Supporting Information for:

Contributions of Transparent Exopolymer Particles by Specific Phytoplankton Groups in the Cosmonaut Sea, East Antarctic

Hu Ji[1, &], Xue Siyou[2, &], Zhao Jun[1, *], Li Dong[1], Fan Gaojing[3], Yu Peisong[1, *], Zhang Haifeng[1], Zhu Changfeng[1], Tao Keyu[1], Yang Xufeng[1], Zhang Cai[1], Pan Jianmin[1], Chen Min[2]

1Key Laboratory of Marine Ecosystem Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China

2College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China

3Polar Research Institute of China, Shanghai 200136, China

*Correspondence to: Zhao Jun (jzhao@sio.org.cn); Yu Peisong (yuppe@sio.org.cn)

&These authors contributed equally to this work and should be considered co-first authors.

Text 1: Linear multiple regression fitting

Table: S1 -S2

Figure: S1-S2

Text 2: CHEMTAX analysis

Linear multiple regression fitting

Multiple linear regression (MLR) is a statistical method employed to examine the relationship between two or more independent variables (explanatory variables) and a dependent variable (response variable). In multiple linear regression, it is assumed that a linear relationship exists, which can be expressed by the following equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \cdots + \epsilon$$

Where Y denotes the dependent variable (response variable). X_1, X_2 ,$\cdots\cdots$, X_n represent the independent variables (explanatory variables). β_0 is the intercept term. β_1, β_2 ,$\cdots\cdots$, β_n are the regression coefficients, which quantify the influence of each independent variable on the dependent variable. ϵ is the error term, accounting for the variability not captured by the model.

The steps for performing multiple linear regression in MATLAB typically include:

1) Data preparation: Ensure that the data format is correct. Usually, the independent and dependent variables need to be numerical arrays.

2) Creating the design matrix: The design matrix includes an intercept term (a column of all 1 s) and all the independent variables.

3) Fitting the model: Conduct multiple linear regression analysis using the fitlm function.

4) Model summary: Use the disp function to view the model summary, which includes coefficients, R-squared value, F-statistic, etc.

5) Prediction: Use the fitted model to make predictions on new data.

6) Result interpretation: Interpret the impact of independent variables on the dependent variable based on the signs and magnitudes of the coefficients.

Data Processing

The surface phytoplankton community generally exhibited a dominance pattern of haptophytes > diatoms > dinoflagellates > other groups, with five stations showing significant deviations: C8-09 (diatoms > dinoflagellates > haptophytes), C9-08 (dinoflagellates > diatoms > haptophytes), C9-11 (haptophytes > chlorophytes > diatoms), C7-06 (dinoflagellates > haptophytes > diatoms), and C7-04 (haptophytes > dinoflagellates > diatoms). These five outlier stations were excluded from subsequent analyses due to their marked departure from the overall community structure. The specific code in Matlab is as follows:

```
% Read the data, please correctly fill in the path where the text is located
data = readtable('D:\The path of the text\filename.txt');
% Extract the dependent variable and independent variables
TEP = data.TEP; % Dependent variable
X = table2array(data(:, 2:end)); % Independent variables, assuming the first column is TEP and the rest are independent variables
% Perform multiple linear regression fitting
mdl = fitlm(X, TEP);
% Display the statistical summary of the model
disp(mdl);
% If you also want to plot the effect of independent variables on the dependent variable
plot(mdl);
```

Results Presentation

The parameters after the model run are shown in Table S 1, and the fitting results are shown in Figure S1.

Table S 1 The first linear fitting result

Linear regression model:				
y ~ 1 + x1 + x2 + x3				
Estimated Coefficients:				
	Estimate	SE	tStat	pValue
Intercept	-1.1427	3.7922	-0.30133	0.76661
x1	0.189	0.040707	4.6429	0.00020226
x2	0.061411	0.032207	1.9067	0.072646
x3	0.031607	0.054937	0.57532	0.57219
Number of observations: 22, Error degrees of freedom: 18				
Root Mean Squared Error: 4.06				
R-squared: 0.831, Adjusted R-Squared: 0.803				
F-statistic vs. constant model: 29.5, p-value = 3.64e-07				

