Dynamics of the bacterial community associated with *Phaeodactylum tricornutum* cultures

Fiona Wanjiku Moejes [1,2], Ovidiu Popa [2,3], Antonella Succurro [2,3], Julie Maguire [1], Oliver Ebenhöh [2,3]

[1] Daithi O'Murchu Marine Research Station, Bantry, Co. Cork, Ireland

[2] Institute for Quantitative and Theoretical Biology, Heinrich-Heine University, Düsseldorf, Germany

[3] CEPLAS (Cluster of Excellence on Plant Sciences), Heinrich-Heine University, Düsseldorf, Germany

Correspondence:

Jun. Prof. Oliver Ebenhöh

Institute for Quantitative and Theoretical Biology

Heinrich-Heine University

Düsseldorf, GERMANY

Tel.: +49 211 81 029 22

Fax: +49 211 81-15563

E-mail: oliver.ebenhoeh@hhu.de

Running title:

Bacterial community profile of diatom cultures

Abstract

The pennate diatom *Phaeodactylum tricornutum* is a model organism able to synthesise industrially-relevant molecules. Large-scale monocultures are prone to bio-contamination, however, little is known about the identity of the invading organisms. To gain insight into the bacterial community associated with diatoms, we translated the complexity of a natural system into reproducible experiments where we investigated the microbiome of *P. tricornutum* cultures. The results revealed a dynamic bacterial community that changed over time and in differing media conditions. We propose a network of putative interactions between *P. tricornutum* and the main bacterial factions, which is translated into a set of ordinary differential equations constituting a computational dynamic model. The proposed mathematical model is able to capture the population dynamics, further supporting the hypothesised interactions. The interdisciplinary approach implemented provides a framework for understanding the dynamics of diatom-associated microbial communities, and provides a foundation for further systematic investigations of host-microbe interactions.

Introduction

Phaeodactylum tricornutum is a diatom first described by Bohlin in 1897 when he found it in samples collected off the coast of Plymouth, United Kingdom. Diatoms belong to the Phylum *Heterokontophyta* and the Class *Bacillariophyceae* (Dangeard, 1933). They are the result of a secondary endosymbiotic event that took place around one billion years ago between a red alga (Rhodophyta) and a heterotrophic eukaryote (Bhattacharya *et al.*, 2007). Unlike most diatoms, which have the distinct ability to precipitate soluble silicic acid to form a silica cell wall, *P. tricornutum* has a poorly silicified cell wall and therefore does not have an obligate requirement for silicic acid (Montsant *et al.*, 2005; Martino *et al.*, 2007). *P. tricornutum* is found in coastal regions such as rock pools and estuaries where aquatic environmental parameters (salinity, temperature) vary greatly as a consequence of tidal changes and solar irradiation (Martino *et al.*, 2011). Its habitual characteristics, peculiar ability to form

45 oval, fusiform, and triradiate cells, as well as its poorly silicified cell wall, have triggered a tremendous
46 increase in scientific research on *P. tricornutum*. The genome sequencing of *P. tricornutum* was
47 completed in 2008, and the subsequent generation of expressed sequence tag (ESTs) databases make
48 *P. tricornutum* an excellent model organism (Montsant *et al.*, 2005; Martino *et al.*, 2007; Bowler *et al.*,
49 2008).

50 Driven by photosynthesis, *P. tricornutum* is able to synthesise a number of commercially relevant
51 molecules, applicable to various industries. In aquaculture, *P. tricornutum* is used as feed for bivalve,
52 echinoderm, crustacean and fish hatcheries (Ryther and Goldman, 1975; Tredici *et al.*, 2009). On
53 average, 18% of the *P. tricornutum* biomass are lipids, making it a potential candidate for biofuel
54 production (Kates and Volcani, 1966; Rebolloso-Fuentes *et al.*, 2001). Furthermore, *P. tricornutum* has
55 the ability to produce the poly-unsaturated fatty acids (PUFA) eicosapentaenoic acid (EPA; 20:5n-3)
56 and docosahexaenoic acid (DHA; 22:6n-3) in high proportions of the total fatty acid content (Siron *et*
57 *al.*, 1989; Rebolloso-Fuentes *et al.*, 2001; Fajardo *et al.*, 2007). Marine-derived EPA and DHA,
58 colloquially known as omega-3 PUFAs, are important in human nutrition with a vast number of health
59 benefits (Yashodhara *et al.*, 2009). *P. tricornutum* is therefore an ideal source of omega-3 PUFAs for
60 the pharma- and nutraceutical industries.

61 To fully exploit the industrial potential of *P. tricornutum* derived products, substantial amounts of
62 microalgal biomass are required, preferably with low production costs. This is achieved by
63 implementation of large-scale cultivation methods such as open raceway ponds and various types of
64 photobioreactors. Microalgal cultivation methods rely on keeping monocultures of the desired
65 species, especially if the final product is a bioactive molecule for human consumption (Mata *et al.*,
66 2010). Photobioreactors (PBRs) are closed systems that allow for the production of monoseptic
67 cultures, fully isolated from potential contamination if cultivation protocols are followed correctly
68 (Grima and Fernández, 1999). However, high operational costs of PBRs would increase production
69 costs. The other option is open raceway ponds, which are simple open-air cultivation systems that
70 have been in use since the 1950s (Chisti, 2007). They are highly susceptible to contamination, and

unless the desired species is a halophile or thermophile (Parmar *et al.*, 2011), it is hard to maintain monocultures. Irrespective of the cultivation method, the establishment of unwanted organisms such as amoeba, ciliates, rotifers, bacteria, viruses, and other photosynthetic organisms in microalgal cultures, is a serious obstacle for large-scale microalgae cultivation (Day *et al.*, 2012; Wang *et al.*, 2013). Although much research is carried out in the field of microalgal culture upscaling, very little is known about the identity and characteristics of these invading organisms, responsible for microalgal culture 'crashes' which lead to loss of biomass, and therefore, loss of revenue.

The establishment of non-target organisms in microalgal cultures should not come as a surprise. Microalgae are not found in monoculture in nature and imposing such an environment is counterintuitive leading to unstable cultures. Rather than looking at these organisms as contaminants, understanding them could allow for the exploration of 'synthetic ecology' as a novel scaling up technique, a concept proposed by Kazamia *et al.*, 2012. The cornerstone of synthetic ecology is the Competitive Exclusion Principle, or Gause's Law, which states 'as a result of competition two species scarcely ever occupy similar niches, but displace each other in such a manner that each takes possession of certain peculiar kinds of food and modes of life in which it has an advantage over its competitor' (Gause, 1934; Hardin, 1960). By 'synthesising' a community of organisms that fills every niche in the ecosystem (i.e. the microalgal culture) supporting the growth of the desired microalgae, we prevent the establishment of other, potentially harmful organisms in the culture, and optimise the utilisation of nutrients (Kazamia *et al.*, 2012).

In order for synthetic ecology to be a legitimate contender as a novel scaling up technique, greater understanding of species-specific interactions is required. Bacteria are present in all of the Earths' biomes (Dykhuizen, 1998), and insight into the microorganisms (plankton) inhabiting our oceans was greatly improved by the three-year study abroad the schooner *Tara*. In May 2015, Sunagawa *et al.* published the metagenomics data from 243 samples collected from 68 unique locations during the *Tara* expedition. The data showed that 58.8% of the sequences belonged to bacteria, even though bacterial densities (10^5 to 10^6 per gram of seawater) in our oceans are orders of magnitudes less than

those found in sediments (10^8 cells per gram), humans (10^{14} cells per gram), or soil (10^9 cells per gram) (Whitman *et al.*, 1998; Amin *et al.*, 2012). The data generated by the *Tara* project shows the sheer amplitude of genetic material belonging to bacteria, coupled with their co-existence with diatoms for more than 200 million years (Amin *et al.*, 2012), fuelled our interest in the microbiome of diatom cultures. Furthermore, in 1958, Provasoli suggested that bacteria can enhance the growth of algae (Provasoli, 1958). In the subsequent decades, species-specific studies have further corroborated his initial idea (Delucca and Mccracken, 1977; Suminto and Hirayama, 1997). Furthermore, Bruckner *et al.* showed an increase in growth of *P. tricornutum* when co-cultured with an Alphaproteobacterium strain as well as when cultured in the spent media of the bacteria (Bruckner *et al.*, 2011). A recent study conducted by Amin *et al.* shows a species-specific interaction between a coastal diatom, *Pseudo-nitzschia multiseries*, and a bacterial *Sulfitobacter* species (SA11), where the bacteria was shown to promote diatom cell division via secretion indole-3-acetic acid IAA, synthesised by the bacterium using diatom secreted and endogenous tryptophan. The IAA and tryptophan act as signalling molecules in this intricate diatom-bacteria relationship (Amin *et al.*, 2015).

With respect to the application in industry, the bacteria act as probiotics for the microalgae culture, just as bacterial probiotics have been successfully implemented in human diet by the pharma- and nutraceutical industries (Parvez and Malik, 2006), poultry industries (Kabir, 2009), and aquaculture industries (Qi *et al.*, 2009), to name a few. By identifying the bacterial community in non-axenic *P. tricornutum* cultures we can start to identify and characterise those that may have a beneficial role in the cultures. Subsequently, a suitable candidate to fill a certain niche in the hypothetical synthetic ecosystem could be chosen.

Results

In order to translate the complexity of a natural system into a reproducible, systematic experimental approach, batch cultures of *Phaeodactylum tricornutum* (CCAP 1052/1B) were cultivated in two media conditions; (1) complete F/2 medium with the addition of sodium metasilicate as the source of silicon,

and (2) minimal media with a source of nitrogen ($NaNO_3$) and phosphorus ($NaH_2PO_4.2H_2O$) at the same concentration as in the F/2 medium recipe. All *P. tricornutum* cultures were obtained from the Culture Collection of Algae and Protozoa (CCAP) based in Oban, Scotland (http://www.ccap.ac.uk/our-cultures.htm). All cultures are obtained non-axenic. Samples were taken at different stages of growth and subsequent barcoded 16S-V6-Next Generation Sequencing carried out. After the implementation of a stringent bioinformatics approach, the identity and abundance of the bacteria present in *P. tricornutum* cultures was revealed. The in the temporal evolution of the relative abundances of bacteria were used to infer a network of interactions between the diatom and the four dominant bacteria families, which was then translated into a mathematical model reproducing the community dynamics.

Characteristics of *Phaeodactylum tricornutum* culture growth

The media composition was shown to have a significant effect on the growth characteristics of *P. tricornutum*. A significant difference (p=0.042, unpaired Wilcoxon signed rank) in the maximal cell density when *P. tricornutum* is cultivated in complete (9.3 x 10^6 cells/mL) or minimal media (11.2 x 10^6 cells/mL) was observed. The growth rates during the exponential phase in both cultures were $\mu_{complete} = 0.43 \pm 0.07\ d^{-1}$ and $\mu_{minimal} = 0.51 \pm 0.04\ d^{-1}$ respectively. In contrast, the death rates when the cultures 'crash' are $\delta_{complete} = 0.09 \pm 0.02\ d^{-1}$ and $\delta_{minimal} = 0.08 \pm 0.04\ d^{-1}$ respectively.

Bacterial community profile of *Phaeodactylum tricornutum* cultures

In order to identify the bacteria present in the *P. tricornutum* cultures, the Ion Torrent™ barcoded Next Generation Sequencing protocol was used to sequence the bacterial gDNA (Quail *et al.*, 2012; Grada and Weinbrecht, 2013). The subsequent 16S rRNA gene sequences were clustered to defined Operational Taxonomic Units (OTUs) using a threshold of ≥97% sequence identity, most of which could be assigned to the genera level (Supplementary Figure S2). Of the 9727 OTUs identified, 8109 corresponded to known sequences in the SILVA database (v.118) (Quast *et al.*, 2013). The OTU

146 abundance at the phylum level showed that 99.97% of all OTUs belonged to Proteobacteria,
147 Bacteroidetes, Actinobacteria and Firmicutes (Figure 1). A comparison of the number of individual
148 reads to the number of unique OTUs showed that the high number of reads per phyla is not the result
149 of a single OTU (Supplementary Figure S3). OTUs with hits to known 16S *P. tricornutum* sequences
150 were discarded.

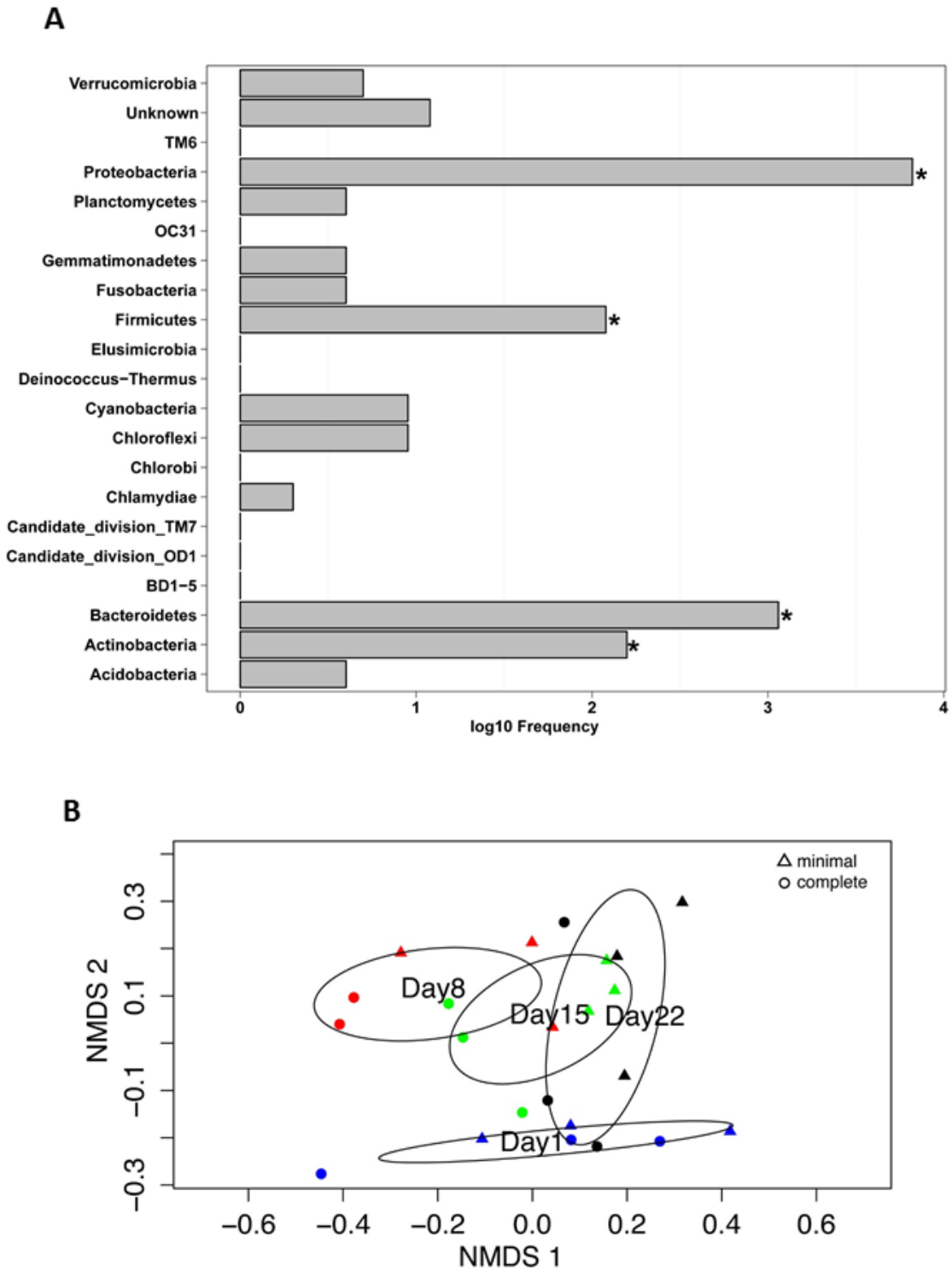
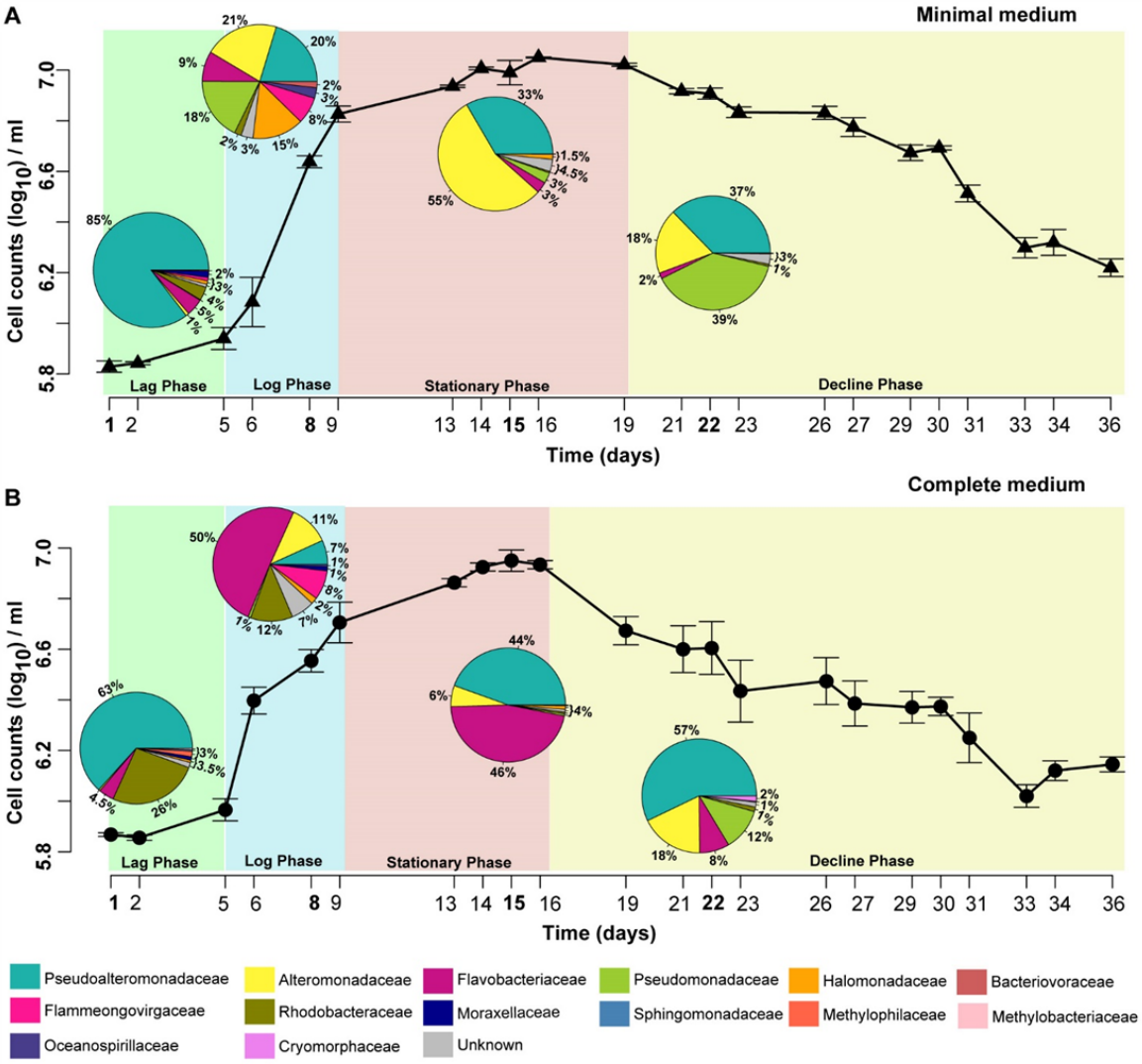


