Supplement of

Controlling factors on the global distribution of a representative marine heterotrophic diazotroph phylotype (Gamma A)

Zhibo Shao[1], Ya-Wei Luo[1]

[1]State Key Laboratory of Marine Environmental Science and College of Ocean and Earth Sciences, Xiamen University, 361102 Xiamen, Fujian, China

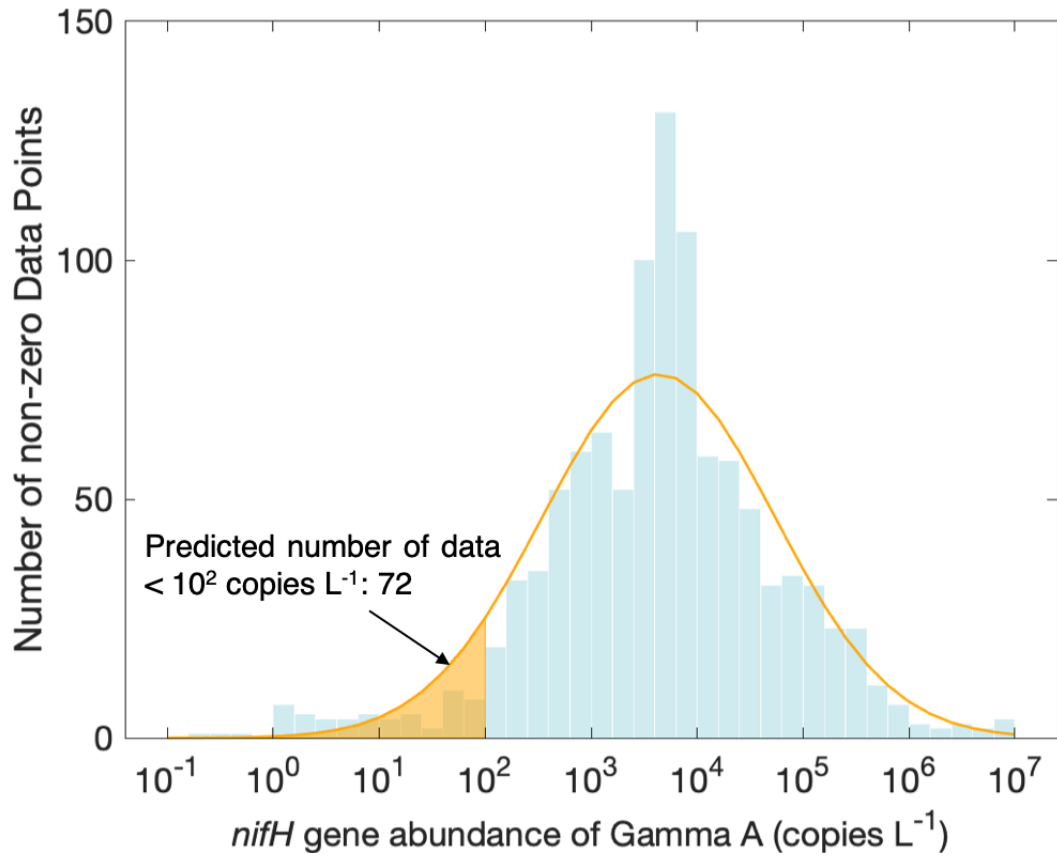*Correspondence to*: Ya-Wei Luo (ywluo@xmu.edu.cn)

