**Supplementary Material**

Ecological equivalence of species within phytoplankton functional groups

By

Crispin M. Mutshinda1, Zoe V. Finkel2 , Claire E. Widdicombe3, Andrew J. Irwin1

1Department of Mathematics and Computer Science, Mount Allison University, Sackville, NB, Canada

2Environmental Science Program, Mount Allison University, Sackville, NB, Canada

3Plymouth Marine Laboratory, Prospect Place, Plymouth, PL1 3DH, UK

1. Summary data for phytoplankton observations

**Table S1.**

List of diatom and dinoflagellate species and morphotypes from Station L4 time series. Columns from left to right: genus and species name, with optional upper range of size class, abbreviated name used in Fig. 4, number of weeks species was observed out of a possible 349 (n), carbon quota (pg C cell–1), and our neutrality index (phi).

|  |
| --- |
| **Diatoms** |
| **Species** | **Abbreviation** | **n** | **C quota** | **phi** |
| *Bacillaria paradoxa* | *B. paradoxa* | 42 | 260 | 0.6 |
| *Chaetoceros affinis* | *C. affinis* | 66 | 25 | 0.62 |
| *Corethron criophilum* | *C. criophilum* | 69 | 900 | 0.64 |
| *Chaetoceros danicus* | *C. danicus* | 90 | 200 | 0.52 |
| *Chaetoceros debilis* | *C. debilis* | 113 | 81 | 0.41 |
| *Chaetoceros decipiens* | *C. decipiens* | 55 | 130 | 0.75 |
| *Chaetoceros densus* | *C. densus* | 80 | 270 | 0.59 |
| *Cerataulina pelagica* | *C. pelagica* | 127 | 520 | 0.67 |
| *Coscinodiscus radiatus* | *C. radiatus* | 28 | 1600 | 0.67 |
| *Chaetoceros simplex* | *C. simplex* | 46 | 8.8 | 0.66 |
| *Chaetoceros socialis* | *C. socialis* | 40 | 12 | 0.49 |
| *Ditylum brightwellii* | *D. brightwellii* | 68 | 1100 | 0.62 |
| *Diploneis crabro* | *D. crabro* | 111 | 160 | 0.58 |
| *Dactyliosolen fragilissimus* | *D. fragilissimus* | 81 | 340 | 0.59 |
| *Detonula pumila* | *D. pumila* | 25 | 500 | 0.53 |
| *Eucampia zodiacus* | *E. zodiacus* | 90 | 270 | 0.62 |
| *Guinardia delicatula* | *G. delicatula* | 197 | 340 | 0.44 |
| *Guinardia flaccida* | *G. flaccida* | 85 | 7800 | 0.52 |
| *Guinardia striata* | *G. striata* | 97 | 140 | 0.58 |
| *Lauderia annulata* | *L. annulata* | 112 | 670 | 0.6 |
| *Leptocylindrus danicus* | *L. danicus* | 104 | 21 | 0.48 |
| *Leptocylindrus mediterraneus* | *L. mediterraneus* | 36 | 11 | 0.78 |
| *Leptocylindrus minimus* | *L. minimus* | 71 | 3.2 | 0.56 |
| *Meuniera membranacea* | *M. membranacea* | 138 | 660 | 0.43 |
| *Nitzschia closterium* | *N. closterium* | 246 | 17 | 0.46 |
| *Navicula distans* | *N. distans* | 99 | 140 | 0.66 |
| *Nitzschia sigmoidea* | *N. sigmoidea* | 92 | 22 | 0.65 |
| *Navicula* sp. | *Navicula* sp. | 74 | 11 | 0.63 |
| *Odontella mobiliensis* | *O. mobiliensis* | 47 | 3600 | 0.51 |
| *Pseudo-nitzschia delicatissima* | *P-n. delicatissima* | 182 | 7.1 | 0.48 |
| *Pseudo-nitzschia pungens* | *P-n. pungens* | 48 | 44 | 0.61 |