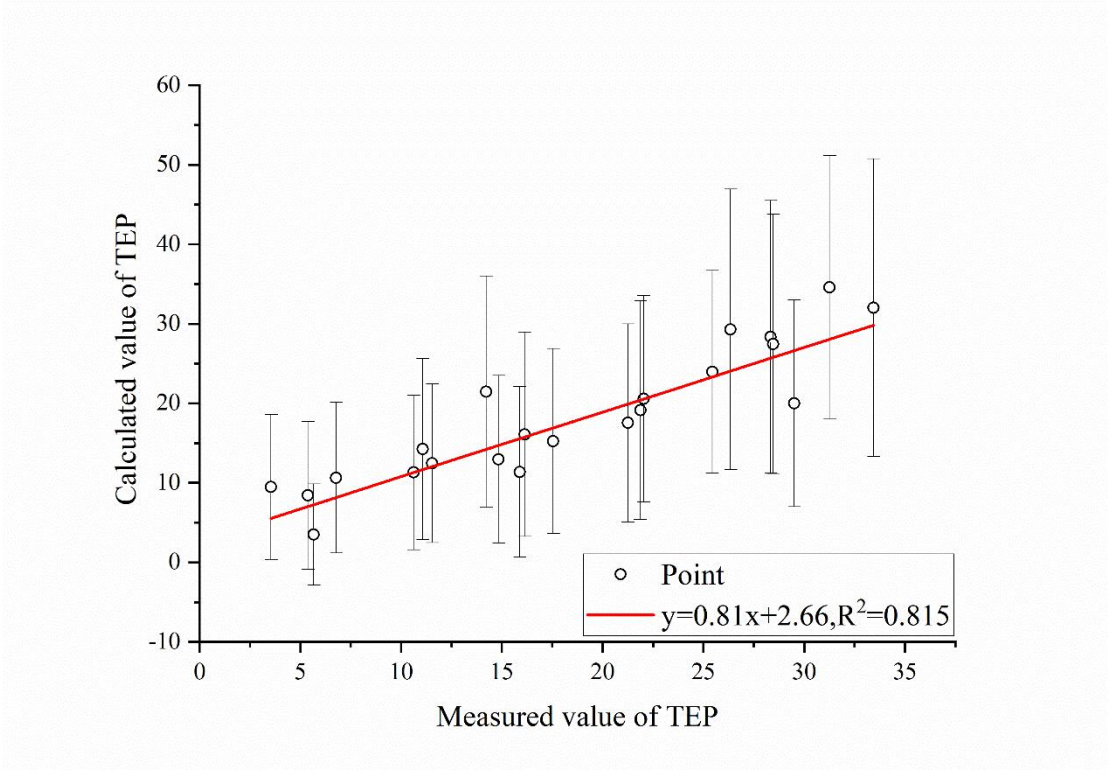

Figure S1 The linear relationship of the first fitting result

MLR fitting Optimization

In Table S 1, we found that the p-values corresponding to the intercept term ε and x3 (dinoflagellates) are relatively large, indicating that the contribution of other factors to TEP may not be significant. We set the intercept term to 0 and excluding β_{Dino} performed the fitting again. The code is as follows:

```
% Read the data, please correctly fill in the path where the text is located
data = readtable(' D:\The path of the text\filename.txt');
% Extract the dependent variable and independent variables
TEP = data.TEP; % Dependent variable
X = table2array(data(:, 2:end)); % Independent variables
```

```
% Perform multiple linear regression fitting, forcing the intercept to be 0
mdl = fitlm(X, TEP, 'Intercept', false);
% Display the statistical summary of the model
disp(mdl);
% Plot the effect of independent variables on the dependent variable
plot(mdl);
% Display R-squared and adjusted R-squared
fprintf('R-squared: %.4f\n', mdl.Rsquared.Ordinary);
fprintf('Adjusted R-squared: %.4f\n', mdl.Rsquared.Adjusted);
% Display regression coefficients
disp(mdl.Coefficients);
% Use the anova function to extract the F-statistic and p-value
anova_table = anova(mdl);
F_statistic = anova_table{2, 4}; % F-statistic is in the 2nd row, 4th column of the ANOVA table
p_value = anova_table{2, 5};     % p-value is in the 2nd row, 5th column of the ANOVA table
% Display the results
fprintf('F-statistic vs. constant model: %.2f, p-value = %.2e\n', F_statistic, p_value);
```

The parameters after the model run are shown in Table S 2, and the fitting results are shown in Figure S2

Table S 2 The final linear fitting result

Linear regression model:
y ~ x1 + x2

Estimated Coefficients:

	Estimate	SE	tStat	pValue
x1	0.20089	0.034204	5.8732	9.5775e-06
x2	0.058154	0.012254	4.7456	0.00012362

Number of observations: 22, Error degrees of freedom: 20
Root Mean Squared Error: 3.89
R-squared: 0.8223, Adjusted R-Squared: 0.8134
F-statistic vs. constant model: 22.52, p-value = 1.24e-04

