Geochemistry, Geophysics, Geosystems

Supporting Information for

Environmental Controls of Size Distribution of Modern Planktonic Foraminifera in the tropical Indian Ocean

Michael B. Adebayo[1], Clara T. Bolton[1], Ross Marchant[2], Franck Bassinot[3], Sandrine Conrod[1], Thibault de-Garidel Thoron[1]

[1] Centre Européen de Recherche et d'Enseignement des Géosciences de l'Environnement (CEREGE) – Aix-Marseille Université, CNRS, IRD, Coll. De France, INRAE – France, [2] School of Electrical Engineering & Robotics – Queensland University of Technology – Australia, [3] Laboratoire des Sciences du Climat et de l'Environnement (IPSL) – CEA-CNRS-UVSQ – Université Paris-Saclay – France

Contents of this file

Introduction

Tables S1 to S7

Figures S1 to S7

Introduction

This supporting information file contains seven tables (Tables S1 to S7) and seven figures (Figures S1 to S7). Table S1 shows the list of R-packages used for data analysis. Six convolutional neural networks (CNNs) were tested for their ability to recognize planktonic foraminifera species accurately and automatically. After five rounds of test, the best performing network was selected based on Accuracy, Precision, and Recall. The outcome of the five iterations is presented in Table S2 and this table contains information about the network type, image input size, batch size, epochs, training time, accuracy, precision, and recall are also provided. Table S3 shows the factor pattern derived from factor analysis of planktonic foraminifera assemblage relative abundance distribution. Only the first two factors are shown because they account for the highest explainable percentage variance in assemblage distribution. In Table S4, we report the modal distribution types observed for the planktonic foraminifera species reported in this work. After investigating the relationship between individual planktonic foraminifera species' $size_{95/5}$ and the environmental parameters tested in this study, the best correlating environmental parameters along with their co-efficient of

determination (R^2) and *p*-values are shown in Table S5. With a focus on annual mean sea surface temperature (SST), we report the relationship between SST and planktonic foraminifera species' $size_{95/5}$ in Table S6. The factor pattern derived from factor analysis of planktonic foraminifera assemblage $size_{95/5}$ is provided in Table S7. Only the first two factors are also shown because they account for the highest explainable percentage variance in assemblage $size_{95/5}$ distribution.

After performing PCA on Tropical Indian Ocean (TIO) surface hydrographic parameters, the factor scores of the first and second principal component axes were projected on a map and the parameters with the highest principal component loadings on both axes are shown in Figure S1. The relative abundances of the two most abundant species in the TIO, *Globigerinita glutinata* and *Globigerinoides ruber albus,* were calculated, and their spatial regional dominance was plotted on a map shown in Figure S2. After calculating the relative abundances of the planktonic foraminifera species encountered in this study, a comparison was made between the results from this work and that of the ForCenS database (ForCenS is a curated database of planktonic foraminifera census counts in marine surface sediment samples) (reference?). Their comparability was examined spatially such that the similarity or dissimilarity in their regions of highest relative abundance can be observed. The spatial distribution maps of each species are shown in Figure S3. Figure S4 shows the relationship between $size_{95/5}$ (of nine planktonic foraminifera species) and SST as recalculated from the Rillo et al. (2020) dataset for the TIO. The Rillo et al. (2020) dataset can be accessed here: https://doi.org/10.5519/0056541. The size spectrum of the 13 most abundant TIO planktonic foraminifera species per core site studied is presented in Figure S5. In Figure S6, we show the relationship between surface carbonate ion concentration, small upwelling species, and large oligotrophic species in the TIO. We show the difference in the relative abundance patterns of *G. glutinata* and *G. bulloides* within the Bay of Bengal, Mozambique Channel, and Arabian Sea in Figure S7.

Table S1. R packages used in this work.

	Packages	**Version**	**References**
1	dplyr	1.0.10	Wickham et al. (2018, 2021)
2	extrafonts	0.18	Chang (2002)
3	forcats	0.5.2	Hyndman et al. (2021)
4	ggaluvial	0.12.3	Brunson & Read (2020)
5	Ggplot2	3.3.6	Wickham (2016)
6	ggpubr	0.4.0	Alboukadel (2020)
7	ggridges	0.5.3	Wilke (2021)
8	hrbrthemes	0.8.0	Rudis et al. (2020)
9	LaplacesDemon	16.1.6	Statisticat (2021)
10	MASS	7.3-58.1	Venables & Ripley (2002)
11	plotly	4.10.0	Sievert (2020)
12	robustbase	0.95-0	Maechler et al. (2022)
13	scales	0.5.2	Wickham et al. (2020)
14	tidyverse	1.3.2	Wickham et al. (2019)
15	tseries	0.10-51	Trapletti & Hornik (2022)
16	vegan	2.6-2	Oksanen et al. (2007)
17	viridis	0.6.2	Garnier et al. (2021)

Table S2. Performance of the "Base Cyclic 16" convolutional neural network (CNN) after five iterations.

Test Round	Image Input Size	Batch Size	Accuracy (%)	Precision (%)	Recall (%)
Round 1	128x128 greyscale	64	88.2	67.3	59.5
Round 2	128x128 greyscale	64	89.4	71.0	62.8
Round 3	128x128 greyscale	64	89.6	79.7	72.1
Round 4	128x128 greyscale	64	89.8	77.7	72.4
Round 5	128x128 greyscale	64	89.9	77.7	72.6

Table S3. *F*-Matrix derived from factor analysis of planktonic foraminifera assemblage relative abundance distribution.

Species	**Factor 1**	**Factor 2**
Globigerina bulloides	– 0.680	– 0.240
Globigerina falconensis	– 0.580	– 0.050
Globigerinella siphonifera	0.415	– 0.100
Globigerinita glutinata	– 0.893	– 0.211
Globigerinoides conglobatus	0.582	0.139
Globigerinoides elongatus	0.107	0.070
Globigerinoides ruber albus	0.398	– 0.120
Trilobatus sacculifer	0.726	– 0.228
Globoquadrina conglomerata	0.536	– 0.245
Globorotalia crassaformis	0.106	0.842
Globorotalia hirsuta	0.215	0.481
Globocoenella inflata	0.032	0.749
Globorotalia menardii	0.177	0.109
Globorotalia scitula	0.181	0.101
Globorotalia truncatulinoides	0.006	0.904
Globorotalia tumida	0.273	0.307
Globorotalia ungulata	– 0.268	– 0.148
Globorotaloides hexagonus	0.491	– 0.424
Globoturborotalita rubescens	– 0.434	– 0.221
Globoturborotalita tenella	– 0.517	– 0.345
Neogloboquadrina dutertrei	0.432	– 0.245
Neogloboquadrina incompta	0.226	0.532
Neogloboquadrina pachyderma	– 0.115	0.310
Orbulina universa	0.194	0.343
Pulleniatina obliquiloculata	0.568	– 0.268
Sphaeroidinella dehiscens	0.271	0.246

Table S4. Species' modal distribution types. We determined the modal distribution of each species' SFD using the "LaplacesDemon" package (version 16.1.6) in R which tests for unimodality, bimodality, and multimodality in a given distribution.