Figure 1. (A) Distribution of Operational Taxonomic Unit (OTU) abundance (LOG scaled) within phyla from complete data set. The bins marked with asterisks correspond to 99.97% of all which belong to Proteobacteria, Bacteriodetes, Actinobacteria and Firmicutes. **(B) Ordination plot of bacterial community in the two media conditions for all sampling points.** To compare the species composition between the different samples (days / media) we used a non-metric multidimensional scaling (NMDS) function based on generalised UniFrac distances (Chen *et al.*, 2012). Triangles and circles correspond to minimal media and complete media conditions, respectively. Blue represents Day 1. Red Day 8. Green Day 15. Black Day 22. The ellipses correspond to the 99% confidence interval to each group centroid.

Rarefaction curves were used to evaluate the alpha diversity in the different media conditions as well as at the different time points (Supplementary Figure S4). Species richness in both minimal and complete media was ~3000. Species richness over time remained between ~2400 and ~2600, with reduced species richness (~1300) on Day 8 (both minimal and complete media) possibly due to elevated levels of 16S *P. tricornutum* chloroplast reads which had to be omitted. Greatest species richness (~3000) was shown on Day 22. Overall, all datasets showed less increase in the number of unique species as the sample size increased, confirming adequate species richness in all culture conditions.

To compare the species composition between the different samples (days/media) we used a non-metric multidimensional scaling (NMDS) function based on generalised UniFrac distances (Chen *et al.*, 2012). We observed a clear divergence in the bacterial community in the two media conditions. Ordination based on the sampling day indicated that the bacterial community was dynamic with a clear divergence visible between Day 1 and the other three sampling days. Day 15 and 22 showed a slight overlap (Figure 2). An adapted version of PermanovaG was used to carry out permutational multivariate analysis of variance using multiple distance matrices which were previously calculated based on the generalised UniFrac distance (Chen *et al.*, 2012). The significance for the test was assessed by 5000 permutations. The results of the PermanovaG tests support the NMDS ordination,

177 confirming a statistically significant effect in the bacterial community profile at the different sampling
178 points and in the two media conditions whereas no significant effect was found in the experimental
179 replicates (Supplementary Figure S5).



180

181 **Figure 2. Bacterial community profile of *Phaeodactylum tricornutum* cultures over time and in**
182 **differing media conditions.** Both panels illustrate the growth of *P. tricornutum* (CCAP 1052/1B)
183 over a 36 day period. The growth curves have been partitioned into lag (green), log (blue),
184 stationary (red), and decline (yellow) phases. The abundance (%) of the 'Top Ten' bacterial families
185 (corresponding colours described in the key) is depicted in pie charts on Days 1, 8, 15 and 22 in
186 both media conditions. The existence of one dominant family at each investigated time point is a
187 peculiar characteristic. In **minimal media (A)**, the lag phase of *P. tricornutum* growth is dominated

by Pseudoalteromonadaceae (85%). However, during the log phase, a wide diversity of bacterial families is observed, with members of the Alteromonadaceae family (21%) beginning to dominate. During the stationary phase, a clear dominance of Alteromonadaceae species (55%) in the community can be observed. The decline phase, however, shows the Pseudomonadaceae (39%) as a dominant family, with Pseudoalteromonadaceae species (37%) increasing in abundance again. In **complete media (B)**, the lag phase is also dominated by Pseudoalteromonadaceae (63%). During the log phase, 50% of the community is composed of members of the Flavobacteriaceae family, with the other 50% distributed amongst a number of different families. Flavobacteriaceae (46%) remain high in abundance during the stationary phase, with Pseudoalteromonadaceae species (44%) beginning to increase in abundance again. As for minimal media (A), Pseudoalteromonadaceae (57%) show clear dominance of the community during the decline phase.

Effect of temporal evolution and media composition on the bacterial community profile

We compared the bacterial community profiles over time and in the different media conditions at the family level so as to avoid diluting the signal of the less abundant genera. Supplementary Figure S6 shows no dynamical difference within the genera that cannot be observed at the family level. By investigating the bacterial community dynamics at the family level, we also include taxonomical information that is unavailable at the genus level.

Overall, the families over-represented in all samples are Pseudoalteromonadaceae, Alteromonadaceae, Flavobacteriaceae and Pseudomonadaceae. Figure 2 illustrates the temporal evolution of the bacterial community in both minimal and complete media with a unique composition at each time point. A remarkable feature is that at all investigated time points there exist one or two dominant families.

Bacterial community in complete media

Members of the Pseudoalteromonadaceae family were highly abundant when *P. tricornutum* cell densities are low (63% and 57% on Day 1 and Day 22, respectively). Flavobacteriaceae species

214 dominated (50%) when the *P. tricornutum* culture is growing exponentially (Day 8). Day 15, when *P.*
215 *tricornutum* cell densities are at their highest, shows co-dominance of both Flavobacteriaceae (46%)
216 and Pseudoalteromonadaceae (44%).

217 **Bacterial community in minimal media**

218 Similarly, in the minimal media, members of the Pseudoalteromonadaceae family were highly
219 abundant when *P. tricornutum* cell densities are low. However, on Day 22 Pseudomonadaceae (39%)
220 and Pseudoalteromonadaceae (37%) are both overrepresented. When the *P. tricornutum* culture is
221 growing exponentially (Day 8) a cluster of Families dominate; namely Alteromonadaceae (21%),
222 Pseudoalteromonadaceae (20%), Pseudomonadaceae (18%), Halomonadaceae (15%) and
223 Flavobacteriaceae (9%). When the cell density of *P. tricornutum* peaks (Day 15), the
224 Alteromonadaceae species take over (55%).

225 The bacterial communities within the two media conditions on Day 1 are more closely related than
226 the communities on days 8 and 15 (see Table S2 for generalised UniFrac distances). As the cultures
227 begin to 'crash' (Day 22), the bacterial communities in the two media conditions increase in similarity
228 again.

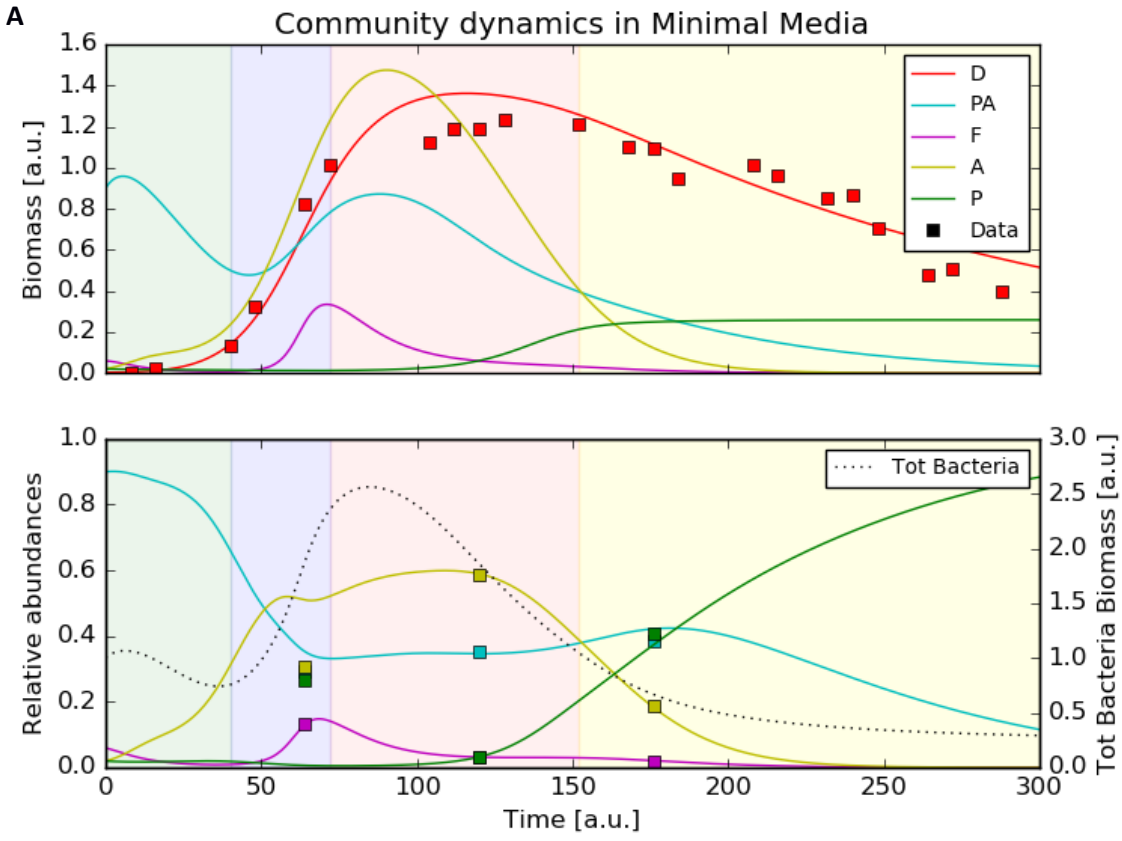
229 In general, the main families identified show a distinct pattern of disappearance and regeneration
230 within the bacterial community. In the complete media, Pseudoalteromonadaceae species start at
231 63% (Day 1), drops in abundance to 7% (Day 8) then recovers to 57% (Day 22). Flavobacteriaceae
232 species, in complete media, start at 4.5% (Day 1), increases in abundance to 50% (Day 8), and then
233 falls back to 8% (Day 22). In the minimal media, Alteromonadaceae species have an abundance of only
234 1% (Day 1), peaks at 55% (Day 15), and decreases down to 18% (Day 22).

235 **Mathematical model**

236 The dynamic changes of the bacterial communities associated with *P. tricornutum* at different growth
237 stages led us to the formulation of a network of bacteria - diatom interactions. In order to test its

238 plausibility, we developed a qualitative mathematical model starting from few key assumptions about
239 nutrients availability and metabolite exchange between the organisms involved, i.e. *P. tricornutum*
240 and general representatives of the four most abundant bacteria families Pseudoalteromonadaceae,
241 Alteromonadaceae, Flavobacteriaceae and Pseudomonadaceae.

242 In Figure 3A and B we show the simulation results from the model performed in complete media and
243 minimal media conditions, respectively, with experimental data superimposed. The top panel shows
244 biomasses of the five organisms (data available only for the diatom), the bottom panel shows relative
245 bacteria abundance versus time (biomass divided by total bacteria biomass). The figures show that
246 the model is able to reproduce the main features of the bacterial community dynamics, like the
247 disappearance and return of Pseudoalteromonadaceae in complete media and the peak of
248 Alteromonadaceae at the end of the diatom's exponential growth phase in minimal media. Because
249 of the qualitative nature of the model, units are arbitrary and the parameters used for simulation do
250 not claim any biological significance (see Supplementary Model Information for more details).



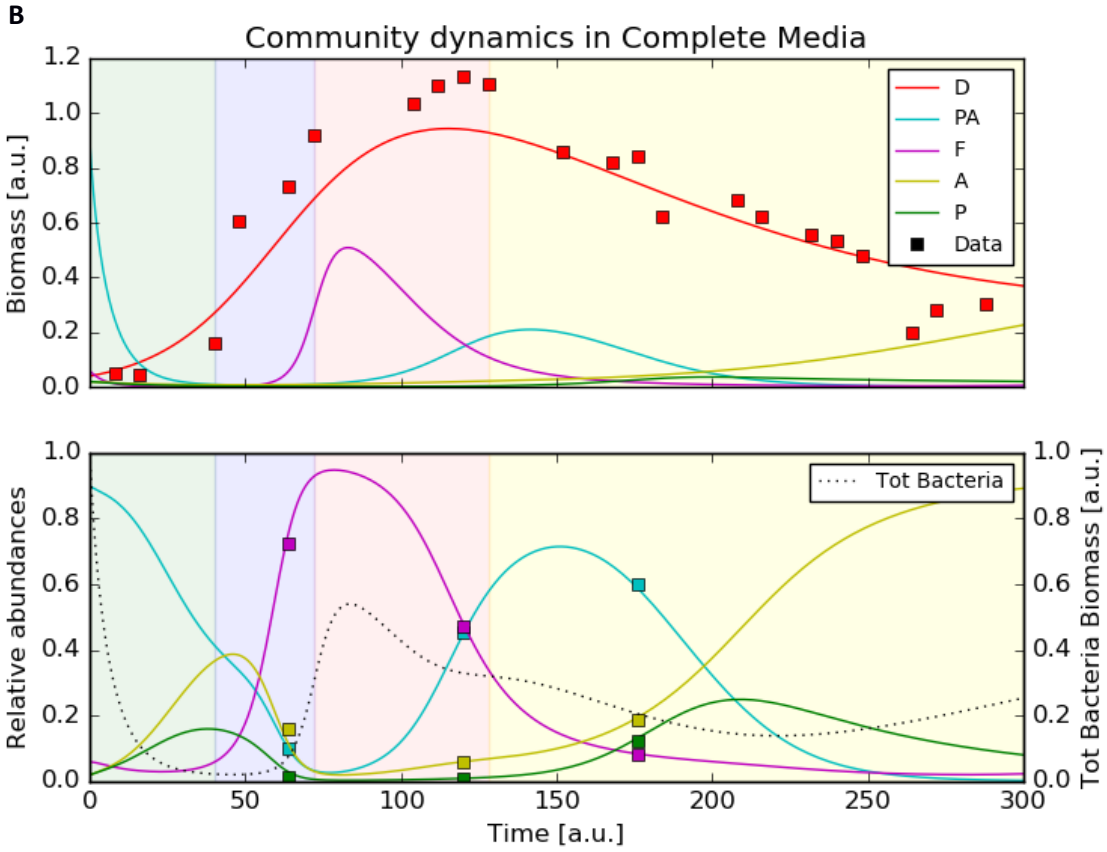


Figure 3A and B. Simulation results (lines) and experimental data (squares) for communities of *P. tricornutum* (D), Pseudoalteromonadaceae (PA), Flavobacteriaceae (F), Alteromonadaceae (A) and Pseudomonadaceae (P) in minimal (A) and complete (B) media conditions. The top panel shows the biomass time course (arbitrary units) for the five organisms and the rescaled data points (squares) for the *P. tricornutum*. The bottom panel shows the variations in relative abundances of the four bacteria (single bacteria biomass/total bacteria biomass) over time and the three sets of data points from the sequencing analysis (the first data point is used as starting condition at time 0). Also shown in the bottom plot (dotted line, right y-axis) is the total bacterial biomass in arbitrary units.