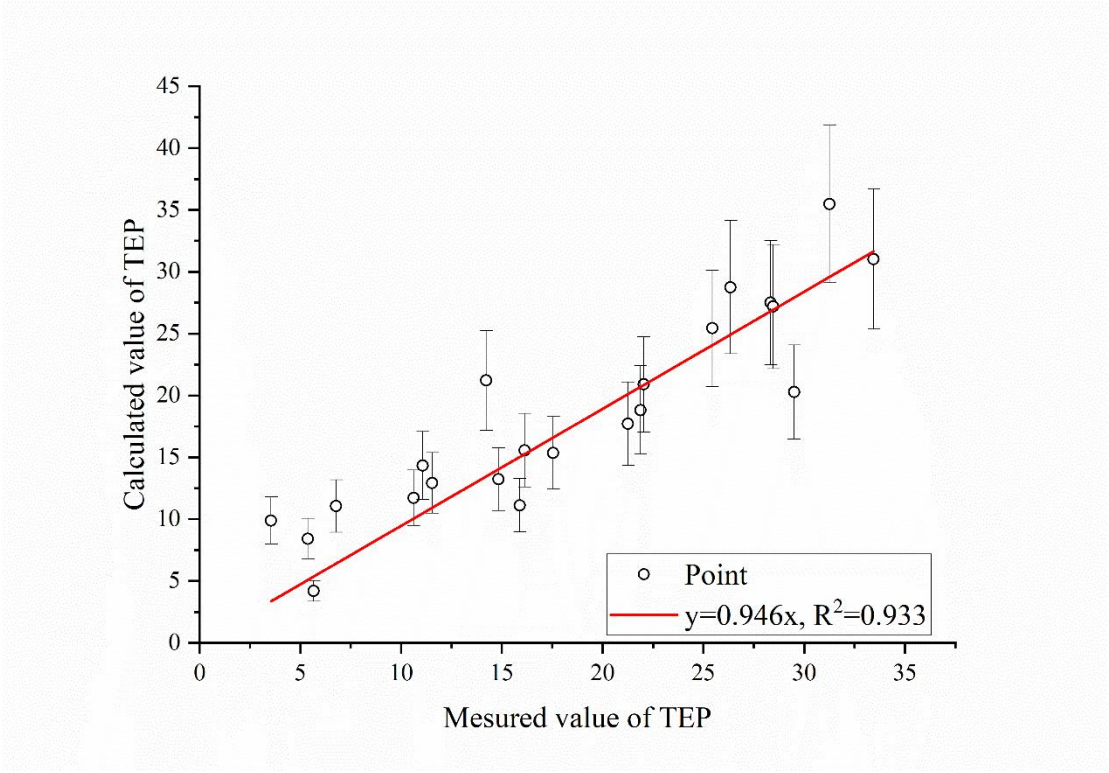

Figure S2 The linear relationship of the final fitting result

CHEMTAX analysis

In the study area, various algal groups are characterized by specific marker pigments and exhibit distinct pigment compositions, defined as ratios relative to Chl *a*. Utilizing the pigment data obtained from this study, along with initial matrices from existing literature and historical references, we identified 7 algal groups present in the surface layer of the Cosmonaut Sea. We established an initial input matrix (Table S3), which led to the final output matrix presented in Table S4. Notably, the characteristic pigments identified for diatoms, haptophytes, cryptophytes, dinoflagellates, and prochlorophytes include fucoxanthin, 19'-acyloxyfucoxanthin, alloxanthin, peridinin, and chlorophyll b, respectively. In the final data processing phase, the two diatom groups (Diat1 and Dita2) and haptophytes (Hapt6HiFe and Hapt6LoFe) were combined into a single category, respectively.

Table S3 Initial matrix of pigments in CHEMTAX

Pigment Selection	1	1	1	1	1	1	1	1	1
Class / Pigment	Chlorophyll C3	Peridinin	Fucoxanthin	Violaxanthin	19-hexanoyloxyfucoxanthin	Alloxanthin	Lutein	Chlorophyll b	Chl *a*
Chloro	0.00	0.00	0.00	0.03	0.00	0.00	0.22	0.18	1.00
Crypto	0.00	0.00	0.00	0.00	0.00	0.22	0.00	0.00	1.00
Diat1	0.00	0.00	0.57	0.00	0.00	0.00	0.00	0.00	1.00
Diat2	0.03	0.00	1.02	0.00	0.00	0.00	0.00	0.00	1.00
Dino1	0.00	0.54	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Hapt6HiFe	0.13	0.00	0.08	0.00	0.40	0.00	0.00	0.00	1.00
Hapt6LoFe	0.27	0.00	0.02	0.00	1.10	0.00	0.00	0.00	1.00

Table S4 final output matrix of pigments in CHEMTAX

Pigment Selection	1	1	1	1	1	1	1	1	1
Class / Pigment	Chlorophyll C3	Peridinin	Fucoxanthin	Violaxanthin	19-hexanoyloxyfucoxanthin	Alloxanthin	Lutein	Chlorophyll b	Chl *a*
Chloro	0.00	0.00	0.00	0.06	0.00	0.00	0.05	0.25	1.00
Crypto	0.00	0.00	0.00	0.00	0.00	0.26	0.00	0.00	1.00
Diat1	0.00	0.00	0.68	0.00	0.00	0.00	0.00	0.00	1.00
Diat2	0.03	0.00	1.22	0.00	0.00	0.00	0.00	0.00	1.00
Dino1	0.00	0.37	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Hapt6HiFe	0.26	0.00	0.10	0.00	0.36	0.00	0.00	0.00	1.00
Hapt6LoFe	0.21	0.00	0.02	0.00	1.08	0.00	0.00	0.00	1.00