	Species	**Modal Distribution Type**
1	*Bella digitata*	Unimodal
2	*Candeina nitida*	Multimodal (Bimodal)
3	*Gallitellia vivans*	Multimodal
4	*Globigerina bulloides*	Unimodal
5	*Globigerina falconensis*	Unimodal
6	*Globigerinella adamsi*	Multimodal
7	*Globigerinella calida*	Multimodal
8	*Globigerinella siphonifera*	Unimodal
9	*Globigerinita glutinata*	Unimodal
10	*Globigerinita uvula*	Unimodal
11	*Globigerinoides conglobatus*	Unimodal
12	*Globigerinoides elongatus*	Unimodal
13	*Globigerinoides ruber albus*	Unimodal
14	*Globoquadrina conglomerata*	Unimodal
15	*Globorotalia crassaformis*	Unimodal
16	*Globorotalia hirsuta*	Unimodal
17	*Globocoenella inflata*	Unimodal
18	*Globorotalia menardii*	Multimodal
19	*Globorotalia scitula*	Unimodal
20	*Globorotalia truncatulinoides*	Unimodal
21	*Globorotalia tumida*	Unimodal
22	*Globorotalia ungulata*	Multimodal
23	*Globorotaloides hexagonus*	Multimodal
24	*Globoturborotalita rubescens*	Unimodal
25	*Globoturborotalita tenella*	Unimodal
26	*Hastigerina pelagica*	Unimodal
27	*Neogloboquadrina dutertrei*	Multimodal
28	*Neogloboquadrina incompta*	Unimodal
29	*Neogloboquadrina pachyderma*	Multimodal (Bimodal)
30	*Orbulina universa*	Multimodal
31	*Pulleniatina obliquiloculata*	Unimodal
32	*Sphaeroidinella dehiscens*	Multimodal
33	*Tenuitella iota*	Unimodal
34	*Trilobatus sacculifer*	Unimodal
35	*Turborotalita humilis*	Unimodal
36	*Turborotalita quinqueloba*	Multimodal

Table S5. Co-efficient of determination *(R^2)* values following a robust regression analysis between individual species size$_{95/5}$ and their best correlating environmental parameters.

Parameter	**Carbonate**			**Temperature**				**Nutrient**		**Oxygen**	**Salinity**
Species	**Carb**	**Carb200**	**ΔCarb500**	**T100**	**T200**	**Wsst**	**Sumsst**	**Phos**	**Phos200**	**ΔOx 200**	**Sal**
G. bulloides	-	-	-	-		-	-	-	0.31	-	-
G. conglobatus	0.19	-	-	-		-	-	-	-	-	-
G. conglomerata	-	-	-	-		-	-	-	-	-	0.22
G. crassaformis	-	-	-	-		-	0.23	-	-	-	-
G. falconensis	-	-	-	-		-	-	-	-	-	0.51
G. glutinata		-	-	-		-	-	0.19	-	-	-
G. hirsuta	-	-	-	-		-	-	-	-	-	0.32
G. inflata	-	-	-	-		0.35	-	-	-	-	-
G. menardii	0.18	-	-	-		-	-	-	-	-	-
G. ruber albus	-	-	-	-		-	-	-	-	0.18	-
G. rubescens	-	-	-	0.10		-	-	-	-	-	-
T. sacculifer	0.10	-	-	-		-	-	-	-	-	-
G. scitula	-	-	-	-	0.16	-	-	-	-	-	-
G. siphonfera	-	-	-	-		0.15	-	-	-	-	-
G. tenella	-	0.23	-	-		-	-	-	-	-	-
G. truncatulinoides	-	-	0.21	-	-	-	-	-	-	-	-
G. tumida	0.29	-	-	-		-	-	-	-	-	-
N. dutertrei	-	-	-	-		-	-	-	-	0.23	-
N. incompta	-	-	-	-		-	-	-	-	-	-
O. universa	-	-	-	0.20		-	-	-	-	-	-
P. obliquiloculata	0.26	-	-	-		-	-	-	-	-	-
N.incompta	None	P > 0.05									

Note. Keys — Carb (Surface carbonate concentration); Carb200 (Carbonate concentration at 200 m); ΔCarb500 (Δ [CO_3^{2-}] between 0 – 500m) ; T100 (Annual mean temperature at 100 m) ; T200 (Annual mean temperature at 200 m) ; Wsst (Winter sea surface temperature), Sumsst (Summer sea surface temperature) ; Phos (Surface phosphate concentration) ; Phos200 (Phosphate concentration at 200 m) ; ΔOx200 (Δ Oxygen concentration between 0 – 200 m) ; and Sal (salinity).

Table S6. Relationship between individual planktonic foraminifera species size$_{95/5}$ and annual mean sea surface temperature (SST).

	SST	
Species	R^2	***p*-value**
Globigerina bulloides	0.07	0.01
Globigerina falconensis	0.08	0.07
Globigerinella siphonifera	0.07	0.002
Globigerinita glutinata	0.003	0.59
Globigerinoides conglobatus	0.002	0.74
Globigerinoides ruber albus	0.02	0.16
Globoquadrina conglomerata	0.13	0.01
Globorotalia crassaformis	0.18	0.01
Globorotalia hirsuta	0.19	0.001
Globocoenella inflata	0.30	2.79e-05
Globorotalia menardii	0.05	0.03
Globorotalia scitula	0.0001	0.93
Globorotalia truncatulinoides	0.02	0.33
Globorotalia tumida	0.02	0.07
Globoturborotalita rubescens	6.51E-06	0.99
Globoturborotalita tenella	0.01	0.15
Neogloboquadrina dutertrei	0.04	0.05
Orbulina universa	0.06	0.06
Pulleniatina obliquiloculata	0.01	0.65
Trilobatus sacculifer	0.05	0.03

Table S7. *F*-Matrix loadings derived from factor analysis of planktonic foraminifera assemblage size.

Species	Factor 1	Factor 2
Globigerina bulloides	0.486	0.191
Globigerina falconensis	0.473	– 0.113
Globigerinella siphonifera	0.651	0.034
Globigerinita glutinata	0.320	0.037
Globigerinoides conglobatus	0.758	0.087
Globigerinoides elongatus	0.760	– 0.106
Globigerinoides ruber albus	0.451	0.180
Globoquadrina conglomerata	0.541	0.475
Globorotalia crassaformis	0.456	– 0.615
Globorotalia hirsuta	0.726	– 0.404
Globocoenella inflata	0.249	– 0.457
Globorotalia menardii	0.785	0.309
Globorotalia scitula	0.695	– 0.345
Globorotalia truncatulinoides	0.333	– 0.735
Globorotalia tumida	0.666	– 0.195
Globorotalia ungulata	0.578	0.025
Globorotaloides hexagonus	0.544	0.310
Globoturborotalita rubescens	0.544	0.057
Globoturborotalita tenella	0.634	0.032
Neogloboquadrina dutertrei	0.650	0.239
Neogloboquadrina incompta	0.707	– 0.030
Neogloboquadrina pachyderma	0.445	– 0.232
Orbulina universa	0.675	0.023
Pulleniatina obliquiloculata	0.755	0.211
Sphaeroidinella dehiscens	0.569	0.054
Trilobatus sacculifer	0.778	0.237