| *Pseudo-nitzschia seriata* | *P-n. seriata* | 118 | 44 | 0.5 |
| *Proboscia alata* | *P. alata* | 25 | 1400 | 0.65 |
| *Proboscia alata* 5µm | *P. alata* 5µm | 112 | 330 | 0.5 |
| *Psammodictyon panduriforme* | *P. panduriforme* | 23 | 160 | 0.85 |
| *Pleurosigma planctonicum* | *P. planctonicum* | 48 | 2800 | 0.61 |
| *Podosira stelligera* | *P. stelligera* | 35 | 2300 | 0.56 |
| *Paralia sulcata* | *P. sulcata* | 156 | 200 | 0.58 |
| *Proboscia truncata* | *P. truncata* | 30 | 6000 | 0.59 |
| Pennate30µm | Pennate30µm | 34 | 61 | 0.65 |
| Pennate50µm | Pennate50µm | 57 | 61 | 0.64 |
| *Pleurosigma* | *Pleurosigma* sp. | 126 | 520 | 0.44 |
| *Rhizosolenia imbricata* 5µm | *R. imbricata* 5µm | 67 | 330 | 0.66 |
| *Rhizosolenia imbricata* 10µm | *R. imbricata* 10µm | 56 | 1000 | 0.57 |
| *Rhizosolenia imbricata* 15µm | *R. imbricata* 15µm | 62 | 1700 | 0.54 |
| *Rhizosolenia setigera* 5µm | *R. setigera* 5µm | 82 | 170 | 0.6 |
| *Rhizosolenia setigera* 25µm | *R. setigera* 25µm | 47 | 3200 | 0.53 |
| *Rhizosolenia styliformis* | *R. styliformis* | 45 | 6000 | 0.56 |
| *Roperia tesselata* | *R. tesselata* | 80 | 610 | 0.67 |
| *Skeletonema costatum* | *S. costatum* | 66 | 6 | 0.58 |
| Small Pennate | Small pennate | 37 | 30 | 0.44 |
| *Thalassionema nitzschioides* | *T. nitzschioides* | 88 | 19 | 0.66 |
| *Thalassiosira punctigera* | *T. punctigera* | 40 | 1600 | 0.6 |
| *Thalassiosira rotula* | *T. rotula* | 70 | 520 | 0.5 |
| *Thalassiosira* 4µm | *Thalassiosira* 4µm | 44 | 2.4 | 0.6 |
| *Thalassiosira* 10µm | *Thalassiosira* 10µm | 103 | 37 | 0.62 |
| *Thalassiosira* 20µm | *Thalassiosira* 20µm | 44 | 200 | 0.58 |
|  |  |  |  |  |
| **Dinoflagellates** |
| ***Species*** | ***Abbreviation*** | **n** | **C quota** | **phi** |
| *Ceratium fusus* | *C. fusus* | 97 | 1200 | 0.68 |
| *Ceratium horridum* | *C. horridum* | 25 | 5600 | 0.76 |
| *Ceratium lineatum* | *C. lineatum* | 117 | 1400 | 0.64 |
| *Ceratium tripos* | *C. tripos* | 40 | 13000 | 0.75 |
| *Dinophysis acuminata* | *D. acuminata* | 81 | 2300 | 0.54 |
| *Gymnodinium* cf*. pygmaeum* | *G. pygmaeum* | 17 | 430 | 0.68 |
| *Gonyaulax spinifera* | *G. spinifera* | 21 | 2100 | 0.8 |
| *Gymnodinium* sp. | *Gymnodinium* sp. | 26 | 74 | 0.87 |
| *Karenia mikimotoi* | *K. mikimotoi* | 124 | 540 | 0.46 |
| *Mesoporos perforatus* | *M. perforatus* | 109 | 710 | 0.7 |
| *Micranthodinium* sp.  | *Micranthodinium* sp.  | 84 | 150 | 0.69 |
| *Prorocentrum balticum* | *P. balticum* | 17 | 56 | 0.6 |
| *Prorocentrum micans* | *P. micans* | 81 | 1400 | 0.52 |
| *Prorocentrum minimum* | *P. minimum* | 76 | 22 | 0.61 |
| *Prorcentrum triestinum* | *P. triestinum* | 31 | 770 | 0.66 |
| *Scripsiella trochoidea* | *S. trochoidea* | 135 | 390 | 0.58 |
| *Scripsiella* sp*. cyst* | *Scripsiella* sp. | 28 | 390 | 0.78 |