Discussion

In nature, *Phaeodactylum tricornutum* is not an isolated sovereign entity impassive to its environment including other organisms. In fact, it is part of a complex ecosystem, which is poorly understood. To

reduce the complexity of a natural system, but nevertheless to gain valuable insights into the dynamics of the bacterial communities associated with diatoms, we investigated here non-axenic cultures of laboratory strains of *P. tricornutum*. Our results showed the trends of the bacterial community dynamics during the batch growth of the *P. tricornutum*.

To progress towards the goal of creating a synthetic community, an in-depth understanding of the naturally occurring diatom-bacterial interactions, which are predominantly based on a 'biological barter trade system' between diatoms and bacteria – where substances such as trace metals, vitamins, and nutrients (nitrate, phosphate, silicate, carbon) are traded – is necessary. Based on our findings and additional insights from previous studies on diatom-bacterial interactions as well as on existing characterisation of known species from each family, we will postulate the role of the particular bacterial families in the *P. tricornutum* cultures. From this we will derive a mathematical model with the aim of reproducing the dynamical evolution of the community composition over time.

The growth dynamics of *P. tricornutum* in the two media conditions showed an accelerated 'culture crash' in the complete media compared to the minimal media, which suggests a more stable culture in the minimal media (Figure 2). Simultaneously, the dynamics of the bacterial community reveals that the community in the minimal media increases in complexity over time. The link between ecosystem complexity and stability based on theoretical and experimental data has been debated by ecologists for over half a century (MacArthur, 1955; Elton, 1958; Gardner and Ashby, 1970; Pimm, 1984). Our observations are in agreement with more recent hypotheses indicating that diversity generally increases the stability of an ecosystem (McCann, 2000).

Prospective role of central bacterial families

The putative roles of each of the dominant families are illustrated in Figure 4. The presence of **Pseudoalteromonadaceae** species is not unexpected as members of this family have been isolated from coastal, open and deep-sea waters, sediments, marine invertebrates, as well as marine fish and algae (Ivanova *et al.*, 2004). The Pseudoalteromonadaceae family has three genera, namely

289 *Pseudoalteromonas*, *Algicola* and *Psychrosphaera* (Rosenberg *et al.*, 2014, 28). Several species of
290 Pseudoalteromonadaceae are reported to possess antibiotic properties with bactericidal effects
291 (Bowman, 2007). For example, concentrated supernatant of a marine bacterium *Pseudoalteromonas*
292 *sp.* strain A28 contained various enzymes including proteases, DNases, cellulases, and amylases,
293 capable of causing the lysis of the diatom *Skeletonema costatum* (Lee *et al.*, 2000). Species of
294 Pseudoalteromonadaceae are also capable of producing cold-adapted enzymes (Venkateswaran and
295 Dohmoto, 2000; Chen *et al.*, 2007; Khudary *et al.*, 2010; Lu *et al.*, 2010; Albino *et al.*, 2012; He *et al.*,
296 2012). Pseudoalteromonadaceae species can produce extracellular polymeric substances allowing
297 them to colonise surfaces, enhancing nutrient uptake whilst limiting diffusion of particular substances
298 across the cell membrane (Holmström and Kjelleberg, 1999). The ability of Pseudoalteromonadaceae
299 species to suppress the growth of competing bacteria could explain the dominance of
300 Pseudoalteromonadaceae in almost all cultures irrespective of media composition, particularly when
301 *P. tricornutum* abundance is limited (Figure 2, Days 1 and 22). *P. tricornutum* on the other hand, may
302 protect other bacterial community members from the bacteriolytic ability of
303 Pseudoalteromonadaceae by producing specific antibacterial compounds themselves. Desbois *et al.*
304 showed that *P. tricornutum* excreted bacteriolytic fatty acids such as eicosapentaenoic acid (EPA;
305 20:5n-3), nucleotides, peptides, and pigment derivatives that can eliminate unwanted competition for
306 nutrients such as organic phosphates from certain bacteria (Desbois *et al.*, 2009).

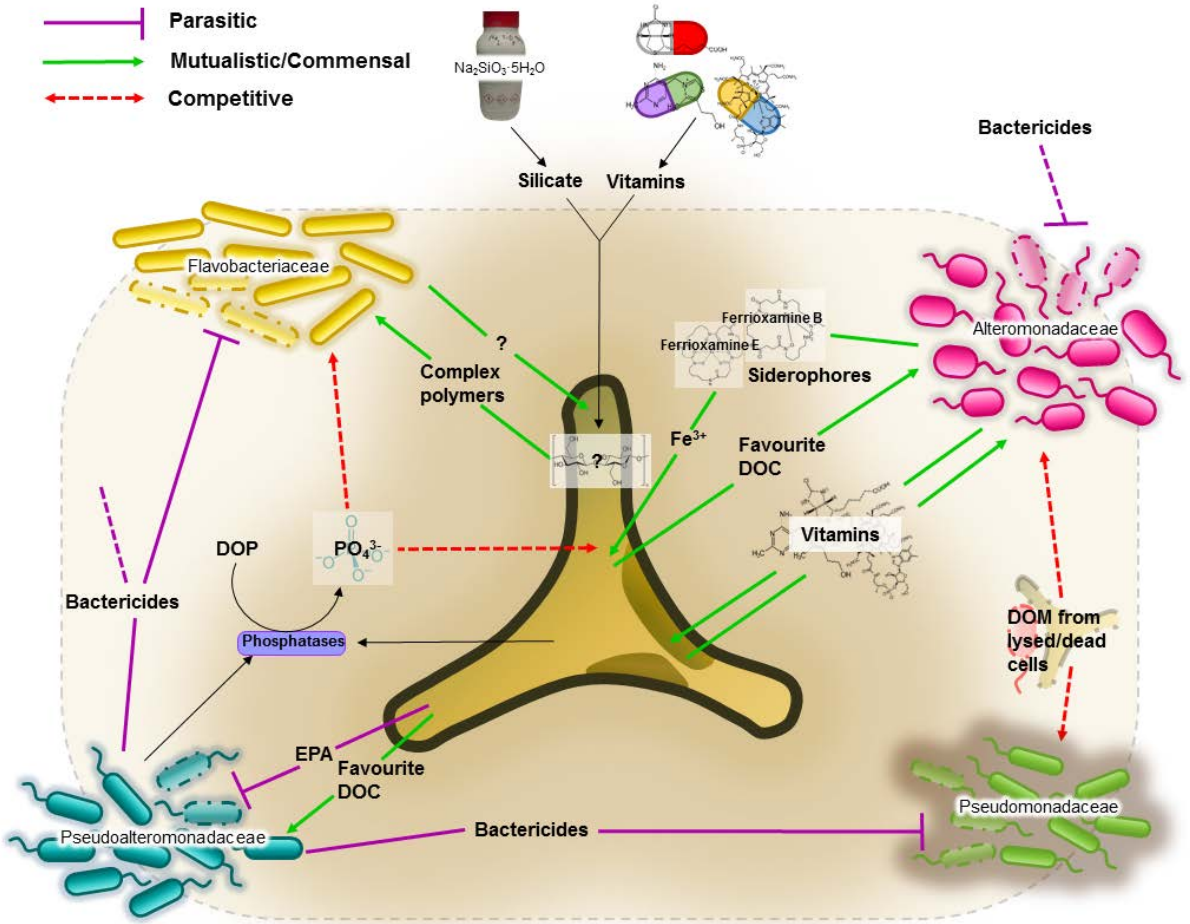


Figure 4. Network of putative interactions between *Phaeodactylum tricornutum* and identified bacterial families. The dotted grey line depicts the **'phycosphere'**; a term coined by Bell and Mitchell in 1972 as an aquatic equivalent of the 'rhizosphere', denoting the region extending outwards from the algal cell in which bacterial growth is stimulated by extracellular products of the alga (Bell and Mitchell, 1972). **Bactericidal Effects.** Several species of the Pseudoalteromonadaceae family have been reported to possess bactericidal effects (Bowman, 2007). *P. tricornutum*, however, can excrete fatty acids (such as eicosapentaenoic acid or EPA), nucleotides, peptides, and pigment derivatives to protect themselves against opportunistic attack or pathogenic damage (Desbois *et al.*, 2009). **Iron.** Siderophores are a group of iron scavengers that act by chelating iron (III). Siderophores are produced and excreted by bacteria, and some cyanobacteria, which then reuptake the siderophores with bound iron (III) via outer-membrane transporters that are siderophore-specific (Vraspir and Butler, 2009). Diatoms are not known to produce siderophores (Soria-Dengg and Horstmann, 1995; Amin *et al.*, 2009). However, based on

genome sequence analyses, the presence of a gene orthologue of a bacterial ferrichrome binding protein suggests the possibility of iron (III)-siderophore utilisation by *P. tricornutum*. Furthermore, it was shown that *P. tricornutum* was able to uptake siderophores ferrioxamines B and E (Soria-Dengg and Horstmann, 1995). **Vitamins.** Prokaryotes are thought to be the main producers of B vitamins (Provasoli, 1963; Provasoli and Carlucci, 1974). Although *P. tricornutum* does not require cobalamin, thiamine and biotin (Croft *et al.*, 2006), production of organic compounds such as EPA can by considerably enhanced by the bioavailability of co-factors such as cobalamin (Yongmanitchai and Ward, 1991). This provides the basis for potential mutualistic interactions. For example, Alteromonadales, dominant in our cultures, are thought to be capable of producing B vitamins (Sañudo-Wilhelmy *et al.*, 2014). **Dissolved Organic Carbon (DOC).** It is estimated that up to 50% of carbon fixed via phytoplankton-mediated photosynthesis is utilised by marine bacteria (Azam *et al.*, 1983), mainly as DOC compounds, defined as the organic material <0.7µm in size (Stocker, 2012). DOC from diatoms originates either from live cells or recently lysed or grazed cells, which determines the type of DOCs available, and therefore likely determining the bacterial consortia associated with the diatom (Amin *et al.*, 2012). **Dissolved Organic Phosphate (DOP).** Both diatoms and bacteria primarily utilise orthophosphate as a source of phosphorus. However, to access phosphate from DOP compounds, both diatoms and bacteria developed mechanisms such as the excretion of enzymes, including phosphatases, to release orthophosphate (PO_4^{3-}) from DOP. The mechanism is not species-specific, which consequently means the 'free' orthophosphates can be acquired by any organism (Persson *et al.*, 1988).

The **Alteromonadaceae** family consists of 16 (yet annotated) named genera (LPSN, 2016a) found predominantly in marine environments (Rosenberg *et al.*, 2014, 5). Members of this family were isolated from nutrient-rich environments such as coastal, open, and deep-sea waters, sediments, marine invertebrates and vertebrates, algae, and temperate and Antarctic marine environments (Ivanova and Mikhaïlov, 2001). They are able to utilise a vast array of compounds as carbon sources; from glucose to glycerol (Rosenberg *et al.*, 2014, 5). Members of this family are known siderophore producers (Reid and Butler, 1991; Holt *et al.*, 2005; Amin *et al.*, 2009). Greek for 'iron carrier',

siderophores are a group of iron scavengers that act by chelating iron (III) that are produced and excreted by bacteria, and some cyanobacteria, which then reuptake the siderophores with bound iron (III) via outer-membrane transporters that are siderophore-specific (Vraspir and Butler, 2009). Iron acquisition is essential for biological processes such as photosynthesis, respiration and nitrogen fixation. Most bioactive trace metals, including iron, exist at nanomolar (10^{-9} M) to picomolar (10^{-12} M) concentrations in our oceans, approximately one-millionth of the intracellular concentration in diatoms (Bruland *et al.*, 1991; Morel and Price, 2003). Diatoms are not known to produce siderophores (Soria-Dengg and Horstmann, 1995; Amin *et al.*, 2009) but previous studies have shown that diatoms can use siderophores as an iron source (Soria-Dengg *et al.*, 2001). However, based on genome sequence analyses, the presence of a gene orthologue of a bacterial ferrichrome-binding protein suggests the possibility of iron (III)-siderophore utilisation by *P. tricornutum* (Soria-Dengg and Horstmann, 1995). No trace metals, including iron (III), were provided to minimal media cultures. However, natural seawater may contain minute traces of bioactive trace metals. The high abundance of Alteromonadaceae in the minimal media suggests a potential supportive role in sequestering traces of iron (III) that may be present in the sterile natural seawater to the *P. tricornutum* (Figure 2). This is further supported by the very low level of Alteromonadaceae in the complete media (11% in complete media compared to 55% in minimal media, both on Day 15) where the culture has been supplied with 11.7 µM of iron (III) chloride hexahydrate.

Flavobacteriaceae are members of the Bacteroidetes phylum and include over 120 genera (LPSN, 2016b) found in soil, sediments and seawater (see (Yoon *et al.*, 2015) for further references). Flavobacteriaceae belong within the Cytophaga-Flavobacterium cluster which has been shown to account for more than 10% of the total bacterial community in coastal and offshore waters (Glöckner *et al.*, 1999; Abell and Bowman, 2005; DeLong *et al.*, 2006). Members of Flavobacteriaceae are proficient degraders of various biopolymers such as cellulose, chitin and pectin (Manz *et al.*, 1996; Kirchman, 2002). They were shown to be omnipresent during phytoplankton blooms, and their

preference for consuming more complex polymers rather than monomers suggests an active role in the processing of organic matter during these blooms (Cottrell and Kirchman, 2000; Pinhassi *et al.*, 2004). Although the exact mechanisms behind them are not perfectly understood, algal blooms are a consequence of exponential growth of phytoplankton (Smayda, 1997). In this respect, the phase of exponential growth of *P. tricornutum* in complete media, when our results showed highest abundance of Flavobacteriaceae, is the artificial equivalent of an algal bloom of *P. tricornutum* (Figure 2). In the minimal media, the abundance of Flavobacteriaceae remains very low; at its maximum on Day 8 it only accounts for 9% of the total bacterial community. Members of the Flavobacteriaceae family could be more demanding than other bacteria that require lower nutrient levels to thrive. It is estimated that up to 50% of carbon fixed via phytoplankton-mediated photosynthesis is utilised by marine bacteria (Azam *et al.*, 1983), mainly as Dissolved Organic Carbon (DOC) compounds, defined as the organic material <0.7µm in size (Stocker, 2012). DOC from diatoms originates either from live cells or recently lysed or grazed cells, which determine the type of DOCs available, and therefore are likely to influence the bacterial consortia associated with the diatom (Amin *et al.*, 2012). This suggests a dynamic complexity within the bacterial consortia based solely on the type of DOC available. Members of the Flavobacteriaceae family might possess the genetic ability to utilise specific DOC produced by *P. tricornutum* grown in complete media.

Pseudomonadaceae are an extraordinarily diverse family of bacteria found in almost all habitats on Earth; in soils, freshwater as well as marine environments, as well as plant and animal-associated pathogens (Starr *et al.*, 1981, 58). Species from the *Pseudomonas* genus are the best studied of the Pseudomonadaceae family, whose sheer genetic diversity explains the ability to thrive in such a wide range of environments (Anzai *et al.*, 2000). Marine isolates from the *Pseudomonas* genus have been shown to produce a wide range of bioactive compounds, many of which exhibit antibacterial as well as antiviral properties (see (Isnansetyo and Kamei, 2009) for further references). Our results, indeed show an elevated level of Pseudomonadaceae OTUs evident on Day 22 of the complete media cultures, and on Days 8 and 22 of the minimal media cultures. The increased presence of

399 Pseudomonadaceae when *the P. tricornutum* culture has 'crashed' could be attributed to its ability to
400 produce antibacterial compounds allowing members of this family to begin to thrive in the community
401 through inhibition of its competitors. Given its exceptional genetic diversity, and thus, its metabolic
402 versatility, allows for members of Pseudomonadaceae to be truly saprophytic; providing a
403 hypothetical explanation of its abundance we could measure when the *P. tricornutum* cultures crash
404 (Figure 2, Day 22 in both media conditions).