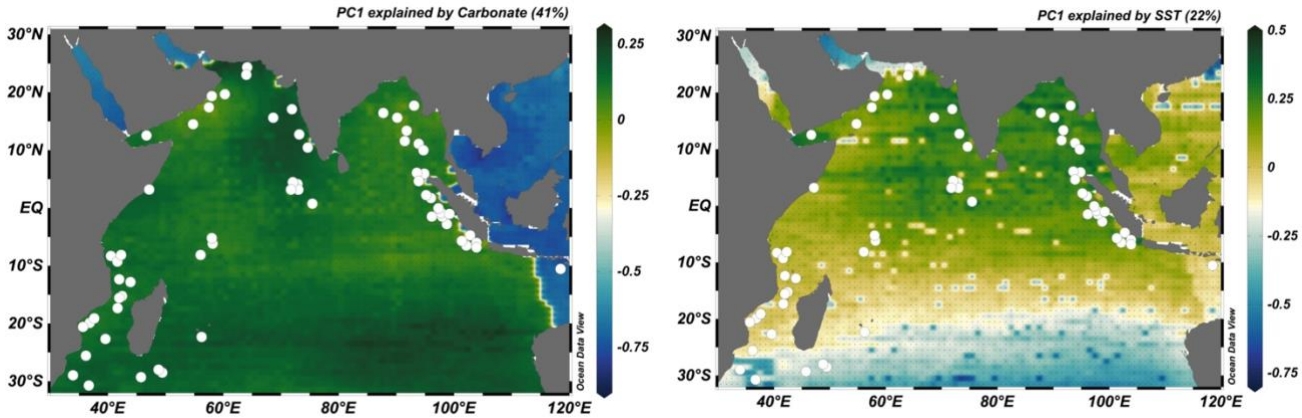


Figure S1. Principal component analysis (PCA) results of surface parameters (sea surface temperature, nitrate, phosphate, silicate, carbonate ion, and oxygen concentration, primary productivity, and salinity) from the study region in the TIO. The values represented in the plots are the scores of each raster from the PCA. The white circles represent our coretop sample sites. NB: the East China Sea, South China Sea, Celebes Sea, Banda Sea, Timor Sea, Arafura Sea and Java Sea areas are blue in color is because we could not obtain the dissolved inorganic carbon and alkalinity data needed to calculate carbonate ion concentration values. These regions are outside of our sampled study region, so this does not affect our results.