Figure S1. Histogram of *nifH* abundance of Gamma A in dataset. Curve shows the predicted number of data points versus abundance, the predicted number of data smaller than 10^2 copies L^{-1} is around 72.
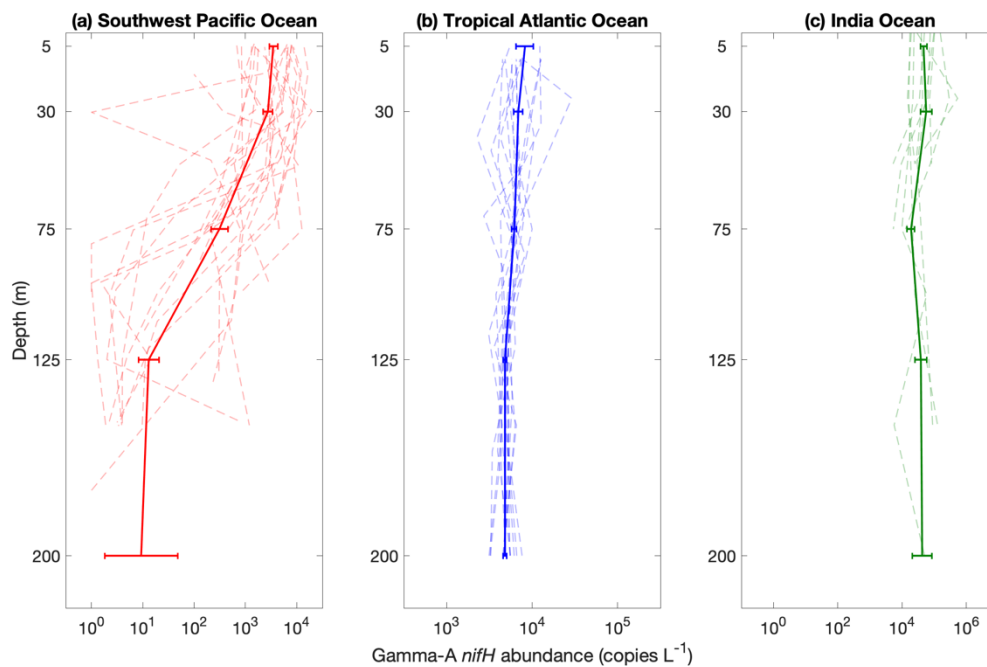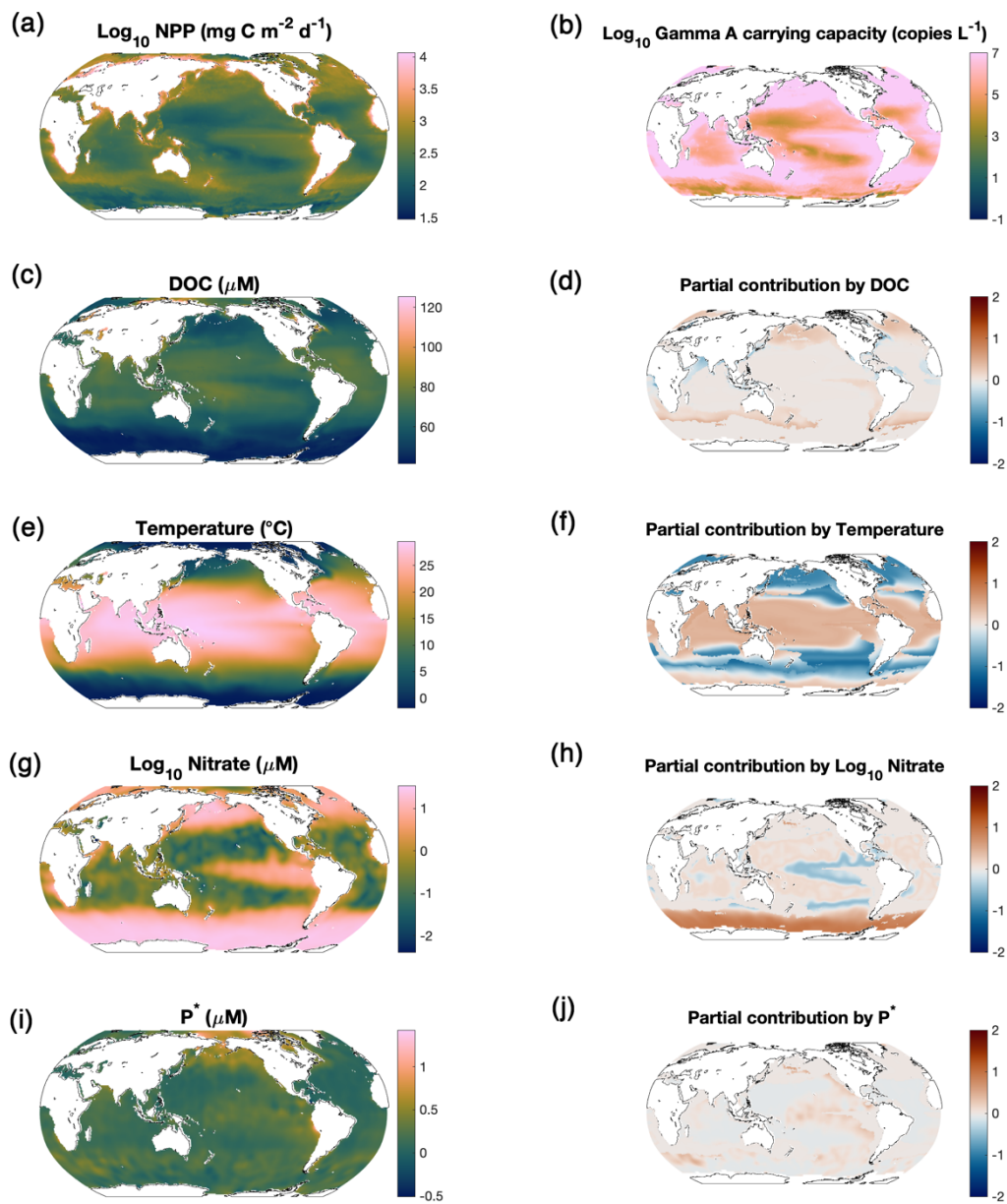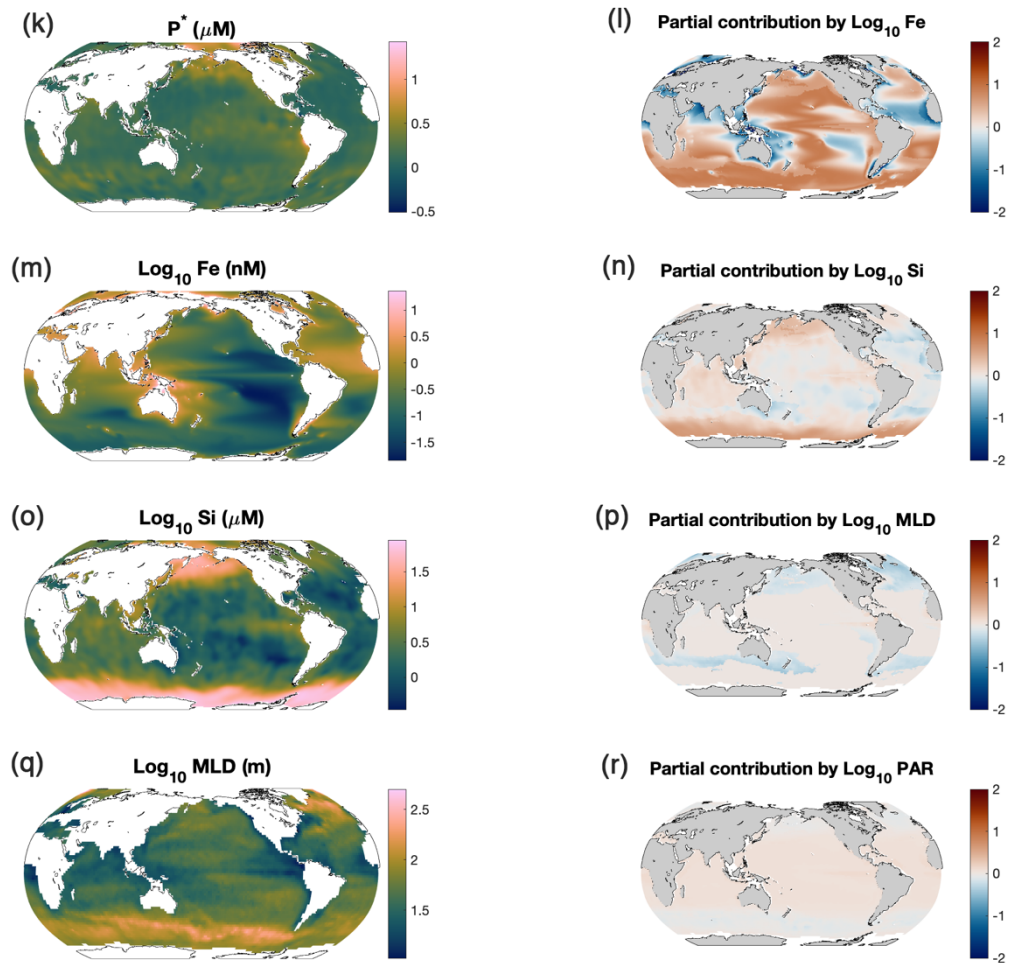
Figure S2. Vertical profiles of Gamma A abundance in (a) Southwest Pacific Ocean, (b) tropical Atlantic Ocean and (c) India Ocean. Dash lines show all the sampled profiles, and solid lines and error bars are the mean and standard error in depth ranges of 0–10 m, 10–50 m, 50–100 m, 100–150 m and 150–250 m.