405 **Mathematical Model**

406 We observed that the bacterial community associated with *Phaeodactylum tricornutum* cultures
407 changed over time, correlating with the growth and subsequent crashing of the diatom cultures. The
408 bioavailability or absence of vitamins, trace metals and silicon, as well as nutrients or bactericidal
409 substances can alter the bacterial community. We built a mathematical model based on simple
410 assumptions extracted from the putative roles we assigned to the dominant bacterial families (see
411 Figure 4) and applied them to standard methods for modelling population dynamics. In particular, we
412 introduced growth limitation from nutrients/micronutrients, as well as from bactericidal-induced
413 death. An ordinary differential equation model cannot, of course, capture mechanisms such as
414 metabolic shifts caused by changes in the environment such as the supplementation of minimal or
415 complete media. Therefore, we did not implement a unique set of parameters for the model in the
416 two conditions. The current qualitative model provides an important proof-of-concept to emphasise
417 the validity of our assumptions, and serves as the motivation for further research bringing the model
418 to a quantitative, predictive level. Indeed, mathematical models are powerful tools towards the goal
419 of synthetic community establishment and control, and the model parameters can be experimentally
420 measured to bring predictive power to the simulations.

421 **Concluding remarks**

422 We postulate that a role within the community can be filled, not by one specific species of bacteria,
423 but rather a number of bacterial species capable of carrying out said role. Which bacteria fill the role

is dependent upon the environmental characteristics and the prevailing needs of the community as a whole at any given time. If a niche is unfilled, bacteria with the ideal metabolic functionality will seize the opportunity and thrive within that niche. The absence of certain micronutrients creates a new niche that can be filled by a certain unique bacterial faction.

Further work is necessary to explore the hypotheses postulated in the Discussion section. This can be achieved by carrying out systematic co-culture experiments with culturable members of the bacterial families of interest. The role of each representative of the bacterial families can be identified by carrying out subsequent –omics studies, which provide a holistic view of the genes (genomics), mRNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics) in a specific biological sample in a non-targeted and non-biased manner (Horgan and Kenny, 2011). The resulting experimental measurements will allow the dynamic model presented here to develop from qualitative to quantitative, providing a powerful predictive tool for culture monitoring such as predicting harvesting point based on the bacterial community.

Materials and methods

Strains and culture conditions

All *Phaeodactylum tricornutum* cultures were obtained from the Culture Collection of Algae and Protozoa (CCAP) based in Oban, Scotland (http://www.ccap.ac.uk/our-cultures.htm). All cultures are obtained non-axenic. Based on previous experimental evidence (unpublished data), the *P. tricornutum* strain CCAP1052/1B displayed optimal growth in 5L cultures. *P. tricornutum* was cultured in Guillard's medium for diatoms (F/2 + Si) in filtered natural seawater chemically sterilised using sodium hypochlorite and sodium thiosulphate pentahydrate. *P. tricornutum* was grown in two media conditions; (1) complete F/2 medium with the addition of sodium metasilicate as the source of silicon, as per Guillard and Ryther, 1962; Guillard, 1975, and (2) minimal media with a source of nitrogen ($NaNO_3$) and phosphorus ($NaH_2PO_4.2H_2O$) at the same concentration as in the F/2 medium recipe.

Recipe was obtained from the Culture Collection of Algae and Protozoa website (see http://www.ccap.ac.uk/pdfrecipes.htm). All cultures were grown in hanging 5L polyethylene bags with a 'V' shaped bottom prepared using a heat sealer (Supplementary Figure S1). All cultures had a modified aeration system provided by a 10ml pipette attached to the main pressurised air supply via 0.2 µm sterile air filters. A modified access port was created to allow for sampling and measurement of environmental parameters. Cultures were kept at 18-20°C and 24hr light at an average of 132.3 µmol m^{-2} s^{-1} using Phillips TL-D 58W 33-640 M9 fluorescent tube lights. All cultures, irrespective of media condition, were inoculated with 250ml from the same 5L stock culture of actively growing non-axenic *P. tricornutum*.

Growth measurements

Growth was monitored every 24 to 48 h using a light microscope and carrying out cell counts of each culture in quadruplicates for each culture. During the cell counts the ratios of the four different morphotypes (oval, fusiform, triradiate and cruciform) were recorded, and descriptions of each culture noted. Samples of each culture were subsequently taken using a sterile 10ml syringe and placed in 50ml Falcon centrifuge tubes and placed in -20°C freezer.

Genomic DNA extraction

All samples from Day 1, 8, 15, and 22 were thawed in a water bath set at 25°C. As per de Gouvion Saint Cyr *et al.*, 2014, samples were centrifuged for 5mins at 2000g to gather the *P. tricornutum* in the pellet while particles such as debris, other organisms, bacteria, and soluble substances remain in the supernatant. Because the bacteria might be attached to the *P. tricornutum* cells in the pellet, the pellet was washed with deionised water and then centrifuged for 5mins at 2000g. This was repeated twice. Genomic DNA extraction was carried out in the Aquaculture and Fisheries Development Centre and University College Cork. The Mo Bio's PowerWater® DNA Isolation Kit (catalogue no. 14900-100-NF) was utilised to carry out the genomic DNA extraction. The protocol provided with the kit was followed.

Presence of gDNA was detected by running a 1% agarose-ethidium bromide gel with 72 wells. The samples were sent on dry ice to Heinrich Heine University, Düsseldorf, for the V6 16S sequencing.

Barcoded 16S-V6-Next Generation Sequencing

Ion Torrent™ barcoded Next Generation Sequencing protocol was used to sequence the bacterial gDNA (Quail *et al.*, 2012; Grada and Weinbrecht, 2013). Amplification of the V6 hyper variable region of 16S rRNA with forward and reverse primers (Supplementary Table S2) was carried out. Ion Reporter™ software assembles all the raw sequencing data and sorts all the reads using the unique sample-specific barcode sequences and removes them from the reads. The outcome is raw FASTQ files which are ready for analysis using bioinformatics tools.

Bioinformatics

A total of 87,077,374 reads were identified. The smallest sample was just over 1 million reads; the largest sample was just under 10 million reads. The sequencing data was subjected to a pipeline adapted and modified from Pylro *et al.*, 2014. Primers were trimmed with fastq-mcf (version 1.04.807) (Aronesty, 2011), the resulting sequences were quality filterted and clustered into OTUs (operational taxonomic units) with usearch (version 8.0.1517; 32Bit – opensource) (Edgar, 2010, 2013). Taxonomy assignment was done by QIIME (version 1.9.0) (Caporaso *et al.*, 2010) with the implemented uclust classifier based on 97% sequence identity to the reference 16S sequences from SILVA 111 database (Quast *et al.*, 2013). Statistical analyses were performed in R (R Development Core Team, 2015).

The complete protocol containing all processing steps is available on https://github.com/QTB-HHU.

Modelling approach

Population dynamics models have been developed since quite some time (Verhulst, 1838; Lotka, 1925; Volterra, 1926) spanning the broad fields of ecology, epidemiology and economics. Starting from our understanding of the organism-to-organism interactions, we developed a dynamic model consisting

of 13 ordinary differential equations and including 56 (55 free) parameters. The parameters are fitted using a genetic algorithm (Mitchell, 1996) which is run in different steps to optimise the fit of *P. tricornutum* growth and/or the bacteria relative abundances to the experimental data in evolving system conditions (see Supplementary Model Information). The model is written in Python (Python Software Foundation, https://www.python.org/) and is available on GitHub (https://github.com/QTB-HHU/communityODE) with instructions and scripts for running.

Acknowledgements

This research was funded by the Marie Curie Initial Training Network project 'AccliPhot' (grant agreement number PITN-GA-2012-316427). Genomic DNA extraction was carried out at the Aquaculture and Fisheries Development Centre, University College Cork, Ireland (funded by Beaufort Marine Research Award in Fish Population Genetics funded by the Irish Government under the Sea Change Programme). Barcoded 16S-V6-Next Generation Sequencing was carried out by the Genomics and Transcriptomics Laboratory at Heinrich-Heine University, Düsseldorf, Germany. OP and OE are funded by Deutsche Forschungsgemeinschaft, Cluster of Excellence on Plant Sciences, CEPLAS (EXC 1028).

Competing interests

To the best of our knowledge, we do not have competing interest to declare.

References

Abell, G. and Bowman, J. (2005) Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean. *FEMS Microbiol. Ecol.* **51**: 265–277.

Albino, A., Marco, S., and Maro, A. Di (2012) Characterization of a cold-adapted glutathione synthetase from the psychrophile Pseudoalteromonas haloplanktis. *Mol. Biosyst.* **8**: 2405–2414.

Amin, S.A, Green, D.H., Hart, M.C., Küpper, F.C., Sunda, W.G., and Carrano, C.J. (2009) Photolysis of iron, siderophore chelates promotes bacterial, algal mutualism. *Proc. Natl. Acad. Sci.* **106**: 17071–17076.

Amin, S.A, Parker, M.S., and Armbrust, E.V. (2012) Interactions between diatoms and bacteria. *Microbiol. Mol.*

Biol. Rev. **76**: 667–84.

Amin, S.A., Hmelo, L.R., van Tol, H.M., Durham, B.P., Carlson, L.T., Heal, K.R., et al. (2015) Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**: 98–101.

Anzai, Y., Kim, H., Park, J., Wakabayashi, H., Oyaizu, H., and The, P. (2000) Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence become a dumping ground for incompletely characterized polarly flagellated , of 128 valid and invalid Pseudomonas species , which included almost valid species of the genus Pseudomona. *Int. J. Syst. Evol. Microbiol.* **50**: 1563–1589.

Aronesty, E. (2011) Command-line tools for processing biological sequencing data. *Expr. Anal. Durham NC.*

Azam, F., Fenchel, T., Field, J., Gray, J., Meyer-Reil, L., and Thingstad, F. (1983) The Ecological Role of Water-Column Microbes in the Sea. *Mar. Ecol. Prog. Ser.* **10**: 257–263.

Bell, W. and Mitchell, R. (1972) Chemotactic and growth responses of marine bacteria to algal extracellular products. *Biol. Bull.* **143**: 265–277.

Bhattacharya, D., Archibald, J.M., Weber, A.P.M., and Reyes-Prieto, A. (2007) How do endosymbionts become organelles? Understanding early events in plastid evolution. *BioEssays* **29**: 1239–1246.

Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008) The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* **456**: 239–244.

Bowman, J. (2007) Bioactive compound synthetic capacity and ecological significance of marine bacterial genus Pseudoalteromonas. *Mar. Drugs* **5**: 220–241.

Bruckner, C.G., Rehm, C., Grossart, H.P., and Kroth, P.G. (2011) Growth and release of extracellular organic compounds by benthic diatoms depend on interactions with bacteria. *Environ. Microbiol.* **13**: 1052–1063.

Bruland, K.W., Donat, J.R., and Hutchins, D.A. (1991) Interactive influences of bioactive trace metals on biological production in oceanic waters. *Limnol. Oceanogr.* **36**: 1555–1577.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**: 335–336.

Chen, J., Bittinger, K., Charlson, E.S., Hoffmann, C., Lewis, J., Wu, G.D., et al. (2012) Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* **28**: 2106–2113.

Chen, X., Xie, B., Lu, J., He, H., and Zhang, Y. (2007) A novel type of subtilase from the psychrotolerant bacterium Pseudoalteromonas sp. SM9913: catalytic and structural properties of deseasin MCP-01. *Microbiology* **153**: 2116–2125.

Chisti, Y. (2007) Biodiesel from microalgae. *Biotechnol. Adv.* **25**: 294–306.

Cottrell, M.T. and Kirchman, D.L. (2000) Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low- and high-molecular-weight dissolved organic matter. *Appl. Environ. Microbiol.* **66**: 1692–7.

Croft, M.T., Warren, M.J., and Smith, A.G. (2006) Algae need their vitamins. *Eukaryot. Cell* **5**: 1175–1183.

Dangeard, P.-A. (1933) Traité d'algologie: introduction à la biologie et à la systématique des algues. Paul Lechevalier & Fils.

Day, J.G., Thomas, N.J., Achilles-Day, U.E.M., and Leakey, R.J.G. (2012) Early detection of protozoan grazers in algal biofuel cultures. *Bioresour. Technol.* **114**: 715–719.

559 DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.-U., et al. (2006) Community genomics
560 among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.

561 Delucca, R. and Mccracken, M.D. (1977) Observations on interactions between naturally-collected bacteria and
562 several species of algae. *Hydrobiologia* **55**: 71–75.

563 Desbois, A.P., Mearns-Spragg, A., and Smith, V.J. (2009) A Fatty Acid from the Diatom Phaeodactylum
564 tricornutum is Antibacterial Against Diverse Bacteria Including Multi-resistant Staphylococcus aureus
565 (MRSA). *Mar. Biotechnol.* **11**: 45–52.

566 Dykhuizen, D. (1998) Santa Rosalia revisited: why are there so many species of bacteria? *Antonie Van*
567 *Leeuwenhoek* **73**: 25–33.

568 Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

569 Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**:
570 996–8.

571 Elton, C. (1958) The Ecology of Invasions by Animals and Plants. Methuen (ed) London.

572 Fajardo, A.R., Cerdán, L.E., Medina, A.R., Fernández, F.G.A., Moreno, P. a. G., and Grima, E.M. (2007) Lipid
573 extraction from the microalga Phaeodactylum tricornutum. *Eur. J. Lipid Sci. Technol.* **109**: 120–126.

574 Gardner, M. and Ashby, W. (1970) Connectance of large dynamic (cybernetic) systems: critical values for
575 stability. *Nature* **228**: 784.

576 Gause, G. (1934) The Struggle for Existence. Williams and Wilkins (ed) Baltimore (MD).

577 Glöckner, F.O., Fuchs, B.M., and Amann, R. (1999) Bacterioplankton compositions of lakes and oceans: a first
578 comparison based on fluorescence in situ hybridization. *Appl. Environ. Microbiol.* **65**: 3721–6.

579 de Gouvion Saint Cyr, D., Wisniewski, C., Schrive, L., Farhi, E., and Rivasseau, C. (2014) Feasibility study of
580 microfiltration for algae separation in an innovative nuclear effluents decontamination process. *Sep. Purif.*
581 *Technol.* **125**: 126–135.

582 Grada, A. and Weinbrecht, K. (2013) Next-Generation Sequencing: Methodology and Application. *J. Invest.*
583 *Dermatol.* **133**: e11–4.

584 Grima, E. and Fernández, F. (1999) Photobioreactors: light regime, mass transfer, and scaleup. *J. Biotechnol.* **70**:
585 231–247.

586 Guillard, R. (1975) Culture of phytoplankton for feeding marine invertebrates. In, W. L. Smith and M. H. Chanley
587 (ed), *Culture of marine invertebrate animals.* Plenum Press, New York, pp. 29–60.

588 Guillard, R. and Ryther, J. (1962) Studies of marine planktonic diatoms: i. *Cyclotella nana* Hustedt, and *Detonula*
589 *confervacea* (Cleve) gran. *Can. J. Microbiol.* **8**: 229-239.

590 Hardin, G. (1960) The competitive exclusion principle. *Science* **131**: 1292–7.

591 He, H., Guo, J., Chen, X., Xie, B., and Zhang, X. (2012) Structural and functional characterization of mature forms
592 of metalloprotease E495 from Arctic sea-ice bacterium Pseudoalteromonas sp. SM495. *PLoS One* **7**:
593 e35442.

594 Holmström, C. and Kjelleberg, S. (1999) Marine Speudoalteromonas species are associated with higter organisms
595 and produce biologically active extracellular agents. *FEMS Microb. Ecol.* **30**: 285–293.

596 Holt, P.D., Reid, R.R., Lewis, B.L., Luther, G.W., and Butler, A. (2005) Iron(III) coordination chemistry of

alterobactin A: A siderophore from the marine bacterium Alteromonas luteoviolacea. *Inorg. Chem.* **44**: 7671–7677.

Horgan, R.P. and Kenny, L.C. (2011) "Omic" technologies: genomics, transcriptomics, proteomics and metabolomics. *Obstet. Gynaecol.* **13**: 189–195.

Isnansetyo, A. and Kamei, Y. (2009) Bioactive substances produced by marine isolates of Pseudomonas. *J. Ind. Microbiol. Biotechnol.* **36**: 1239–1248.

Ivanova, E.P., Flavier, S., and Christen, R. (2004) Phylogenetic relationships among marine Alteromonas-like proteobacteria: Emended description of the family Alteromonadaceae and proposal of Pseudoalteromonadaceae fam. nov., Colwelliaceae fam. nov., Shewanellaceae fam. nov., Moritellaceae fam. nov., Ferri. *Int. J. Syst. Evol. Microbiol.* **54**: 1773–1788.