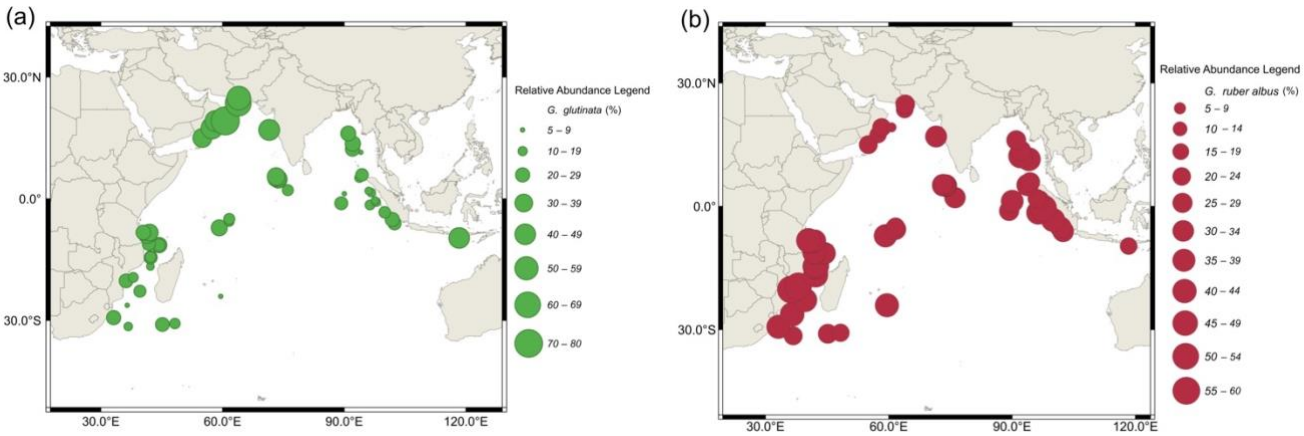
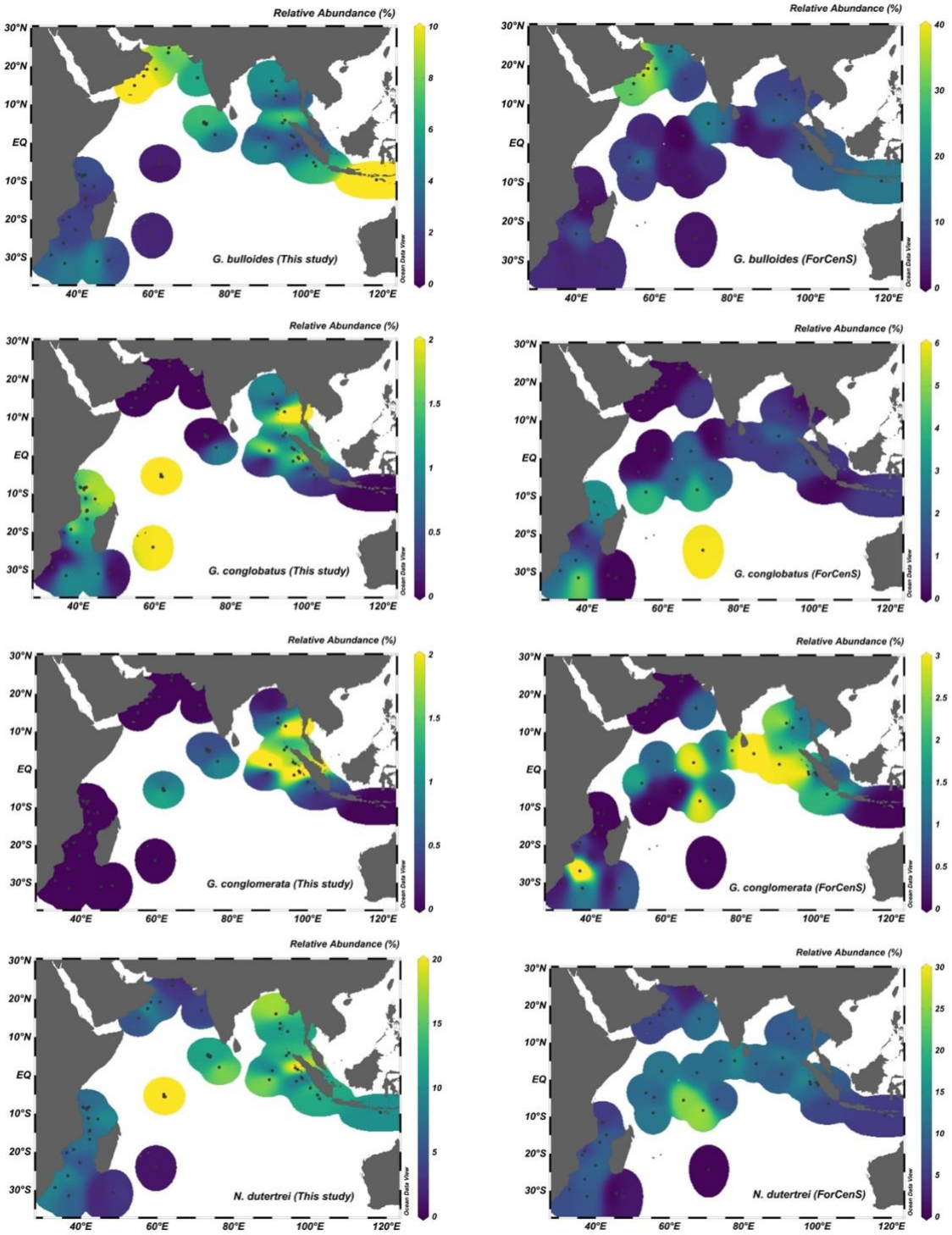
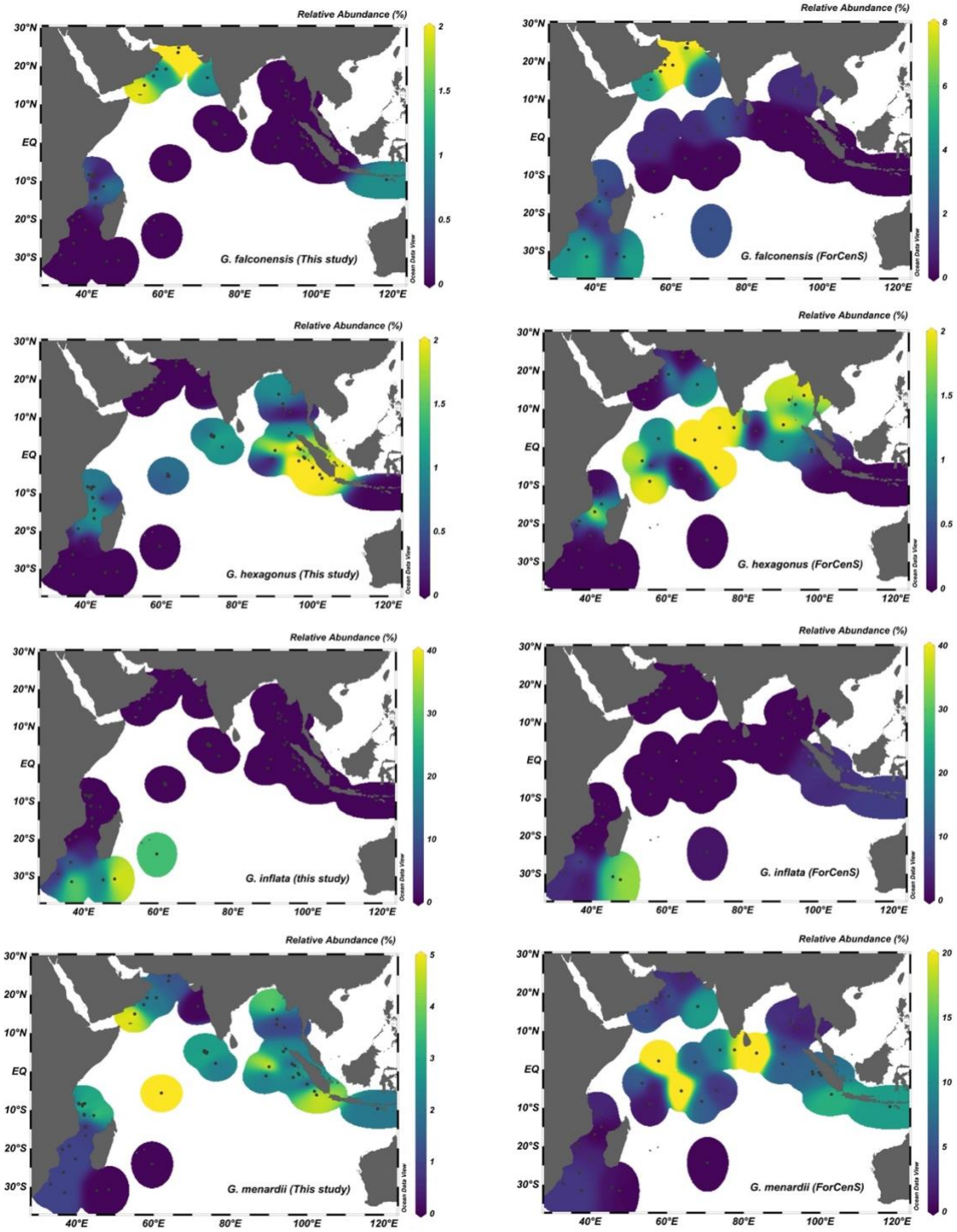
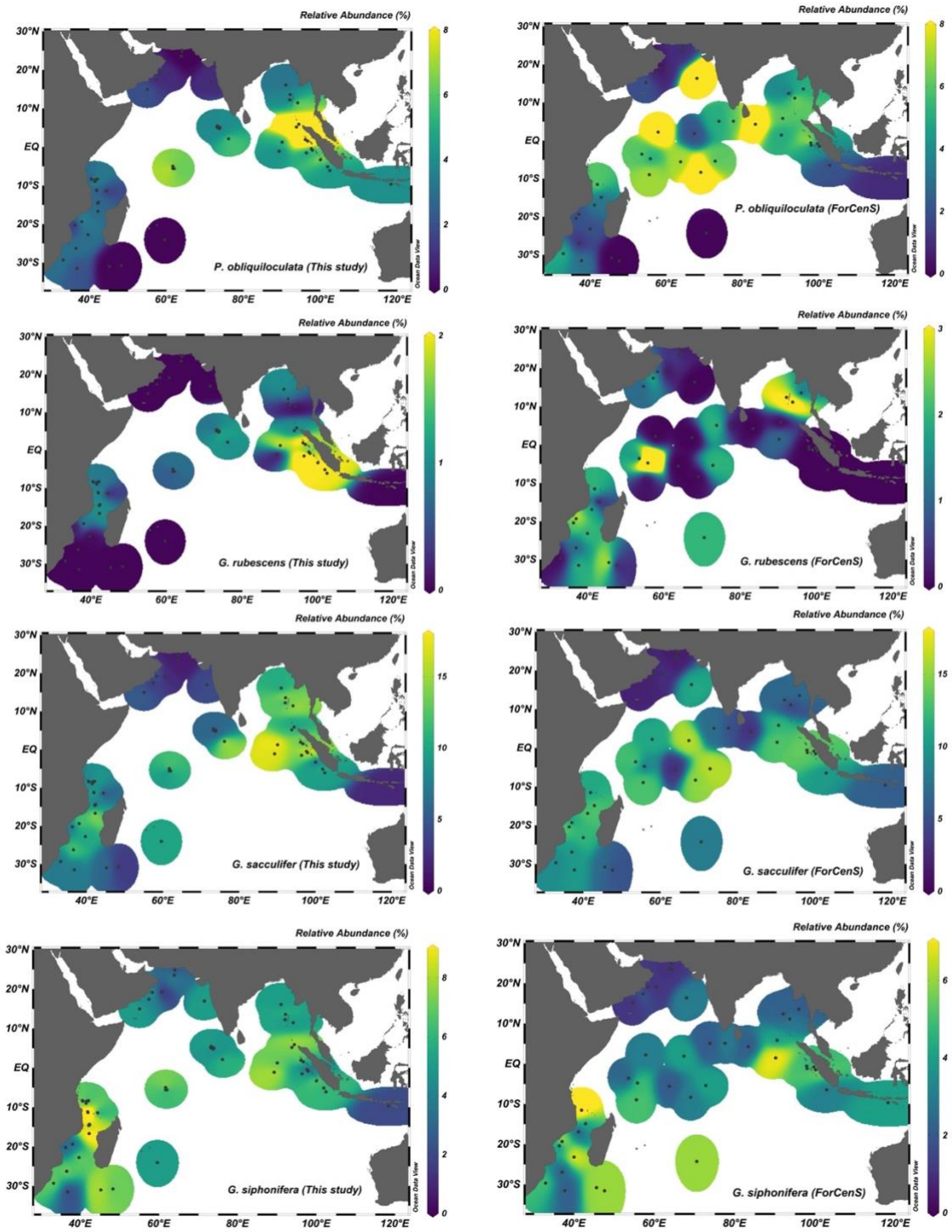