(a) Log$_{10}$ NPP (mg C m^{-2} d^{-1})

(b) Log$_{10}$ Gamma A carrying capacity (copies L^{-1})

(c) DOC (μM)

(d) Partial contribution by DOC

(e) Temperature (°C)

(f) Partial contribution by Temperature

(g) Log$_{10}$ Nitrate (μM)

(h) Partial contribution by Log$_{10}$ Nitrate

(i) P* (μM)

(j) Partial contribution by P*

30

Figure continues on the next page

4

Figure S3. Maps of net primary production (NPP) and environmental parameters and their contribution to the prediction of Gamma A abundance. The data are the annual means in the surface 0–25 m ocean. The contribution includes the NPP-predicted carrying capacity and the partial contribution to the generalized additive model by the environmental parameters. The scientific color map batlow and vik are used in several figures to prevent visual distortion of the data and exclusion of readers with color-vision deficiencies (Crameri et al., 2020).
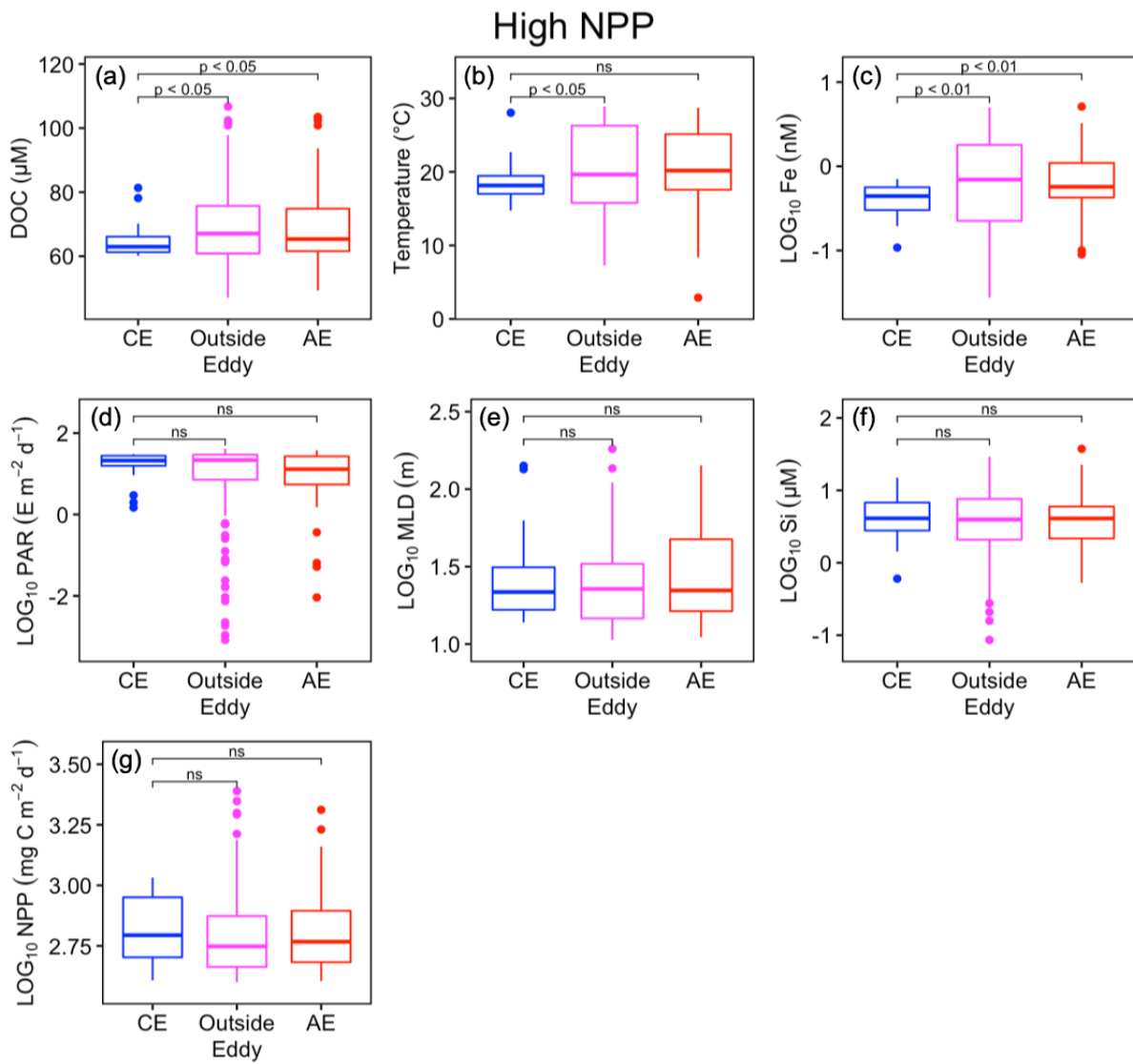
High NPP