Ivanova, E.P. and Mikhaĭlov, V. V (2001) A new family of Alteromonadaceae fam. nov., including the marine proteobacteria species Alteromonas, Pseudoalteromonas, Idiomarina i Colwellia. *Mikrobiologiia* **70**: 15–23.

Kabir, S. (2009) The role of probiotics in the poultry industry. *Int. J. Mol. Sci.* **10**: 3531–3546.

Kates, M. and Volcani, B.E. (1966) Lipid components of diatoms. *Biochim. Biophys. Acta* **116**: 264–278.

Kazamia, E., Aldridge, D.C., and Smith, A.G. (2012) Synthetic ecology – A way forward for sustainable algal biofuel production? *J. Biotechnol.* **162**: 163–169.

Khudary, R. Al, Venkatachalam, R., and Katzer, M. (2010) A cold-adapted esterase of a novel marine isolate, Pseudoalteromonas arctica: gene cloning, enzyme purification and characterization. *Extremophiles* **14**: 273–285.

Kim, S.M., Jung, Y., and Kwon, O. (2012) A Potential Commercial Source of Fucoxanthin Extracted from the Microalga Phaeodactylum tricornutum. *Appl. Biochem. Biotechnol.* **166**: 1843–1855.

Kirchman, D.L. (2002) The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol. Ecol.* **39**: 91–100.

Lee, S., Kato, J., Takiguchi, N., Kuroda, A., and Ikeda, T. (2000) Involvement of an Extracellular Protease in Algicidal Activity of the Marine Bacterium Pseudoalteromonas sp. Strain A28. *Appl. Environ. Microbiol.* **66**: 4334–4339.

Lotka, A. (1925) Elements of Physical Biology Williams and Wilkins (ed).

LPSN (2016a) List of Prokaryotic names with Standing in Nomenclature - Alteromonadaceae.

LPSN (2016b) List of Prokaryotic names with Standing in Nomenclature - Flavobacteriaceae.

Lu, M., Wang, S., Fang, Y., Li, H., Liu, S., and Liu, H. (2010) Cloning, expression, purification, and characterization of cold-adapted α-amylase from Pseudoalteromonas arctica GS230. *Protein J.* **29**: 591–597.

MacArthur, R. (1955) Fluctuations of animal populations and a measure of community stability. *Ecologycology* **36**: 533–536.

Manz, W., Amann, R., Ludwig, W., Vancanneyt, M., and Schleifer, K.H. (1996) Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum cytophaga-flavobacter-bacteroides in the natural environment. *Microbiology* **142**: 1097–1106.

Martino, A. De, Bartual, A., Willis, A., Meichenin, A., Villazán, B., Maheswari, U., and Bowler, C. (2011) Physiological and Molecular Evidence that Environmental Changes Elicit Morphological Interconversion in

636 the Model Diatom Phaeodactylum tricornutum. *Ann. Anat.* **162**: 462–481.

637 Martino, A. De, Meichenin, A., Shi, J., Pan, K., and Bowler, C. (2007) Genetic and phenotypic characterization of
638 Phaeodactylum tricornutum (Bacillariophyceae) accessions 1. *J. Phycol.* **43**: 992–1009.

639 Mata, T., Martins, A., and Caetano, N. (2010) Microalgae for biodiesel production and other applications: a
640 review. *Renew. Sustain. Energy Rev.* **14**: 217–232.

641 McCann, K.S. (2000) The diversity-stability debate. *Nature* **405**: 228–33.

642 Mitchell, M. (1996) An introduction to genetic algorithms MIT Press, Cambridge, MA.

643 Montsant, A., Jabbari, K., Maheswari, U., and Bowler, C. (2005) Comparative genomics of the pennate diatom
644 Phaeodactylum tricornutum. *Plant Physiol.* **137**: 500–513.

645 Morel, F.M.M. and Price, N.M. (2003) The biogeochemical cycles of trace metals in the oceans. *Science* **300**: 944–
646 7.

647 Owens, T.G. and Wold, E.R. (1986) Light-harvesting function in the diatom Phaeodactylum tricornutum: II.
648 Distribution of excitation energy between the photosystems. *Plant Physiol.* **80**: 732–738.

649 Parmar, A., Singh, N., and Pandey, A. (2011) Cyanobacteria and microalgae: a positive prospect for biofuels.
650 *Bioresour. Technol.* **102**: 10163–10172.

651 Parvez, S. and Malik, K. (2006) Probiotics and their fermented food products are beneficial for health. *J. Appl.*
652 *Microbiol.* **100**: 1171–1185.

653 Peng, J., Yuan, J.P., Wu, C.F., and Wang, J.H. (2011) Fucoxanthin, a marine carotenoid present in brown seaweeds
654 and diatoms: Metabolism and bioactivities relevant to human health. *Mar. Drugs* **9**: 1806–1828.

655 Persson, G., Jansson, M., and Kluwer, C. (1988) Phosphate uptake and utilizaton by bacteria and algae.
656 *Hydrobiologia* **170**: 177–189.

657 Pimm, S. (1984) The complexity and stability of ecosystems. *Nature* **307**: 321–326.

658 Pinhassi, J., Sala, M.M., Havskum, H., Peters, F., Guadayol, O., Malits, A., and Marrasé, C. (2004) Changes in
659 bacterioplankton composition under different phytoplankton regimens. *Appl. Environ. Microbiol.* **70**:
660 6753–66.

661 Provasoli, L. (1958) Nutrition and ecology of Protozoa and Algae. *Annu. Rev. Microbiol.* **12**: 279–308.

662 Provasoli, L. (1963) Organic regulation of phytoplankton fertility. In, Hill,M. (ed), *The sea: ideas and observations*
663 *on progress in the study of the seas*. Wiley-Interscience, New York, pp. 165–219.

664 Provasoli, L. and Carlucci, A. (1974) Vitamins and growth regulators. In, Stewart,W. (ed), *Algal Physiology and*
665 *Biochemistry, Botanical monographs, 10*. Blackwell Scientific Publications, California, pp. 741–787.

666 Pylro, V.S., Roesch, L.F.W., Morais, D.K., Clark, I.M., Hirsch, P.R., and Tótola, M.R. (2014) Data analysis for 16S
667 microbial profiling from different benchtop sequencing platforms. *J. Microbiol. Methods* **107**: 30–37.

668 Qi, Z., Zhang, X.-H., Boon, N., and Bossier, P. (2009) Probiotics in aquaculture of China — Current state, problems
669 and prospect. *Aquaculture* **290**: 15–21.

670 Quail, M., Smith, M.E., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., et al. (2012) A tale of three next
671 generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq
672 sequencers. *BMC Genomics* **13**: 341.

673 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013) The SILVA ribosomal RNA gene
674 database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**: D590–D596.

675 R Development Core Team (2015) R: A language and environment for statistical computing. *R Found. Stat.*
676 *Comput. Vienna, Austria.* http://www.R–project.org/.

677 Rebolloso-Fuentes, M.M., Navarro-Pérez, A., Ramos-Miras, J.J., and Guil-Guerrero, J.L. (2001) Biomass nutrient
678 profiles of the microalga Phaeodactylum tricornutum. *J. Food Biochem.* **25**: 57–76.

679 Reid, R. and Butler, A. (1991) Investigation of the mechanism of iron acquisition by the marine bacterium
680 Alteromonas luteoviolaceus: Characterization of siderophore production. *Limnol. Oceanogr.* **36**: 1783–
681 1792.

682 Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., and Thompson, F. eds. (2014) The Prokaryotes Springer
683 Berlin Heidelberg, Berlin, Heidelberg.

684 Ryther, J. and Goldman, J. (1975) Microbes as food in mariculture. *Annu. Rev. Microbiol.* **29**: 429–443.

685 Sañudo-Wilhelmy, S. a., Gómez-Consarnau, L., Suffridge, C., and Webb, E. A. (2014) The Role of B Vitamins in
686 Marine Biogeochemistry. *Ann. Rev. Mar. Sci.* **6**: 339–367.

687 Siron, R., Giusti, G., and Berland, B. (1989) Changes in the fatty acid composition of Phaeodactylum tricornutum
688 and Dunaliella tertiolecta during growth and under phosphorus deficiency. *Mar. Ecol. Prog. Ser.* **55**: 95–
689 100.

690 Smayda, T. (1997) Harmful algal blooms: their ecophysiology and general relevance to phytoplankton blooms in
691 the sea. *Limnol. Oceanogr.* **42**: 1137–1153.

692 Soria-Dengg, S. and Horstmann, U. (1995) Ferrioxamines B and E as iron sources for the marine diatom
693 Phaeodactylum tricornutum. *Mar. Ecol. Prog. Ser.* **127**: 269–277.

694 Soria-Dengg, S., Reissbrodt, R., and Horstmann, U. (2001) Siderophores in marine coastal waters and their
695 relevance for iron uptake by phytoplankton:experiments with the diatom Phaeodactylum tricornutum.
696 *Mar. Ecol. Prog. Ser.* **220**: 73–82.

697 Starr, M.P., Stolp, H., Trüper, H.G., Balows, A., and Schlegel, H.G. eds. (1981) The Prokaryotes Springer Berlin
698 Heidelberg, Berlin, Heidelberg.

699 Stocker, R. (2012) Marine Microbes See a Sea of Gradients. *Science* **338**: 628–633.

700 Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., et al. (2015) Structure and function
701 of the global ocean microbiome. *Science* **348**: 1–10.

702 Suminto and Hirayama, K. (1997) Application of a growth-promoting bacteria for stable mass culture of three
703 marine microalgae. *Hydrobiologia* **358**: 223–230.

704 Tredici, M.R., Biondi, N., Ponis, E., Rodolfi, L., Zittelli, G.C., Burnell, G., and Allan, G. (2009) Advances in microalgal
705 culture for aquaculture feed and other uses. In, Burnell,G. and Allan,G. (eds), *New Technologies in*
706 *Aquaculture: Improving Production Efficiency, Quality and Environmental Management.* Woodhead
707 Publishing Ltd, Cambridge, pp. 610–676.

708 Venkateswaran, K. and Dohmoto, N. (2000) Pseudoalteromonas peptidolytica sp. nov., a novel marine mussel-
709 thread-degrading bacterium isolated from the Sea of Japan. *Int. J. Syst. Evol. Microbiol.* **50**: 565–574.

710 Verhulst, P. (1838) Notice sur la loi que la population suit dans son accroissement. correspondance
711 mathématique et physique publiée par a. *Quetelet* **10**: 113–121.

712 Volterra, V. (1926) Fluctuations in the abundance of a species considered mathematically. *Nature* **118**: 558–560.

713 Vraspir, J.M. and Butler, A. (2009) Chemistry of marine ligands and siderophores. *Ann. Rev. Mar. Sci.* **1**: 43–63.

714 Wang, H., Zhang, W., Chen, L., Wang, J., and Liu, T. (2013) The contamination and control of biological pollutants
715 in mass cultivation of microalgae. *Bioresour. Technol.* **128**: 745–750.

716 Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci.*
717 *U. S. A.* **95**: 6578–6583.

718 Yashodhara, B.M., Umakanth, S., Pappachan, J.M., Bhat, S.K., Kamath, R., and Choo, B.H. (2009) Omega-3 fatty
719 acids: a comprehensive review of their role in health and disease. *Postgrad. Med. J.* **85**: 84–90.

720 Yongmanitchai, W. and Ward, O.P. (1991) Growth of and omega-3 fatty acid production by Phaeodactylum
721 tricornutum under different culture conditions. *Appl. Environ. Microbiol.* **57**: 419–425.

722 Yoon, J., Jo, Y., Kim, G.J., and Choi, H. (2015) Gramella lutea sp. nov., a Novel Species of the Family
723 Flavobacteriaceae Isolated from Marine Sediment. *Curr. Microbiol.* **71**: 252–8.

724

Supplementary Material

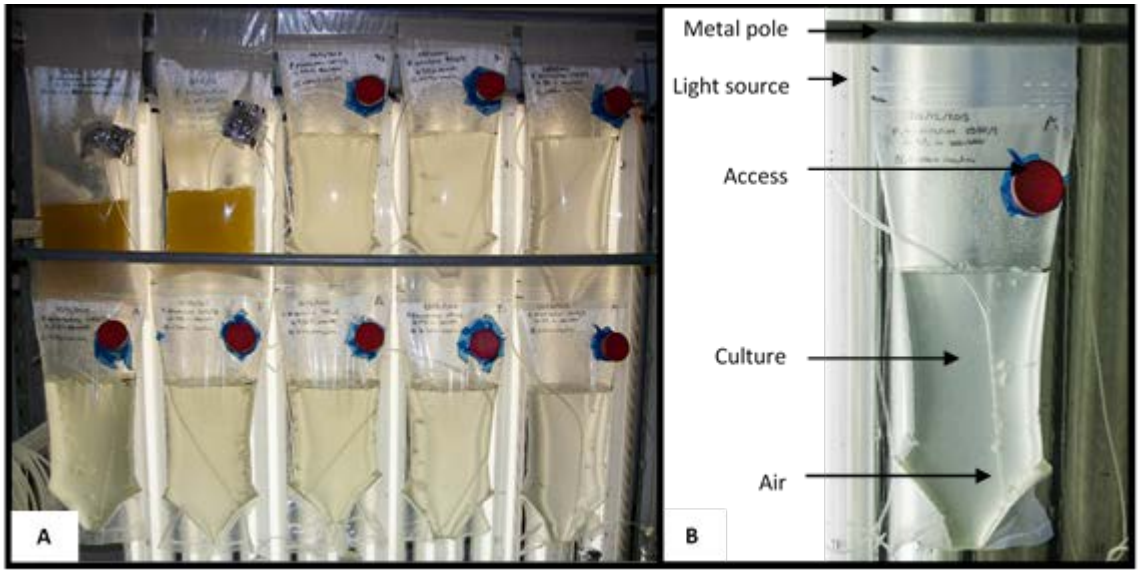
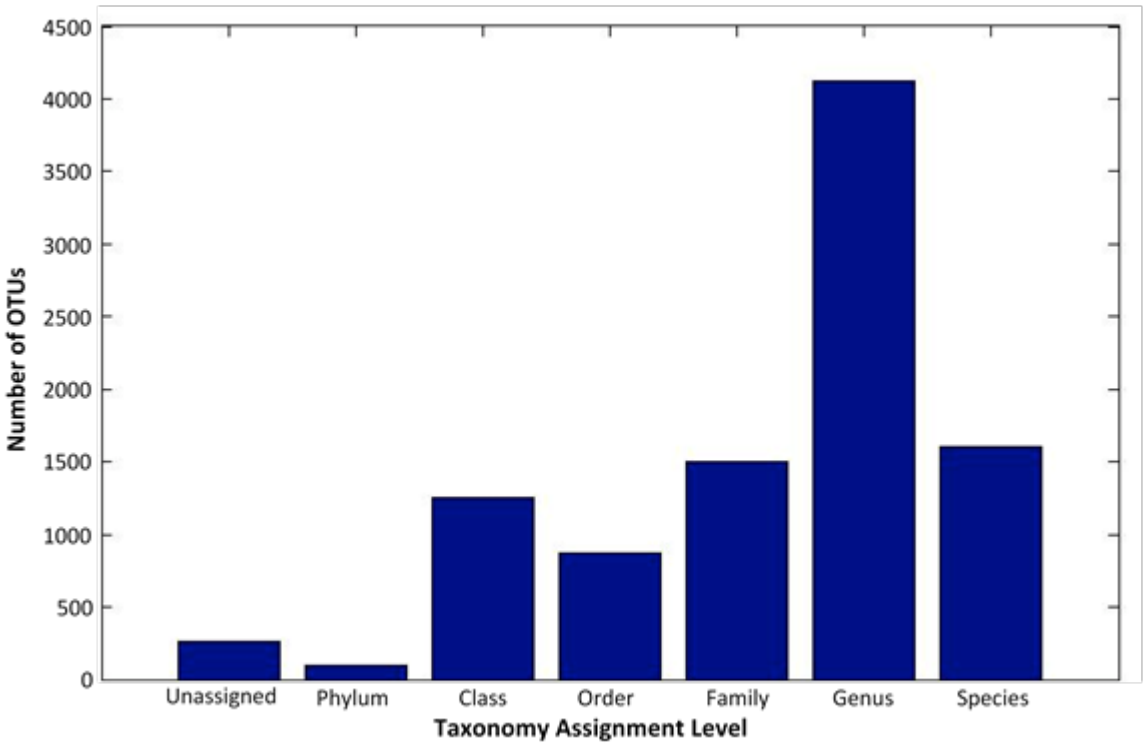