Figure S2. Map showing the relative abundance of *G. glutinata* and *G. ruber albus* in the tropical Indian Ocean (TIO). *G. glutinata* (denoted in green) dominates in the northern TIO while *G. ruber albus* (denoted in red) dominates the southern TIO.


Relative Abundance (%)
G. bulloides (This study)
G. bulloides (ForCenS)
G. conglobatus (This study)
G. conglobatus (ForCenS)
G. conglomerata (This study)
G. conglomerata (ForCenS)
N. dutertrei (This study)
N. dutertrei (ForCenS)
30°N
20°N
10°N
EQ
10°S
20°S
30°S
40°E
60°E
80°E
100°E
120°E
Ocean Data View


Relative Abundance (%)
G. falconensis (This study)
G. falconensis (ForCenS)
G. hexagonus (This study)
G. hexagonus (ForCenS)
G. inflata (this study)
G. inflata (ForCenS)
G. menardii (This study)
G. menardii (ForCenS)
30°N
20°N
10°N
EQ
10°S
20°S
30°S
40°E
60°E
80°E
100°E
120°E
Ocean Data View


Relative Abundance (%)
P. obliquiloculata (This study)
P. obliquiloculata (ForCenS)
G. rubescens (This study)
G. rubescens (ForCenS)
G. sacculifer (This study)
G. sacculifer (ForCenS)
G. siphonifera (This study)
G. siphonifera (ForCenS)
30°N
20°N
10°N
EQ
10°S
20°S
30°S
40°E
60°E
80°E
100°E
120°E
Ocean Data View

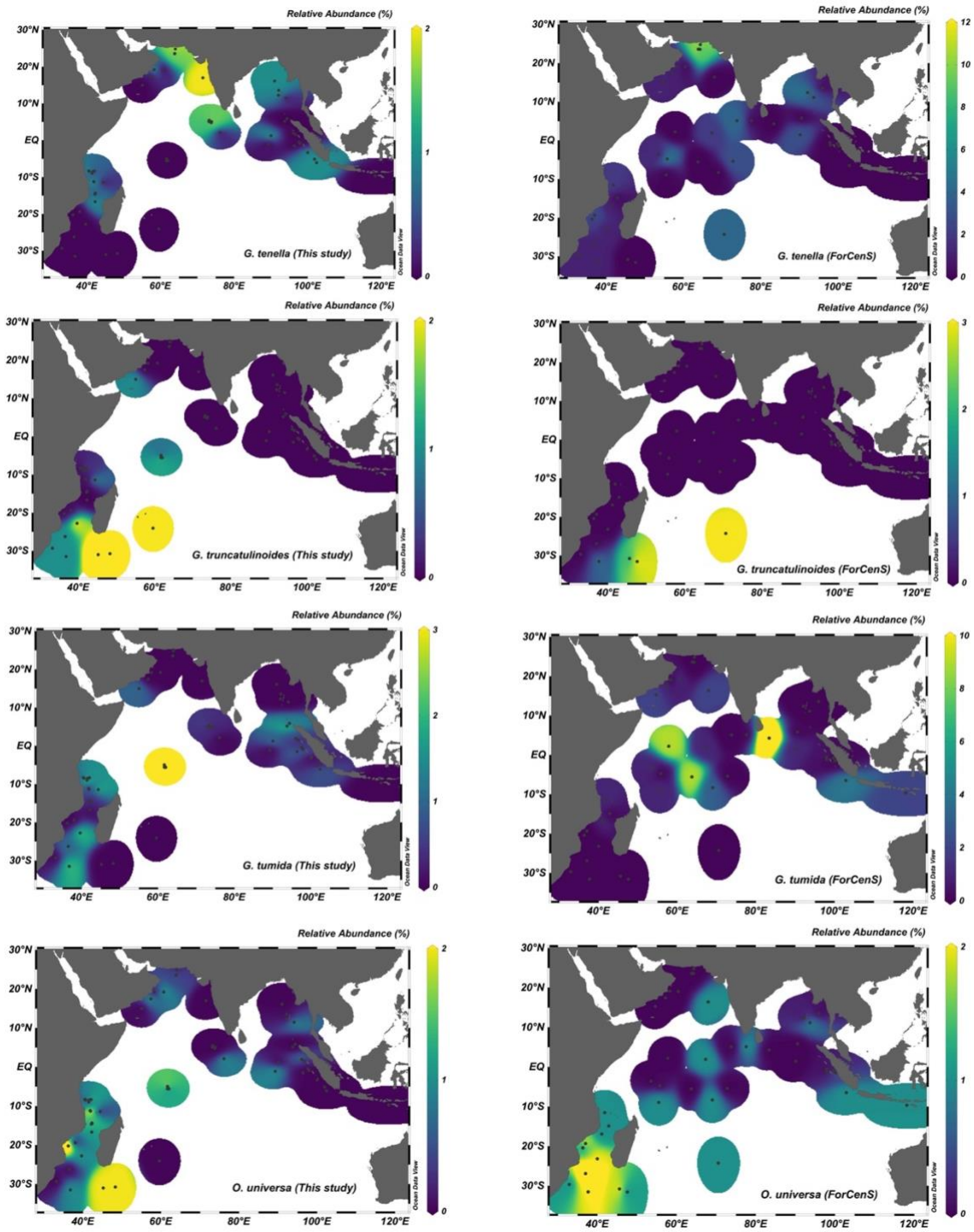


Figure S3. Similarities and differences in the relative abundance distributions of some of the species recorded in this study vs in the ForCens surface sediment database (Siccha & Kucera, 2017).

