0.005303	-2.566502	0.010749

Table S12Backward Stepwise Regression of linear models for *Pseudanabaena*

Parameter	VIF
temp	2.025278
weekPAR	1.449414
TN	1.734340
TP	1.148326
NH4	1.179282
wind	1.119961

Table S13VIF analysis of selected predictors of *Planktothrix* and *Pseudanabaena* models

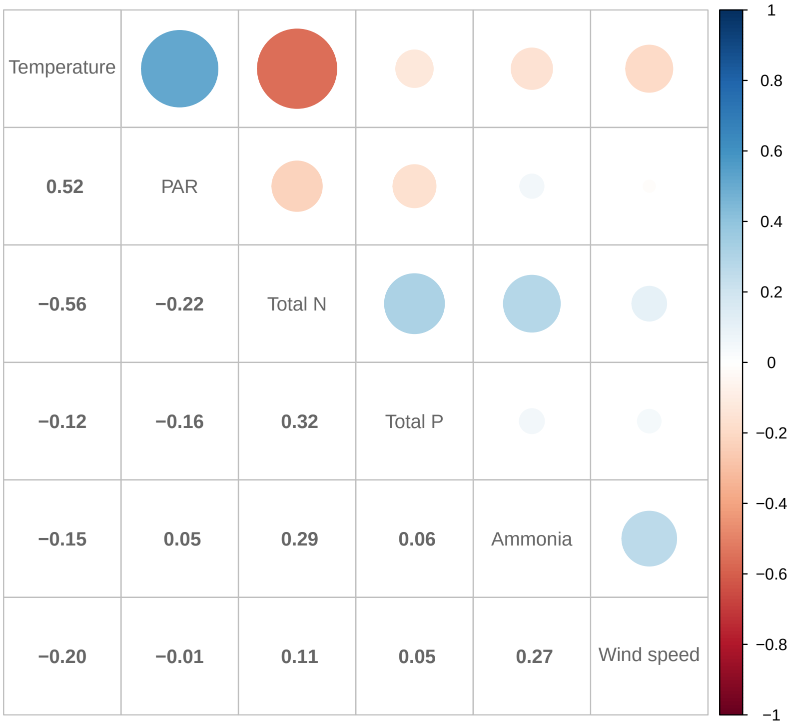

Fig. S8Correlation between environmental factors

Term	edf	ref.df	Statistic	p-value
t2(temp, weekPAR)	2.999974	2.999974	8.441657	2.12287e-05
t2(TN, NH4)	2.999898	2.999898	2.800880	0.04014
s(TP)	1.000008	1.000008	0.759490	0.38418
s(wind)	1.000000	1.000000	3.944025	0.04794

Table S14Summary of GAM model with 6 predictors for *Planktothrix* (named as GAM1)

Term	edf	ref.df	Statistic	p-value
t2(temp, weekPAR)	5.733091	5.733091	9.604014	0
t2(TN, NH4)	2.999999	2.999999	0.964346	0.41001
s(TP)	2.236544	2.236544	2.650483	0.1095
s(wind)	4.787511	4.787511	3.886710	0.00265

Table S15Summary of GAM model with 6 predictors for *Pseudanabaena* (named as GAM2)

Term	edf	ref.df	Statistic	p-value
t2(temp, weekPAR)	2.999993	2.999993	8.439392	1.912541e-05
t2(TN, NH4)	2.999870	2.999870	4.088651	0.00689

Table S16Summary of GAM model with 4 predictors for *Planktothrix* (named as GAM3)

Term	edf	ref.df	Statistic	p-value
t2(temp, weekPAR)	5.870002	5.870002	9.200651	0
s(TP)	2.222784	2.222784	3.073369	0.06404
s(wind)	4.665106	4.665106	3.782337	0.00378

Table S17Summary of GAM model with 4 predictors for *Pseudanabaena* (named as GAM4)