Figure S1. Non-axenic *Phaeodactylum tricornutum* culture set up. 5L polyethylene bags with a 'V' shaped bottom were created using the heat sealer machine. The bags were then rinsed and filled with 5L of filtered seawater. Afterwards each bag was sealed and hung approximately 30 cm from the light source. A small incision was made to insert the aeration tubing. This consists of a 10ml pipette attached to silicon tubing which is attached to a sterile air filter connecting it to the main air supply. A modified access port was created to take samples and measure the environmental parameters (Photographs courtesy of Maria Rubio Bernal)



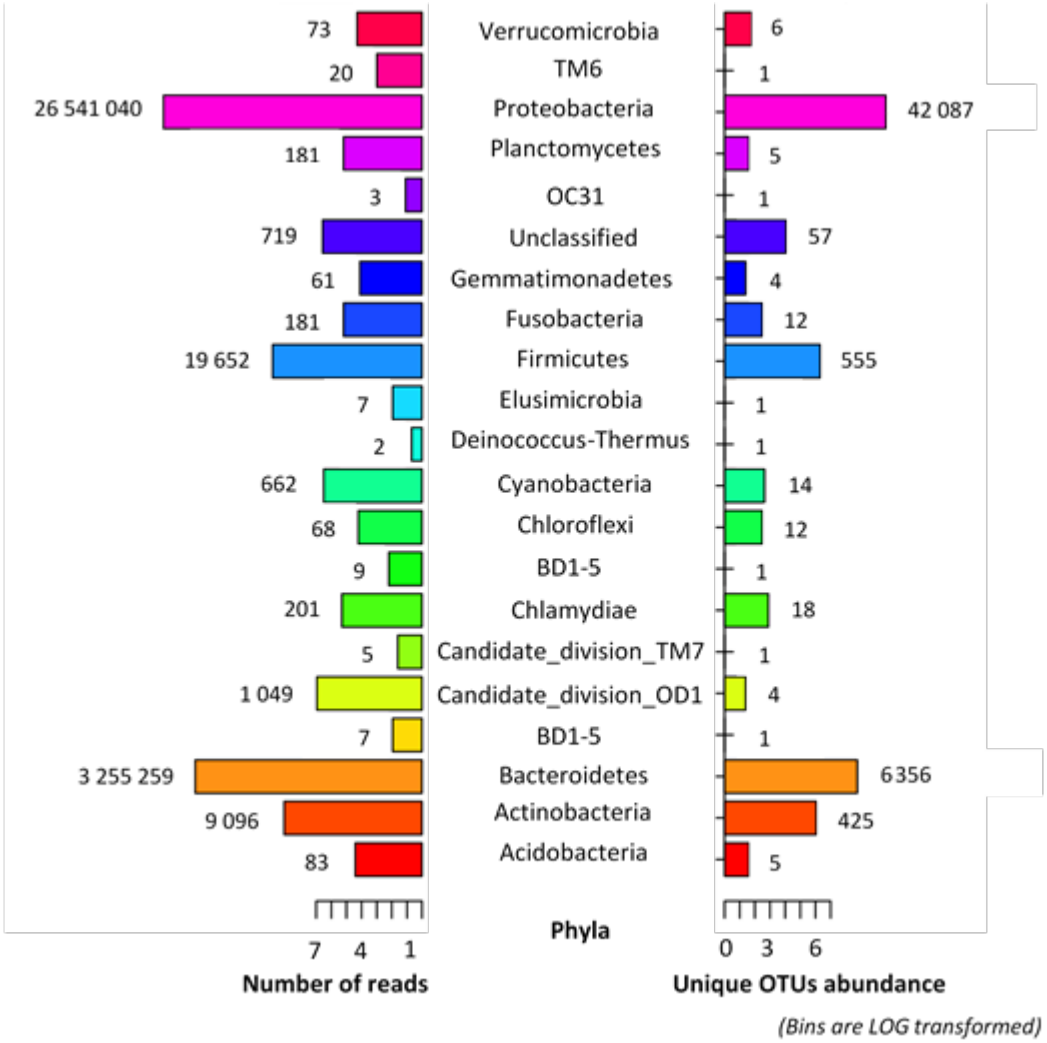
735

736 **Figure S2. Operational Taxonomic Unit (OTU) Taxonomy Assignment Level.** The 16S rRNA gene

737 sequences were clustered to defined Operational Taxonomic Units (OTUs) at ≥97% sequence identity.

738 Most OTUs could be assigned to the genera level, using the SILVA database (v.118) (Quast *et al.*, 2013).

739



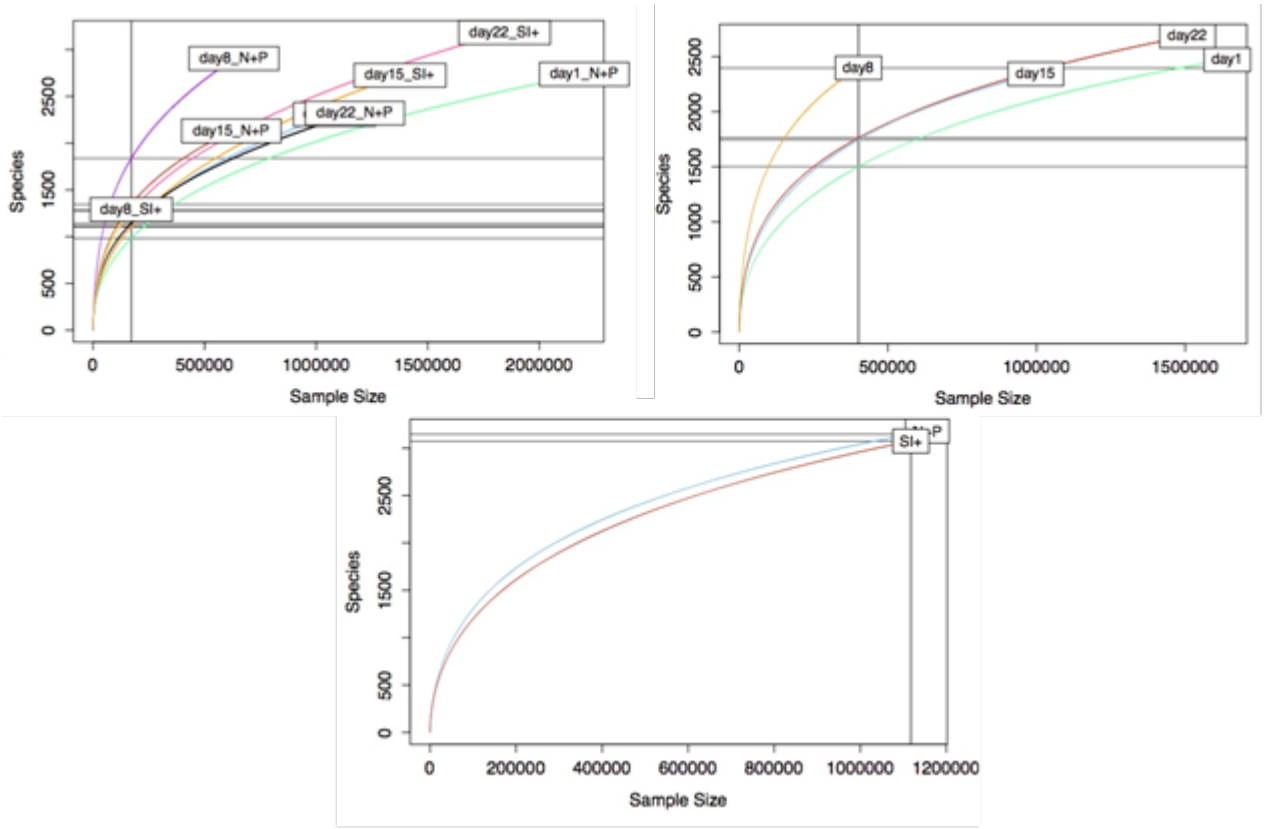
740

741 **Figure S3. Number of reads per unique OTU abundance (at the phylum level).** A comparison of the

742 number of individual reads to the number of unique OTUs shows that phyla with high number of reads

743 do not result in single OTUs.

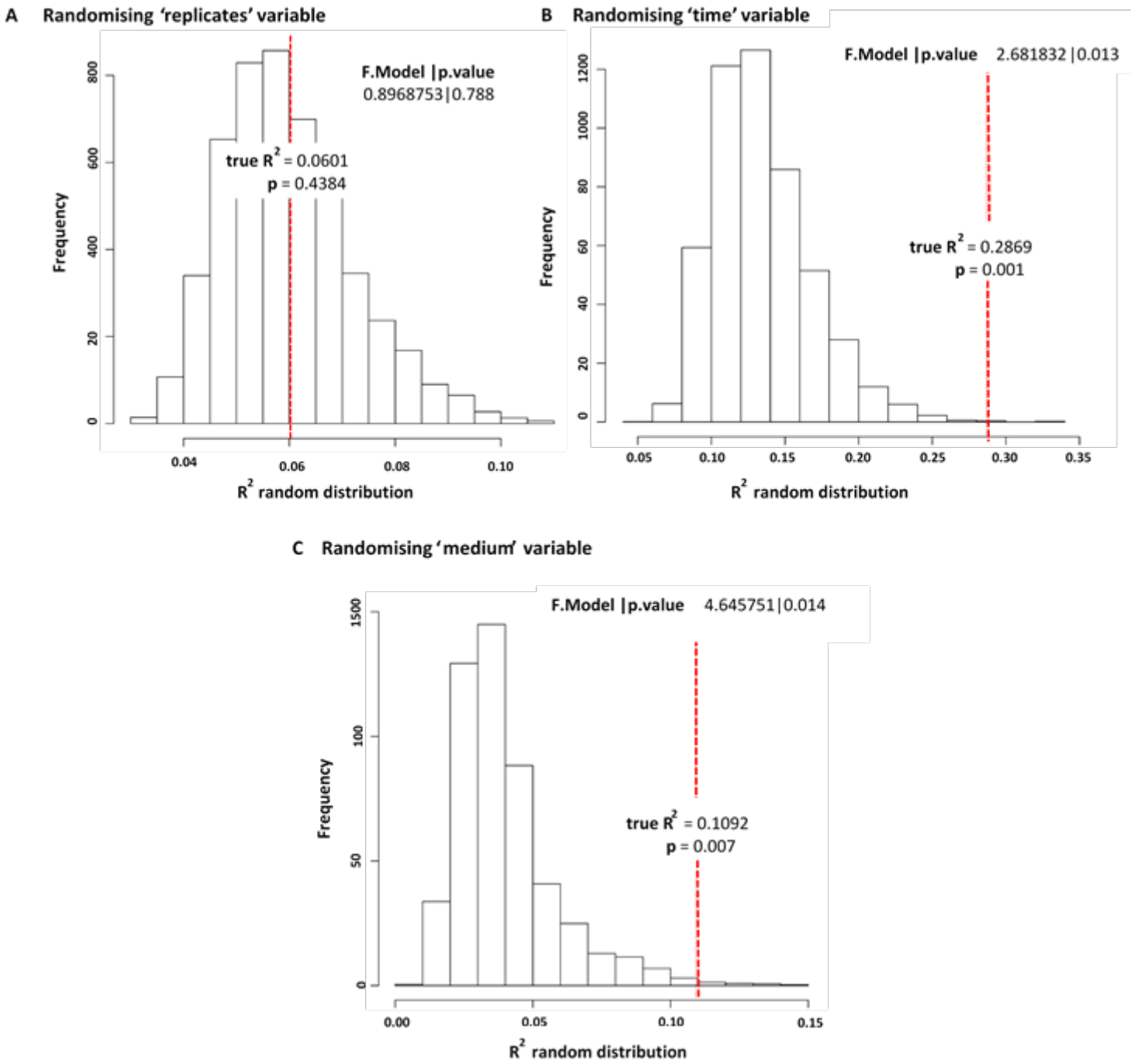
744



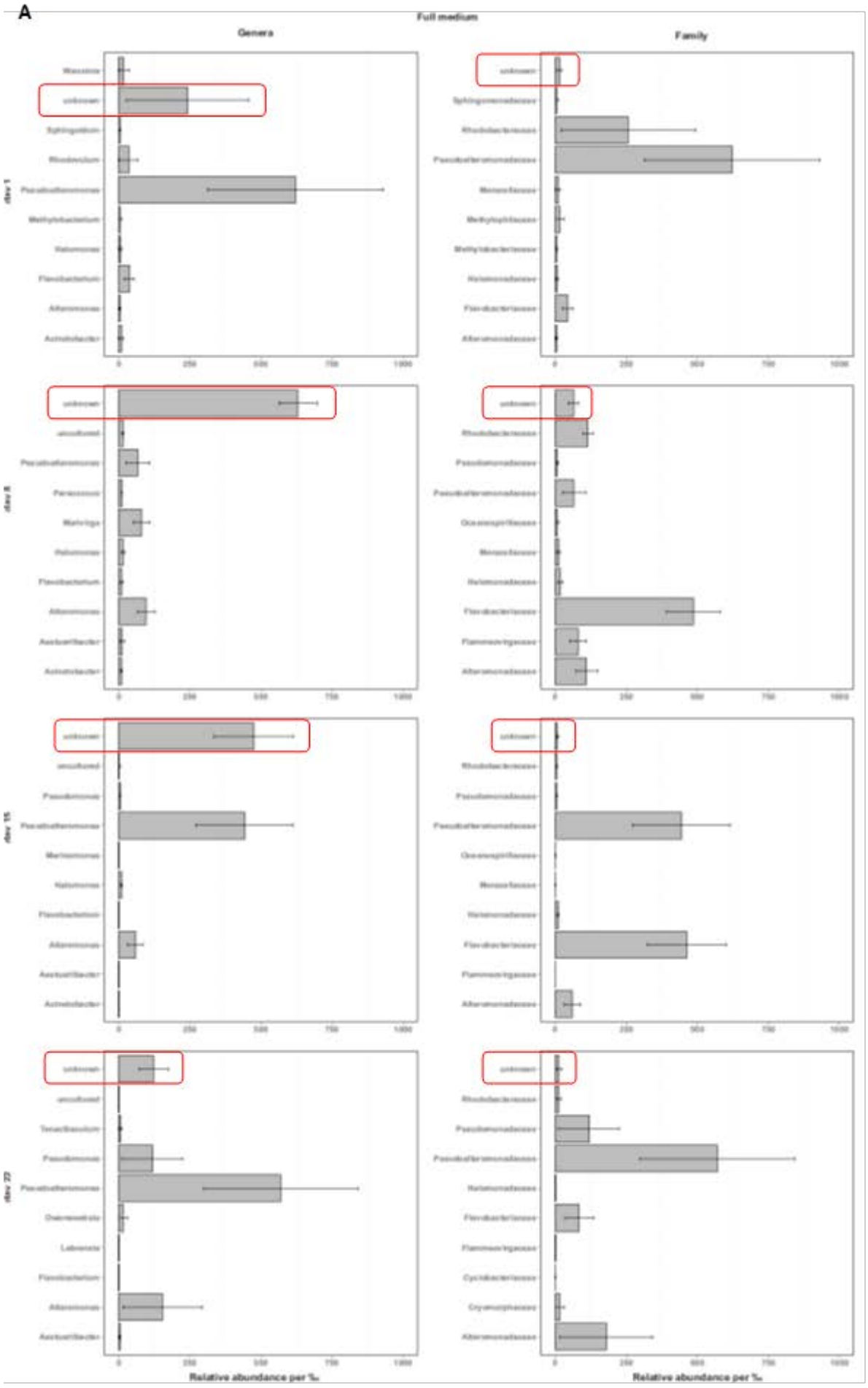
745

746 **Figure S4. Alpha diversity.** Rarefaction curves were used to evaluate the Alpha diversity in the
747 different media conditions as well as at the different time points. Species richness in both minimal and
748 complete media was ~3 000. Species richness over time remained between ~2 400 and 2 600, with
749 reduced species richness (~1 300) on Day 8 (both minimal and complete media) possibly due to
750 elevated levels of 16S *P. tricornutum* chloroplast reads which had to be omitted. Greatest species
751 richness (~ 3 000) was shown on Day 22.