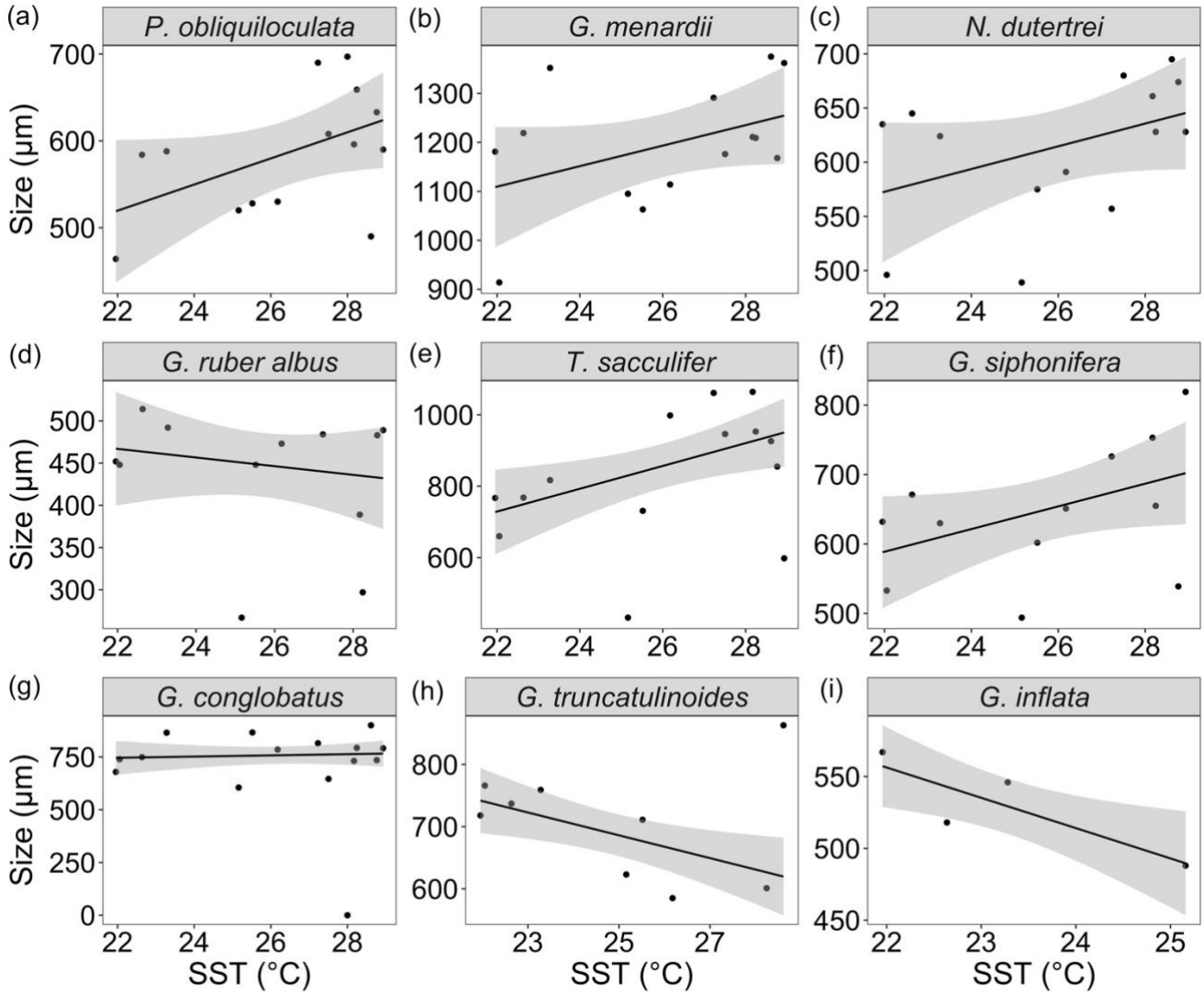


Figure S4. Relationship between size (extracted from the Rillo et al., 2020 dataset for the tropical Indian Ocean) and annual mean sea surface temperature (SST).