2. Functional group biomass is log normally distributed

The histogram of the diatom functional group biomass on the natural log scale is roughly bell-shaped and the overlaid normal density curve fits the histogram well (Fig. S1, top-left panel). The log-normality of diatom biomass is further illustrated by the linear normal QQ-plot (top middle panel) and the symmetric boxplot of the diatom log biomass (top-right panel). We found the same results for the dinoflagellate biomass (bottom panels). All these diagnostic plots support the log-normality assumption for the diatom and dinoflagellate biomass. This normality assumption on the functional group log-biomass is also formally corroborated by the two-sample Kolmogorov-Smirnov goodness-of-fit test (Scheffé 1943, Smirnov 1948). For each sample, we simulated a sample of size 10,000 from the hypothesized distribution the normal distribution centred at the empirical mean and standard deviation given by the empirical standard deviation of the sample of interest. We then used the 2-sample Kolmogorov-Smirnov test in R to assess the agreement of the distributions of the two samples. The results of the Kolmogorov-Smirnov test are shown in Table S2, with all p-values much larger than 0.05 cut-off, implying failure to reject the null hypothesis that the samples are log-normally distributed.

|  |  |  |
| --- | --- | --- |
| *Sample* | *D Statistics* | *p-value* |
| Diatom Log-Biomass | 0.0409 | 0.6269 |
| Dinoflagellate Log-Biomass | 0.0567 | 0.3755 |

**Table S2:** Results of Kolmogorov-Smirnov log-normality goodness-of-fit test on the observed total diatom and total dinoflagellate log biomasses over the study period.



**Figure S1**: Graphical diagnostic for the normality the total diatom (top) and total dinoflagellate (bottom) log-biomasses observed at Station L4 over the study Period. (Left) histogram of log-biomasses with normal densities overlaid, (center) normal QQ-plot of log-biomasses, and (right) boxplot of log-biomasses.

**References**

Scheffé, H. (1943) Statistical inference in the non-parametric case. Annals of Mathematical Statistics 14, 305-332.

Smirnov, N. (1948) Goodness of fit of empirical distributions. Annals of Mathematical Statistics 19, 279-281.

3. Model description

The next section contains the OpenBUGS (Thomas *et al.* 2006; Lunn *et al.* 2009) code for fitting the Bayesian model described in “Ecological equivalence of species within phytoplankton functional groups” by Mutshinda *et al.*

The model describes the neutral drift of a species’ biomass, *S*, within its functional group biomass envelope, *G*, on the natural logarithmic scale with *g* and *s* denoting the natural logarithms of *G* and *S*, respectively. The functional group log-biomass is modelled as a function of temperature (*Temp*), photosynthetically active radiation (*PAR*) and density-dependence regulation, whose effects on the functional group log-biomass are denoted by beta[1], beta[2], and delta, respectively. The proportion of the functional group biomass attributable to the focal species during week w under neutral drift and the observed counterpart are denoted by gamma[w] and p[w], respectively. The probability, eta[w], that gamma[w] is larger than p[w] is computed in BUGS through the step(.) function as eta[w]<-step(gamma[w]-p[w]) and the model R2 representing the proportion of the variation in the functional group biomass accounted for by the model is also computed within OpenBUGS.

We assessed the convergence of the MCMC by visual inspection of traceplots and autocorrelation function. Figs. S2 and S3 show traceplots with 3 Markov chains started from dispersed initial values, as well as autocorrelation functions for parameters of the diatom and dinoflagellate functional group models, respectively.

The model accounted for 96% and 98% of the variation in the diatom and dinoflagellate functional group biomasses, respectively, with model predictions close to the observed data (Fig. S4). The model residual were concentrated around zero with no trend and no serial correlation (Fig. S5), implying that our functional group biomass model assumptions are sensible and that the variation in the data, in particular the observed seasonal cycles, can largely be explained by fluctuations in irradiance and temperature.



**Figure S2:** Traceplots (left) and autocorrelations functions (ACF, right) for parameters of the diatom functional group biomass dynamics model: the intrinsic growth rate (*r*, top), the temperature effect (**, middle), and the irradiance effect (**,bottom). 50,000 iterations of three Markov chains were run starting from dispersed initial values, and a thinning factor of 5 was applied to the MCMC samples. The plotted results are based on the last 1000 post-thinning MCMC samples.