752

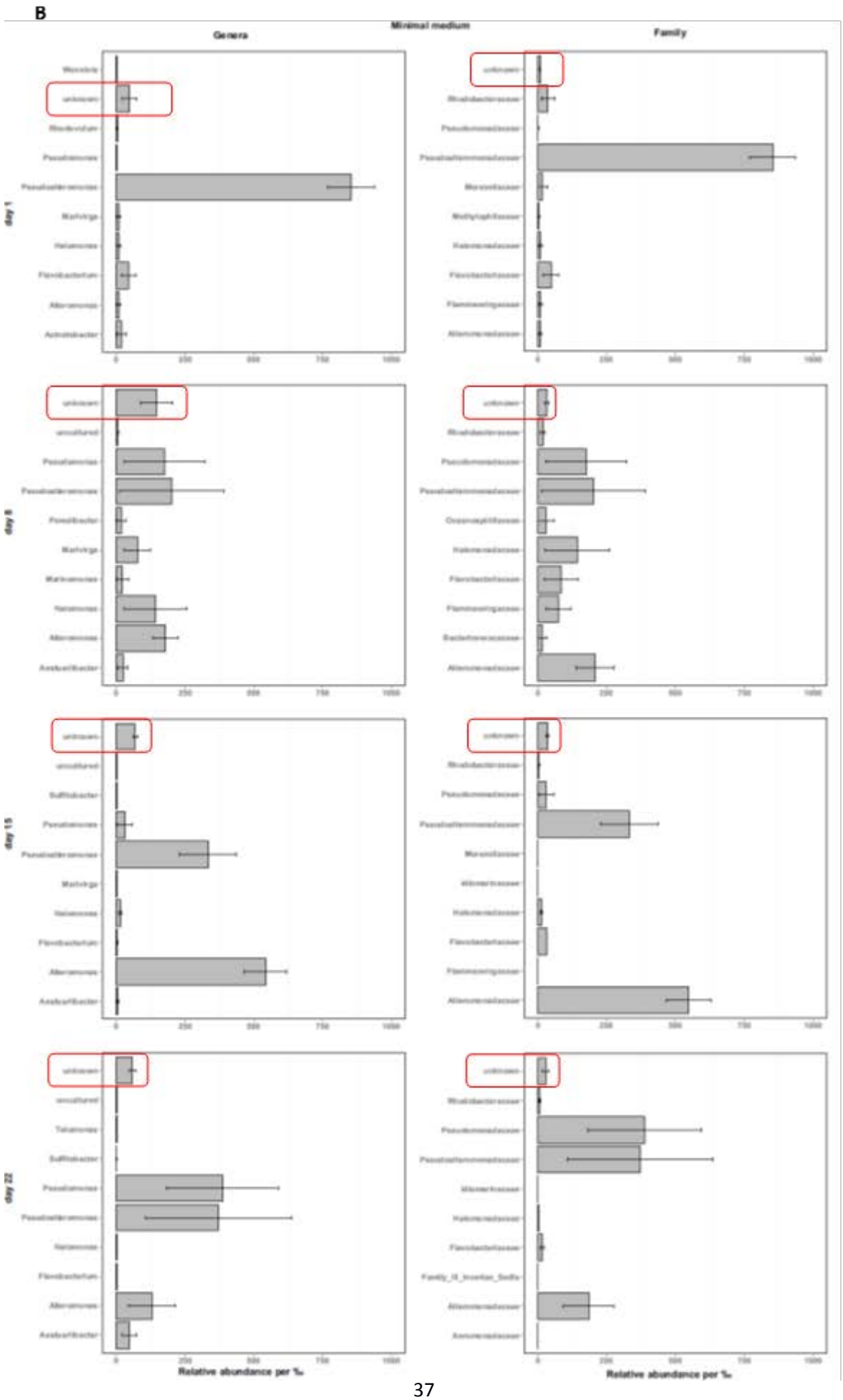


753 **Figure S5. Beta diversity.** A modified version of PermanovaG was used to carry out permutational
754 multivariate analysis of variance using multiple distance matrices. The distance matrices [24x24] were
755 previously calculated based on the generalised UniFrac distance (Chen *et al.*, 2012), weighted UniFrac
756 and unweighted UniFrac (Lozupone and Knight, 2005) distance. The significance for the test was
757 assessed by 5000 permutations. **A** shows no significant effect between the replicates (p-value of
758 0.4384). **B** shows a significant effect for the time variable (p-value of 0.001). **C** shows also shows a
759 significant effect for the medium variable (p-value of 0.007)



760

761



762 **Figure S6. Comparison between bacterial community at genera level and family level. A** in complete
763 media. **B** in minimal media. We show no dynamical difference within the genera that cannot be
764 observed at the family level. Encircled in red, there are a greater number of OTUs that could not be
765 assigned a taxonomy ('unknowns') at the genera level than at the family level. By investigating the
766 bacterial community dynamics at the family level, we also include taxonomical information that is
767 unavailable at the genus level.

768

Table S1. 16S V6 rRNA primer sequences. 'Max' is the complete media. 'Min' is the minimal media. 'A', 'B', and 'C' are the three replicates.

Description	Barcode Name	Adaptor A Sequence	Barcode Sequence	Forward V6 Primer Sequence	Reverse V6 Primer Sequence
Day 1_Max_A	IonXpress_071	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGAGGCTCCGAC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 1_Max_B	IonXpress_072	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGAAGGCCACAC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 1_Max_C	IonXpress_073	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCTGCCTGTC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 1_Min_A	IonXpress_074	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGATCGGTTC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 1_Min_B	IonXpress_075	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCAGGAATAC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 1_Min_C	IonXpress_076	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGGAAGAACCTC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 8_Max_A	IonXpress_077	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGAAGCGATTC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 8_Max_B	IonXpress_078	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CAGCCAATTCTC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 8_Max_C	IonXpress_079	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CCTGGTTGTC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 8_Min_A	IonXpress_080	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCGAAGGCAGGC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 8_Min_B	IonXpress_081	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CCTGCCATTCGC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 8_Min_C	IonXpress_082	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TTGGCATCTC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 15_Max_A	IonXpress_083	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTAGGACATTC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 15_Max_B	IonXpress_084	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTTCCATAAC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 15_Max_C	IonXpress_085	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CCAGCCTCAAC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 15_Min_A	IonXpress_086	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTTGGTTATTC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 15_Min_B	IonXpress_087	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TTGGCTGGAC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 15_Min_C	IonXpress_088	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CCGAACACTTC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 22_Max_A	IonXpress_090	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTAACCACGGC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 22_Max_B	IonXpress_091	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGGAAGGATGC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 22_Max_C	IonXpress_092	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTAGGAACCGC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 22_Min_A	IonXpress_093	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTTGTCCAATC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 22_Min_B	IonXpress_094	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCCGACAAGC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 22_Min_C	IonXpress_095	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGGACAGATC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT

770 **Table S2. Generalised UniFrac distances of bacterial communities in complete and minimal media**
771 **over time.** Generalised UniFrac distance contains an extra parameter α controlling the weight on
772 abundant lineages so the distance is not dominated by highly abundant lineages. α = 0.5 has overall
773 the best power.

		DAY 1		DAY 8		DAY 15		DAY 22	
		Minimal	Complete	Minimal	Complete	Minimal	Complete	Minimal	Complete
DAY 1	Minimal	0	0.5158104	0.7195151	0.8637909	0.6668939	0.7504169	0.6655156	0.6608732
	Complete	0.5158104	0	0.7020723	0.7203046	0.7176411	0.7125498	0.7019796	0.6733559
DAY 8	Minimal	0.7195151	0.7020723	0	0.6641379	0.5844515	0.6637369	0.5905962	0.497263
	Complete	0.8637909	0.7203046	0.6641379	0	0.8558383	0.5896856	0.8340209	0.7723285
DAY 15	Minimal	0.6668939	0.7176411	0.5844515	0.8558383	0	0.6605545	0.4569922	0.5042935
	Complete	0.7504169	0.7125498	0.6637369	0.5896856	0.6605545	0	0.630498	0.5720828
DAY 22	Minimal	0.6655156	0.7019796	0.5905962	0.8340209	0.4569922	0.630498	0	0.4597439
	Complete	0.6608732	0.6733559	0.497263	0.7723285	0.5042935	0.5720828	0.4597439	0

774

775

776

777

778 **References for Supplementary Figures**

779 Chen, J., Bittinger, K., Charlson, E.S., Hoffmann, C., Lewis, J., Wu, G.D., et al. (2012) Associating
780 microbiome composition with environmental covariates using generalized UniFrac distances.
781 *Bioinformatics* **28**: 2106–2113.

782 Lozupone, C. and Knight, R. (2005) UniFrac : a New Phylogenetic Method for Comparing Microbial
783 Communities UniFrac : a New Phylogenetic Method for Comparing Microbial Communities. *Appl.*
784 *Environ. Microbiol.* **71**: 8228–8235.

785 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013) The SILVA ribosomal
786 RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.*
787 **41**: D590–6.

Supplementary Model Material: Mathematical model of the population dynamics in *P. tricornutum* associated bacterial community

Oliver Ebenhöh*

October 14, 2016

Contents

1 Model description

In order to test our understanding emerging from the analysis of the experimental data we built a Ordinary Differential Equations (ODE) model to simulate how the bacterial community develops in time. Considering the limited information that can be extracted from the experimental data, the model is purely qualitative and provides a proof of concept that a quantitative model can be constructed if dedicated experiments are designed for calibration. We base the organism-to-organism interactions on the production/consumption of metabolites. Metabolites include nutrients, micronutrients and toxins.

1.1 Introduction

We develop a dynamic model represented by a set of 13 ODEs. Five ODEs describe the variation in time of the populations of *P. tricornutum (D)*, *Pseudoalteromonas (PA)*, *Flavobacterium (F)*, *Alteromonas (A)* and *Pseudomonas (P)*. The other eight ODEs describe the production and consumption of the metabolites we consider as mainly contributing to drive the community dynamics: the dissolved organic carbons of preference for PA and A (DOC_{PA} and DOC_A, respectively), the complex polymers (COP) consumed by F, generic vitamins (Vit) and iron (Fe) needed by D and produced by A, bactericidial molecules (EPA and Bac, produced by D and by PA respectively) and the dissolved organic matter (DOM).

The model is built from the following working hypotheses:

1. the growth γ of each population follows a standard Verhulst equation parametrized with a carrying capacity CC and scaled by Michaelis-Menten-like terms that encode the dependence on necessary nutrients as a scaling factor $0 < \varepsilon < 1$;

2. the death δ of each population is inversely proportional to $(1+\gamma)$ to account for the fact that cells during replication (high growth rate) are healthier;

3. additional contributions to population death is given by the presence in the environment of noxious elements like bactericidal substances;

*oliver.ebenhoeh@hhu.de

4. changes in metabolite concentrations are in general directly proportional to the growth γ of the consumers and producers;

5. for the DOC_A and COP metabolites we introduce the hypothesis that, in the event of micronutrient scarcicity, the diatom D will secrete more organic carbons favorited by those bacteria (A in our case) able to provide the needed micronutrients (Fe and Vit in our model);

Despite its simplicity and the minimal amount of assumptions made to build it, this model has 55 unknown free parameters (5 carrying capacities CC, 34 maximal rates v, 15 "Michaelis-Menten like" constants K, the fraction of DOC_A-dependent growth $\epsilon_{\mathrm{DOC}_A}$).

1.2 Equations

Five ODEs describe the variation in time of the populations of organism O, with γ^O and δ^O being its growth and death rate:

$$\frac{dD}{dt} = \gamma^D D - \delta^D D \tag{1}$$
$$\frac{dPA}{dt} = \gamma^{PA} PA - \delta^{PA} PA \tag{2}$$
$$\frac{dF}{dt} = \gamma^F F - \delta^F F \tag{3}$$
$$\frac{dA}{dt} = \gamma^A A - \delta^A A \tag{4}$$
$$\frac{dP}{dt} = \gamma^P P - \delta^P P \tag{5}$$

Eight ODEs describe the variation in time of the metabolites J, with $v_J^{\mathrm{prod/cons}(O)}$ being the maximal production/consumption rate of J by organism O:

$$\frac{d\mathrm{Vit}}{dt} = v_{\mathrm{Vit}}^{\mathrm{prod(A)}} \gamma^A A - v_{\mathrm{Vit}}^{\mathrm{cons(D)}} \gamma^D D \tag{6}$$
$$\frac{d\mathrm{Fe}}{dt} = v_{\mathrm{Fe}}^{\mathrm{prod(A)}} \gamma^A A - v_{\mathrm{Fe}}^{\mathrm{cons(D)}} \gamma^D D \tag{7}$$
$$\frac{d\mathrm{DOC}_{PA}}{dt} = v_{\mathrm{DOC}_{PA}}^{\mathrm{prod(D)}} \gamma^D D - v_{\mathrm{DOC}_{PA}}^{\mathrm{cons(PA)}} \gamma^{PA} PA \tag{8}$$
$$\frac{d\mathrm{DOC}_A}{dt} = (v_{\mathrm{DOC}_A}^{\mathrm{prod(D)}} + \phi) \gamma^D D - v_{\mathrm{DOC}_A}^{\mathrm{cons(A)}} \gamma^A A \tag{9}$$
$$\frac{d\mathrm{COP}}{dt} = (v_{\mathrm{COP}}^{\mathrm{prod(D)}} + \psi) \gamma^D D - v_{\mathrm{COP}}^{\mathrm{cons(F)}} \gamma^F F \tag{10}$$
$$\frac{d\mathrm{EPA}}{dt} = v_{\mathrm{EPA}}^{\mathrm{prod(D)}} \gamma^D D - v_{\mathrm{EPA}}^{\mathrm{deg}} \mathrm{EPA} \tag{11}$$
$$\frac{d\mathrm{Bac}}{dt} = v_{\mathrm{Bac}}^{\mathrm{prod(PA)}} \gamma^{PA} PA - v_{\mathrm{Bac}}^{\mathrm{deg}} \mathrm{Bac} \tag{12}$$
$$\frac{d\mathrm{DOM}}{dt} = v_{\mathrm{DOM}}^{\mathrm{prod(D)}} \delta^D D - v_{\mathrm{DOM}}^{\mathrm{cons(A)}} \gamma^A A - v_{\mathrm{DOM}}^{\mathrm{cons(P)}} \gamma^P P \tag{13}$$

ϕ and ψ are additional terms for DOC_A and COP production respectively (see Section 1.3). v_J^{deg} is the degradation rate of the bactericidal substances. Organism O growth and death rates depend in general on carrying capacity CC^O, maximal rates $v_{\gamma/\delta}^O$ and on metabolites concentrations J with Michaelis-Menten-like

constants K_J^O and eventually maximal rates v_J^O:

$$\gamma^D = v_\gamma^D \cdot \frac{\text{Vit}}{\text{Vit} + K_{\text{Vit}}^D} \frac{\text{Fe}}{\text{Fe} + K_{\text{Fe}}^D} (1 - \frac{D}{CC^D}) \tag{14}$$

$$\delta^D = v_\delta^D \frac{1}{1 + \gamma^D} \tag{15}$$

$$\gamma^{PA} = v_\gamma^{PA} \frac{\text{DOC}_{PA}}{\text{DOC}_{PA} + K_{\text{DOC}_{PA}}} (1 - \frac{PA}{CC^{PA}}) \tag{16}$$

$$\delta^{PA} = v_\delta^{PA} (1 + \frac{v_{\text{EPA}}^{PA} \cdot \text{EPA}}{\text{EPA} + K_{\text{EPA}}}) \frac{1}{1 + \gamma^{PA}} \tag{17}$$

$$\gamma^F = v_\gamma^F \frac{\text{COP}}{\text{COP} + K_{\text{COP}}} (1 - \frac{F}{CC^F}) \tag{18}$$

$$\delta^F = v_\delta^F (1 + \frac{v_{\text{Bac}}^F \cdot \text{Bac}}{\text{Bac} + K_{\text{Bac}}^F}) \frac{1}{1 + \gamma^F} \tag{19}$$

$$\gamma^A = \gamma_{\text{DOC}_A}^A + \gamma_{\text{DOM}}^A \tag{20}$$

$$\gamma_{\text{DOC}_A}^A = v_\gamma^A \frac{\epsilon_{\text{DOC}_A} \cdot \text{DOC}_A}{\text{DOC}_A + K_{\text{DOC}_A}^A} (1 - \frac{A}{CC^A}) \tag{21}$$

$$\gamma_{\text{DOM}}^A = v_\gamma^A \frac{(1 - \epsilon_{\text{DOC}_A}) \cdot \text{DOM}}{\text{DOM} + K_{\text{DOM}}^A} (1 - \frac{A}{CC^A}) \tag{22}$$

$$\delta^A = v_\delta^A (1 + \frac{v_{\text{Bac}}^A \cdot \text{Bac}}{\text{Bac} + K_{\text{Bac}}^A}) \frac{1}{1 + \gamma^A} \tag{23}$$

$$\gamma^P = v_\gamma^P \frac{\text{DOM}}{\text{DOM} + K_{\text{DOM}}^P} (1 - \frac{P}{CC^P}) \tag{24}$$

$$\delta^P = v_\delta^P (1 + \frac{v_{\text{Bac}}^P \cdot \text{Bac}}{\text{Bac} + K_{\text{Bac}}^P}) \frac{1}{1 + \gamma^P} \tag{25}$$

In the case of A, where growth is thought to be sustained by two different complementary nutrients, the final growth γ can be represented as the sum of two terms $\gamma_{\text{DOC}_A}^A$ and γ_{DOM}^A (Equations 21 and 22), with the parameter $0 < \epsilon_{\text{DOC}_A} < 1$.