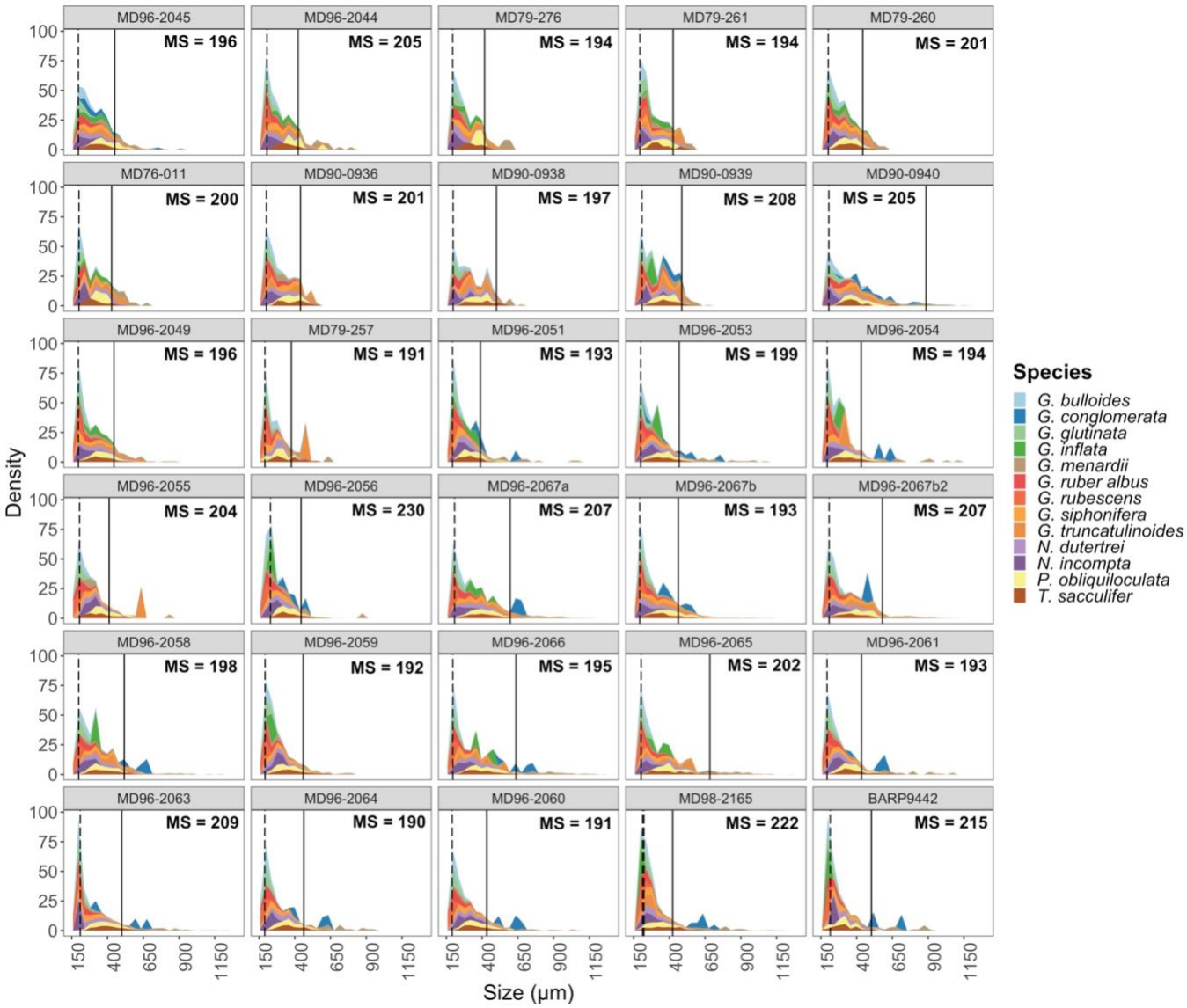


Figure S5a. Size spectra of the 13 most abundant foraminifera species in each core-top. Dashed lines represent the modal size, and the black non-dashed lines represent the $size_{95/5}$. "MS" means modal size and the corresponding values per site are written inset. The title of each facetted plot is the Core-ID of the samples used for this analysis.

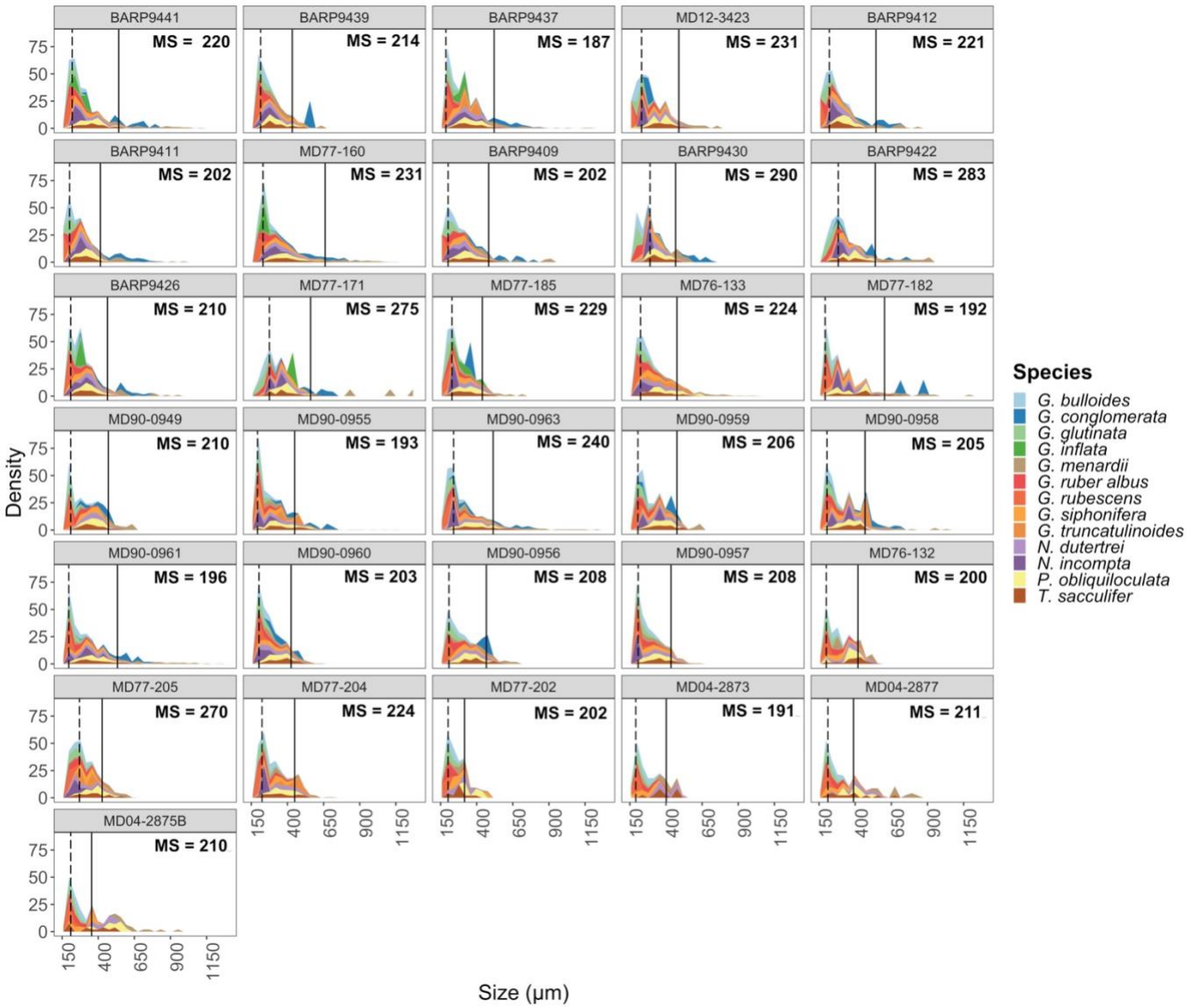


Figure S5b. Size spectra of the 13 most abundant foraminifera species in each core-top. Dashed lines represent the modal size, and the black non-dashed lines represent the $size_{95/5}$. "MS" means modal size and the corresponding values per site are written inset. The title of each facetted plot is the Core-ID of the samples used for this analysis.