**Figure S3:** Traceplots (left) and autocorrelations functions (ACF, right) for parameters of the dinoflagellate functional group biomass dynamics model: the intrinsic growth rate (*r*, top), the temperature effect (**, middle), and the irradiance effect (**,bottom). 50,000 iterations of three Markov chains were run starting from dispersed initial values, and a thinning factor of 5 was applied to the MCMC samples. The plotted results are based on the last 1000 post-thinning MCMC samples.



**Figure S4:** Observed against predicted diatom (left) and dinoflagellate (right) functional group log-biomasses.



**Figure S5:** Model residuals (observed-predicted functional group biomasses) over the study period (left) and associated autocorrelation functions (ACFs, right). The lag in the ACFs is the time difference (in weeks). There is no trend in the residuals with all values clustering around zero, and no serial correlation, implying that our assumed functional group biomass model structure is sensible, and that the seasonal cycles in the data are largely explained by fluctuations in irradiance and temperature.

**References**

Thomas A., O'Hara R.B., Ligges U., Sturtz S. (2006) Making BUGS open. R News. 6, 12–17.

Lunn, D.; Spiegelhalter, D.; Thomas, A.; Best, N. (2009). "The BUGS project: Evolution, critique and future directions". *Statistics in Medicine* 28, 3049–3067.

4. OpenBUGS code

model {

# PROCESS MODEL

 for (t in 2:T) {

 mu[t] <- r+g[t-1]+delta\*g[t-1]+beta[1]\*Temp[t]+beta[2]\*PAR[t]

 g[t] ~ dnorm(mu[t], tau\_proc.fg)

 G[t] <- exp(g[t])

 m[t] <- log(p[t-1]\*G[t])

 s[t] ~ dnorm(m[t], tau.sp[t])

 tau.sp[t] ~ dgamma(w1, w2)

 Dem.prec[t] <- tau.sp[t]/exp(s[t-1])

 S[t] <- exp(s[t])

 gamma[t] ~ dbeta(alpha1, alpha2)

 # gamma[t]: neutral model prediction of species relative biomass

 # Residuals

 erf[t] <- y[t]-g[t]

 ers[t] <- x[t]-s[t]

 # SAMPLING MODEL

 y[t] ~ dnorm(g[t], tau\_obs1)

 x[t] ~ dnorm(s[t], tau\_obs2)

 X[t] <- exp(x[t])

 Y[t] <- exp(y[t])

 # PREDICTION

 y\_pred[t] ~ dnorm(g[t], tau\_obs1)

 x\_pred[t] ~ dnorm(s[t], tau\_obs2)

 # Prediction error

 pred\_erf[t] <- y[t] - y\_pred[t]

 pred\_ers[t] <- x[t] - x\_pred[t]

 p[t] <- X[t] / Y[t]

 # p[t]: observed species relative biomass

 eta[t] <- step(gamma[t]-p[t])

 }

 # PARAMETER MODEL AND INITIAL DISTRIBUTIONS

 r ~ dnorm(0, 0.1)

 delta ~ dnorm(0, 1)

 beta[1] ~ dnorm(0, 0.01)

 beta[2] ~ dnorm(0, 0.01)

 alpha1 ~ dexp(1)

 alpha2 ~ dexp(1)

 w1 ~ dexp(1)

 w2 ~ dexp(1)

 tau\_proc.fg ~ dgamma(1, 1)

 var\_proc.fg <- 1/tau\_proc.fg

 tau\_obs1 ~ dgamma(1, 1)

 tau\_obs2 ~ dgamma(1, 1)

 var\_obs1 <- 1/tau\_obs1

 var\_obs2 <- 1/tau\_obs2

 # INITIAL STATES DISTRIBUTIONS

 g[1] ~ dnorm(0,1.0E-3)I(0,)

 s[1] ~ dnorm(0,1.0E-3)I(,g[1])

 y[1] ~ dnorm(g[1],tau\_obs1)I(0,)

 x[1] ~ dnorm(s[1],tau\_obs2)I(,y[1])

 S[1] <- exp(s[1])

 G[1] <- exp(g[1])

 gamma[1] <- S[1]/G[1]

 p[1] <- exp(x[1])/exp(y[1])

 # COMPUTING R-Square

 for (t in 2:T) {

 numerator[t] <- (g[t]-mean(y[]))\*(g[t]-mean(y[]))

 denominator[t] <- numerator[t]+(y[t]-g[t])\*(y[t]-g[t])

 }

 R2 <- sum(numerator[2:349])/sum(denominator[2:349])

}