1.3 DOC$_A$ and COP production

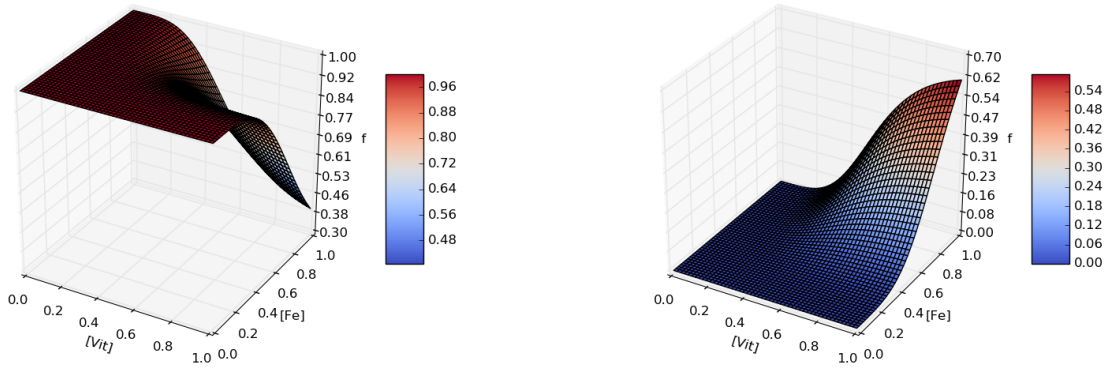
When D is grown in minimal media conditions, the emergence of A is observed over F. From this observation we hypothesise that D can produce extra organic carbons for either A or F depending on the scarcicity of micronutrients to favor the growth of A if more Vit or Fe is needed. We model the production of DOC$_A$ and COP (Equations 9 and 10) introducing the functions ϕ and ψ defined as:

$$\phi = v_{\text{DOC}_A\text{COP}}^D \cdot (1 - \xi) \tag{26}$$

$$\psi = v_{\text{DOC}_A\text{COP}}^D \cdot \xi \tag{27}$$

$$\xi = \frac{\text{Vit}^4}{\text{Vit}^4 + K_{\text{Vit}}'^D} \frac{\text{Fe}^4}{\text{Fe}^4 + K_{\text{Fe}}'^D} \tag{28}$$

where $v_{\text{DOC}_A\text{COP}}^D$ is the maximal additional production rate and $0 < \xi < 1$ depends on Vit and Fe with fourth order Hill equations terms parametrised with $K_{\text{Vit}}'^D$ and $K_{\text{Fe}}'^D$ (see Figure 1).



(a) (b)

Figure 1: Example for DOC_A ((a), $1-\xi$) and COP ((b), ξ) additional production rates dependent on Vit and Fe availability in the media. Here $K'^{D}_{\mathrm{Vit}}= 0.1$, $K'^{D}_{\mathrm{Fe}}= 0.5$.

2 Parameters choice

The model has 56 parameters, of which 55 are free parameters (see Table 1). Being a qualitative model, we do not aim at interpreting the absolute parameter values in a biological sense. We can however draw considerations from relative values and stability tests.

Parameter sub-set	$\mathcal{P}(D)$	$\mathcal{P}(PA)$	$\mathcal{P}(A)$	$\mathcal{P}(F)$	$\mathcal{P}(P)$	degradation
Sub-set size	15	9	14	8	8	2

Table 1: Total number of parameters for each parameter set. The dependent parameter is $\epsilon_{\mathrm{DOM}} = 1 - \epsilon_{\mathrm{DOC}_A}$ in the sub-set of A parameters $\mathcal{P}(A)$.

2.1 Parameter fitting

The available data that can be used to fit the model parameters are the diatom biomass growth in two media conditions and four time points with bacteria relative abundances again in two media conditions. We can therefore fit the diatom biomass D evolution and the four relative bacteria i abundances $B_i / \sum_j B_j$ time-course.

We implement as general strategy a genetic algorithm, where an "individual" i is a full set of 56 parameters $\mathcal{P}_i$, a "population" is an *ensemble* of parameter sets $\{\mathcal{P}_i\}$, a population at a certain evolution step is a "generation" and "evolution" goes as:

1. the first generation $\{\mathcal{P}_i\}^0$ is populated by extracting the parameters from random uniform distributions within user-chosen ranges;
2. for each $\mathcal{P}_i$ the ODE system is solved and a fitness score (see Section 2.1.1) is computed;
3. the most fit 10% individuals are retained as parents for the next generation;
4. the remaining individuals have a probability $P = 0.05$ to be also selected as parents;
5. parents are crossed to obtain enough children to reach the original population size;
6. crossing means randomly pick a parameter sub-set from one parent or the other;
7. each children has a probability $P = 0.3$ to randomly mutate one parameter;
8. the process is repeated from step 2. until generation $\{\mathcal{P}_i\}^{G_{max}}$.

2.1.1 Fitness score

Fitness scores are computed in a different way when fitting the diatom growth or the bacteria relative abundances. When fitting to the diatom biomass data we compute the score as a simple euclidean distance:

$$s = \sqrt{\sum_{\mathrm{t}} (x_t - X_t)^2} \tag{29}$$

where the sum over time extends over 22 time points, x_t is the D biomass at time t and X_t is the biomass data at time t. The lower s, the better the fit. This score definition works well to fit the measurements of diatom biomass, but presents a big problem when used with bacteria relative abundances. A relative abundance is a number between 0 and 1, and we observe high variations including bacteria population going from very close to 0 to high abundance. Having only three time points to fit (the first 16S measurement is used as initial point), it can happen that constantly low abundant population are kept by the algorithm. We therefore define for the fit of bacteria relative abundances the following score:

$$s = \sum_{\mathrm{t}} \sqrt{\sum_{\mathrm{o}} \left(1 - e^{\frac{r_{o_t} - R_{o_t}}{r_{o_t}}}\right)^2} \tag{30}$$

where the sum over time extends over 3 time points and the sum over organisms over the 4 bacterial species, r_{o_t} is the relative abundance from the ODEs system solution for organism o at time t and R_{o_t} is it the corresponding experimental relative abundance. This score definition allows to penalize the event of population extinction: when r is 0, the exponential term is 0 and the score is 1, while when $r = R$ the exponential term is 1 and the score is 0 (see Figure **??**).

T	8	16	40	48	64	72	104	112	120	128	152
MM	0.004	0.021	0.133	0.325	0.820	1.012	1.121	1.187	1.192	1.233	1.209
CM	0.050	0.044	0.162	0.605	0.733	0.919	1.037	1.099	1.134	1.108	0.859
T	168	176	184	208	216	232	240	248	264	272	288
MM	1.104	1.096	0.951	1.015	0.965	0.851	0.869	0.704	0.481	0.504	0.394
CM	0.821	0.844	0.624	0.682	0.624	0.556	0.535	0.478	0.199	0.282	0.303

Table 2: Datasets used to fit diatom growth in minimal and complete media (MM and CM respectively). Time is scaled (1/3 of a day) to fit reasonably the growth phases (lag-log-exp-decay) using parameters $\mathcal{O}(1)$. For the same reason cell counts are scaled to bring the lower count close to 0, but not feature-scaled to avoid loosing information on differences among MM and CM conditions. Only average values, and not experimental errors, are taken into account.

	Complete Media				Minimal Media			
t	PA	F	A	P	PA	F	A	P
64	0.101	0.724	0.159	0.014	0.294	0.132	0.308	0.264
120	0.453	0.474	0.061	0.010	0.351	0.031	0.585	0.031
176	0.600	0.084	0.189	0.126	0.385	0.020	0.187	0.406

Table 3: Relative abundances of the four bacterial families at three intermediate time points (days 8, 15 and 22). The abundances were scaled from the experimental values (where more families were present) to add to unity.

2.1.2 Results of the genetic algorithm

The chosen population size is 200 and the algorithm stops either after non significant increase in fitness or at generation number 50. The algorithm can be run to fit six scenarios:

- D-MM: D Biomass in Minimal Media;
- D-CM: D Biomass in Complete Media;
- B-MM: Bacteria relative abundances in Minimal Media;
- B-CM: Bacteria relative abundances in Complete Media;
- D*B-MM: D Biomass and Bacteria relative abundances in Minimal Media;
- D*B-CM: D Biomass and Bacteria relative abundances in Complete Media;

For D-type fits, the fitness score of Eq. 29 is used. For B-type fits, the fitness score of Eq. 30 is used. For D*B-type fits, the fitness score is the product of the two scores. We will refer to D-fit, B-fit and D*B-fit in the following if media is not to be specified.

Considering the fact that a simple ODE model cannot capture metabolic readjustment, we do not expect to obtain the same parameters for CM and MM conditions. The fitting is therefore performed separately in the two conditions and in the following steps:

1. B-fit is run 20 times varying all 55 parameters in $\mathcal{O}(1)$ ranges
2. The parameters from the best B-fits are kept ($\mathcal{P}_{MM1}$ and $\mathcal{P}_{CM1}$)
3. After checking the effect of varying the different parameters sets (see Figure 2 and Section **??**), different variation ranges are chosen to perform refits
4. D*B-CM is run 5 times varying $\mathcal{P}(D, deg)_{CM1} \pm 50\%$, $\mathcal{P}(A, F, P)_{CM1} \pm 20\%$, $\mathcal{P}(PA)_{CM1} \pm 10\%$
5. D*B-MM is run 5 times varying $\mathcal{P}_{MM1} \pm 50\%$, and the best parameters are kept ($\mathcal{P}_{MM2}$)
6. D*B-MM is run again 5 times varying $\mathcal{P}(D)_{MM2} \pm 5\%$, $\mathcal{P}(A, F, P, PA, deg)_{MM2} \pm 80\%$

Stability analysis (see Figure 2 and Section **??**) shows that the only parameters from other sub-sets influencing the biomass growth curve in CM are δ^A, $v_{\mathrm{DOC}_A}^{\mathrm{cons(A)}}$ and $v_{\mathrm{Fe}}^{\mathrm{prod(A)}}$, but we ignore them for these first iterations.

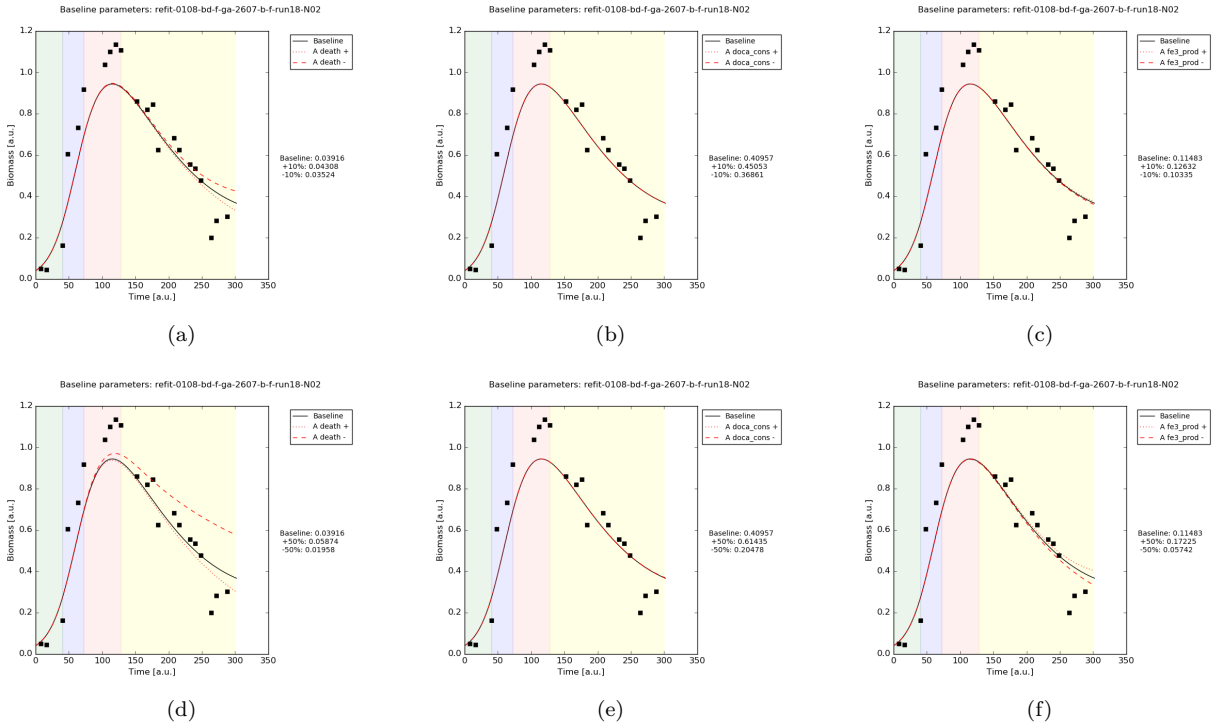


Figure 2: Diatom growth in MM simulation results. The parameters δ^A, $v_{\mathrm{DOC}_A}^{\mathrm{cons(A)}}$ and $v_{\mathrm{Fe}}^{\mathrm{prod(A)}}$, are varied by $\pm 10\%$ (a, b, c respectively) and by $\pm 50\%$ (d, e, f respectively).

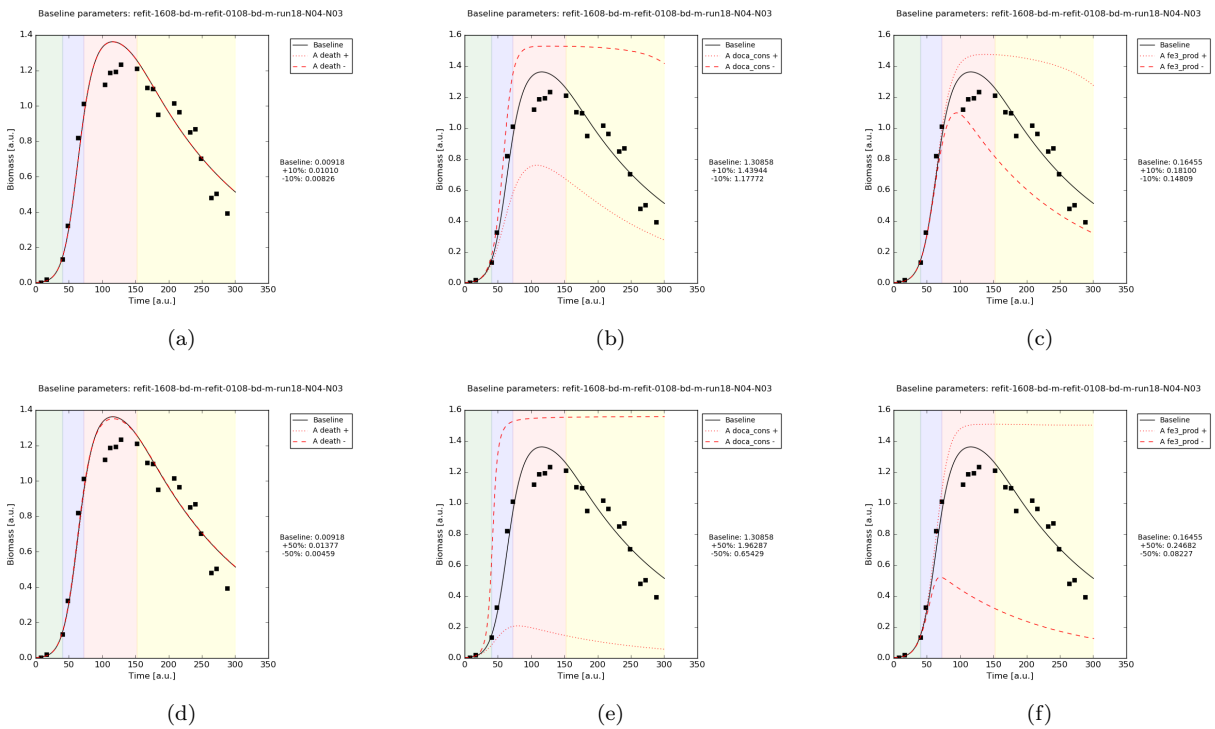


Figure 3: Diatom growth in CM simulation results. The parameters δ^A, $v_{\mathrm{DOC}_A}^{\mathrm{cons(A)}}$ and $v_{\mathrm{Fe}}^{\mathrm{prod(A)}}$, are varied by $\pm 10\%$ (a, b, c respectively) and by $\pm 50\%$ (d, e, f respectively).