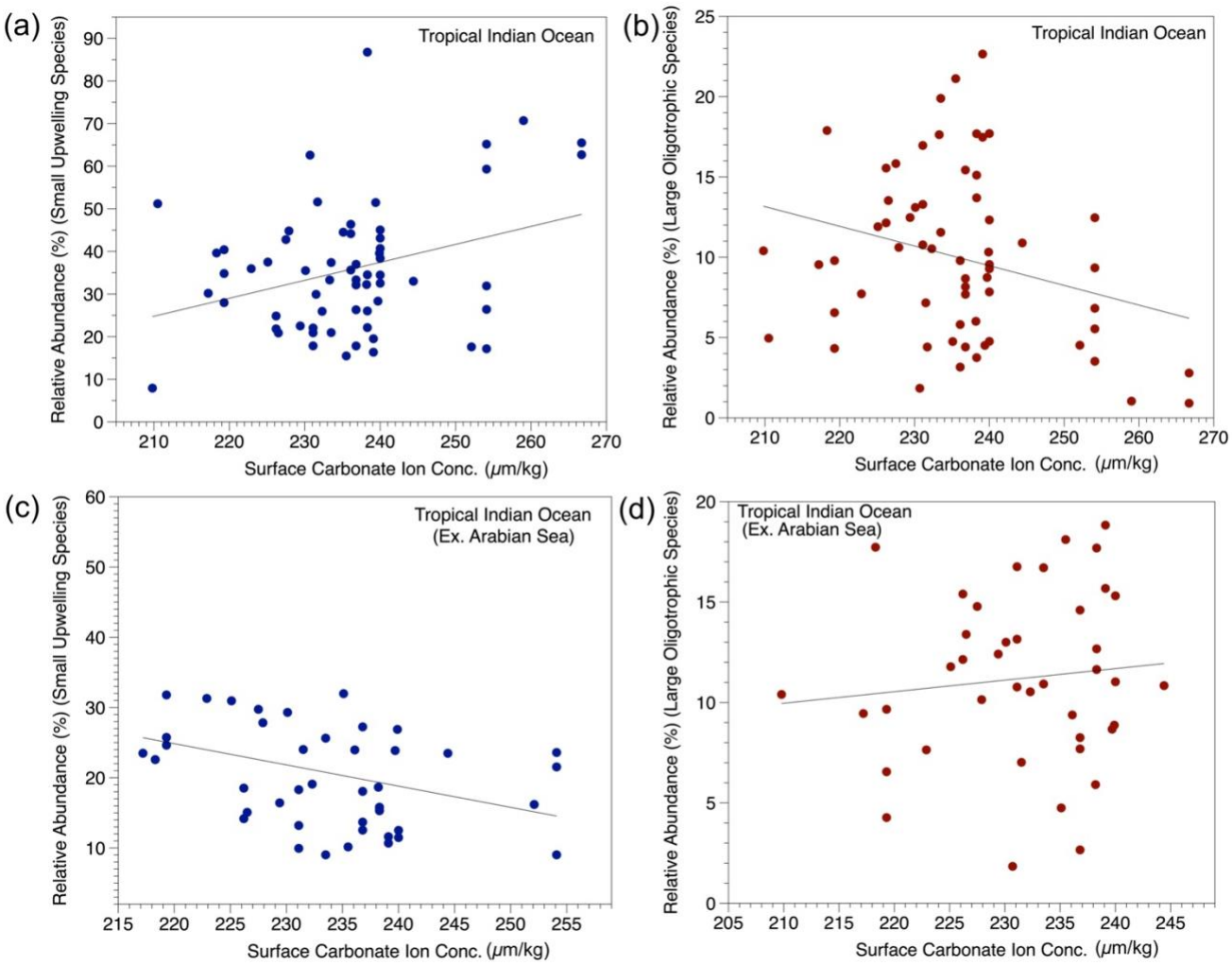


Figure S6. Relationship between surface carbonate ion concentration and (a) Relative abundance of small upwelling species (*G. bulloides*, *G. glutinata* and *N. incompta*) including all sites from all four regions defined in this study. (b) Relative abundance of large oligotrophic species (*O. universa* and *T. sacculifer*) including all sites from all four regions defined in this study. (c) Relative abundance of small upwelling species excluding Arabian Sea sites (d) Relative abundance of large oligotrophic excluding Arabian Sea sites.

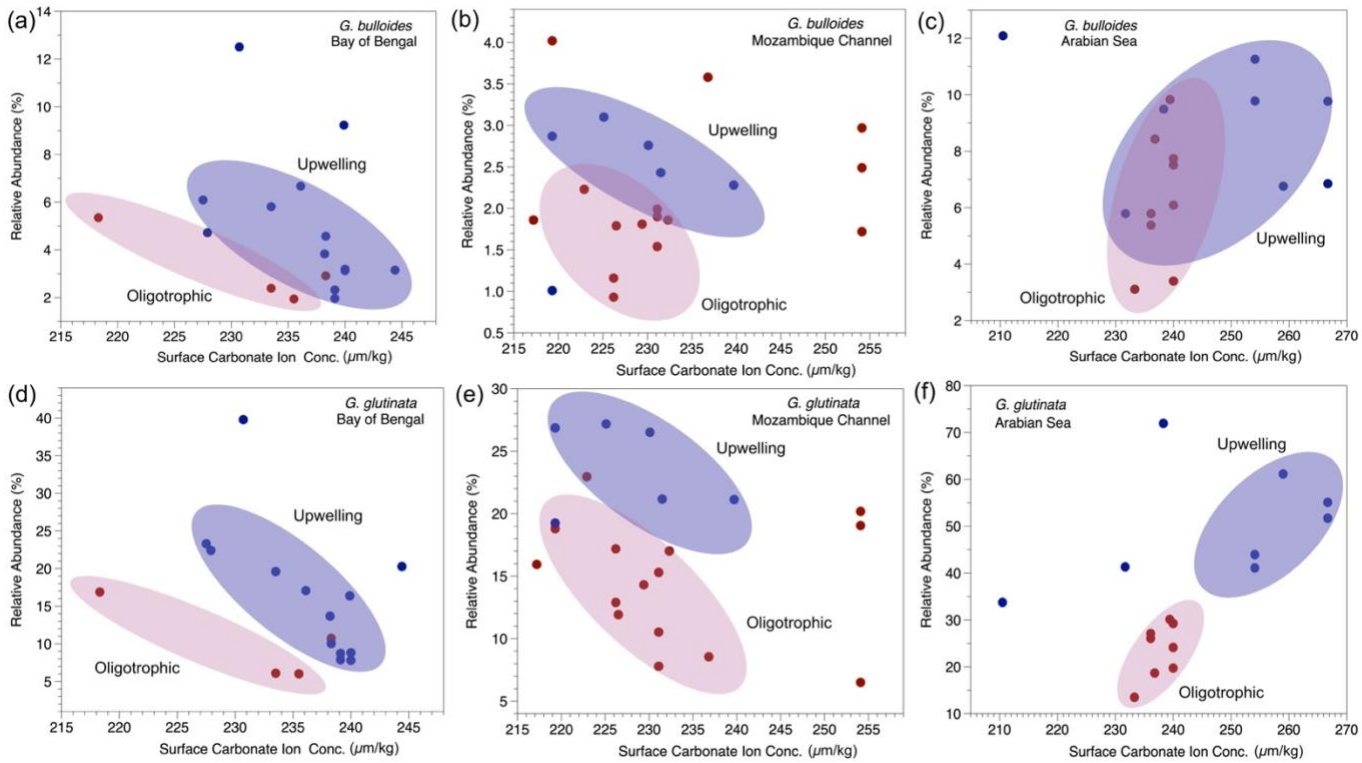


Figure S7. Relationship between surface carbonate ion concentration and (a) Relative abundance of *G. bulloides* in the Bay of Bengal, (b) Relative abundance of *G. bulloides* in the Mozambique Channel, (c) Relative abundance of *G. bulloides* in the Arabian Sea, (d) Relative abundance of *G. glutinata* in the Bay of Bengal, (e) Relative abundance of *G. glutinata* in the Mozambique Channel, and (f) Relative abundance of *G. glutinata* in the Arabian Sea. The light blue spheres are used to denote the general direction and trend among upwelling clusters sites while the light red spheres are used to denote the general direction and trend among oligotrophic cluster sites. We show here that the direction and trend of the association between the relative abundances of both *G. bulloides* and *G.* glutinata (two small upwelling species) and surface carbonate ion concentration changes in the Arabian Sea. Upwelling sites were discriminated from oligotrophic sites based on the chlorophyll-*a* and primary productivity concentrations at each site.