Figure S4. Climatological data in high NPP area grouped by types of mesoscale eddies (CE: cyclonic eddy; AE: anticyclonic eddy) or outside eddies. Values shown on the brackets are p-value of t-test.
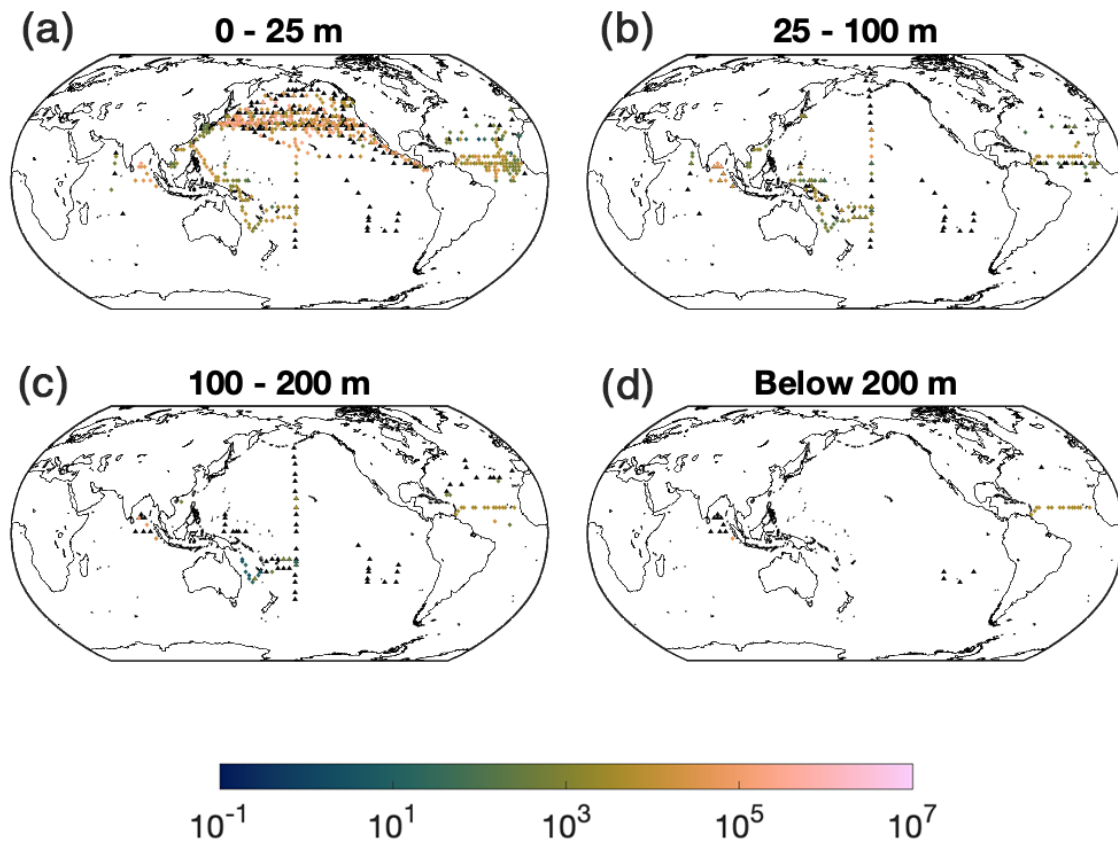
45

Figure S5. Gamma A abundance including zero-value points (*nifH* copies L^{-1}). The panels show data in depth ranges of (a) 0-25 m, (b) 25-100 m, (c) 100-200 m and (d) below 200 m. For clear demonstration, data are binned to 2° × 2° and geometric means in each bin are shown. Zero data were denoted as black triangles.

Reference

65   Crameri, F., Shephard, G. E., and Heron, P. J.: The misuse of colour in science communication, Nat. Commun., 11, 5444, 10.1038/s41467-020-19160-7, 2020.